Universidade Estadual de Campinas

Faculdade de Engenharia Elétrica e de Computação

Dimensão Topológica e Mapas Auto Organizáveis de Kohonen

Autora: Sarajane Marques Peres

Orientador: Prof. Dr. Márcio Luiz de Andrade Netto

Tese de Doutorado apresentada à Faculdade de Engenharia Elétrica e de Computação como parte dos requisitos para obtenção do título de Doutor em Engenharia Elétrica. Área de concentração: **Engenharia de Computação**.

Banca Examinadora

Fernando Antônio de Campos Gomide, Dr. ..... DCA/FEEC/Unicamp

Fernando José Von Zuben, Dr. ..... DCA/FEEC/Unicamp

Maurício Fernandes Figueiredo, Dr. ..... DC/UFSCar

Agma Juci Machado Traina, Dra. ..... ICMC/USP-São Carlos

Clodoaldo Aparecido de Moraes Lima, Dr. ..... DCA/FEEC/Unicamp

Campinas, SP

Setembro/2006

FICHA CATALOGRÁFICA ELABORADA PELA

BIBLIOTECA DA ÁREA DE ENGENHARIA E ARQUITETURA - BAE - UNICAMP

Peres, Sarajane Marques Peres

P415d Dimensão topológica e mapas auto organizáveis de Kohonen / Sarajane Marques Peres. – Campinas, SP:[s.n.], 2006.

Orientador: Márcio Luiz de Andrade Netto.
Tese (doutorado) - Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação.

1. Inteligência artificial. 2. Redes Neurais (Computação). 3. Fractais. 4. Lógica Difusa. I. Netto Andrade, Márcio Luiz de. II. Universidade Estadual de Campinas. Faculdade de Engenharia Elétrica e de Computação. III. Título

Título em Inglês: Topological dimension and self organizing maps

Palavras-chave em Inglês: Artificial intelligence, Artificial neural Networks, Self organizing maps, Fractal theory, Fuzzy approximated reasoning

Área de concentração: Engenharia da Computação

Titulação: Doutora em Engenharia Elétrica

Banca examinadora: Fernando Antônio de Campos Gomide, Fernando José Von Zuben, Maurício Fernandes Figueiredo, Agma Juci Machado Traina e Clodoaldo Aparecido de Moraes Lima

Data da defesa: 19/09/2006

Resumo

Redes Neurais Artificiais Auto-Organizáveis (RNA-AO), introduzidas por Teuvo Kohonen na década de 60, constituem uma poderosa ferramenta para análise de dados, mais especificamente para análise de agrupamentos, visualização e aproximação de superfícies.

Nesta tese definiu-se uma nova forma para determinar a dimensão topológica do espaço de saída da RNA-AO a partir da análise do conjunto de dados a ser explorado pela rede, realizada com o apoio combinado da Teoria de Fractais e do Raciocínio Aproximado Fuzzy. Ao combinar essas duas teorias, concebeu-se uma nova medida de dimensão fractal, a medida de Dimensão Fractal Fuzzy Significativa (DFFS) de um conjunto de dados.

Tanto o processo de determinação da DFFS quanto sua aplicação como inferência da dimensão topológica para a RNA-AO foram validados neste trabalho. O primeiro por meio de sua aplicação ao problema de Tendência a Agrupamentos e o segundo por meio da análise de qualidade das RNAs-AO projetadas segundo tal inferência.

Palavras-chave: Inteligência Artificial, Redes Neurais Artificiais, Mapas Auto Organizáveis de Kohonen, Teoria de Fractais, Raciocínio Aproximado Fuzzy.

Abstract

Self Organizing Maps (SOM), introduced by Teuvo Kohonen during the decade of 1960's, is a powerful tool for data analysis, mainly for clustering analysis and surface approximation.

In this thesis, we have defined a new way to determine the output space topological dimension of the SOM using the analysis of the dataset to be explored by the map. This analysis is carried out with the combined support of the Fractal Theory and the Fuzzy Approximated Reasoning, deriving a new fractal dimension measure: the Meaningful Fractal Fuzzy Dimension - DFFS (of the Portuguese "Dimensão Fractal-Fuzzy Significativa").

The DFFS determination process and its application as an inference to the SOM topological dimension have been both validated in this work. The former has been carried out through its application to the Clustering Tendency Analysis and the latter through the quality analysis of the SOM designed by such inference.

Keywords: Artificial Intelligence, Artificial Neural Networks, Self-Organizing Maps, Fractal Theory, Fuzzy Approximated Reasoning.

Agradecimentos

Agradeço à minha família, meu bem maior. A meus pais tenho que agradecer por tudo: o que hoje sou, o que hoje posso, o que hoje faço. Ao Marcelo, além de agradecer todo o amor e companheirismo, preciso dizer que meu amor por ele só se fortaleceu diante da sua capacidade de me fazer entender, incontáveis vezes, que apesar de ser um árduo caminho, eu chegaria lá. Ao meu irmão, à minha cunhada, aos meus sobrinhos ... nossa ... quantas vezes estive ausente (festas de aniversários, festas da escola, reuniões de família) e eles sempre compreenderam.

Ao meu orientador sou grata pela orientação, muitas vezes referentes a assuntos que nada tinham a ver com minha tese. Sou grata pelas conversas descontraídas de final de reunião. Sou grata pela compreensão quando as tarefas eram tantas que refletiam em uma evolução um pouco mais lenta dos trabalhos do doutorado. Sou grata! Sou grata! Sou grata!

Agradeço aos professores das disciplinas, pela vontade em disseminar o conhecimento, aos professores da FEEC que ora durante um café na "cantina da mecânica", ora durante as reuniões da CPG, ora nos corredores, estão sempre transmitindo, de alguma forma, as suas experiências. E aos professores componentes desta banca agradeço a disposição para analisar e criticar este trabalho.

À Tia Cida (que apesar de ser "minha" tia, é hoje chamada pelos meus amigos de Tia Cida, o que me deixa bastante, porém positivamente, enciumada) por ter me acolhido em sua casa durante os três anos em que morei em Campinas. Aprendi muito morando com ela. Ela é realmente incrível!

Agradeço ao meu grande amigo Clodis, amigo de rir, amigo de chorar, amigo de morrer de rir pra não chorar durante nossas longas a deliciosas conversas via internet, onde discutíamos sobre tudo ... inclusive sobre nossas impressões, dúvidas e conclusões mediante a nossa comum experiência em cursar um doutorado.

Agradeço a Mábia (com quem aprimorei minha capacidade analítica), a Rúbia (com quem aprimorei minha capacidade em estar sempre disposta) e ao Plínio (com quem aprimorei minha

capacidade de ter paciência), que conheci durante minha estada na FEEC e que terei, ao menos em meu coração, como amigos para sempre.

Aos colegas da Universidade Estadual do Oeste do Paraná e aos da Universidade Estadual de Maringá que sempre me estimularam, me incentivaram, e me "cobraram" a conclusão do trabalho. E às próprias instituições: a primeira por ter me liberado, por três anos, para dedicação exclusiva a este trabalho e à segunda por ter me acolhido durante os dois últimos anos de desenvolvimento deste trabalho.

E, por último, mas não com menos importância, agradeço a Deus por ter tido a oportunidade de fazer todos os agradecimentos acima.

Sumário

Lista de Figuras

Lista de Tabelas

Glossário

- RNA-AO - Redes Neurais Artificiais Auto-Organizáveis: modelos matemáticos que representam o desenvolvimento da auto-organização de projeções topográficas do cérebro natural e são utilizadas em diversas aplicações computacionais que necessitam da interpretação e processamento de dados.

- MRCM - Multiple Reduction Copy Machine: sistema copiador composto por mais de uma lente, em que cada uma delas faz uma cópia reduzida de uma imagem sobre uma mesma folha de papel (em locais diferentes da mesma folha), e que uma vez acionado não finaliza o processo enquanto não lhe for ordenado.

- SFI - Sistemas de Funções Iteradas: um SFI consiste de um espaço métrico completo combinado com um conjunto finito de mapeamentos auto-contrativos, com respeito a um determinado fator de contratividade.

- DCF - Dimensão de Correlação Fractal: uma das variações de dimensão Fractal.

- BC - Box-Counting: uma das variações de dimensão Fractal, comumente utilizadas em experimentos computacionais.

- BMU - do inglês *Best Matching Unit*: neurônio da RNA-AO vencedor da competição para representar um determinado dado de entrada.

- DFFS - Dimensão Fractal Fuzzy Significativa: variação da dimensão Box-Counting introduzida nesta tese. Caracterizada por medir a real porção do espaço onde residem os dados de um conjunto de dados, considerando sua organização em grupos.

- PCA - do inglês *Principal Components Analysis*: técnica estatística que modela uma matriz como uma combinação de componentes principais ortogonais. Esses componentes principais formam um conjunto de variáveis que definem projeções que representam o montante máximo de variação nas dimensões de um conjunto de dados.

Lista de Símbolos

- n - Tamanho de amostras
- d - Número de atributos de um conjunto de dados e número de dimensões do espaço vetorial dos dados
- $\Re^2$ - Espaço Euclidiano
- C - Espaço Esférico
- $(X, dist)$ - Espaço Métrico
- $dist$ - Medida de Distância
- Λ - Atrator (conjunto) sob análise quanto à dimensão fractal
- $(H(X), h)$ - Espaço cujos pontos são subconjuntos compactos de X e onde vale a métrica h
- E - Dimensão de Imersão
- I - Dimensão Intrínseca
- T - Dimensão Topológica
- D - Dimensão Fractal
- log/log - Gráfico formado com pontos especificados pelo logaritmo de suas coordenadas
- $O(f(n))$: notação big-Oh utilizada para representar a definição de um limite superior para a função de complexidade de um algoritmo. Nesse trabalho $f(n)$ assumirá:

  - $O(N)$ - Complexidade algoritmica linear em relação a N - número de dados
  - $O(N * D * i)$ - Complexidade algoritmica em relação a N - número de dados, D - número de atributos descritivos e i - número de iterações de um algoritmo
  - $O(2^E)$ - Complexidade algoritmica em relação a E - dimensão de imersão de um conjunto de dados

- A - Conjunto *Fuzzy*
- $\overline{A}$ - Conjunto *Fuzzy* Complemento de A
- μ_A - Função de pertinência de um conjunto *fuzzy* A
- X - Universo de Discurso
- t-norma e s-norma - Normas Triangulares
- min - Operador do tipo t-norma
- max - Operador do tipo s-norma
- X, T(X), **X**, G, M - Definição de uma variável lingüística: sendo X o nome da variável; $T(X)$ o conjunto de termos de X cujos elementos são rótulos de valores lingüísticos de X; G uma gramática para gerar os nomes de X; M uma regra semântica para associar com cada rótulo $L \in T(X)$ ao seu significado $M(L)$ que é um conjunto fuzzy no universo **X**.
- L_2 - Lógica Clássica ou Lógica Bi-valorada
- L_n - Lógica N-valorada
- L_∞ - Lógica de Infinitos Valores
- L_1 - Lógica Padrão de Lukasiewicz
- τ - Valor Verdade de uma Proposição
- $p \Rightarrow q$ - Operação de Implicação Lógica

- W - Vetores de Peso ou Conjunto de Neurônios de um Rede Neural
- $w_{i,j}$ - Peso que conecta um neurônio da camada de entrada e um neurônio da camada de saída de uma rede neural. Também interpretado como uma das coordenadas do vetor que representa o neurônio a camada de saída.
- $D(s_j)$ - Distâncias de um neurônio j, da camada de saída, aos neurônios da camada de entrada de uma rede neural
- (A) - Espaço de saída de uma Rede Neural Artificial Auto Organizável
- (V) - Espaço de entrada de uma Rede Neural Artificial Auto Organizável
- α - Taxa de aprendizado
- $h(r, s)$ - Função de ajuste de vizinhança
- σ - Desvio padrão da distribuição normal
- μ - Média da distribuição normal
- E_q - Erro de Quantização
- ξ - Erro Topológico
- P - Produto Topográfico
- n_k^V - O $k - ésimo$ vizinho mais próximo do neurônio j no espaço de entrada (V) do mapeamento
- n_k^A - O $k - ésimo$ vizinho mais próximo do neurônio j no espaço de saída (A) do mapeamento
- d_A - Métrica para o cálculo da distância entre dois pontos de uma matriz
- d_V - Métrica para o cálculo da distância entre dois vetores
- ρ - Coeficiente de Spearman

- Medida Z - Medida de "Zrehen"
- $C(\Omega)$ - Medida C
- $\Phi(k)$ - Função Topográfica
- seg - Segmento de reta
- αseg - Inclinação do segmento de reta
- j_{corte} - Ponto de corte da curva Box Counting para cálculo da DFFS
- H_{hpq} - Índice de Hopkins
- U_j e W_j - Distâncias utilizadas no cálculo do Índice de Hopkins
- $\theta(n^2 * d)$ - Limite superior da função de complexidade para o processo de cálculo do Índice de Hopkins
- **U** - Representação da variável lingüística "uniformidade"
- **A** - Representação da variável lingüística "agrupamentos"
- **N** - Representação da variável lingüística "normalidade"
- **P** - Representação da variável lingüística "pontos próximos"
- $P+$ - Menor Produto Topográfico não negativo - Critério 1
- $\mid P \mid$ - Menor Produto Topográfico em módulo - Critério 2
- ξ_2 - Menor Erro Topológico com precisão de duas casas decimais
  - ξ_2^d - considerando a vizinhança em diamente - Critério 3a
  - ξ_2^r - considerando a vizinhança retangular - Critério 3b
- $2\xi_2$ - Segundo menor Erro Topológico com precisão de duas casas decimais
  - $2\xi_2^d$ - considerando a vizinhança em diamente - Critério 4a

  - $2\xi_2^r$ - considerando a vizinhança retangular - Critério 4b

- ξ_4 - Menor Erro Topológico com precisão de quatro casas decimais
  - ξ_4^d - considerando a vizinhança em diamente - Critério 5a
  - ξ_4^r - considerando a vizinhança retangular - Critério 5b
- $2\xi_4$ - Segundo menor Erro Topológico com precisão de quatro casas decimais
  - $2\xi_4^d$ - considerando a vizinhança em diamente - Critério 6a
  - $2\xi_4^r$ - considerando a vizinhança retangular - Critério 6b
- $< E_q$ - Menor Erro de Quantização - Critério 7
- $\mid P \mid \succ \mu E_q$ - Produto Topográfico $\succ$ Erro de Quantização - Critério 8
- $ß^d \succ \mu E_q$ - Erro Topológico $\succ$ Erro de Quantização - Critério 9
- $< E_q \succ P$ - Erro de Quantização $\succ$ Produto Topográfico
  - $< E_q \succ P+$ - considerando o menor valor não negativo para o Produto Topográfico - Critério 10a
  - $< E_q \succ \mid P \mid$ - considerando o menor valor em módulo para o Produto Topográfico - Critério 10b
- $< E_q \succ \xi$ - Erro de Quantização $\succ$ Erro Topológico.
  - $< E_q \succ \xi^d$ - considerando o menor Erro Topológico com vizinhança em diamante - Critério 11a
  - $< E_q \succ 2\xi^d$ - considerando o segundo menor Erro Topológico com vizinhança em diamante - Critério 11b
- $\mid P \mid \succ \pm E_q$ - Produto Topográfico $\succ$ Erro de Quantização Intermediário - Critério 12
- $\xi^d \succ \pm E_q$ - Erro Topológico $\succ$ Erro de Quantização Intermediário - Critério 13

Capítulo 1

Introdução

> "...para expandir o nosso conhecimento temos de nos aventurar além das fronteiras da realidade, o que muitas vezes requer grande coragem intelectual. Nós não enfrentamos dragões de sete cabeças ou feiticeiros malignos, mas batalhamos contra a nossa própria ignorância e limitações." O fim da Terra e do Céu: o apocalipse na ciência e na religião - *Marcelo Gleiser*.

Um dado ou objeto é, substancialmente, diferente de uma informação. Quando analisado isoladamente é um simples conjunto de atributos descritivos (identificadores, adjetivos e medidas) que o diferencia dos demais elementos de sua categoria. Uma informação, ou um dado ou objeto em um sentido mais amplo, num âmbito informacional, possui um conjunto de relacionamentos estabelecidos entre ela(e) e os demais elementos presentes no espaço onde residem.

Esses relacionamentos são capazes de agregar valor aos objetos, produzindo um conhecimento mais refinado, que é significativo quando se pretende analisar o espaço onde os objetos estão localizados. Por exemplo, uma pesquisa de sensoriamento pode constatar o aumento na densidade populacional de uma região geográfica (região A) e a diminuição em outra (região B), enquanto constata um aumento nos investimentos industriais da região A. Cada um dos conjuntos de dados obtidos não leva, isoladamente, a nenhuma conclusão, mas o relacionamento entre "aumento de investimentos industriais" e "aumento da densidade populacional" na região

A pode indicar um relacionamento entre o exôdo da região B e a falta de investimentos em empregos nessa região.

Exemplos mais genéricos, onde os objetos são representados por pontos (vetores) num espaço d-dimensional (com $d > 0$ e principalmente com $d \gg 0$), são utilizados por este trabalho para fomentar a necessidade da exploração e melhoria dos métodos de análise de dados. Os métodos mais utilizados neste tipo de tarefa de análise estão aprisionados na necessidade de visualização dos resultados, reduzindo os espaços de observação a uma dimensão, necessariamente, ≤ 3. No entanto, apesar de freqüentemente este limite dimensional ser suficiente, esta tese advoga que, em alguns casos, este limite pode ser baixo para que os resultados sejam satisfatórios.

Um conjunto que possui n dados com d atributos descritivos é considerado um conjunto residente no espaço d-dimensional. A alta dimensão desse espaço impede a análise visual direta dos mesmos. Sendo assim, esta análise deve ser realizada por meio de técnicas especiais desenvolvidas para este fim (veja GRINSTEIN; TRUTSCHL; CVEK (2001)) ou por meio de técnicas que resultam em informações referentes à estrutura apresentada pelo conjunto de dados: as técnicas de análise de dados.

1.1 Análise de dados

A análise de dados pode ser vista sob diversos aspectos para os quais são desenvolvidas áreas de estudo completas. Classificação e agrupamento de dados são exemplos dessas áreas, onde estudam-se algoritmos e métodos para realização de tais tarefas. A diferença entre as duas está na existência de *rótulos* que fazem a identificação dos objetos *a priori*, presentes nas tarefas de classificação (ou análise discriminante) e ausentes nas tarefas de agrupamento (ou análise de grupos).

Cada uma delas são exploradas por diferentes óticas (estatística, combinatória, otimização linear e não-linear, teoria dos grafos e de conjuntos), gerando as mais diversas formas de realização das tarefas de análise de dados. O contexto deste trabalho abrange a área de agrupamento de dados e esta seção compõe uma revisão suscinta dessa área, com o objetivo de nortear a aplicabilidade das contribuições propostas.

Segundo ARABIE; HUBERT; DE SOETE (1996), agrupamento é a forma de análise de dados mais madura e utilizada, caracterizada por métodos que, de alguma forma, identificam grupos homogêneos de objetos, baseados em dados disponíveis, sejam quais forem esses dados. Para JAIN; DUBES (1988) a prática de agrupar objetos de acordo com suas similaridades é a base para muitas ciências e um dos modos fundamentais de compreensão e aprendizado sobre um conjunto de dados.

O processo de descoberta de agrupamentos realiza, naturalmente, uma redução do conjunto de dados em conjuntos menores e de observação mais fácil. Os conjuntos menores são constituídos de dados (ou itens) similares, escolhidos e agrupados por meio de algum critério capaz de medir essa similaridade considerando os objetivos e usos previamente projetados. Diferentes objetivos de uso para o mesmo conjunto de dados podem levar à determinação de diferentes medidas de similaridade, que são válidas para cada domínio de aplicação.

Um grupo pode ser entendido como um conjunto de elementos que são similares, sendo que diferentes grupos são formados por objetos que não possuam qualquer similaridade. Entretanto, para algumas aplicações, tal entendimento pode ser insuficiente e outros podem ser elencados: o primeiro define um grupo como um agregado de objetos tal que a distância entre quaisquer dois pontos no grupo (distância intra-grupo) seja menor do que a distância entre qualquer ponto do grupo e qualquer ponto fora do grupo (distância inter-grupos); o segundo descreve um grupo como uma região que contém uma densidade relativamente alta de pontos, separados de outras regiões por uma região com uma densidade relativamente baixa de pontos e o terceiro, definido por TASDEMIR; MERéNYIT (2005), advoga que um ponto em um grupo está mais próximo de algum ponto no mesmo grupo do que a qualquer outro de outro grupo. .

Historicamente, os esforços relacionados à atividade de agrupamento desenvolvem-se por caminhos diferentes, estabelecendo uma série de divisões que especificam métodos distintos de tratamento do problema. A "taxonomia" atribuída a essas diferentes formas assume variações na literatura específica que, de um modo geral, estão citadas aqui (ARABIE; HUBERT; DE SOETE (1996), COSTA (1999)):

- Métodos de agrupamento hierárquico: cuja saída é freqüentemente representada em um dendograma. Utilizam-se de uma métrica de distância específica e seguem a noção intui-

tiva de que graus relativos de semelhança entre objetos poderiam ser visualizados em uma representação hierárquica. Podem ser classificados em agrupamento hierárquico aglomerativo ou divisivo, dependendo de como a decomposição hierárquica é formada, se de forma *bottom-up* ou *top-down*.

- Métodos de agrupamento por particionamento: surgiram como uma alternativa aos métodos hierárquicos, utilizando algoritmos que associam cada ponto do conjunto analisado a exatamente um grupo, maximizando a homogeneidade dentro de cada grupo e maximizando a heterogeneidade entre eles. São divididos em duas grandes categorias, os algoritmos baseados em centróides e os baseados em medóides. Os algoritmos baseados em centróides representam cada conjunto usando o centro de gravidade dos exemplos. Os algoritmos que utilizam medóides representam cada conjunto por meio dos exemplos mais próximos ao centro de gravidade;

- Métodos baseados em densidade: neste método os grupos crescem de acordo com a densidade da vizinhança dos objetos ou de acordo com alguma função de densidade especificada. Os grupos descobertos podem ter formatos arbitrários.

- Métodos baseados em grades: aqui uma estrutura de dados em grade de multiresolução é utilizada para quantizar o espaço em um número finito de células que formam a estrutura em grade onde todas as operações de agrupamento são realizadas.

- Métodos baseados em modelos: tentam tentam otimizar o ajuste entre os dados de entrada e algum modelo matemático. Seguem duas abordagens principais: abordagem estatística ou de agrupamento conceitual, em que dado um conjunto de objetos não nominados, é produzido esquema de classificação sobre os objetos (encontrando descrições características de cada grupo, que representa um conceito ou uma classe); abordagem neural, que tenta representar cada grupo por meio de um vetor, que atua como um "protótipo" de grupo.

Neste contexto, introduz-se a adequabilidade da utilização de Redes Neurais Artificiais Auto Organizáveis (RNA-AO), introduzidas por KONONEN (2000) e também conhecidas como Redes de Kohonen ou *Self Organizing Maps* - SOM. Trata-se de um modelo matemático baseado

numa propriedade biológica cerebral, fortemente relacionada com ordenamento espacial das unidades de processamento do cérebro (neurônios). Esta propriedade é necessária para possibilitar a formação de estruturas cerebrais capazes de modelar informações que são apresentadas ao cérebro com o intuito de serem interpretadas e assimiladas. Cada um destes modelos (ou mapas) é capaz de descrever relações topológicas entre as informações usando uma representação N-dimensional.

Geralmente, estas representações têm uma dimensão muito menor do que a real dimensão do espaço onde as informações foram obtidas. Assim, as estruturas cerebrais realizam uma redução de dimensão que é capaz de criar abstrações que refletem características suficientes à interpretação de informações específicas.

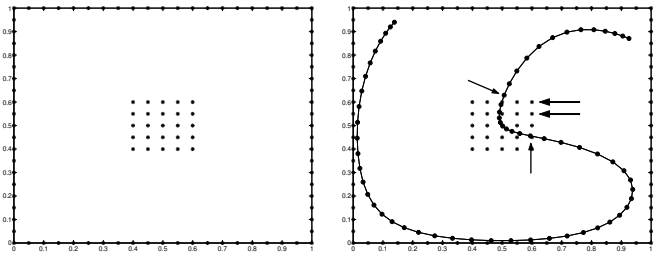
As RNAs-AO são capazes de realizar as tarefas de redução de dimensão e de criar as abstrações citadas acima, além de proporcionar a sumarização do objeto analisado, isto é, a quantização dos dados. Essa tarefa é realizada por meio da construção de um mapeamento entre os dados originais e um mapa, com características quantizadoras e topográficas, que os representa de forma simplificada.

1.2 Motivação

As RNAs-AO são modelos paramétricos de análise de dados. A determinação dos parâmetros deste modelo insere um nível de complexidade na modelagem que pode influenciar no seu uso adequado. Algumas destas dificuldades têm sido estudadas e atenuadas por trabalhos desenvolvidos nesta área de conhecimento (veja ERWIN; OBERMAYER; SCHULTEN (1992a), ERWIN; OBERMAYER; SCHULTEN (1992b), RIESENHUBER; BAUER; GEISEL (1996), COSTA; NETTO (2001) e BAUER; VILLMANN (1995)), mas muito ainda precisa ser explorado. Por exemplo, a redução de dimensão que será aplicada a um conjunto de dados, analisado por uma RNA-AO, é determinada pelo parâmetro de dimensão topológica do mapa de saída da rede e pode gerar distorções topográficas que tornam as relações de vizinhança obtidas no mapeamento diferentes das originais (aquelas presentes no objeto analisado). Estabelecer adequadamente este parâmetro pode significar melhores resultados de representação

da informação.

Na Figura 1.1 é ilustrado um dos problemas de distorção de topologia e geração de falsas relações entre os objetos analisados. Um conjunto de dados localizado no espaço 2-dimensional é mostrado na Figura 1.1(a) e este conjunto está representado por um mapeamento 1-dimensional na Figura 1.1(b). Note que, nessa última figura, existem objetos que possuem uma relação de vizinhança no espaço 2-dimensional em que residem (apontados pelas duas flechas paralelas) que não são corretamente refletidas no mapeamento 1-dimensional (como mostram as outras duas flechas). Esse e outros tipos de problemas de distorção de topologia serão discutidos neste trabalho.



(a) (b)

Fig. 1.1: Exemplo didático de distorção de topologia: (a) conjunto de dados 2-dimensional e (b) mapeamento 1-dimensional.

É fato que em aplicações reais onde os conjuntos de dados possuem um alto número de atributos, a redução de dimensão pode levar à perda de informação importante. Entretanto, redução de dimensão sem perda de informação relevante é possível e justificável. De acordo com KANERVA (1993), a representação vetorial de um conceito, percepção, *momentum* ou pedaço de informação em um contexto de alta dimensão não exige uma forte acuidade e isso advém da natureza das distribuições de alta dimensão, porque qualquer dado em um espaço de altas dimensões está relativamente longe da maior parte deste espaço e de qualquer outro dado nele situado (para um estudo específico sobre este problema veja BEYER et al. (1999)). Esta distância e a grande quantidade de atributos descritivos dos dados tornam possível uma representação inexata, isto é, a omissão de alguns dos atributos não causará problemas de representação. Neste sentido, vetores de altas dimensões são tolerantes a falhas e a ferramenta

ou modelos que faz uso dessa característica é mais robusta que outras. KANERVA (1993) explica essa teoria com o exemplo: "... Os sinais recebidos por sistemas de visão em dois momentos diferentes dificilmente são idênticos, e ainda assim os sistemas podem identificar a fonte do sinal como um indivíduo, objeto, cena, lugar ou coisa específica".

Isto posto, evidencia-se a possibilidade e os benefícios em trabalhar com conjuntos de dados menores em dimensão e em tamanho, contudo não é interessante que, com o alcance desta redução do conjunto de dados, se percam as principais relações existentes entre os dados. As RNAs-AO são capazes de executar tarefas de redução de dimensão e de criar as abstrações citadas acima, porém construindo uma representação dos dados que mantenha, em algum nível de abstração, as relações de vizinhança topológica existentes no conjunto de dados original.

Neste ponto é interessante enfatizar os dois principais objetivos das RNAs-AO que visam facilitar a análise de um conjunto de dados e apresentar um dilema relacionado a eles: o primeiro objetivo está relacionado à redução da dimensão de um conjunto de dados preservando suas relações topológicas; o segundo objetivo trata da criação de abstrações por meio da quantização do espaço; e o dilema: geralmente, quanto menor é o erro (distorção) topológico, maior é o erro de quantização (quantização fraca do espaço). Mais especificamente, quando a dimensão do conjunto de dados é mais alta que a dimensão do mapa de saída da RNA-AO, ou seja, quando existe a redução da dimensão, preservar relações de vizinhança topológica e quantizar o espaço tornam-se metas contraditórias.

1.3 Objetivos

O objetivo principal deste trabalho é estudar o problema de tomada de decisão sobre qual seria a redução de dimensão máxima possível que permita uma boa quantização do espaço sem cair em problemas como: distorções topográficas, sobre-ajuste (fenômeno caracterizado pela representação fiel dos dados por um mapeamento, ocasionando perda na capacidade de geração de abstrações) ou sub-ajuste (fenômeno caracterizado pela rigidez de um mapeamento que não consegue atingir uma flexibilidade suficiente para mapear os dados do conjunto analisado).

A resposta para essa tomada de decisão resulta em uma medida de dimensão onde suposta-

mente seria possível alcançar satisfatoriamente os dois objetivos da RNA-AO. Propõe-se, neste trabalho, o uso desta medida de dimensão como determinante da dimensão do espaço de saída da RNA-AO.

Como objetivos marginais pretende-se:

- modelar, automatizar e testar uma abordagem para realizar a tomada de decisão;
- analisar os efeitos desta tomada de decisão em relação a diferentes perspectivas de análise de qualidade dos mapeamentos das RNAs-AO;
- estudar o resultado desta tomada de decisão quando aplicada sobre conjuntos de dados que apresentam diferentes características, com o intuito de fornecer indicações de tipos de conjunto de dados para os quais ela é mais adequada.

1.4 Metodologia

Para alcançar o objetivo deste estudo, propõe-se a exploração das características fractais do conjunto de dados que será submetido à análise via RNA-AO. Essa exploração resulta na análise da complexidade do conjunto de dados por meio da medição de sua Dimensão Fractal, ou seja, verificando se existe um número mínimo de atributos capazes de descrever o conjunto de dados (TRAINA; TRAINA; FALOUSTOS (1999)), ou verificando qual é a real porção do espaço que é, efetivamente, ocupada pelo conjunto. Supôs-se possível o uso desta informação para determinar a dimensão ideal para o mapeamento e a verificação desta hipótese, sob a concepção de uma nova medida de dimensão fractal, constitui a principal contribuição deste trabalho.

Existem diferentes formas para encontrar a Dimensão Fractal de um objeto e, entre elas, está o Algoritmo Box-Counting, que tem como resultado a medida de dimensão fractal Box-Counting, uma variação de Dimensão Fractal. Este algoritmo é o mais adequado para implementações computacionais segundo a literatura (PEITGEN; JURGENS; SAUPE (1992), BARNSLEY (1988)) e mais comumente encontrado em trabalhos computacionais. O processo

de execução deste algoritmo fornece dados que são explorados neste trabalho para suportar a geração de conhecimento sobre um conjunto de dados.

Tal conhecimento viabiliza a inferência da dimensão para o projeto do espaço de saída de uma RNA-AO a ser aplicada para analisar o conjunto de dados, e também proporciona a realização de algumas considerações sobre a existência de agrupamentos no conjunto de dados, constituindo, essa última, uma contribuição mais geral deste trabalho para a área de análise de dados.

A automação do processo de inferência é proporcionada por meio da concepção de uma modificação no critério de parada do processo Box-Counting, realizada com os recursos da Teoria de Raciocínio Aproximado. Essa teoria proporciona uma análise refinada das informações geradas durante o próprio processo, concluindo quando este deve ser interrompido, dando origem à nova medida de dimensão fractal.

O resultado dessa automação foi avaliado mediante a sua execução para resolver um problema clássico de análise de dados - Tendência a Agrupamentos - e também por meio da comparação de resultados com aqueles obtidos na aplicação de uma solução clássica para o mesmo problema. Só então, a partir da obtenção de resultados que mostraram a verossimilhança da automação, é que ela foi aplicada para atender ao objetivo principal deste trabalho. O resultado dessa aplicação foi analisado sob diversos critérios de avaliação, os quais foram projetados de forma a satisfazer diferentes interpretações das medidas de qualidade (critério de validação internos[1] e relativos[2]) aplicadas às RNAs-AO construídas durante os testes.

1.5 Organização deste trabalho

Este trabalho é apresentado em 8 capítulos (incluindo esta introdução), referências bibliográficas e 4 apêndices. A introdução teve o objetivo de situar o leitor no contexto do trabalho assim como apresentar brevemente os problemas envolvidos na área de Análise de Dados para motivar a sua exploração.

Para o alcance do objetivo principal desta tese utilizou-se alguns recursos de duas im-

[1] Medem a qualidade da solução utilizando apenas os próprios dados.
[2] Decidem qual, dentre algumas estruturas, é a melhor para representar os dados.

portantes áreas de estudo matemático: Teoria de Fractais e Raciocínio Aproximado *Fuzzy*, apresentadas no Capítulo 2 e no Capítulo 3, respectivamente. Em ambas as apresentações está-se prestando atenção especial aos conceitos utilizados para o desenvolvimento do trabalho. Mais especificamente, em Teoria de Fractais dá-se ênfase à questão de Dimensão Fractal em sua versão Dimensão Box-Counting; e em Raciocínio Aproximado Fuzzy aos conhecimentos referentes a Lógica Fuzzy.

O Capítulo 4 introduz a teoria das Redes Neurais Artificiais Auto Organizáveis (RNAs-AO) salientando os aspectos relativos às medidas de qualidade de mapeamento referentes à distorção topológica e quantização vetorial. A otimização do projeto de mapeamento, avaliada por meio dos quesitos de qualidade, é o principal objetivo da tese.

As contribuições originais do trabalho, obtidas a partir da modelagem de uma abordagem Fractal Fuzzy de apoio à tomada de decisão, são apresentadas no Capítulo 5. Trata-se de uma abordagem baseada em regras fuzzy aplicada sobre informações originárias do processo de obtenção da Dimensão Box-Counting de um conjunto de dados[3]. A partir do resultado deste modelo, é proposta uma nova medida de dimensão fractal aqui chamada de "Dimensão Fractal Fuzzy Significativa" (DFFS). Com ela é possível decidir sobre a dimensão do espaço de saída da RNA-AO e também de inferir outras informações que podem ser úteis como "conhecimento *a priori*" para trabalhos de descoberta de agrupamentos.

A automação da abordagem descrita no Capítulo 5 é validada no Capítulo 6. A validação é feita por meio da aplicação da Abordagem Fractal Fuzzy ao problema de Tendência a Agrupamentos e por meio da comparação das soluções apresentadas por ela às soluções obtidas com a Abordagem de Hopkins - um modelo clássico para a resolução desse problema. Testes, resultados, comparações e análises também estão descritos neste capítulo.

A fim de demonstrar os resultados obtidos com a utilização da abordagem Fractal Fuzzy para alcançar o principal objetivo deste trabalho, discute-se, no Capítulo 7, a determinação da dimensão ideal para o espaço de saída de uma RNA-AO por meio da aplicação dessa abordagem como forma de medir a real dimensão do conjunto de dados sob análise. Neste capítulo são apresentados os testes realizados sobre diferentes conjuntos de dados, bem como as análises dos

[3]Conjunto de dados que se pretende analisar via RNA-AO.

resultados obtidos.

Finalmente, no Capítulo 8, as considerações finais são discutidas. Virtudes e limitações do uso da abordagem proposta são pontuadas, bem como sugestões de melhoramentos e novas oportunidades de aplicação são apresentadas, indicando caminhos a serem pesquisados futuramente.

Nos apêndices A, B C e D são apresentados, respectivamente, os conjuntos de dados 2 e 3-dimensionais utilizados nos testes referentes à validação da Abordagem Fractal Fuzzy, uma descrição resumida dos conjuntos de dados reais utilizados nos testes, alguns exemplos de tabelas de dados nas quais se organizam informações de análise para obtenção dos resultados apresentados no Capítulo 7, e na forma digital (CD), informações adicionais referentes aos conjuntos de dados, implementações e testes realizados nesta tese, .

Capítulo 2

Teoria de Fractais

> Os cientistas não estudam a natureza porque ela é útil, mas porque nela os homens deleitam-se e o fazem porque ela é bonita. Se a natureza não fosse bonita, não valeria a pena conhecê-la e se a natureza não fosse digna de ser conhecida, a vida não seria digna de ser vivida. (Henri Poincaré **apud** PEITGEN; JURGENS; SAUPE (1992))

Geometria Euclidiana, trigonometria, cálculo e outras áreas da matemática foram descobertas por causa da necessidade que a humanidade sentiu de entender, modelar e interpretar as formas observadas no mundo real. A modelagem dessas formas se deu, e ainda se dá, por meio de linhas retas, curvas, parábolas e outras formas simples. Geometria Euclidiana e funções elementares tais como seno, co-seno e polinômios são a base dos métodos tradicionais de análise experimental de dados que podem fornecer aproximações da estrutura física dos objetos e, em um nível menos preciso, das tais formas naturais.

No entanto, tais teorias clássicas não são suficientes ou adequadas para modelar estruturas complexas comumente encontradas na natureza. Nuvens, costas, vegetais, assumem formas irregulares que, para fins de estudo e análise, são vistas como **fractais**, objeto de estudo da Geometria Fractal ou Geometria da Natureza como foi tratada por Mandelbrot, um dos pesquisadores responsáveis pelo nascimento desta teoria (PEITGEN; JURGENS; SAUPE (1992)). A Geometria Fractal é vista como uma extensão da Geometria Clássica e fornece uma

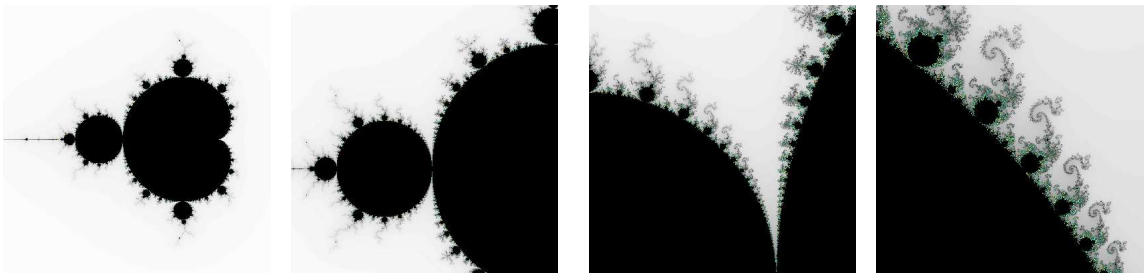
nova linguagem para tratamento de estruturas físicas.

Neste capítulo é fornecida uma visão suscinta da Teoria de Fractais com ênfase na Geometria Fractal e nas questões referentes à sub-área Dimensão Fractal. O método de análise Box-Counting, que fornece a dimensão Box-Counting do objeto analisado, é tratado com maiores detalhes por ser o método escolhido para aplicação ao problema tratado nesta tese. Além disso, alguns conceitos definidos aqui serão úteis para o estudo de topologia de SOM, discutidos no Capítulo 4.

2.1 Introdução

O termo **fractal** (do latim "*fractus*") significa "irregular" ou "quebrado" e é usado para designar objetos irregulares que em diferentes escalas de observação possuem uma estrutura detalhada a ponto de não poderem ser explicados dentro da Geometria Tradicional (ou Geometria Euclidiana). Para descrevê-los é necessário o uso de uma geometria específica, a **Geometria Fractal**.

Muitos objetos classificados como fractais são compostos de partes que, de alguma forma, são similares ao todo, como se observa na Figura 2.1 onde em (a) o Conjunto de *Mandelbrot* está mostrado em sua totalidade, e nos demais quadros, foram realizadas ampliações gradativas para mostrar o padrão de repetição presente na estrutura.



(a) (b) (c) (d)

Fig. 2.1: Conjuntos de Mandelbrot[2]: (a) em tamanho natural e (b), (c) e (d) em três ampliações.

[2]Imagem obtida em `http://www.mat.uc.pt/~jaimecs/imagfrac.html` em 26 de setembro de 2002.

Algumas definições para esses objetos são encontradas na literatura, entretanto, todas se mostram insatisfatórias, já que não conseguem abranger todo o espaço ocupado e descrito pela geometria fractal e excluem casos interessantes. A definição original de fractais dada por *Mandelbrot* (FALCONER (1990)) define um fractal como um conjunto com dimensão *Hausdorff* [3] estritamente maior que sua dimensão topológica.

Contudo, existe atualmente um consenso que Fractais não são definidos em pequenos "*statements*" mas por muitas figuras e contextos. Sendo assim, é preferível entender fractais como um conjunto F que, tipicamente, possui as seguintes propriedades (FALCONER (1990)):

- tem uma estrutura fina, isto é, detalhada em pequenas escalas definidas arbitrariamente;
- é suficientemente irregular para ser descrito em uma linguagem geométrica tradicional, tanto localmente quanto globalmente;
- freqüentemente tem alguma forma de auto-similaridade;
- usualmente possui uma dimensão fractal (definida de alguma maneira) que é maior do que sua dimensão topológica;
- em muitos casos é definido de forma simples e recursiva.

Existem basicamente três classes de objetos fractais:

- fractais geométricos: que repetem, continuamente, um padrão idêntico ou com algum grau de similaridade e são construídos a partir de transformações, por exemplo, o Triângulo de *Sierpinsk*;
- fractais algébricos: que são gerados pela aplicação recursiva de uma expressão algébrica. Exemplos desta classe são: Conjuntos de *Mandelbrot*, Conjuntos de *Julia* e os atratores estranhos de *Barnley* e *Lorenz*;
- fractais estocásticos: que são gerados pela distribuição de objetos no espaço, segundo uma distribuição de probabilidade e independente da escala considerada (Figura 2.2). A

[3] Taxonomia de dimensões é discutida na Seção 2.3

maioria dos objetos naturais, bem como seus fenômenos, podem ser descritos por esta classe de fractais (TRAINA; TRAINA; FALOUSTOS (1999)).

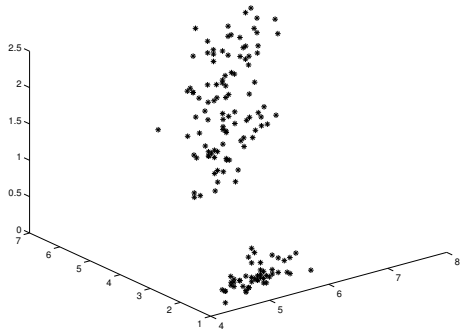


Fig. 2.2: Exemplo de fractal estocástico.

2.2 Conceitos Básicos

A Teoria Fractal estuda subconjuntos "complicados" de espaços simples. Nesta teoria, os focos de atenção são os subconjuntos do espaço gerados por transformações simples do espaço dentro dele mesmo. Fractais são objetos residentes em espaços métricos completos, como o $\Re^2$ (Euclidiano) ou o C (Esférico), denotados por $(X, dist)$.

Um espaço métrico $(X, dist)$ é a combinação de um espaço X e uma função $dist : X \times X \rightarrow \Re$, no qual as medidas de distâncias d entre pares de pontos x e y em X obedecem aos seguintes axiomas:

1. $dist(x, y) = dist(y, x), \forall x, y \in X$
2. $0 < dist(x, y) < \infty, \forall x, y \in X, x \neq y$
3. $dist(x, x) = 0, \forall x \in X$
4. $dist(x, y) \leq dist(x, z) + dist(z, y), \forall x, y, z \in X$

Uma transformação simples é aquela facilmente explicada com o uso de um pequeno conjunto de parâmetros. Assim, define-se uma transformação simples da seguinte forma: seja

$(X, dist)$ um espaço métrico. Uma transformação em X é uma função $f : X \to X$, a qual associa exatamente um ponto $f(x) \in X$ para cada ponto $x \in X$.

Um espaço métrico $(X, dist)$ é completo se toda seqüência de pontos $\{x_n\}_{n=1}^{\infty}$ em X, que converge para um ponto $x \in X$ (é uma seqüência Cauchy), tem um limite $x \in X$.

O conceito de distância entre dois pontos num espaço é dependente da métrica utilizada, a qual determina a estrutura geodésica do espaço. Geodésica numa esfera são grandes círculos; no plano, com a métrica Euclidiana, são linhas. Então, a noção de distância em um espaço N-dimensional tem, ou deve ter, uma interpretação relacionada ao que se mede.

Essa noção leva a outros dois conceitos importantes: equivalência métrica e homeomorfismo. Duas métricas $dist_1$ e $dist_2$ em um espaço X são equivalentes se existem constantes $0 < c_1 < c_2 < \infty$ nesse espaço, de modo que $c_1 dist_1(x, y) \leq dist_2(x, y) \leq c_2 dist_1(x, y)\ \forall\ (x, y) \in X \times X$. Assim, qualquer par de métricas equivalentes fornece a mesma noção de quais pontos estão próximos e quais estão distantes. Isto leva ao conceito de espaços métricos equivalentes: dois espaços métricos $(X_1, dist_1)$ e $(X_2, dist_2)$ são equivalentes se existe uma função $h : X_1 \to X_2$ que é um-para-um e inversível, tal que a métrica $\bar{dist}_1$ em X_1 definida por $\bar{dist}_1(x, y) = dist_2(h(x), h(y))\ \forall\ x, y \in X_1$ é equivalente a $dist_1$. E por fim, se a função que mapeia um espaço para outro e sua inversa são contínuas, esta função é um homeomorfismo entre X_1 e X_2 e dizemos que estes espaços são homeomórficos.

Como ilustração considere o exemplo (BARNSLEY (1988)): dado um par de pontos x e y em uma região $\diamond \subset \Re$, tome a distância Euclidiana entre este pontos $(dist_1(x, y))$. Imagine um pedaço de borracha sobre $\diamond$, a qual é esticada várias vezes, levando cópias dos pontos x e y para novas localizações. A distância Euclidiana entre estes pontos deslocados é dita $dist_2(x, y)$. Observando a Figura 2.3 é possível verificar que as distâncias Euclidianas entre x e y medidas antes e depois da deformação são equivalentes. Estas métricas permanecem equivalentes se a deformação do espaço não leva a dobras, sobreposições, rasgos ou a "esticamentos infinitos". A Figura 2.4 mostra dois espaços que não são metricamente equivalentes (suas geometrias são diferentes) mas são homeomórficos (têm a mesma topologia[4]).

Um exemplo clássico de um conjunto com características de fractais é o Conjunto de *Cantor*

[4]Em topologia, linhas retas podem se transformar em curvas e círculos podem se transformar em triângulos ou quadrados.

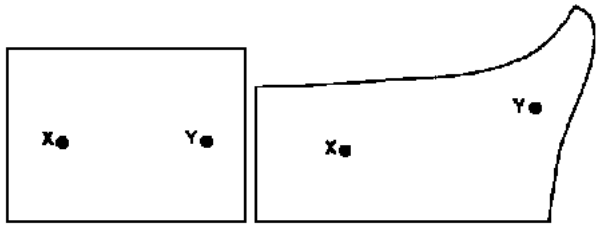


Fig. 2.3: Exemplos de espaços de mesma geometria (BARNSLEY (1988)).

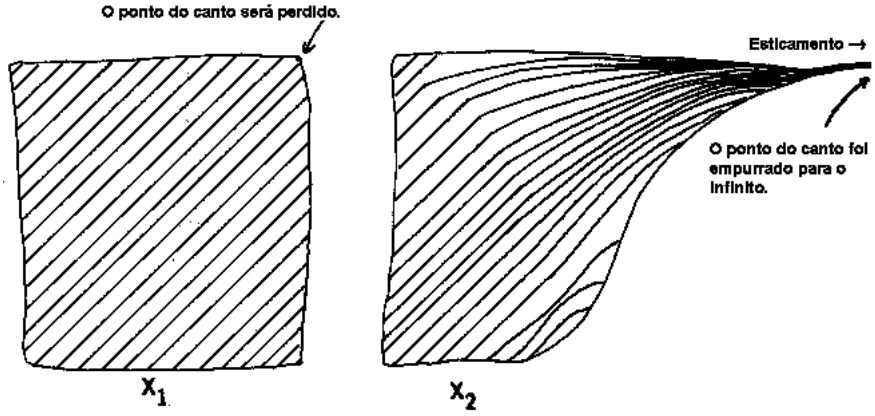


Fig. 2.4: Exemplo de espaços com mesma topologia e geometrias diferentes. Adaptado de BARNSLEY (1988).

(PEITGEN; JURGENS; SAUPE (1992))[5]. Ele é composto por um conjunto infinito de pontos no intervalo $[0, 1]$, por exemplo $0, 1, 1/3, 2/3, 1/9, 2/9, 7/9, ...$ Plotando estes e todos os pontos pertencentes ao conjunto não se obteria nada além de "pontos". Entretanto, plotando todos os pontos que não pertencem a esse conjunto obter-se-iam linhas verticais de mesmo comprimento cujos pontos base são exatamente todos os pontos que não pertencem ao Conjunto *Cantor*.

A construção do Conjunto *Cantor* inicia com o intervalo $[0, 1]$. Em seguida, o intervalo $(1/3, 2/3)$ é removido gerando dois novos, $[0, 1/3]$ e $[2/3, 1]$ de comprimento $1/3$ cada um. Repetindo-se este passo de forma recursiva e infinita obtém-se, como limitantes dos intervalos, o conjunto infinito de pontos que compõe o Conjunto *Cantor* (Figura 2.5). Observe que mesmo repetindo-se o passo infinitamente, os pontos limitantes de intervalos permanecem no conjunto. O Conjunto *Cantor* é uma transformação simples dele mesmo, como pode ser observado no

[5]Outros exemplos de fractais clássicos e a descrição dos respectivos processos de construção podem ser encontrados em PEITGEN; JURGENS; SAUPE (1992).

seu processo de construção. Esta transformação é definida como $\{R; w1, w2\}$ onde $w1 = \frac{1}{3x}$ e $w2 = \frac{1}{3x} + \frac{2}{3}$.

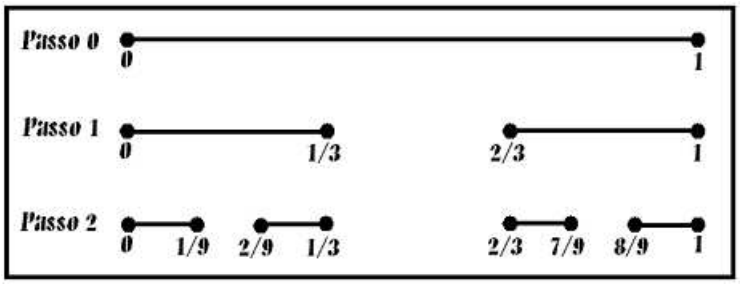


Fig. 2.5: Pontos limitantes dos intervalos - Conjunto Cantor.

Para entender melhor o mundo descrito pela teoria dos fractais é necessário explorar também, os conceitos de similaridade e auto-similaridade. A idéia de similaridade pode ser observada quando a partir de uma imagem se produz uma outra por meio de um processo recursivo de redução da primeira (Figura 2.6). Após um número suficiente de ciclos, a imagem inicial tende a reduzir-se a um ponto. Um exemplo típico seria o processo de obter-se fotocópias reduzidas de uma imagem impressa. Teoricamente, as cópias geradas no processo descrito acima (transformação de similaridade ou similitude) são ditas similares à imagem original (PEITGEN; JURGENS; SAUPE (1992)).

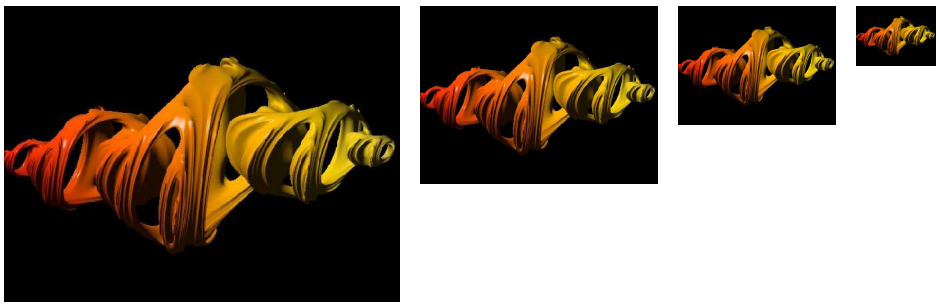


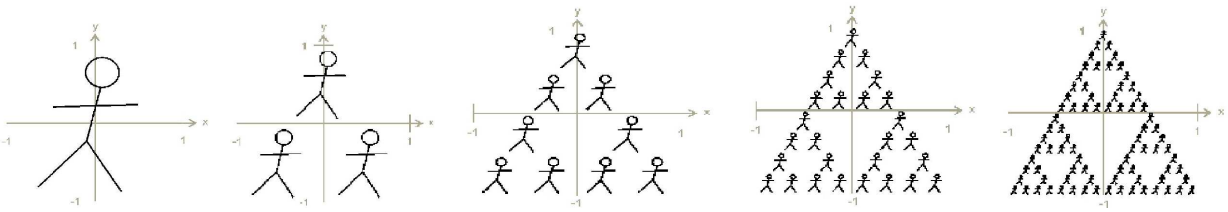
Fig. 2.6: Iterações do processo de redução sobre o Conjunto de *Julia*[6].

Um processo semelhante a esse, conhecido como *Multiple Reduction Copy Machine (MRCM)*, ilustra como os conceitos de similaridade e iteratividade se relacionam à construção de fractais.

[6]Imagem obtida em `http://www.geocities.com/CapeCanaveral/2854/` em 26 de setembro de 2002.

O processo MRCM, descrito e analisado em PEITGEN; JURGENS; SAUPE (1992), é um sistema copiador composto por mais de uma lente, em que cada uma delas faz uma cópia reduzida de uma imagem sobre uma mesma folha de papel (em locais diferentes da mesma folha), e que uma vez acionado não finaliza o processo enquanto não lhe for ordenado.

A Figura 2.7 mostra o resultado obtido após algumas iterações do processo MRCM de três lentes, ilustrando que figuras completas são formadas a partir de figuras iguais a elas mesmas numa escala de observação diferente. O objeto em formação no processo ilustrado na Figura 2.7 é o Triângulo de *Sierpinsk*.



(a) (b) (c) (d) (e)

Fig. 2.7: Resultado da ativação de um sistema MRCM[7].

.

Um objeto é considerado auto-similar quando contém partes que ao serem removidas recursivamente e comparadas com o todo, apresentam formatos similares (senão iguais), porém menores. Numa idealização teórica, um fractal pode possuir a propriedade de auto-similaridade de forma infinita.

No caso de ser necessário estabelecer uma definição única para fractais, a literatura da área geralmente trabalha com a definição de fractais determinísticos. Um fractal dessa natureza é um ponto fixo de uma transformação contratora de $(H(X), h)$, em que $H(X)$ denota o espaço cujos pontos são subconjuntos compactos de X e onde vale a métrica h, sendo que o mapeamento de contração deve ser de construção simples e facilmente especificado.

Seja $w : X \to X$ uma transformação contratora, então:

1. w é contínuo;

[7]Imagem obtida em `http://www.cs.bris.ac.uk/~pclark/ifs/` em 26 de setembro de 2002

2. w mapeia $H(X)$ dentro dele mesmo;

3. se w vale em $X \to X$ com um fator de contração, então vale em $H(X) \to H(X)$ com o mesmo fator de contração.

Fractais determinísticos podem ser gerados por Sistemas de Funções Iteradas (SFI). Um SFI consiste de um espaço métrico completo (X, d) combinado com um conjunto finito de mapeamentos de contração $w_n : X \to X$, com respeito ao fator de contratividade s_n, para $n = 1, 2, ..., N$. Existe um ponto $A \in H(X)$, que sob determinadas condições (BARNSLEY (1988)) é chamado atrator SFI, o qual constitui o fractal.

Como exemplo, note o SFI que determina o Conjunto Cantor: $\{R; w1, w2\}$ onde $w1(x) = 1/3x$ e $w2(x) = 1/3x + 2/3$, com fator de contração $s = 1/3$ e intervalo inicial $[0, 1]$. Em (BARNSLEY (1988)) é ilustrada a forma de computação de um SFI.

2.3 Dimensão Fractal

O conceito **dimensão** é um grande ponto de discussão dentro da Matemática e, a partir de estudos e discussões acerca deste tema, estabeleceram-se diversas noções de dimensão que estão, de alguma forma, relacionadas. Para cada contexto ou situação onde seja necessário o conceito de dimensão uma ou mais noções existentes se adequam e fazem sentido. A taxonomia utilizada para os diferentes conceitos de dimensão pode variar e, até mesmo, uma troca de nomenclatura pode ocorrer causando confusões[8].

Segundo PEITGEN; JURGENS; SAUPE (1992), o conceito de dimensão pode ser intuitivamente explicado como o número de parâmetros (coordenadas) independentes requeridos para descrever unicamente um ponto num espaço. Entretanto, este conceito pode fornecer informações sobre um objeto de diversas maneiras.

A dimensão **de imersão** E de um objeto ou conjunto de dados é a dimensão do espaço onde ele reside. Um vetor residente num espaço vetorial possui componentes que o localizam em cada eixo do espaço, assim o número de componentes do vetor ou o número de eixos do

[8]Neste trabalho, a taxonomia descrita a seguir será utilizada mesmo quando se tratar de referência a trabalhos que utilizam outra taxonomia. Neste caso, toma-se o cuidado de analisar o conceito envolvido para localizá-lo na taxonomia adotada.

espaço onde reside quantificam E. Formalmente, seja v um vetor descrito pelas coordenadas $x1, x2$ e $x3 \in \Re$, o espaço no qual v reside tem dimensão E igual a 3 (três) ($\Re^3$). Outra forma de exemplificar o conceito de dimensão de imersão é considerando um conjunto de dados que é descrito por uma lista de atributos; diz-se que E, para esse conjunto, é o número de atributos.

A **dimensão intrínseca** I de um objeto ou conjunto de dados é o número mínimo de parâmetros livres necessários para gerar ou representar o objeto ou os dados analisados.

A **dimensão topológica** T de um objeto pode ser vista como a mesma dimensão de um segundo objeto topologicamente equivalente ao primeiro, mas que tenha dimensão I inteira. Por exemplo, uma curva qualquer é topologicamente equivalente a uma reta, cuja dimensão I é igual a 1. Diz-se que um ponto possui dimensão I igual a 0 e uma reta possui dimensão I igual a 1, um plano tem dimensão I igual a 2, um cubo tem dimensão I igual a 3. No entanto, esses objetos não representam a diversidade de objetos existentes e a dimensão fractal preenche esta lacuna.

A **dimensão fractal** D de um conjunto de dados é a dimensão do objeto representado pelo conjunto de dados, desconsiderando-se a dimensão E, ou seja, um conjunto de dados pode representar um objeto que tem dimensão menor ou igual à dimensão do espaço no qual ele reside. Esta medida é geralmente usada para comparar fractais já que, segundo BARNSLEY (1988), ela pode quantificar, de forma subjetiva, a percepção que se tem sobre o quão densamente um fractal ocupa o espaço métrico onde está definido. Uma curva, por exemplo, apesar de residir num espaço de dimensão E igual a 2, pode ter dimensão D menor ou igual a 2.

O uso do conceito de dimensão fractal tem sido estimulado em diversas áreas da ciência, visto que, por meio dele, acredita-se poder descobrir uma nova forma de interpretar a organização das estruturas e dos fenômenos complexos que compõem o mundo e poder analisá-los com maior precisão. Diferentes formas de interpretar o que significa dimensão fractal são encontradas na literatura, destacando-se:

- Em BARNSLEY (1988): a dimensão fractal pode ser entendida como uma tentativa de quantificar a densidade de um fractal em relação ao espaço métrico em que ele reside.
- Em BARBARá; CHEN (2003): a dimensão fractal mede o número de dimensões preenchidas pelo objeto representado pelo conjunto de dados.

Assim, a dimensão D pode ser utilizada para extração de informação de um conjunto de dados. Ela mensura o grau de complexidade de um objeto ou o espaço que é efetivamente ocupado por ele, podendo indicar redundâncias ou similaridades de representação.

O desenvolvimento das noções de dimensão foi influenciado pela descoberta das curvas de preenchimento espacial (*space-filling curves*) e do estudo de procedimentos para medição de seu comprimento). Inicialmente essas curvas eram consideradas objetos 1-dimensionais, mas percebeu-se que elas podiam preencher um plano (objeto 2-dimensional) e então iniciou-se um estudo sobre sua real medida de dimensão. Alguns exemplos de curvas de preenchimento espacial são mostrados na Figura 2.8 (conhecidas como curvas de *Hilbert* e curvas de *Sierpinsk*, respectivamente).

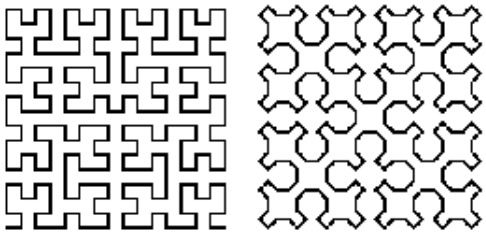


Fig. 2.8: Curvas de preenchimento espacial.

Uma das vantagens de se usar a medida de dimensão fractal é o fato dela poder ser obtida numericamente. Por exemplo, considere o experimento de determinação da dimensão fractal de conjuntos do mundo real, descrito em BARNSLEY (1988), reproduzido aqui de forma resumida: na Figura 2.9 existe uma nuvem de pontos na paisagem. Para medir a dimensão fractal desta nuvem inicia-se por cobrir a nuvem por discos de raio ϵ. Essa cobertura é realizada várias vezes, para uma variação de $\epsilon = 3cm$ a $\epsilon = 0.3cm$ e para cada caso conta-se o número de discos necessários $\aleph(\Lambda, \epsilon)$ para cobrir toda a nuvem (Λ - o atrator). Os valores obtidos neste procedimento estão listados na Tabela 2.1 na forma log/log, tomando ϵ como uma medida de precisão, e estão plotados na Figura 2.10. A dimensão fractal da nuvem de pontos é então calculada como sendo a inclinação da linha reta que melhor aproxima os pontos projetados, sendo determinada como $\approx 1,2$.

Esse experimento ilustra um procedimento para o cálculo da dimensão fractal. Existem

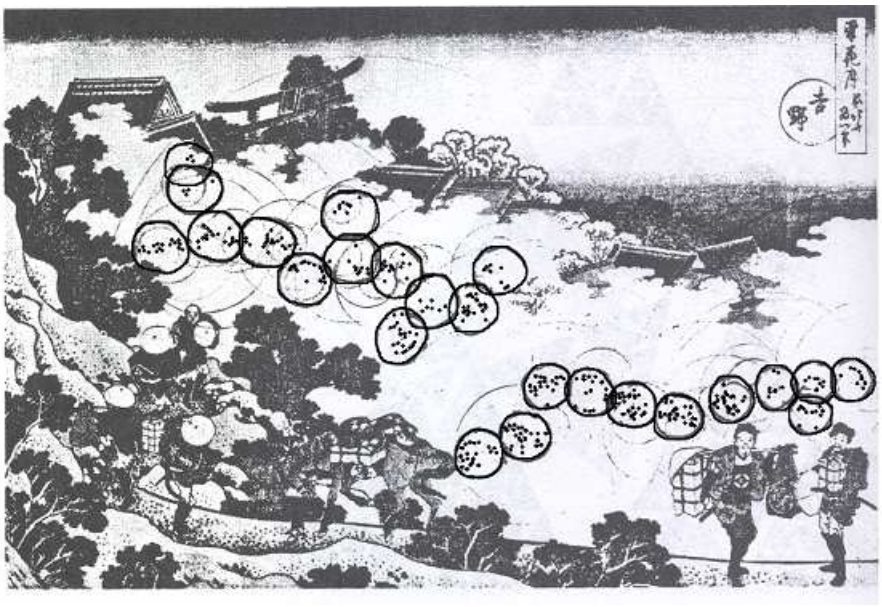


Fig. 2.9: Cobertura da nuvem de pontos por discos de raio $\epsilon > 0$ (BARNSLEY (1988)).

diferentes formas de calcular esta medida e cada uma delas origina medidas diferentes que são casos especiais de dimensão fractal (dimensão de auto-similaridade, dimensão de compasso, dimensão Box-Counting entre outras) usados em diferentes contextos.[9].

Uma outra forma de calcular a dimensão fractal é ilustrada no experimento de medição do comprimento da costa de um país. A curva que descreve a costa de um país é uma curva fractal e, para efeitos práticos, faz-se necessário medi-la. Entretanto, é possível encontrar na literatura especializada diferentes resultados de medição para a costa de um mesmo país e isso não significa, necessariamente, que algumas dessas medidas estejam incorretas, apenas a forma de medição ou a escala de medição para cada um dos resultados é diferente.

Para elucidar essa questão apresenta-se o seguinte exemplo (PEITGEN; JURGENS; SAUPE (1992)): " ... a medida da costa no mapa geográfico da Grã-Bretanha. Toma-se um conjunto de passos de uma largura determinada. Por exemplo, se a escala do mapa é $1 : 1.000.000$ e a largura do passo é $5cm$, a distância verdadeira correspondente a este passo é $50km$. Caminha-se ao longo da costa contando o número de passos necessários para percorrê-la (Figura 2.11). Multiplica-se este número pelo tamanho do passo e obtém-se a medida da costa. Repetindo o

[9]Da mesma forma como acontece com as definições de dimensão, a taxonomia de tipos de dimensões fractais pode apresentar diferenças na literatura.

Tab. 2.1: Valores usados na obtenção da dimensão fractal da nuvem de pontos da Figura 2.9 (BARNSLEY (1988)).

$log(\frac{1}{\epsilon})$	$log(\aleph(\Lambda, \epsilon))$
-1.1	0,69
-0.69	1,09
-0.405	1,39
-0.182	1,79
0	1,95
0,29	2,30
0,693	2,77
0,916	3,13
1,204	3,43
4,2	5,59

experimento para diferentes larguras de passo, diferentes medidas são obtidas."

Analisando o experimento percebe-se que quanto menor o passo mais precisa é a medição. Um gráfico contrastando a medida obtida e o tamanho do passo utilizado mostra a relação entre a diminuição do passo e o aumento da precisão. Para tal, aconselha-se o uso de um diagrama log/log devido a variação das medidas nos dois eixos, onde o eixo horizontal representa a determinação do passo (considerando o passo como uma medida de precisão) e o eixo vertical representa o comprimento obtido.

Os pontos plotados na Figura 2.12 simulam o resultado da medição de curva da costa britânica, mas é necessário extrair deste gráfico um número que represente tal medida para que ela faça sentido. Observe que os pontos não se localizam exatamente ao longo de uma reta, e nem se espera que isso aconteça visto a natureza do experimento, mas eles expressam um tipo de "previsão" das mudanças que ocorrem nos resultados obtidos de acordo com o aumento da precisão do processo de medição. Usando um método estatístico, por exemplo, mínimos quadrados, é possível aproximar a reta que melhor define a relação de evolução da medição e assim, calculando a inclinação da reta, obtém-se um número que pode ser interpretado como a relação entre as medidas resultantes e o aumento de precisão obtido com cada medição, ou como o grau de complexidade da curva estudada, ou seja, a medida de sua dimensão fractal.

Especificando um pouco mais esse processo chega-se à formulação 2.1, que representa como

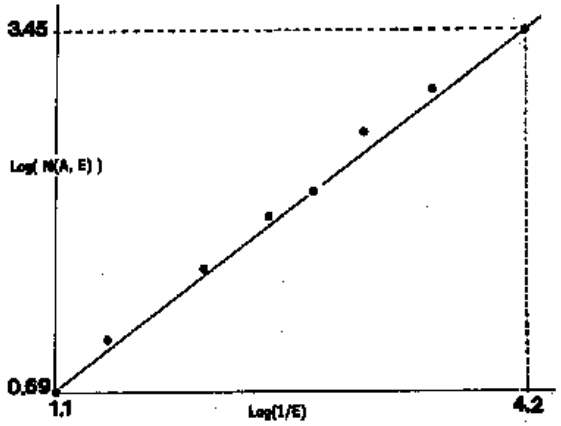


Fig. 2.10: Dados obtidos da cobertura da nuvem de pontos por discos de raio $\epsilon > 0$ (BARNSLEY (1988)).

o comprimento muda quando muda o passo de medição, assumindo que em um gráfico log/log as medidas caem em uma linha reta (PEITGEN; JURGENS; SAUPE (1992)):

$$log(u) = \partial.log(\frac{1}{r}) + b \tag{2.1}$$

em que u é o comprimento em número de passo, r é o tamanho do passo, b corresponde ao logaritmo do comprimento da curva obtido sob a escala $r = 1$ (o ponto em que a reta intercepta o eixo y) e ∂ corresponde à inclinação da reta que aproxima a curva relacionada. Assim, ∂ e b caracterizam a lei de crescimento da reta.

Segundo PEITGEN; JURGENS; SAUPE (1992) a idéia fundamental desse processo é assumir que as duas quantidades envolvidas - comprimento ou superfície ou volume por um lado e escala por outro - não variam arbitrariamente, e sim, relacionam-se por meio de uma lei de potência, representada na Equação 2.2 que permite computar uma quantidade a partir da outra.

$$u = b.(\frac{1}{r})^{\partial} \tag{2.2}$$

De forma similar pode-se analisar uma lei de potência que relacione o número de pedaços a obtidos sobre um fator de redução de r na análise de outras estruturas e expresse o grau de auto-similaridade dessas. Na Tabela 2.2 são mostradas as medições feitas sobre os objetos:

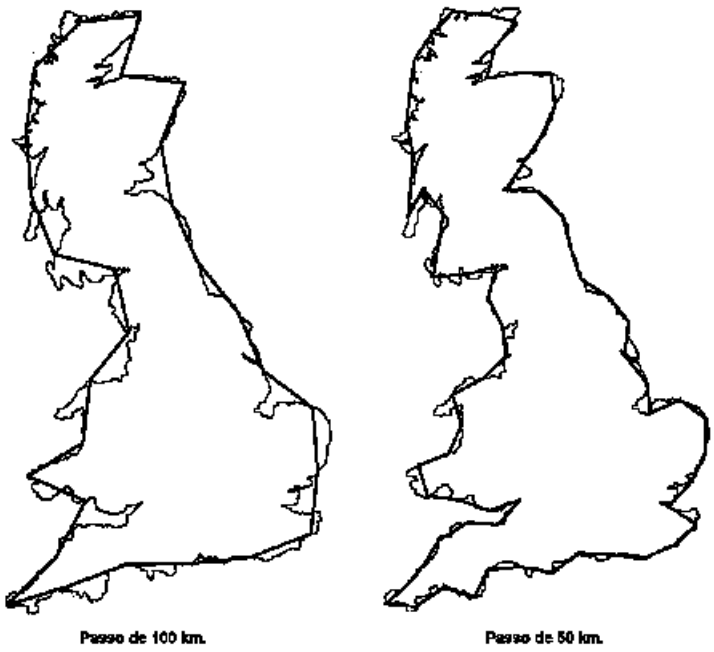


Fig. 2.11: Representação do processo de medição da Costa Britânica: (a) passo de $100km$; (b) passo de $50km$ (PEITGEN; JURGENS; SAUPE (1992)).

linha, quadrado, cubo e curva de Koch (PEITGEN; JURGENS; SAUPE (1992)). A lei de potência obtida desse processo de análise é dada pela Equação 2.3.

$$a = (\frac{1}{r})^D \tag{2.3}$$

em que D = 1 para a linha, D = 2 para o quadrado e D = 3 para o cubo. O expoente coincide com os números que expressam a dimensão topológica da linha, do quadrado e do cubo (PEITGEN; JURGENS; SAUPE (1992)). Para o caso da curva de Koch o cálculo pode ser feito como segue: $4 = 3^D$, aplicando logaritmo $log(4) = D.log(3) \approx 1,2619$. Assim, a Equação 2.3 é equivalente à Equação 2.4, a qual expressa a dimensão de auto-similaridade (um tipo de dimensão fractal).

$$D_s = \frac{log(a)}{log(\frac{1}{r})} \tag{2.4}$$

De forma geral, a dimensão fractal pode ser calculada a partir da Equação 2.5:

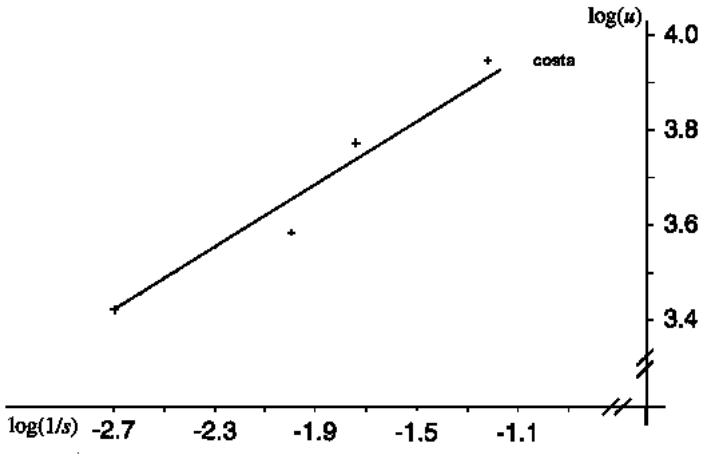


Fig. 2.12: Medida da curva fractal: Costa Britânica. O coeficiente angular da reta de aproximação é a medida da curva fractal (PEITGEN; JURGENS; SAUPE (1992)).

$$D_q = \frac{1}{q-1}.\lim_{r\to 0}\frac{Log(\sum_i (C_i(r)^q))}{Log(r)} \tag{2.5}$$

em que r é o tamanho de um lado de uma célula de um hiper-quadriculado imposto sobre o espaço onde o objeto analisado se localiza, o qual atua como os passos da medição de curvas fractais ou como as divisões em pedaços utilizadas no cálculo da dimensão de auto-similaridade, e $C_i(r)$ indica o número de pontos encontrados em cada célula (assim como u em 2.2 ou a em 2.3).

Diferentes valores de q fornecem os diferentes valores para a dimensão fractal D_q gerando medidas diferentes. Por exemplo, se $q = 0$ então D_0 é a dimensão de *Hausdorff* (TRAINA; TRAINA; FALOUSTOS (1999)); se $q = 2$, a medida gerada equivale à dimensão de Correlação Fractal DCF (SCHUSTER (1984)).

2.4 Dimensão Box-Counting

A dimensão Box-Counting é especialmente utilizada para mensurar fractais estocásticos mas é também útil para fractais de outras classes. Ela propõe um método sistemático de medida aplicável a qualquer estrutura em qualquer dimensão.

Tab. 2.2: Medições feitas sobre os objetos: linha, quadrado, cubo e curva de Koch (PEITGEN; JURGENS; SAUPE (1992)).

Objeto	Número de pedaços	Fator de redução
linha	3	1/3
linha	6	1/6
linha	173	1/173
quadrado	$9 = 3^2$	1/3
quadrado	$36 = 6^2$	1/6
quadrado	$29929 = 173^2$	1/173
cubo	$27 = 3^3$	1/3
cubo	$216 = 6^3$	1/6
cubo	$5177717 = 173^3$	1/173
curva de Koch	4	1/3
curva de Koch	16	1/6
curva de Koch	4^k	$1/3^k$

O conceito de dimensão Box-Counting está relacionado ao conceito de auto-similaridade, sendo que várias vezes, os resultados do cálculo da dimensão de auto-similaridade e da dimensão *Box-Counting* são os mesmos. Entretanto, para casos de estruturas onde a auto-similaridade não está explícita, como no caso de grande parte dos fractais estocásticos, a dimensão Box-Counting apresenta-se como uma alternativa útil para análise dimensional. Esse é o caso da Figura 2.13. Note que não existe uma curva definida que possa ser medida por meio de um conjunto de passos, assim como não existe auto-similaridade aparente[10].

Para computar a dimensão Box-Counting de uma estrutura é necessário situá-la em uma grade regular de células (hipercubos) de mesmo tamanho, contar quantas dessas células são ocupadas pela estrutura (por pontos que compõem a estrutura) e relacionar esta quantidade (N) ao fator que determina o tamanho das células (r). Esse procedimento deve ser executado diversas vezes, sendo que a cada iteração, o fator r deve ser diminuído modificando N (visto que o cálculo de N é dependente de r, a notação $N(r)$ será usada no lugar de N). Este processo iterativo deve ser repetido até que cada célula ocupada contenha apenas um ponto. A partir dos valores de r e $N(r)$ obtidos em cada uma dessas iterações, um diagrama log/log deve

[10]Em alguma escala de observação é possível encontrar algum grau de similaridade entre o conjunto de pontos que se encontra na parte inferior da figura citada com alguma parte do conjunto de pontos situado mais acima

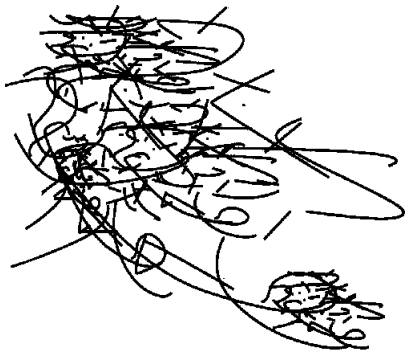


Fig. 2.13: Fractal Estocástico.

ser construído (um diagrama $\frac{log(N(r))}{log(\frac{1}{r}))}$) seguindo as diretrizes de medição de curvas fractais. Sobre estes pontos, uma polinomial de grau 1 deve ser aproximada cuja medida de inclinação representa a medida da dimenção fractal da estrutura explorada. A Figura 2.14 ilustra duas iterações deste processo.

Segundo a literatura pesquisada, para aplicações computacionais aconselha-se o uso de uma seqüência de grades cujo tamanho das células em cada iteração seja reduzido por um fator de $\frac{1}{2}$ em relação à célula da iteração anterior.

Formalmente, a obtenção da dimensão Box-Counting é regida pelo seguinte teorema (BARNSLEY (1988)): seja $\Lambda \in H(X)$, em que Λ é um atrator e $(X, dist)$ é um espaço métrico. Cubra X com uma grade de caixas quadradas fechadas com lado de comprimento $(\frac{1}{2n})$. Seja $Nn(\Lambda)$ o número de caixas com lado de comprimento $(\frac{1}{2n})$ que intersectam o atrator. Se

$$D = \lim_{n \to \infty} \frac{Log(Nn(\Lambda))}{Log(2)^n} \tag{2.6}$$

então A tem dimensão fractal D.

O denominador da função considera o comprimento do lado como uma medida de precisão - quanto menor, mais preciso - assim, o fator $Ln(2n)$ é a simplificação de $Ln(\frac{1}{(\frac{1}{2n})})$.

Segundo PEITGEN; JURGENS; SAUPE (1992), a dimensão Box-Counting é uma das medidas mais usadas em todas as ciências e um dos motivos para tal sucesso é a facilidade de implementação computacional, mesmo quando a estrutura a ser medida está situada em um espaço de alta dimensão, sendo que nesses casos, a célula da grade tem a forma de um hipercubo.

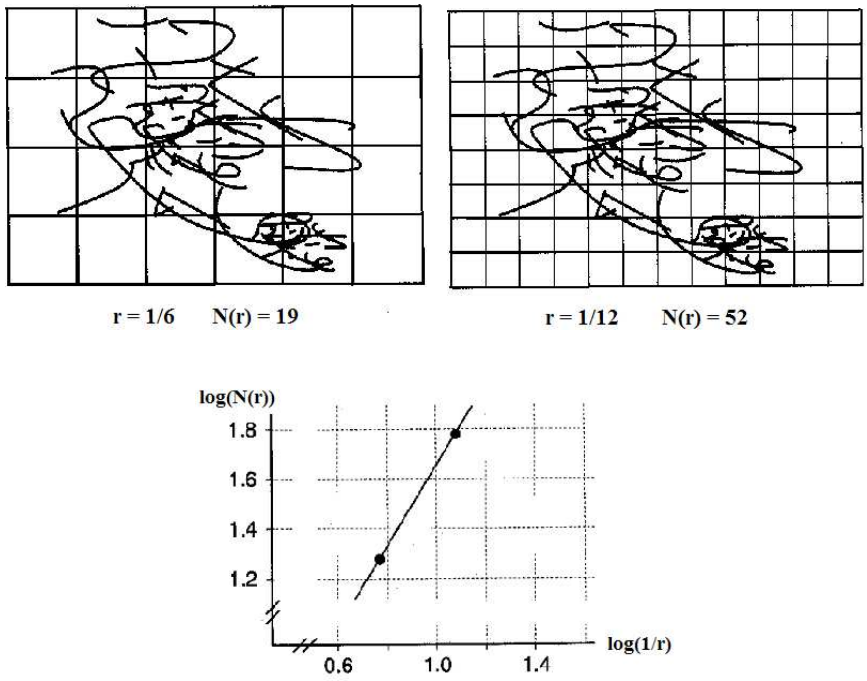


Fig. 2.14: Iterações do processo de cálculo da Dimensão Box-Counting (PEITGEN; JURGENS; SAUPE (1992)).

2.4.1 Alguns exemplos da aplicação do algoritmo Box-Counting

PEITGEN; JURGENS; SAUPE (1992) descreve, de forma suscinta, os resultados obtidos a partir da aplicação do algoritmo de cálculo da dimensão Box-Counting sobre o mapa da costa britânica. Na Figura 2.15 são mostradas duas iterações do processo: a primeira com a cobertura de hipercubos com lados de tamanho calculado a partir de um fator de redução de 1/24 e, a segunda sobre um fator de redução de 1/32 (assumindo uma normalização inicial que fez com que a primeira cobertura se desse com um hipercubo de lado de tamanho unitário). A contagem de hipercubos ocupados resultou em 194 e 283, respectivamente, para a primeira e segunda coberturas mostradas na Figura 2.15. A partir destes poucos dados já é possível derivar a medida de dimensão Box-Counting. Inserindo essas medidas num diagrama *log/log* e calculando a inclinação da reta que melhor aproxima esses pontos tem-se $\approx 1,31$.

Em BARNSLEY (1988) é descrito um experimento que revela uma das utilidades da determinação da dimensão fractal de fenômenos físicos. O objetivo é comparar dois conjuntos de dados obtidos por diferentes meios, de um mesmo sistema físico (uma combustão). Os dados

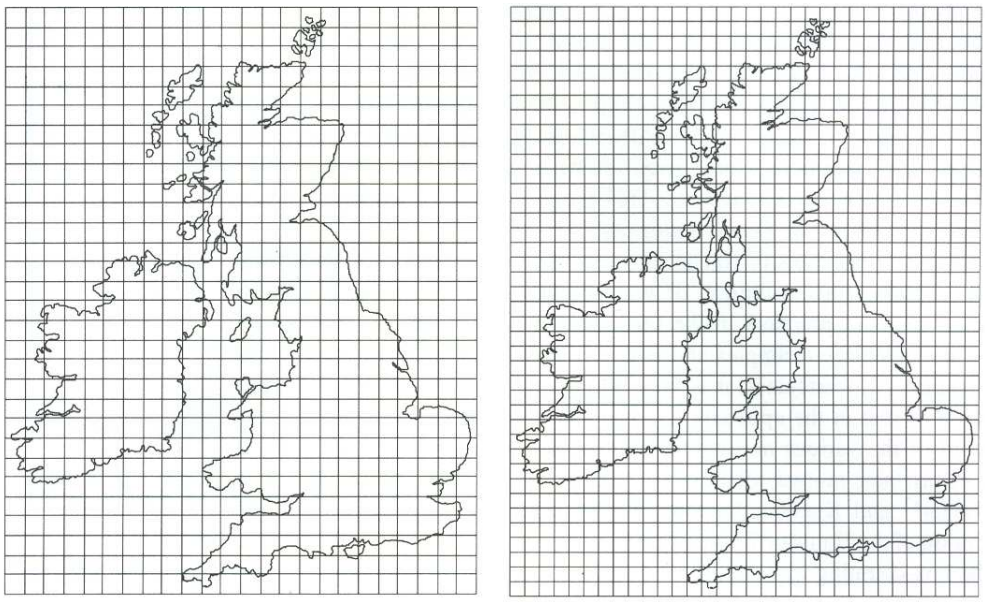


Fig. 2.15: Iterações do processo de cálculo da dimensão Box-Counting para o mapa da Costa Britânica (PEITGEN; JURGENS; SAUPE (1992)).

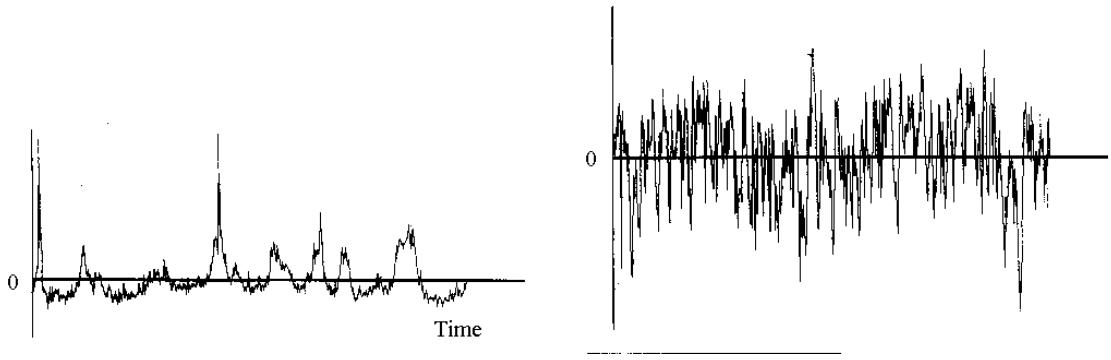
são séries temporais para a temperatura e velocidade de dois pontos diferentes em um jato de fogo.

O experimento é conduzido da seguinte forma: a chama é analisada a) via laser e b) via um metal condutor, ambos apropriados para realizar a tarefa com alta precisão, acuidade e sensibilidade[11]. As saídas obtidas nos dois experimentos são mostradas na Figura 2.16 como um gráfico de *voltagem/tempo*.

Visualmente as figuras são bastante diferentes. A dimensão Box-Counting dos gráficos das duas séries temporais são calculadas a partir da aplicação do método Box-Counting com a utilização dos mesmos parâmetros. Ambos os cálculos resultam no mesmo valor: $\approx 1{,}5$. Isto sugere que, apesar das aparências diferentes, existe uma fonte comum de geração de dados e devido a natureza caótica do experimento, a dimensão fractal pode ser vista como uma forma de quantificação do caos.

O comportamento do algoritmo Box-Counting também pode ser ilustrado por meio da análise de conjuntos de dados simples. Seguem alguns exemplos da aplicação deste processo em

[11] Veja BARNSLEY (1988) para detalhes técnicos sobre o funcionamento de cada um desses mecanismos de medição.



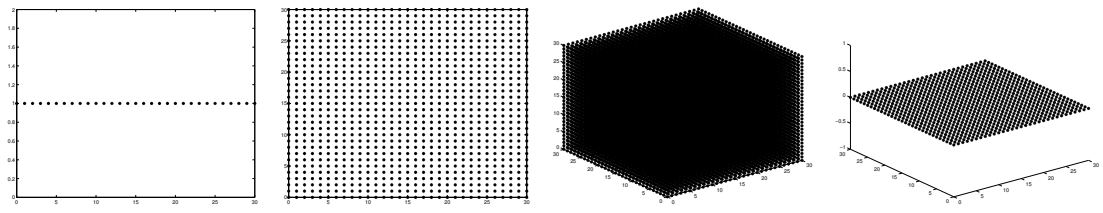
(a) (b)

Fig. 2.16: Gráficos de voltagem/tempo resultantes da análise de uma combustão: (a) via laser e (b) via um metal condutor (BARNSLEY (1988)).

tais conjuntos, os quais foram escolhidos de maneira a formar um grupo de exemplos didáticos e abrangem distribuições uniformes, distribuições normais e distribuições com agrupamentos.

Distribuições uniformes

Na Figura 2.17 são mostrados quatro conjuntos de dados: a) 1-dimensional com 30 pontos; b) 2-dimensional com $30x30$ pontos; c) 3-dimensional com $30x30x30$ pontos e d) 3-dimensional com $30x30x30$ pontos sendo que uma das dimensões é fixa no ponto 0.



(a) (b) (c) (d)

Fig. 2.17: Conjuntos de dados com distribuição uniforme.

Os gráficos *log/log* gerados pelo algoritmo Box-Counting para cada um desses conjuntos estão mostrados na Figura 2.18, respectivamente. Observe que em todos eles os pontos incidem em uma reta e que os valores das dimensões calculados pelo algoritmo se aproximam de um

valor inteiro: a) $0,9935$ b) $1,9869$ e c) $2,9804$. Isso ocorre porque a distribuição uniforme ocupa o espaço de uma forma muito similar a um objeto de dimensão D inteira. Por exemplo, o conjunto 1-dimensional uniformemente distribuído se aproxima a uma reta, o conjunto 2-dimensional se aproxima a um plano e o conjunto 3-dimensional a um cubo. O último conjunto tem $D = 1,9869$ (como o conjunto localizado no espaço 2-dimensional), pois se aproxima de um plano, apesar de localizar-se no espaço 3-dimensional.

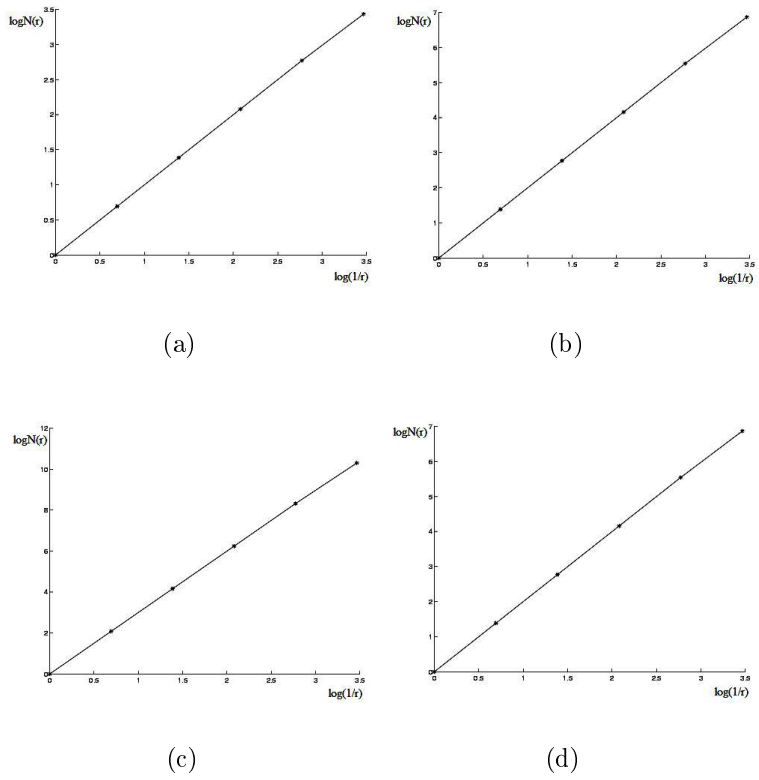
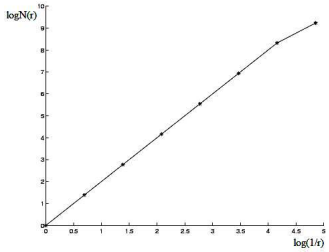


Fig. 2.18: Gráficos *log/log* gerados pelo processo Box-Counting para conjuntos de dados uniformemente distribuídos. Variação no eixo x: [0;3,5]. Variações nos eixos y, respectivamente: [0;3,5], [0;7], [0;12] e [0;7].

Um conjunto de dados uniformemente distribuídos, localizado no espaço 2-dimensional e formado por $100X100$ pontos e tem dimensão fractal $D = 1,9430$ (veja Figura 2.19) , segundo o cálculo do algoritmo Box-Counting, um pouco mais baixa que o caso do conjunto da Figura 2.17, porque a curva resultante do processo de cálculo da dimensão no caso da figura mais "densa" sofre um decaimento no último segmento, causado pela sensibilidade do algoritmo aos pontos que estão na borda do conjunto, efeito esse evidenciado pelo quantidade de pontos

deste conjunto.

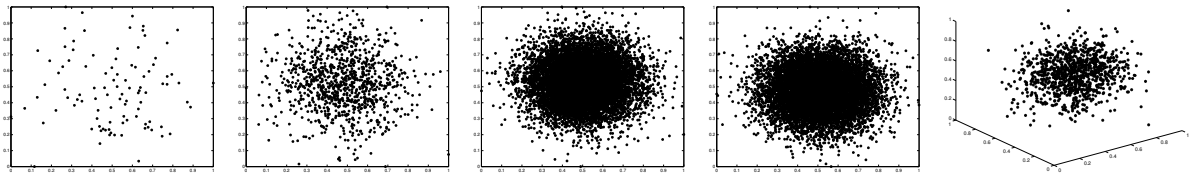


(a)

Fig. 2.19: (a) conjunto de dados com distribuição uniforme com 100 pontos e (b) gráfico *log/log* obtido da aplicação do processo Box-Counting.

Distribuições normais

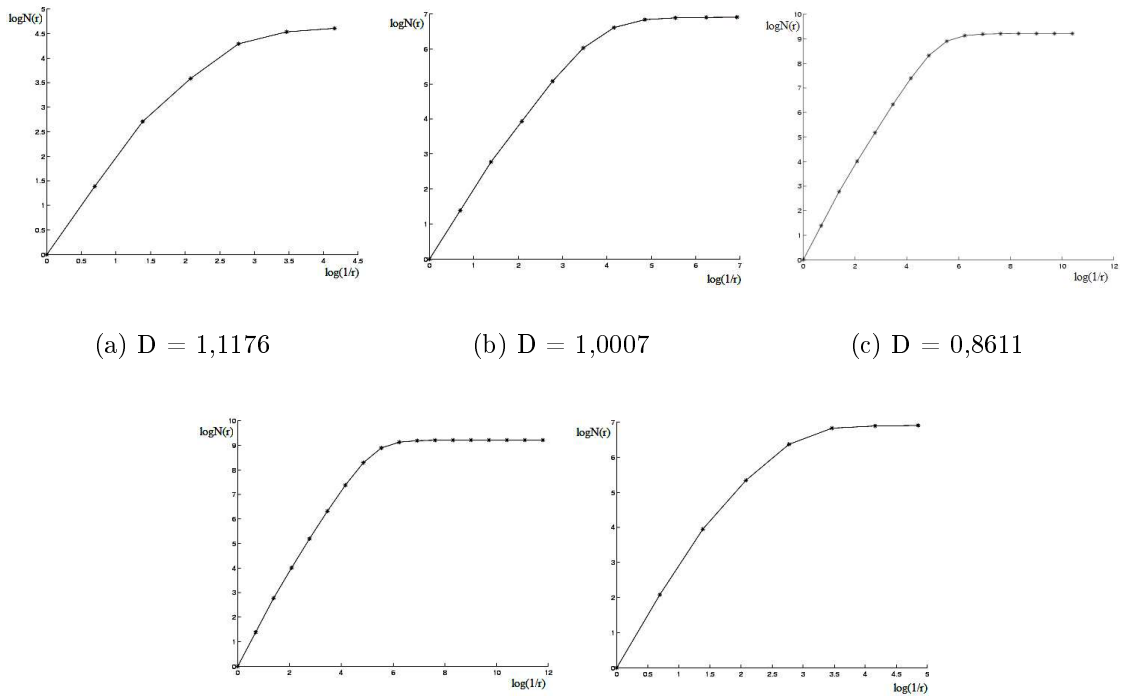
Na Figura 2.20 estão ilustrados conjuntos formados por distribuições normais com diferentes valores para a média e a variância e com diferentes número de pontos. As curvas formadas pela execução do algoritmo Box-Counting, com as respectivas retas de aproximação e valores de dimensões calculados, são mostrados na Figura 2.21.



(a) (b) (c) (d) (e)

Fig. 2.20: Conjuntos de dados com distribuição normal (Gaussiana).

Observe que para distribuições uniformes com um volume maior de número de pontos, o algoritmo executa um maior número de iterações e, nas últimas iterações, a taxa de ocupação de hipercubos (representada pelo eixo y) tem pouca variação. Isso faz com que a reta de aproximação dos pontos tenha uma inclinação mais baixa e a dimensão calculada também fique mas baixa. Tal fato é uma das motivações para a medida de dimensão proposta nesta tese e é explorado no Capítulo 5.



(a) D = 1,1176 (b) D = 1,0007 (c) D = 0,8611

(d) D = 0,7196 (e) D = 1,4098

Fig. 2.21: Gráficos *log/log* gerado pelo processo Box-Counting para conjuntos de dados normalmente distribuídos. Variação nos eixos: [0-4.5;0-5], [0-7;0-7], [0-12;0-10], [0-12;0-10] e [0-5;0-7].

Distribuições com agrupamentos

Para ilustrar o processo de obtenção da dimensão fractal via algortimo Box-Counting sobre conjuntos de dados que possuem agrupamentos escolheu-se alguns conjuntos 2-dimensionais simples. Na Figura 2.22 são mostrados os conjuntos escolhidos para ilustração e, na Figura 2.23, os resultados dos cálculos das dimensões para cada um deles e os gráficos *log/log* resultantes.

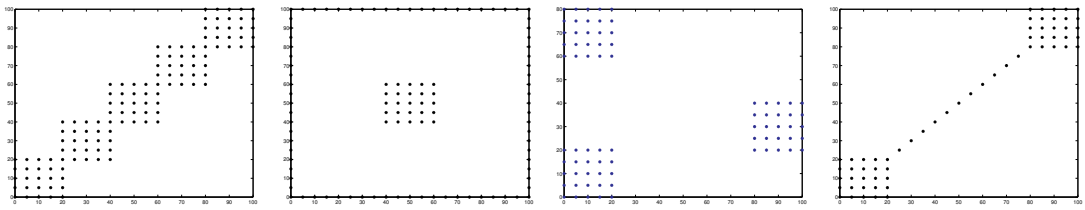
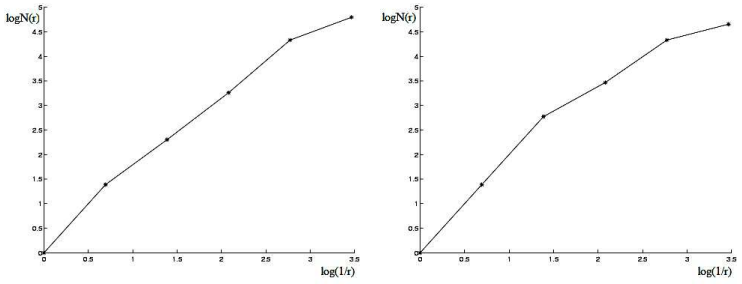
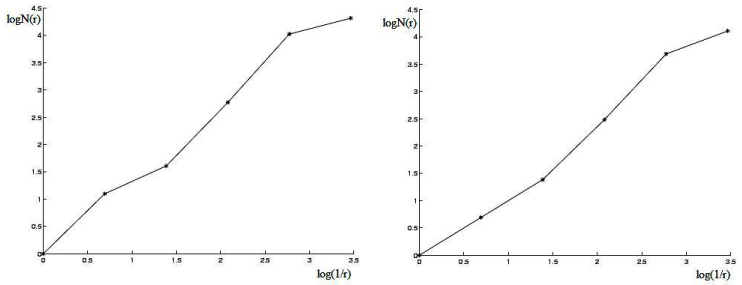


Fig. 2.22: Conjuntos com agrupamentos.



(a) $D = 1,3919$ (b) $D = 1,3519$



(c) $D = 1,2997$ (d) $D = 1,2630$

Fig. 2.23: Gráficos *log/log* para as conjuntos com agrupamentos da Figura 2.22. Variação dos eixos: [0-5;0-3,5], [0-5;0-3,5], [0-4,5;0-3,5] e [0-4,5;0-3,5].

2.4.2 Implementações para o algoritmo Box-Counting

O procedimento utilizado na obtenção da medida da dimensão Box-Counting vem sendo explorado pela pesquisa computacional com o intuito de obter implementações que possuam graus de complexidade algoritmica que permitam sua aplicação para conjuntos de dados grande e de alta dimensão. Dentre as diversas formas possíveis de se conceber uma implementação eficiente, três são destacadas neste texto. Outras implementações são discutidas em BARTH; BAUMANN; NONNENMACHER (1992), BLOCK; BLOH; SCHELLNHUBER (1990), HOU et al. (1990), LIEBOVITCH; TOTH (1989), CHACHERE (1992), GRASSBERGER (1993) e MOLTENO (1993) *apud* FEENY (2000).

Implementação com estrutura de grades multi-níveis e alocação dinâmica

Em TRAINA et al. (2000), os autores descrevem uma implementação que computa a dimensão Box-Counting em tempo linear ($O(N)$, sendo N o número de pontos no conjunto analisado) em relação ao número de pontos do conjunto de dados. O objetivo principal deste trabalho é utilizar a dimensão fractal de um conjunto de dados, em sua variação Box-Counting, para excluir atributos que não contribuem, significativamente, para caracterizar o comportamento do conjunto de dados.

A implementação utiliza uma estrutura de grades multi-níveis que representa os hipercubos utilizados para cobrir o espaço do conjunto de dado analisado (veja Seção 2.4) e em que cada nível possui um raio correspondente à metade do raio do nível anterior. A profundidade desta estrutura é igual ao número de pontos no gráfico *log/log*. A Figura 2.24 mostra como a estrutura de grades é representada e criada na memória principal do computador, de tal forma que o número de pontos no gráfico não ultrapasse o limite de memória disponível.

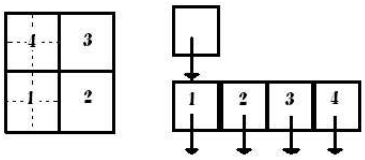


Fig. 2.24: Representação gráfica da grade de hipercubos (à esquerda) e da estrutura de dados (à direita) que a representa para o caso 2-dimensional (adaptado de TRAINA et al. (2000)).

A complexidade algorítmica desta implementação é $O(N * D * i)$, em que N é o número de pontos do conjunto de dados, D é a dimensão do espaço no qual o conjunto está localizado e i é o número de pontos no gráfico *log/log*. A complexidade relativamente baixa desta implementação se deve principalmente aos seguintes fatores: para cada resolução da estrutura de grades é possível manter, para fins de cálculo das somas das ocupações, somente as células que possuem pelo menos um ponto; e cada ponto do conjunto de dados é diretamente associado a células em cada nível da estrutura.

Tal algoritmo pode ser implementado utilizando-se estruturas estáticas (*arrays*) ou dinâmicas (alocação de memória sob demanda), sendo a última a mais adequada quando o conjunto de dados está localizado em espaços de altas dimensões. Na Figura 2.25 mostra-se uma repre-

sentação gráfica da estrutura de dados formada para um exemplo em que a estrutura de grades assume três níveis e o conjunto de dados analisado possui 5 pontos e está localizado em um espaço 2-dimensional. Note que novas células são adicionadas sob demanda à estrutura e que, apenas aquelas células que possuem pontos são expandidas no próximo nível da estrutura de grades. Para chegar ao final do processo, no exemplo da Figura 2.25, o algoritmo teria que realizar mais refinamentos, até que cada célula contivesse apenas um ponto do conjunto de dados.

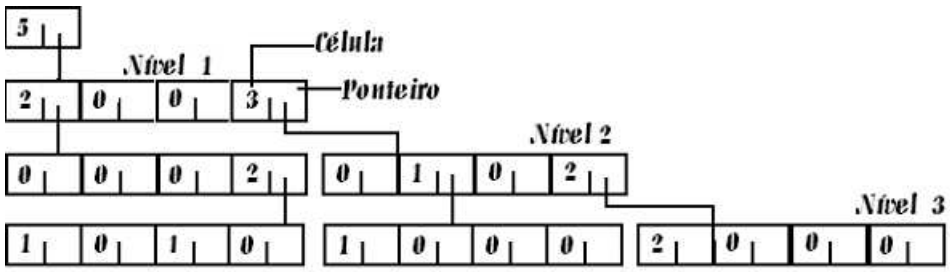


Fig. 2.25: Instância da estrutura de dados em um momento de execução do algoritmo Box-Counting (adaptado de TRAINA et al. (2000)). O número na célula indica o número de pontos existentes nela.

Implementação recursiva

Em FEENY (2000), o autor apresenta um método recursivo para a obtenção da medida da dimensão fractal de um conjunto de dados, com velocidade de computação $O(N)$, onde N é o número de pontos no conjunto.

O autor implementa o processo do algoritmo Box-Counting de uma maneira recursiva, considerando que cada imposição da grade de hipercubos corresponde à uma ordem do fractal. A primeira é a ordem 0 (zero), em que um único hipercubo cobre todo o conjunto. A seguinte, na segunda imposição da grade, cada hipercupo ocupado faz parte da cobertura de primeira ordem do fractal. Cada hipercubo ocupado pode ser visto isoladamente como uma região de ordem zero original que será dividida novamente em quatro hipercubos.

Em sua implementação recursiva, a idéia descrita acima forma a base de desenvolvimento do algoritmo. O término da recursão é gerenciado por um número limite de granularidade de

observação e a saída do algoritmo é uma lista de tamanhos de hipercubos espaçados logaritmicamente e uma lista de funções de partições (definida como $Z(q,r) = (\frac{m_i}{n})^q$, nas quais q é a ordem da subregião i, r é o comprimento da aresta do hipercubo, m é o número de pontos na subregião, n é o número de pontos no conjunto de dados) para cada tamanho de hipercubo, as quais reduzem o número de hipercubos na cobertura para 0. Em FEENY (2000) o autor descreve o algoritmo de forma simplificada para um conjunto 2-dimensional como segue:

- Leia e normalize o conjunto de dados de modo que ele esteja localizado em um quadrado unitário (IxI). Armazene os n pontos do conjunto em um *array* DATA. Inicialize um *array* INDEX de tamanho n. Inicialize uma variável LEVEL com 0. Inicialize a região (primeira imposição de grade de hipercubos) de ordem 0 como $R_0 = IxI$.

- Subdivida essa região em quatro sub-regiões iguais por meio da diminuição de seu comprimento por um fator de 2.

- Usando uma subrotina ZOOM, tome cada sub-região como se ela fosse a primeira, considerando apenas os dados que se encontram na região "mãe" correspondente (utilize o ARRAY de índices para controlar essa informação - para a região de ordem 0 todos os índices de INDEX possuem o mesmo valor). Incremente LEVEL.

- Na atual sub-região, atualize INDEX armazenando-o em uma nova variável NEWINDEX definida dentro da subrotina. O tamanho desta nova variável indica o número de pontos existentes na subregião correspondente. Atualize a função de partição.

- Se no nível máximo de granularidade de observação não foi alcançado e se NEWINDEX não é vazio, repita o procedimento a partir do passo 2, com NEWINDEX fazendo o papel de INDEX.

O algoritmo recursivo desenvolvido e descrito em FEENY (2000) possui um aninhamento de *loopings* em que cada um corresponde a uma dimensão do conjunto de dados. Assim, a complexidade algoritmica é determinada como uma função da dimensão E do espaço do conjunto de dados analisados, da forma $O(2^E)$.

Implementação com correção da quantização do espaço

KRUGER (1996) apresenta uma revisão da abordagem de HOU et al. (1990) para implementação do algoritmo Box-Counting - *fast Box-Counting*. Nessa revisão o autor adaptou a abordagem para diminuir efeitos de quantização do espaço, necessários para implementações computacionais, e que causam uma diminuição na acuidade do resultado obtido. A adaptação possibilita que a curva resultante do algoritmo seja melhor representada, isto é, possua mais pontos em sua formação.

A questão de quantização do espaço está relacionada à finitude de resolução imposta pela implementação computacional do algoritmo Box-Counting para o cálculo da dimensão fractal, visto que a abordagem comum para implementá-lo utiliza-se da plotagem de $log((r))$ por $log(1/r)$ e da inclinação aproximada desta curva para também aproximar o valor da dimensão fractal (veja 2.4). Além disso, existe a questão da finitude do número de pontos disponíveis para representar o fractal analisado (imposição de fatores experimentais). Esta quantização leva a alterações do processo de análise do fractal que afetam o número de pontos que devem compor a curva usada para determinar a medida de dimensão.

Basicamente, o algoritmo original é suportado pelo uso de um esquema de endereçamento de memória flexível que torna a implementação mais adequada à análise de conjuntos de dados multi-dimensionais, e apresenta a complexidade algoritmica de $O(ENlogN)$, em que E é a dimensão do espaço onde os dados estão inseridos e N é o número de dados no conjunto analisado. O algoritmo é implementado sobre o conceito de "intercalação de bits". A alteração proposta modifica a forma de concepção das *strings* de *bits* gerando mais pontos para a construção da curva. Uma breve descrição da implementação bem como dos testes realizados sobre ela pode ser analisada em KRUGER (1996).

2.5 Considerações finais

Um fractal, e mais especificamente para o escopo deste trabalho, um fractal estocástico geralmente é constituído de muitas partes ou pontos que estão organizados de uma forma que dificulta, quando não impossibilita, a sua descrição em relação a que cada parte ou ponto

significa. Desta forma, estes objetos são melhor descritos ou analisados por meio de suas inter-relações ou relações com o espaço onde residem.

Este capítulo pretendeu fornecer os conceitos necessários e suficientes para o entendimento de como a Teoria de Fractais é utilizada no desenvolvimento desta tese. O conceito de Dimensão Fractal é uma das sub-áreas desta teoria que fornece um ferramental para o estudo de como o fractal está relacionado ao espaço onde ele reside e, esta informação será a base para o estudo da distribuição de conjuntos de dados e de como é possível adequar uma RNA-AO para produzir melhores resultados.

Capítulo 3

Raciocínio Aproximado Fuzzy

> "Se cada dia cai, dentro de cada noite há um poço onde a claridade está prisioneira. Há que sentar-se na borda do poço da sombra e pescar a luz caída com paciência". Últimos Sonetos - *Pablo Neruda.*

Instituições sociais, políticas e educacionais utilizam-se de classificações para dividir a sociedade em categorias, de acordo com a sua renda familiar. Esta rotulação varia de sociedade para sociedade, de época para época e admite diferentes rótulos em diferentes contextos. Modelar essa situação não é fácil. Por exemplo, considere a seguinte categorização realizada com base na renda familiar brasileira: uma família que recebe menos de $R\$100,00$ por mês e por membro é considerada "pobre"; por outro lado, é considerada "rica" aquela família que recebe acima de $R\$2.500,00$ por membro por mês. Sobre classificações deste tipo desenvolvem-se programas de assistência social, tabelas de tarifas de impostos, entre outras modelagens que podem afetar direta ou indiretamente a vida dos cidadãos.

A questão que se levanta desta situação não é, ao contrário do que possa parecer, se os limites de $R\$100,00$ e $R\$2.500,00$ estão de acordo com a realidade, mas sim, se os cidadãos que se encontram às margens destes limites estão sendo adequadamente tratados pelos modelos que os representa.

Neste capítulo é tratada uma teoria que pretende adequar os modelos de descrição e trata-

mento das situações do mundo real às reais necessidades de representação. Essa teoria, a teoria de conjuntos fuzzy, é descrita suscintamente neste trabalho, com ênfase à discussão de sua capacidade de representação e de manipulação da incerteza em seus diferentes aspectos. A sub-área lógica fuzzy, e sua aplicação em sistemas de raciocínio aproximado, é descrita com maior atenção por ser utilizada, nesta tese, como uma das ferramentas na solução do problema aqui tratado. Para um estudo mais completo sobre essa teoria recomenda-se o acesso à literatura especializada, como KLIR; YUAN (1995) e PEDRYCS; GOMIDE (1998), nas quais este capítulo está baseado.

3.1 Introdução

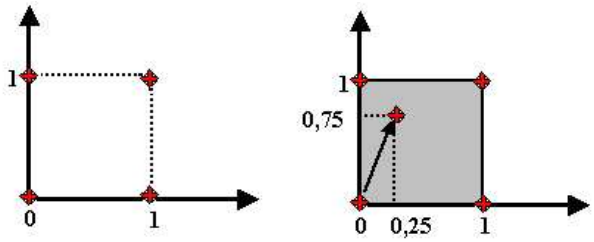
A teoria de conjuntos fuzzy foi introduzida por Loft Zadeh, na década de 60, época em que a visão tradicional sobre como se deveria trabalhar com informação científica (evitando incertezas) estava começando a ser revista. A essência dessa teoria está na aceitação da "incerteza" como um fato das situações reais, que deve ser apropriadamente modelado e tratado por métodos matemáticos quando da intenção desses em interpretar e lidar com situações do mundo real.

A idéia de "incerteza" passa pelas dificuldades em lidar com informações imprecisas, falta de especificidade, conceitos vagos entre outras características que podem fazer com que uma situação não pareça clara ou nítida o suficiente para ser analisada pela teoria de conjuntos *crisp*.

O objetivo da teoria de conjuntos fuzzy é tratar os fenômenos naturais, ou as situações reais, como elas são para que se tenha uma modelagem muito próxima do que é real. Segundo KLIR; YUAN (1995), a possibilidade de estudo da incerteza tende a reduzir a complexidade dos modelos e a aumentar a sua credibilidade.

A teoria de conjuntos fuzzy é uma generalização da teoria de conjuntos que nos liberta da dicotomia imposta por esta última e aumenta o poder de expressividade dos conjuntos (veja a Figura 3.1). Na primeira teoria, um elemento pode pertencer ou não a um conjunto em determinados graus e, mais do que isso, pode pertencer, com graus apropriados, a diferentes

conjuntos definidos sobre um mesmo universo de discurso.



(a) (b)

Fig. 3.1: Interpretação geométrica do poder de expressividade da teoria de conjuntos (a) *crisp* e (b) fuzzy (PEDRYCS; GOMIDE (1998)).

Isso permite resolver a questão apresentada no início deste capítulo. Uma modelagem utilizando a teoria de conjuntos fuzzy poderia representar o conjunto de pessoas que pertecem à classe "pobre" como aquelas que possuem uma renda per capita em até $R\$100,00$ e, utilizando-se de uma função decrescente, modelar a pertinência parcial daquelas que possuem uma renda per capita de $R\$100,00$ ou mais. Dessa forma, a modelagem poderia expressar que uma pessoa vivendo em uma família cuja renda per capita é $R\$110,00$ é "pobre", porém não tanto quanto aquela que pertence a uma família cuja renda por pessoa é $R\$98,00$.

Como consequência, é possível definir também uma generalização da lógica Aristotélica, permitindo criar modelos computacionais que lidem com expressões de linguagem natural, capazes de resolver problemas com um grau de precisão avançado e de lidar com conceitos de acordo com o que eles significam no contexto em que estão inseridos. Essa computação é embasada pela teoria de raciocínio aproximado, sub-área da teoria de conjuntos fuzzy que é o foco de uso desta área nesta tese.

3.2 Conceitos Básicos

O que caracteriza a pertinência de um elemento a um conjunto é a função característica que define esse conjunto. No caso de conjuntos *crisp* tal função associa o valor 0 ou 1 a cada elemento de um universo de discurso, de forma a discriminá-lo como pertencente ou não ao

conjunto em questão.

Essa função pode ser generalizada de forma a associar a cada elemento do universo de discurso um valor de um intervalo, normalmente entre 0 e 1, indicando o grau de pertinência desses elementos ao conjunto. A função característica generalizada assume um papel de **função de pertinência** que define um **conjunto fuzzy**. Assim define-se a função de pertinência de um conjunto fuzzy A como μ_A (3.1).

$$\mu_A : X \rightarrow [0, 1]. \tag{3.1}$$

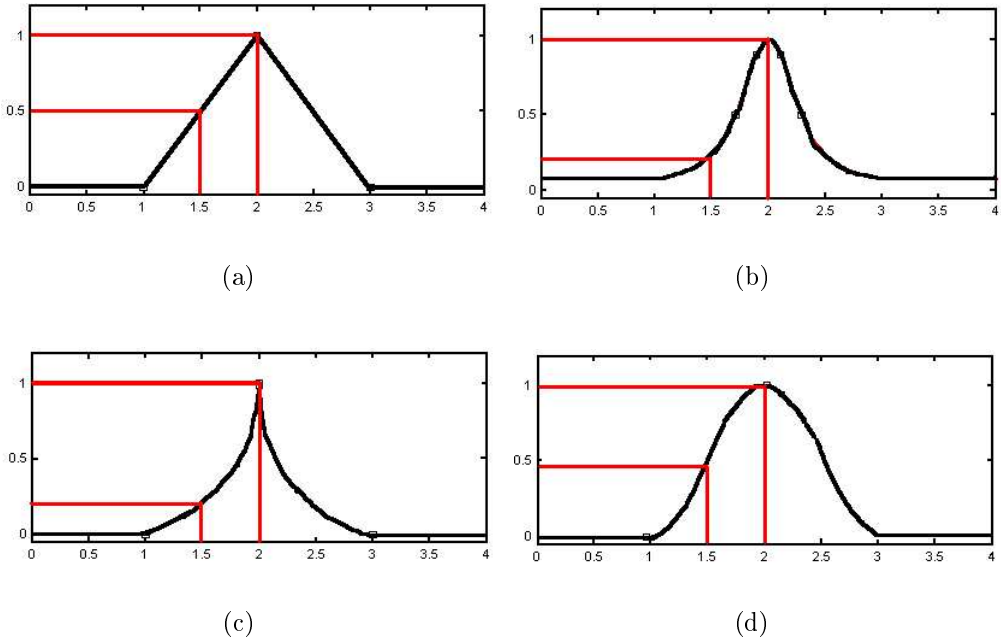
em que X é o universo de discurso (um conjunto de elementos *crisp*).

O conjunto fuzzy pode assumir diversas formas. Como exemplo, observe as representações gráficas da Figura 3.2 (KLIR; YUAN (1995)). Cada um dos conjuntos fuzzy expressa a idéia geral de uma classe de números reais próximos ao número 2. Segundo o conjunto expresso na Figura 3.2(a), o número real $1,5$ pertence ao conjunto de números próximos ao número 2 com um grau de pertinência de $0,5$. Segundo o conjunto ilustrado na Figura 3.2(b) o número real $1,5$ pertence ao conjunto de números reais próximos ao número 2 com um grau de pertinência de $0,3$.

Tais conjuntos possuem diferenças, mas como representam o mesmo conceito, acabam por apresentar algumas similaridades: em todos eles o número real 2 tem grau de pertinência 1 (pertinência total); para todos os outros números reais o grau de pertinência é < 1; a função de pertinência é simétrica com respeito a $x = 2$; os graus de pertinência decrescem monotonicamente de 1 para 0 de acordo com o aumento da diferença $|2 - x|$ e os números reais que estão fora do intervalo $[1, 3]$ não pertencem aos conjuntos ou pertencem com graus de pertinências desprezíveis (KLIR; YUAN (1995)).

Assim como a definição de conjuntos *crisps* é generalizada para conjuntos fuzzy, também o são as operações entre conjuntos. As três principais operações entre conjuntos são: complemento, interseção e união.

A operação de complemento aplicada sobre um conjunto fuzzy A, resultando em $\overline{A}$, com respeito a um conjunto universo X é definida para todos os elementos $x \in X$ pela equação 3.2:



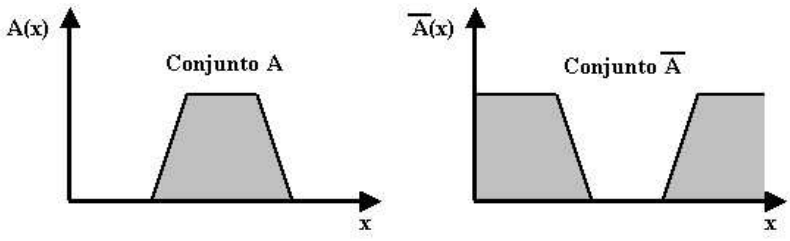
(a) (b)

(c) (d)

Fig. 3.2: Quatro formas diferentes de representar o conceito de números reais próximos a 2 por meio de conjuntos fuzzy. Adaptado de KLIR; YUAN (1995).

$$\overline{A}(x) = 1 - A(x). \tag{3.2}$$

onde $\overline{A}(x)$ pode ser interpretado como o grau para o qual x pertence a $\overline{A}$ ou como o grau para o qual x não pertence a A (KLIR; YUAN (1995)). Graficamente esta operação pode ser representada como na Figura 3.3.



(a)

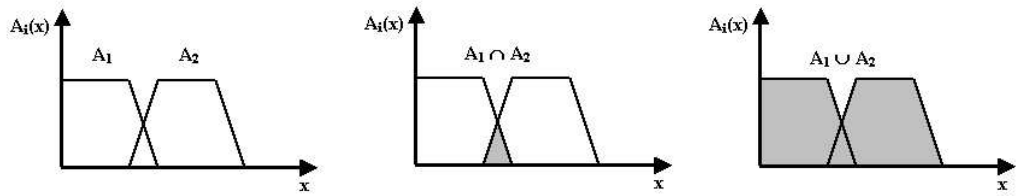
Fig. 3.3: Complemento $\overline{A}$ de um conjunto fuzzy A.

A generalização das operações de interseção e união pode ser feita sob mais de um modelo

(ou forma). Esses modelos são embasados em normas triangulares aplicáveis à generalização da operação de interseção (t-norma) e de união (s-norma).

As normas triângulares t-norma e s-norma são operações binárias ($t : [0,1]^2 \rightarrow [0,1]$) e ($s : [0,1]^2 \rightarrow [0,1]$), respectivamente, que satisfazem as propriedades de comutatividade, associatividade, monotonicidade e condições de limite ($0tx = 0, 1tx = x$ e $xs0 = x, xs1 = 1$).

O operador *min* ($\wedge$) é uma t-norma e o operador *max* ($\vee$), uma s-norma. Outros exemplos de normas triângulares podem ser encontrados em PEDRYCS; GOMIDE (1998) e KLIR; YUAN (1995). A execução das operações de interseção e união por meio das normas *min* e *max* está ilustrada na Figura 3.4.



(a) (b) (c)

Fig. 3.4: Interseção e união de conjuntos fuzzy baseadas nos operadores *min* e *max*.

Outro conceito importante dentro da teoria de conjuntos fuzzy é o conceito de variável lingüística. Segundo PEDRYCS; GOMIDE (1998), pode-se definir uma variável lingüística, informalmente, como uma variável cujos valores são palavras ou sentenças, ao invés de números. Formalmente, PEDRYCS; GOMIDE (1998) definem uma variável lingüística caracterizada por uma quíntupla definida como em 3.3:

$$\langle X, T(X), \mathbf{X}, G, M \rangle \tag{3.3}$$

sendo X o nome da variável; $T(X)$ o conjunto de termos de X cujos elementos são rótulos de valores lingüísticos de X; G uma gramática para gerar os nomes de X; M uma regra semântica para associar cada rótulo $L \in T(X)$ ao seu significado $M(L)$ que é um conjunto fuzzy no universo $\mathbf{X}$, cuja variável base é x.

Para exemplificar os mesmos autores consideram uma variável lingüística chamada *tem-*

peratura. X = *temperatura* com temperaturas variando no intervalo $\mathbf{T} = [0, 50]$ e variável base $t \in \mathbf{T}$. O conjunto de termos associados com essa variável pode ser, entre outros, $T(temperatura)$ = *muito baixa, baixa, média, alta, não baixa e não muito alta, muito alta* em que cada termo em $T(temperatura)$ é um rótulo de um valor lingüístico de *temperatura*. O significado $M(T)$ de um rótulo $T \in T(temperatura)$ é definido como uma restrição $T(t)$ na variável base t imposta pelo nome de T. Assim $M(T)$ é um conjunto fuzzy de T, cuja função de pertinência $T(t)$ cobre a semântica do nome de T.

Os termos *muito baixa, não baixa e não muito alta* e *muito alta* podem ter suas funções de pertinências derivadas da aplicação de operações de interseção ou complemento e da aplicação de modificadores sobre as funções de pertinências dos termos *baixa* e *alta*, caracterizando a "computação com variáveis e termos lingüísticos". Para maiores detalhes sobre esse processo veja PEDRYCS; GOMIDE (1998).

Outro conceito importante para as definições pertinentes a este trabalho se refere às relações fuzzy. Essas relações são uma generalização do conceito clássico de relações, definido em 3.4, pois admitem a noção de associação parcial entre pontos num universo de discurso (PEDRYCS; GOMIDE (1998)).

$$R : \mathbf{X} \times \mathbf{Y} \longrightarrow \{0, 1\} \tag{3.4}$$

em que $\mathbf{X}$ e $\mathbf{Y}$ são dois universos de discurso e R é definida em $\mathbf{X} \times \mathbf{Y}$ como qualquer subconjunto do produto Cartesiano desses dois universos. Exemplos de uma relação clássica e de uma relação fuzzy são, respectivamente, R e R'.

$$R = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

sendo que $\mathbf{X} = \{a, b, c\}$ constitui as dimensões "linha" e $\mathbf{Y} = \{x, y, z, w\}$ constitui as dimensões "coluna".

$$R' = \begin{bmatrix} 0.3 & 0.5 & 1.0 \\ 0.3 & 0.9 & 0.3 \\ 0.1 & 1.0 & 0.1 \end{bmatrix}$$

sendo que $\mathbf{X} = \{a, b, c\}$ constitui as dimensões "linha" e $\mathbf{Y} = \{x, y, z\}$ constitui as dimensões "coluna".

3.3 Lógica Fuzzy

Lógica é o estudo de métodos e princípios de raciocínio em todas as suas possíveis formas (KLIR; YUAN (1995)). A lógica clássica trabalha com proposições que podem assumir um de dois valores, $falso$ ou *verdadeiro*. A lógica é dividida em diversas sub-áreas e, neste trabalho, está-se interessado na sub-área de lógica proposicional.

A lógica proposicional trabalha com combinações de variáveis (variáveis lógicas) que ao assumirem valores verdadeiros podem produzir novas variáveis como função destas combinações (funções lógicas, chamadas de operações lógicas ou primitivas lógicas quando utilizam uma ou duas variáveis). Conjunção, negação, implicação, disjunção são exemplos de primitivas lógicas.

Um conjunto de primitivas é dito completo se qualquer função lógica de variáveis $v_1, v_2, ..., v_n$ (para qualquer n finito) pode ser composta por um número finito dessas primitivas. Como exemplo de um conjunto de primitivas completo tem-se: <*negação, conjunção e disjunção*> e <*negação e implicação*>. A combinação de, por exemplo, *negação, conjunção e disjunção* em uma expressão algébrica apropriada é dita uma fórmula lógica e, a partir dela, qualquer outra função lógica pode ser obtida.

Quando a variável representada por uma fórmula lógica é sempre verdadeira, independentemente de valores verdadeiros serem associados às variáveis participantes da fórmula, tem-se uma tautologia. As tautologias são a base do raciocínio dedutivo e as mais comuns são:

$$(a \wedge (a \Longrightarrow b)) \Longrightarrow b \ (modus\ ponens) \tag{3.5}$$

$$(\bar{b} \wedge (a \Longrightarrow b)) \Longrightarrow \bar{a} \ \textit{(modus tollens)} \tag{3.6}$$

$$((a \Longrightarrow b) \wedge (b \Longrightarrow c)) \Longrightarrow (a \Longrightarrow c) \ \textit{(silogismo hipotético)} \tag{3.7}$$

A dicotomia da lógica clássica, ou lógica bi-valorada (L_2), assim como a dicotomia dos conjuntos *crisp*, pode representar uma limitação de raciocínio. Para suprir essa limitação, desenvolveram-se diferentes lógicas conhecidas como lógicas multi-valoradas: lógica tri-valorada, lógica n-valorada (L_n) e lógica de infinitos valores (L_∞) ou lógica padrão de Lukasiewicz (L_1).

Segundo KLIR; YUAN (1995) é possível fazer uma relação entre teoria de conjuntos fuzzy e L_1. Os graus de pertinência $A(x)$ para $x \in X$, para os quais o conjunto fuzzy A é definido no universo de discurso X, pode ser interpretado como os valores verdade da proposição "x é um membro do conjunto A" em L_1. Assim como os valores verdade para todos $x \in X$ de qualquer proposição "x é P" em L_1, podem ser interpretados como graus de pertinência $P(x)$ pelos quais o conjunto fuzzy caracterizado pela propriedade P é definido em X. Tem-se então a idéia de proposições fuzzy que é o conceito que caracteriza a lógica fuzzy. A lógica fuzzy, por sua vez, pode ser vista como um "veículo para computação com valores verdade lingüísticos" (PEDRYCS; GOMIDE (1998)).

Uma proposição clássica requer ser verdadeira ou falsa, já em proposições fuzzy, a veracidade ou falsidade é expressa em termos de graus. Caso se assuma que a verdade é expressa pelo valor 1 e a falsidade pelo valor 0, os graus de verdade de uma proposição fuzzy são expressos por números no intervalo unitário $[0, 1]$ ou por rótulos lingüísticos (verdadeiro ou muito verdadeiro, por exemplo).

Em PEDRYCS; GOMIDE (1998), uma proposição fuzzy é vista como uma representação de um pedaço de conhecimento e é definida na forma "X é A", em que X é o nome de um objeto e A é o nome de um conjunto fuzzy no universo de discurso $\mathbf{X}$; esta proposição é associada a dois conjuntos fuzzy: um conjunto fuzzy A definido no universo de discurso $\mathbf{X}$, e o conjunto fuzzy definido em $[0, 1]$ que representa que o valor verdade τ de X é A. Para exemplificar considere a proposição "Pressão é pequena é muito verdadeiro": o universo de discurso $\mathbf{X}$ é o

universo da variável lingüística "pressão", "pequena" é um termo lingüístico representado por um conjunto fuzzy correspondente a A e "muito verdadeiro" corresponde ao segundo conjunto fuzzy que representa o valor verdade da proposição.

Proposições do tipo do exemplo anterior são ditas atômicas e podem ser combinadas usando conectivos lógicos (disjunções, conjunções, negações e implicações), aumentando o seu poder de expressão. KLIR; YUAN (1995) classificam as proposições fuzzy simples em quatro tipos:

1. Proposições fuzzy não-condicionais e não-qualificadas:

   Este tipo de proposição $p1$ pode ser expressa conforme 3.8, em que V é uma variável que assume valores ν de um conjunto universo $\mathbf{V}$, e F é um conjunto fuzzy em $\mathbf{V}$ que representa um predicado fuzzy.

   $$p1 : V \text{ is } F. \tag{3.8}$$

   Segundo essa proposição, dado um valor ν de V, o valor pertence a F com grau de pertinência $F(\nu)$, e esse é interpretado como o grau de verdade da proposição. Por exemplo (KLIR; YUAN (1995)), "seja V a temperatura do ar e seja F a função de pertinência que representa, num dado contexto, o predicado (ou termo linguístico) *alto*, a proposição fuzzy equivalente a situação é:

   $$p1' : \text{ temperatura } (V) \text{ is alta } (F). \tag{3.9}$$

   O grau de verdade da proposição $p1$' depende do valor atual da temperatura e da definição do predicado *alto*."

2. Proposições fuzzy não-condicionais e qualificadas:

   Este tipo de proposição $p2$ é caracterizado por 3.10 ou por 3.11 ,

   $$p2a : \{V \text{ is } F\} \text{ is } S, \tag{3.10}$$

$$p2b: \text{ Pro } \{V \text{ is } F\} \text{ is } P, \tag{3.11}$$

em que V e F têm o mesmo significado que em $p1$, Pro $\{V$ is $F\}$ é a probabilidade do evento fuzzy "V" ser F, S é um qualificador verdadeiro fuzzy e P é um qualificador de probabilidade fuzzy. Tanto S quanto P são reprsentados por conjuntos fuzzy definidos em [0,1].

Para exemplificar 3.10, considere a proposição qualificada-verdadeira (KLIR; YUAN (1995)): *Tina é jovem é muito verdade*, em que o predicado *jovem* e o qualificador de verdade *muito verdade* são representados por conjuntos fuzzy. Assumindo que o valor 26 anos tem grau de pertinência $0,87$ ao conjunto fuzzy *jovem* e que o valor $0,87$ tem grau de pertinência $0,76$ ao conjunto fuzzy *muito verdade* , a proposição exemplo pertence ao conjunto de proposições que são muito verdade com grau de pertinência $0,76$, e esse é o seu grau de verdade.

Para exemplificar 3.11, considere a variável V como a média da temperatura diária t em F^o para algum lugar no planeta durante um certo mês. Então, a proposição de probabilidade qualificada "$p2b$' : Pro{temperatura t (em um dado lugar e tempo) é por volta de 75^oF é provável" caracteriza um aspecto climático em um dado lugar e tempo. "por volta de 75^oF" deve ser representado por um conjunto fuzzy definido em $\Re$ e o qualificador "provável" é expresso por um conjunto fuzzy definido em [0,1]. Assumindo uma distribuição de probabilidade $t : f(t)$ é possível encontrar o grau de verdade $T(p2b')$ usando 3.12 e 3.13.[1]

$$\text{Pro}\{V \text{ is } F\} = \sum_{t \in \Re} f(t) \cdot F(v) \tag{3.12}$$

$$T(p2b') = P(\sum_{t \in \Re} f(t) \cdot F(v)) \tag{3.13}$$

3. Proposições fuzzy condicionais e não-qualificadas:

[1] Veja (KLIR; YUAN (1995)) para o desenvolvimento numérico deste exemplo.

Tal tipo de proposição pode ser representada por 3.14, em que X e Y são variáveis cujos valores estão nos conjuntos X e Y, respectivamente; e A e B são conjuntos fuzzy em X e Y, respectivamente.

$$p3 : \text{Se } (X) \text{ é } A, \text{então } (Y) \text{ é } B, \tag{3.14}$$

Essas proposições também podem ser vistas como $\langle X, Y \rangle$ é R, em que R é uma relação fuzzy em $\mathbf{X} \times \mathbf{Y}$ que é determinada para cada $x \in \mathbf{X}$ e cada $y \in \mathbf{Y}$ por $R(x, y) = \vartheta[A(x), B(y)]$ onde ϑ é um operador de implicação fuzzy (discutido na Seção 3.3.1).

Regras do tipo *if-then* são um tipo de combinação de proposições que dão origem às regras de inferência utilizadas em raciocínio aproximado fuzzy.

4. Proposições fuzzy condicionais e qualificadas

   Proposições deste tipo podem ser representadas por 3.15 ou por 3.16:

$$p4a : \text{Se } (X) \text{ é } A, \text{então } (Y) \text{ é } B \text{ é } S, \tag{3.15}$$

$$p4a : \text{Pro } \{(X) \text{ é } A | (Y) \text{ é } B\} \text{ is } P, \tag{3.16}$$

   em que Pro$\{(X)$ é $A|(Y)$ é $B\}$ é uma probabilidade condicional.

3.3.1 Implicação Fuzzy

De forma geral, uma implicação fuzzy é uma função da forma $\vartheta : [0, 1] \times [0, 1] \rightarrow [0, 1]$, que para qualquer valor verdade a, b de proposições fuzzy p, q respectivamente, define o valor verdade $\vartheta(a, b)$ da proposição condicional "se p então q" (KLIR; YUAN (1995)).

Na lógica clássica a operação de implicação $p \Rightarrow q$ pode ser definida de diferentes, porém equivalente, formas. Mas as extensões fuzzy dessas formas não são equivalentes, o que gera diferentes classes de implicações fuzzy.

Por exemplo, na lógica clássica uma implicação ϑ pode ser definida como 3.17 ou como 3.18, para todo a e $b \in \{0, 1\}$.

$$\vartheta(a, b) = \bar{a} \vee b. \tag{3.17}$$

$$\vartheta(a, b) = max\{x \in 0, 1 | a \wedge x \leq b\}. \tag{3.18}$$

Na lógica fuzzy, as extensões destas fórmulas são 3.19 e 3.20, para todo a e $b \in [0, 1]$, em que a disjunção é uma $s - norma$ denotada por u, a negação é o complemento fuzzy denotado por c e a conjunção é uma $t - norma$ denotada por i.

$$\vartheta(a, b) = u(c(a), b) \tag{3.19}$$

$$\vartheta(a, b) = sup\{x \in [0, 1] | i(a, x) \leq b\} \tag{3.20}$$

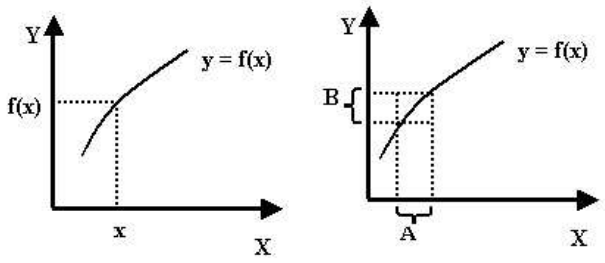
As fórmulas 3.17 e 3.18 são equivalentes, mas 3.19 e 3.20 não o são. Assim, operadores de implicação fuzzy específicos são obtidos com a escolha de s-normas, t-normas e complementos fuzzy específicos. Dado que as diferentes operações s-norma, t-norma e complementos também diferenciam entre si, as implicações fuzzy seguirão a mesma diferenciação (veja KLIR; YUAN (1995), para consultar as implicações fuzzy mais comuns na literatura).

3.3.2 Inferência fuzzy

Na lógica clássica, as regras de inferência são baseadas em tautologias. Da mesma forma acontece na lógica fuzzy. As regras de inferência fuzzy são baseadas nas generalizações das tautologias: *modus ponens, modus tollens* e *silogismo hipotético*. E essas generalizações, por sua vez, são baseadas em regras de inferência composicionais.

A Figura 3.5 ilustra o relacionamento expresso por uma função $y = f(x)$, de variáveis X e Y pertencentes ao universo de discuros $\mathbf{X}$ e $\mathbf{Y}$. No caso (a) a variável base X é igual a x e é possível inferir que $Y = f(x)$. No caso (b), a variável base X está no conjunto A, então

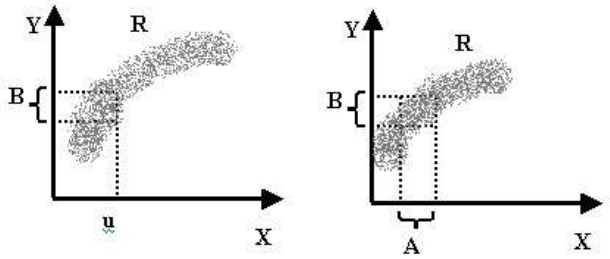
infere-se que Y está no conjunto $B = \{y \in \mathbf{Y} | y = f(x), x \in A\}$.



(a) (b)

Fig. 3.5: Relacionamento funcional entre duas variáveis (KLIR; YUAN (1995)).

Se as variáveis estiverem relacionadas por meio de uma relação $\mathbf{X} \times \mathbf{Y}$, então dado $X = u$ e uma relação R, é possível inferir que $Y \in B$, em que $B = \{y \in \mathbf{Y} | \langle x, y \rangle \in R\}$. Da mesma forma, se $X \in A$ então infere-se que $Y \in B$, em que $B = \{y \in \mathbf{Y} | \langle x, y \rangle \in R, x \in A\}$. Na Figura 3.6 são ilustradas essas situações.

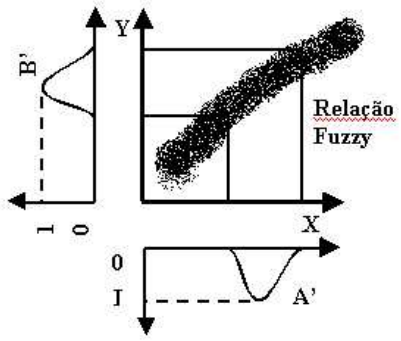


(a) (b)

Fig. 3.6: Relacionamento entre variáveis expresso por uma relação: (a) quando $X = u$; (b) quando $X \in A$ (KLIR; YUAN (1995)).

Assumindo que R é uma relação fuzzy em $\mathbf{X} \times \mathbf{Y}$ e A' e B' são conjuntos fuzzy definidos em $\mathbf{X}$ e $\mathbf{Y}$, então, se R e A' são dados, B pode ser inferido por meio de 3.21, que é chamada de regra de inferência composicional (veja Figura 3.7).

$$B'(y) = sup_{x \in \mathbf{X}} min[A'(x), R(x, y)]. \tag{3.21}$$



(a)

Fig. 3.7: Regra de inferência composicional (KLIR; YUAN (1995)).

O uso de uma regra de inferência composicional para concluir que "Y é B'" a partir de uma relação obtida das proposições "p : Se X é A, então Y é B" e "$q : X$ é A" é chamada de *modus ponens generalizado*.

Um procedimento similar gera o *modus tollens generalizado*, sendo que para esta situação muda a regra de inferência composicional para $A'(y) = sup_{y \in \mathbf{Y}} min[B'(x), R(x, y)]$.

A generalização do *silogismo hipotético* é baseada em duas proposições fuzzy condicionais: p_a : Se X é A, então Y é B e p_b : Se Y é B, então Z é C. Como conclusão tem-se que p_c : Se X é A, então Z é C. Para cada uma destas proposições uma relação fuzzy é determinada e a regra de inferência composicional que possibilita tal tautologia é $R_3(x, z) = sup_{y \in \mathbf{Y}}$ min $[R_1(x, y), R_2(y, z)]$.

3.4 Raciocínio Aproximado Fuzzy

O raciocínio aproximado fuzzy trabalha com raciocínio baseado em regras de produção ou regras de inferência fuzzy. Regras, num contexto geral, podem ser vistas como um esquema de representação que transforma o conhecimento em informação manipulável por sistemas de computação. Regras fuzzy são capazes de representar o conhecimento por meio do uso de variáveis lingüísticas e sistemas baseados em regras fuzzy podem então computar com palavras. Essa computação se dá como em 3.22

$$\begin{array}{l} \textit{Regra 1:}\ \text{If } X \text{ é } A_1,\ \text{então } Y \text{ é } B_1 \\ \textit{Regra 2:}\ \text{If } X \text{ é } A_2,\ \text{então } Y \text{ é } B_2 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \textit{Regra n:}\ \text{If } X \text{ é } A_n,\ \text{então } Y \text{ é } B_n \\ \textit{Fato: } X \text{ é } A' \\ \hline \text{Conclusão: } Y \text{ é } B' \end{array} \tag{3.22}$$

em que $A^{'}$, $A_j \in F(X)$, $B^{'}$, $B_j \in F(Y)$ para todo $j \in N_n$, e $\mathbf{X}$, $\mathbf{Y}$ são conjuntos de valores de variáveis X e Y. Ou seja, dado n regras do tipo "Se-então" e um fato em relação à variável X, é possível concluir um valor para a variável Y. Para KLIR; YUAN (1995), a forma mais comum de determinar o valor da variável Y é pelo método da interpolação. Na Figura 3.8 é mostrada uma interpretação geométrica do método da interpolação. Note que o fato é um conjunto fuzzy. Este método consiste de dois passos (KLIR; YUAN (1995)):

1. Calcular o grau de consistência $r_j(A^{'}) = h(A^{'} \cap A_j)$, entre o fato dado e o antecedente de cada regra "se-então" j em termos da altura da interseção dos conjuntos associados $A^{'}$ e A_j. Assim, para cada $j \in N_n$, e usando a interseção fuzzy min,

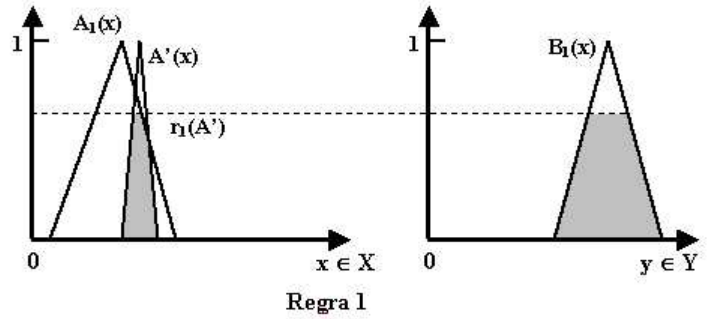
$$r_j(A^{'}) = sup_{x\ in X} min[A^{'}(x), A_j(x)]. \tag{3.23}$$

2. Calcular a conclusão $B^{'}$ truncando cada conjunto B_j no valor $r_j(A^{'})$, a qual expressa o grau para o qual o antecedente A_j é compatível com um dado fato $A^{'}$, e executando a união dos conjuntos truncados. Ou seja,

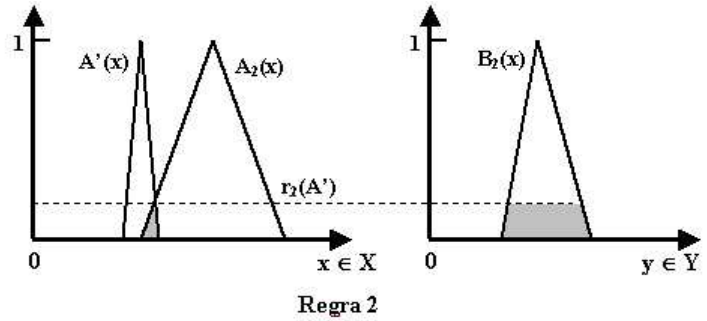
$$B^{'}(y) = sup_{j \in N_n} min[r_j(A^{'}, B_j(y)] \tag{3.24}$$

para todo $y \in Y$.

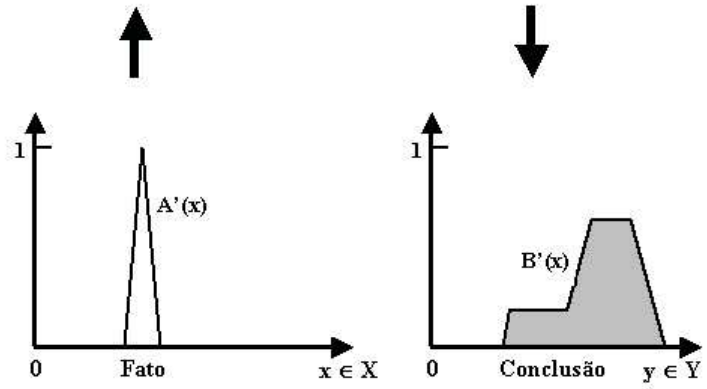
Na Figura 3.9 é mostrada uma interpretação geométrica para o desenvolvimento do raciocínio aproximado para regras com premissas compostas. Neste exemplo, as regras são do tipo



(a)



(b)



(c)

Fig. 3.8: Interpretação geométrica do método de interpolação KLIR; YUAN (1995).

Se X é A_i e Y é B_i então Z é C_i e é necessário compor as premissas antes de processar o grau de ativação da conclusão. Note que o fato é um valor "crisp".

Como pode ser observado na Figura 3.9, geralmente a variável base que ativa os conjuntos fuzzy tem um valor "crisp". O mapeamento de valores "crisp" para um grau de pertinência a um conjunto fuzzy, ou seja, o projeto do conjunto fuzzy, é conhecido como processo de "fuzificação".

Além disso, pode ser observado nas Figuras 3.8 e 3.9 que a conclusão obtida no processo de combinação das regras de inferência é um conjunto fuzzy (formado a partir das combinações adequadas das ativações das conclusões das regras mediante as ativações das premissas das regras). Tal conjunto representa, de maneira fuzzy, as respostas do processo de inferência à ocorrência de um ou mais fatos no sistema modelado. Para que essa resposta tenha um sentido prático no universo de discurso modelado, é necessário, na maioria das vezes, extrair um valor "crisp" do domínio onde as conclusões são modeladas, que melhor expresse esse conjunto e que possa indicar uma resposta final ou uma ação a ser tomada no universo de discurso. Para tal, é necessário aplicar um método de defuzificação.

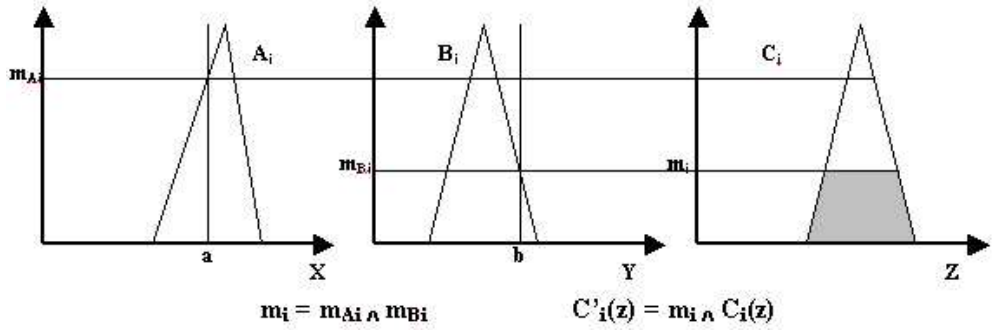
Existem vários métodos de defuzificação e cada um deles é suportado por justificativas que lhe dão alguma racionalidade (veja KLIR; YUAN (1995) e PEDRYCS; GOMIDE (1998)). O método do centróide foi o escolhido para ser aplicado nesta tese e por isso está delhado nesta seção.

O método do centróide

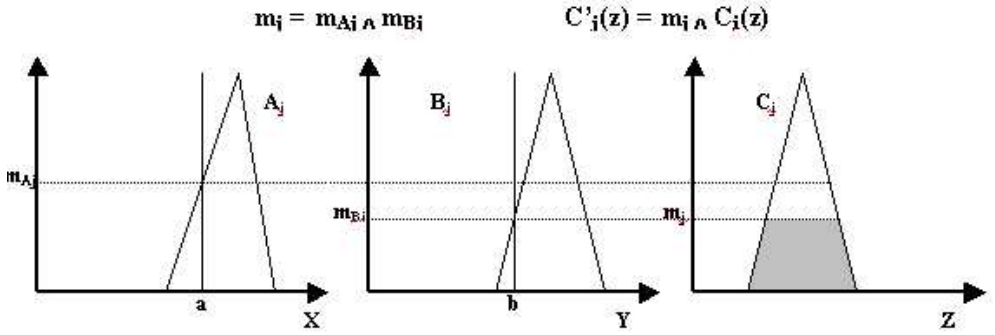
No método do centróide, também conhecido como "centro de área" ou "centro de gravidade", o valor defuzificado $d_c(U)$ é definido como um valor dentro do domínio da variável y para o qual a área sob o gráfico da função de pertinência U é dividida em duas subáreas iguais.

Para o caso discreto, no qual U é definido como um conjunto universo finito $(u_1, u_2, ..., u_n)$, a fórmula de cálculo de U é definida em 3.25. Caso $d_c(U)$ não seja igual a nenhum valor no conjunto universo, toma-se o valor mais próximo a ele pertencente a esse conjunto.

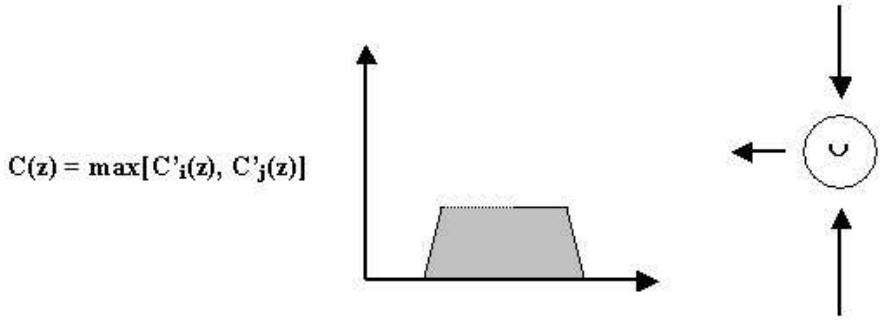
$$d_c(U) = \frac{\sum_{k=1}^{n} U(u_k)u(k)}{\sum_{k=1}^{n} U(u_k)} \tag{3.25}$$



(a)



(b)



(c)

Fig. 3.9: Interpretação geométrica da composição sup-min, adaptado de PEDRYCS; GOMIDE (1998).

3.5 Considerações Finais

" ... uma semente não constitue uma pilha, nem duas, nem três ... por outro lado todos concordam que 100 milhões de sementes constituem uma pilha. Qual é então o limite apropriado? Nós podemos dizer que 325.647 sementes não constituem uma pilha, mas que 325.648 constituem? (BOREL (1950) **apud** PEDRYCS; GOMIDE (1998)). Tratar problemas em que os limites de intervalos não devem ser definidos de forma tão rígida é um dos principais objetivos da teoria de conjuntos fuzzy.

Nesta tese utilizou-se de uma subárea desta teoria, o Raciocínio Aproximado Fuzzy, com a intenção de relaxar a definição de intervalos de modelagem para alcançar maior precisão de resposta. Neste capítulo pretendeu-se fornecer os conceitos básicos para o entendimento da teoria utilizada na abordagem desenvolvida nesta tese. Dentre os conceitos estudados, deu-se ênfase à lógica fuzzy e ao processo de inferência fuzzy, constituintes do Raciocínio Aproximado Fuzzy.

Capítulo 4

Redes Neurais Artificiais Auto Organizáveis

"... qualquer dado se torna importante se está conjugado a outro. A conexão muda a perspectiva. Induz a pensar que todos os aspectos do mundo, todas as vozes, toda palavra escrita ou dita não tem o sentido que parece, mas nos fala um Segredo. O critério é simples: suspeitar, suspeitar sempre. Podemos mesmo até ler o que está por trás de uma placa de sentido proibido."
Belbo a Causabon em O pêndulo de Foucault - *Umberto Eco*.

O cérebro humando é responsável por controlar o funcionamento de todo o corpo humano, desde suas funções vitais até a execução de ações que são "aprendidas" durante a vida. Por meio de suas 10^{11} unidades básicas, os neurônios, e por meio das conexões existentes entre eles, o cérebro codifica e decodifica estímulos gerando informações sobre emoções, pensamentos, percepção e cognição.

Para HAYKIN (1999), o cérebro biológico, ou rede neural, é um processador massivamente distribuído e paralelo, composto por unidades simples de processamento, as quais são capazes de estocar conhecimento experimental e torná-lo disponível para uso. Ainda, segundo o autor, existem diversas propriedades e capacidades que compõem ou são oferecidas por este recurso.

Entre elas está a propriedade de "auto organização", capaz de oferecer capacidade de abstração sobre conceitos do mundo real. Esta tese está especialmente relacionada a essa propriedade e a essa capacidade.

Neste capítulo são apresentadas as Redes Neurais Artificiais Auto Organizáveis (RNAs-AO), principal objeto de estudo desta tese. Primeiramente discute-se, rapidamente, a propriedade de auto organização de um sistema e as propriedades cerebrais que caracterizam um comportamento auto organizável. Em seguida são fornecidos os aspectos teóricos básicos, concernentes à concepção das RNAs-AO, dando ênfase às medidas de avaliação das suas capacidades quantizadora e de preservação topológica. Dentre as medidas de avaliação dá-se destaque a uma delas - o Produto Topológico - por ser uma das medidas utilizadas nesta tese e também pela oportunidade de, discutindo-a, fornecer subsídio para um melhor entendimento sobre questões referentes a conceitos de topologia e vizinhança topológica, essenciais para o entendimento dos propósitos aqui idealizados.

4.1 Auto Organização e Redes Neurais

A "auto organização" consiste, basicamente, de repetidas modificações em um sistema ou ambiente, em resposta a estímulos externos ou internos e em concordância com as leis que regem as possíveis ações executadas sobre suas variáveis ou componentes. Esse processo de modificação ou adaptação deve ocorrer no sistema até que uma configuração final seja alcançada, via convergência ou via outro critério de parada.

Segundo HAYKIN (1999), a questão-chave de um sistema auto organizável consiste em mostrar como uma configuração útil pode surgir a partir da auto organização. A resposta é baseada em uma observação (TURING (1950) *apud* HAYKIN (1999)): "A ordem global pode surgir de interações locais".

Essa afirmação é particularmente válida para o cérebro biológico, onde um conjunto de interações locais e originalmente aleatórias entre os neurônios da rede cerebral pode resultar em um estado de ordem global capaz de gerar comportamento coerente, como o reconhecimento de padrões espaciais ou temporais. Para redes neurais biológicas ou para modelos artificiais de

tais redes, existem alguns princípios de auto organização HAYKIN (1999), estabelecidos por MALSBURG (1990), que servem de embasamento para a afirmação acima.

O primeiro desses princípios ("Modificações nos pesos sinápticos tendem a se auto amplificar"), está relacionado com o postulado de Hebb (HEBB (1949) *apud* HAYKIN (1999)), o qual afirma que "quando o axônio de uma célula A está próximo o bastante de uma célula B e contribui repetidamente e persistentemente para o disparo dessa última, um processo de crescimento ou mudança metabólica acontece nas duas células, de forma que a eficiência de A como uma das células que disparam B é amplificada".

Dessa forma, mesmo fazendo parte de um processo maior, a amplificação da força ou do peso sináptico de um neurônio se dá mediante eventos e estados de variáveis locais. A força referente a uma sinapse específica de uma rede neural é dependente da correlação existente entre a força do efeito pré-sináptico e a força do efeito pós-sináptico, caracterizando um ação de auto definição.

Entretanto, apesar do caráter de auto definição, a rede neural, biológica ou artificial, constitui um sistema que deve chegar a um estado de equilíbrio. Além disso, os recursos envolvidos nos sistemas, tais como quantidade de energia ou quantidade de estímulos, são limitados e devem ser compartilhados por seus elementos. Sendo assim, tais elementos entram em um processo de competição por recursos que justifica o segundo princípio: "Limitação de recursos leva à competição entre sinapses e, portanto, à seleção daquela mais adequada". Sinapses que se auto fortalecem tendem a permanecer no sistema, cada vez mais fortes, em detrimento de outras que com a impossibilidade de fortalecimento, desaparecem.

Dentro do processo de auto definição e competição, emerge um comportamento de cooperação entre alguns conjuntos de sinapses que ativam um neurônio, formando um ciclo de cooperação e fortalecimento dentro do sistema. O terceiro princípio ("Modificações nos pesos sinápticos tendem a cooperar entre si") postula exatamente esse fato.

Por fim, o último princípio se refere às condições dos estímulos do sistema neural, que fomentam todo o processo de auto organização. Esse princípio diz que "a ordem e a estrutura presentes nos padrões de ativação representam informação redundante que é absorvida pela rede neural na forma de conhecimento, o que é um pré-requisito para o aprendizado auto

organizável".

Os princípios brevemente comentados acima formam uma base neurobiológica para a concepção de algoritmos de sistemas auto organizáveis como os Mapas Neurais Auto Organizáveis de Kohonen, descritos e discutidos na seqüência deste capítulo.

4.2 Redes Neurais Artificiais Auto Organizáveis

As RNAs-AO são modelos matemáticos que representam o desenvolvimento da auto organização de projeções topográficas do cérebro natural e são utilizadas em diversas aplicações computacionais que necessitam da interpretação e do processamento de dados.

Esse modelo de inspiração biológica combina aspectos de quantização vetorial com a propriedade de continuidade de uma função. Os dois aspectos possuem também uma justificativa biológica ao levar-se em consideração a eficiência do modelo inspirador, o sistema cerebral.

Uma correspondência do modelo natural com o modelo artificial pode ser encontrada em BAUER; HERRMANN; VILLMANN (1999): "um bom exemplo de mapas neurais no cérebro biológico é o mapa de orientação do sistema visual de muitas espécies. Nesses mapas, os neurônios individuais possuem campos receptivos que os fazem responder somente a um pequeno conjunto de estímulos, onde cada subconjunto está localizado dentro do espaço da retina ou no espaço dos ângulos de orientação dos estímulos. Essa propriedade de resposta de cada neurônio pode ser comparada à resposta de cada vetor quantizador de um mapa neural artifical. Além disso, os neurônios biológicos são dispostos, no mapa de orientação, de tal forma que neurônios vizinhos respondem a subconjuntos de estímulos com orientações e posições similares no espaço da retina, como acontece com os vetores de quantização. O arranjo topográfico de propriedades de resposta forma o elemento de continuidade do mapa neural".

Um mapa neural[1] ω associa entradas e, pertencentes a um conjunto E, para saídas s, pertencentes a um conjunto S. As entradas são unidades de recebimento de estímulos neurais; os estímulos são os dados sob análise comumente representados como em 4.1. As saídas são as representações da entrada, adequadas durante o mapeamento e podem ser chamadas de

[1] A relação representada em ω pode ser $1 - para - 1$ ou $Muitos - para - 1$. Neste trabalho, está-se particularmente interessado na relação $Muitos - para - 1$.

"neurônios".

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1d} \\ x_{21} & x_{22} & \cdots & x_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nd} \end{bmatrix} \tag{4.1}$$

sendo n o número de exemplos no conjunto de dados sob análise e d o número de atributos que descreve um dado (d é a dimensão do espaço onde os dados estão inseridos). Ou seja, cada estímulo neural (dado) é codificado por meio de um vetor, onde cada componente representa a coordenada daquela entrada em um dos eixos do espaço no qual ela está localizada. Da mesma forma, os neurônios de saída são representados por vetores (geralmente em quantidade menor do que n) denominados "vetores de pesos" (W), cujos componentes representam pesos que devem ser ajustados de acordo com as entradas da rede. Esses vetores devem estar arranjados em um *array* N-dimensional (determinando a dimensão do espaço onde os neurônios estão inseridos - espaço de saída) segundo uma geometria de vizinhança, que determina a topologia desse espaço.

Nesse conjunto, formado pelas unidades de entrada, os neurônios do mapa e seus parâmetros de configuração dentro de um "mapa topográfico" formam a arquitetura de uma RNA-AO. Um exemplo de uma RNA-AO é mostrado na Figura 4.1.

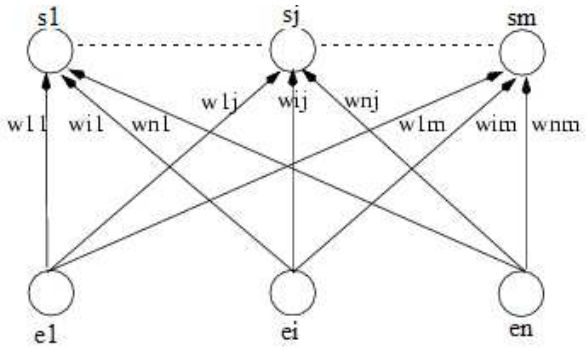


Fig. 4.1: Exemplo da arquitetura de um mapa neural auto organizável. As linhas tracejadas representam as relações de vizinhança. Adaptado de FAUSETT (1994).

Numa RNA-AO os mapeamentos são gerados por meio de um processo de aprendizado

competitivo e estocástico, o qual adapta os vetores de peso de tal forma que eles possam, em algum nível de abstração de dimensão, representar as entradas fornecidas ao mapa. O algoritmo para a construção desse aprendizado ou, como é comumente conhecido, o algoritmo de treinamento, é descrito pelo Algoritmo 1.

Antes de iniciar o processo de treinamento da RNA-AO, ou o processo de construção do mapa, é necessário inicializar o conjunto de pesos que definem o posicionamento dos neurônios. A inicialização localiza os neurônios no espaço de entrada aleatoriamente dentro de um delimitação referente ao espaço ocupado pelos dados, ou linearmente em relação à distribuição dos dados, ou ainda, utilizando algum outro conhecimento "a priori" sobre a estrutura desses.

Algorithm 1 Treinamento seqüêncial de uma Rede Neural Artificial Auto Organizável.

Inicialize os pesos $w_{i,j}$. Determine os parâmetros da vizinhança topológica. Determine os parâmetros da taxa de aprendizado.
while Condição de parada é falsa **do**
  for Cada vetor de entrada e em E **do**
    for Para cada neurônio s em S, compute **do**
      $D(s_j) = \sum_i (w_{ij} - e_i)^2$
      Encontre o índice (j) tal que $D(s_j)$ seja mínima.
      for Todas as unidades j dentro de uma vizinhança específica de s_j **do**
        for Para todo i **do**
          $w_{ij}(novo) = w_{ij}(velho) + \alpha[e_i - w_{ij}(velho)]$
        end for
      end for
    end for
  end for
  Atualize a taxa de aprendizado.
end while

Os dados de entrada são aleatoriamente ou sequencialmente apresentados ao mapa durante o processo de treinamento. O processo de treinamento pode ocorrer de forma seqüencial ou "em lote". No processo sequencial, a cada apresentação de um dado de entrada, o neurônio de localização vetorial mais próxima a ele é escolhido para adaptar-se, aproximando-se do dado mediante uma determinada métrica aplicada aos seus pesos (como no Algoritmo 1). Esse neurônio é dito "vencedor"[2] e, além dele, seus vizinhos no espaço de saída também são ajustados ao dado apresentado, porém com uma intensidade menor (veja a Figura 4.2) Este aspecto caracteriza a cooperação entre os neurônios no processo de aprendizado (ou de treinamento).

[2]O neurônio vencedor é comumente chamado de BMU - *Best Matching Unit*.

No processo de treinamento em lote, todos os dados de entrada são apresentados ao mapa e o ajuste de pesos é realizado mediante uma média das necessidades de ajuste verificada a cada apresentação de um dado.

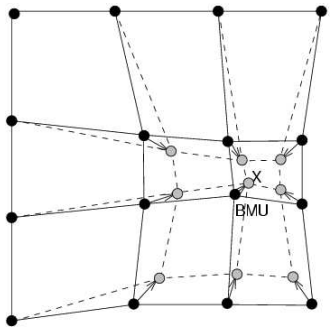


Fig. 4.2: Ajuste de pesos do neurônio vencedor e dos seus vizinhos (VESANTO (1997)). Os círculos escuros representam os neurônios antes do ajuste. Os círculos claros representam os neurônios após o ajuste.

Na Figura 4.3 observa-se a estrutura de vizinhança de um neurônio. A organização em um mapa neural 1-dimensional e em dois mapas neurais 2-dimensionais (com *lattice* retangular e hexagonal) são mostradas nas Figuras 4.3(a), (b) e (c), respectivamente. O neurônio ao qual se refere o esquema de vizinhança é denotado por % e as vizinhanças de níveis 0, 1 e 2 possuem limites representados, respectivamente, por $*$, $+$ e $\#$. Em estruturas simples (para qualquer dimensão do espaço de saída) os neurônios localizados próximos às arestas do mapa possuem menos unidades em sua vizinhança[3].

De acordo com as relações de vizinhança definidas na arquitetura da rede têm-se estabelecidas as configurações do espaço de saída (A) do mapa - o relacionamento topológico dos neurônios. Por exemplo, uma vizinhança 2-dimensional, como mostrado na Figura 4.3(b) ou (c), especifica a dimensão 2 do espaço de saída do mapa neural, onde as distâncias entre as unidades ou neurônios são medidas matricialmente (espaço matricial).

O espaço de entrada do mapa (V) é formado pelas relações de distância vetorial entre as unidades de saída, ou neurônios do mapa. Como cada neurônio é representado por um vetor com número de componentes igual ao número de componentes dos vetores que representam

[3]Esse "efeito de borda" pode ser corrigido se assim se desejar.

(a) (b) (c)

Fig. 4.3: Estrutura de vizinhança em um mapa neural auto organizável: (a) mapa neural de topologia 1-dimensional; (b) mapa neural de topologia 2-dimensional com vizinhança retangular; (c) mapa neural de topologia 2-dimensional com vizinhança hexagonal. Adaptado de FAUSETT (1994).

cada dado analisado, a dimensão do espaço de entrada do mapa é determinada pela dimensão do espaço do conjunto de dados sob análise.

No processo de treinamento o ajuste de peso é realizado mediante uma regra denominada "regra de aprendizado", a qual atua como forma de minimização de erros, realizada sob uma variação $\triangle$. Em 4.2, é apresentada uma fórmula para determinar essa variação, onde o termo $(w_r - e)$ ajuda a minimizar erros.

$$\triangle = \alpha h(r, s)(w_r - e) \tag{4.2}$$

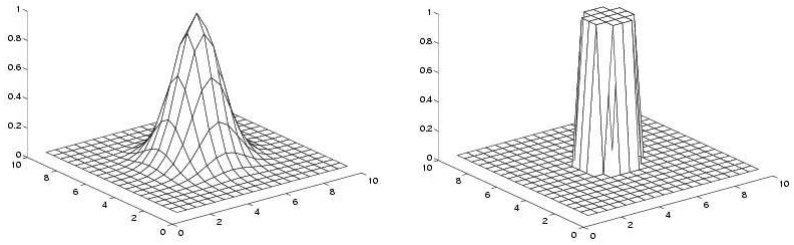
em que w_r é o vetor de peso a ser ajustado, r é o neurônio vizinho, e é o dado de entrada, s é o neurônio vencedor, $h(r, s)$ é uma função de ajuste de vizinhança e α é a taxa de aprendizado. A taxa de aprendizado pode ser reduzida, durante o processo de treinamento, por meio de uma função linear ou geométrica do tempo, de modo a realizar uma espécie de refinamento ao final do processo. Experimentos relatados na literatura, como em FAUSETT (1994), mostram que os resultados obtidos com os dois tipos de função são similares.

Um ajuste ainda mais fino do processo de treinamento é realizado com o uso da função de vizinhança que a ajusta durante o processo de treinamento. A função de vizinhança h pode ser formulada para forçar o alinhamento dos neurônios vizinhos durante o processo de adaptação, induzindo, indiretamente, à preservação da topologia. Um exemplo de uma função de ajuste

de vizinhança é apresentado em 4.3.

$$h(r,s) = exp(-\frac{\| r-s \|^2}{2\sigma^2(t)}) \tag{4.3}$$

em que σ é o fator que determina a largura da função de vizinhança de acordo com o tempo (t) de treinamento, atuando como um fator de escala no ordenamento topográfico. A representação gráfica da vizinhança constituída por meio dessa função está na Figura 4.4(a). Na Figura 4.4(b) é mostrado o gráfico de uma função de vizinhança mais simples.



(a) (b)

Fig. 4.4: Dois tipos de função de vizinhança (VESANTO (1997)): (a) função de vizinhança Gaussiana; (b) função de vizinhança "Bubble".

Durante o treinamento do mapa neural, a taxa de aprendizado e a função de vizinhança devem ser reduzidas, buscando a estabilização dos ajustes dos pesos ou convergência do mapeamento. A formação do mapeamento ocorre em duas fases: a ordenação inicial (fase mais rápida) e a convergência final (fase mais demorada pois ajustes finos são realizados). Durante a fase de ordenação os neurônios estabelecem seus lugares no mapa topográfico e durante a fase de convergência a acuidade estatística na localização dos neurônios é realizada para construir uma representação de dados de maior qualidade. Para uma discussão mais profunda sobre os parâmetros funcionais que influenciam a fase de ordenação e de convergência em uma RNA-AO, veja ERWIN; OBERMAYER; SCHULTEN (1992a) e ERWIN; OBERMAYER; SCHULTEN (1992b).

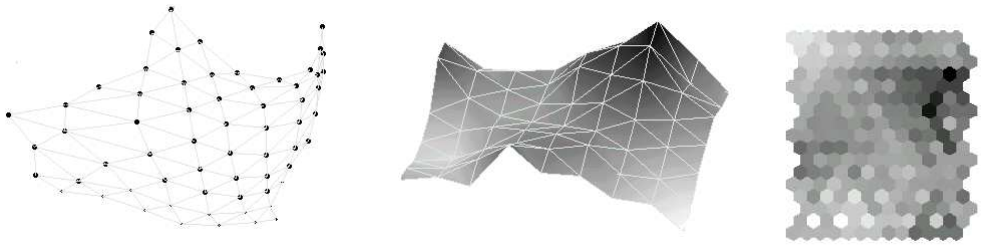
Parâmetros como a largura ou altura da função de vizinhança podem influenciar no resultado topográfico do mapeamento uma vez que exercem influência sobre as escalas de comprimento da camada de saída, ou seja, influenciam a forma de discretização das unidades neuronais e

levam a diferentes formas de cobertura de espaço.

Por exemplo, BAUER; HERRMANN; VILLMANN (1999) analisou um mapa 2-dimensional com um formato linear e um mapa 1-dimensional e concluiu que se a largura da função de vizinhança é grande quando comparada à largura da linha formada pela camada de saída do mapa 2-dimensional, este será semelhante a um mapa 1-dimensional, enquanto que para larguras menores, outras distorções topológicas poderão ser observadas.

Enfim, uma RNA-AO é caracterizada pela formação de um mapa topográfico dos padrões de entrada, no qual a localização espacial (coordenadas) dos neurônios no *lattice* é indicativo das características estatísticas contidas nos padrões (HAYKIN (1999)). Nesse mapeamento topográfico, cada informação é mantida em seu próprio contexto e os neurônios que representam pedaços da informação permanecem localizados proximamente no mapa, de forma que eles possam interagir por meio de pequenas conexões sinápticas.

O mapa topográfico precisa ser analisado para que conclusões sobre os dados sejam inferidas, e essa análise pode se dar por meio de observação direta do mapa ou por meio da observação de estruturas dele derivadas, por exemplo: matriz de distâncias e Matriz-U. Na Figura 4.5 são mostrados: (a) um mapeamento obtido por meio do treinamento de uma RNA-AO 3-dimensional para o conjunto de dados Íris; (b) a respectiva matriz de distâncias; (c) a respectiva Matriz-U. Para mais detalhes sobre formas de visualização e análise do mapeamento resultante em uma RNA-AO, veja VESANTO (1997) e COSTA (1999).



(a) (b) (c)

Fig. 4.5: Visualização dos mapeamentos resultantes da aplicação de uma RNA-AO 3-dimensional sobre o conjunto de dados Íris: (a) o mapa topográfico; (b) a matriz de distâncias; (c) a Matriz-U. Adaptado de ALHOUNIEMI et al. (2002)

4.3 Avaliação da Qualidade de uma Rede Neural Artificial Auto Organizável

A análise de um mapeamento resultante da aplicação de uma RNA-AO sobre um conjunto de dados exige tempo de trabalho. Assim, é interessante que se aplique técnicas de avaliação, expressa por medidas específicas, capazes de avaliar a qualidade de um mapeamento antes que ele seja submetido a uma análise mais detalhada. Em HAYKIN (1999) são apresentadas quatro propriedades que um mapeamento topográfico que descreve características de um conjunto de dados deve apresentar. A qualidade desse mapeamento pode ser medido por meio da análise do quão bem o mapeamento satisfaz estas propriedades, que são:

- Aproximação do espaço de entrada: um dos objetivos de uma RNA-AO é representar um conjunto grande de vetores de entrada, por meio de um conjunto menor de vetores localizados em um espaço de dimensão mais baixa. Ou seja, realizar a quantização do espaço e a redução de dimensão;

- Ordenação Topológico: esta propriedade é consequência do processo de adaptação (alterações de pesos sinápticos), que força a aproximação do neurônio vencedor ao vetor de entrada analisado. Além disso, as adaptações nos neurônios da vizinhança do vencedor também contribuem para o estabelecimento de uma ordenação que reflete a topologia dos dados de entrada;

- Correspondência de densidade: esta propriedade afirma que, caso em uma região do espaço de entrada existam muitos dados, a região correspondente no espaço de saída deverá conter também uma maior concentração de neurônios. A tarefa de identificação de agrupamentos, possível de ser realizada através da uma análise mais detalhada do mapeamento resultante, está relacionada a esta propriedade;

- Seleção de características: o mapa auto organizável é capaz de identificar as principais características de um conjunto de dados distribuído de forma não linear no espaço de entrada. Esta capacidade é equivalente ao resultado de um Processo de Análise de Componentes Principais Não-Linear.

Os dois objetivos que uma RNA-AO se propõe a resolver, quantização vetorial e mapeamento topográfico, podem ser mensurados. A reconstrução do erro constitui uma forma importante para avaliação dos resultados obtidos com o mapeamento conforme o primeiro objetivo. Os resultados obtidos quanto ao segundo objetivo são avaliados por meio da análise da topografia do mapa. Relações topográficas entre as unidades neuronais distinguem esse tipo de mapa neural e o algoritmo de formação de mapa de outros esquemas de quantização vetorial (como o algoritmo de agrupamento *k-means*). A topografia de um mapa neural pode e deve ser explorada com vários objetivos (BAUER; HERRMANN; VILLMANN (1999)): redução das consequências de ruído em uma trasmissão de dados quantizados, interpolação, melhoraria da visualização de dados localizados em espaços de altas dimensões, e outras aplicações que fazem uso da relação de distância e vizinhança entre pontos.

A avaliação topográfica caracteriza uma das formas mais eficientes de avaliar a qualidade de um mapeamento realizado por uma RNA-AO. Como os dados analisados podem não residir em um espaço de dimensão igual a escolhida para o espaço onde reside a camada de saída do mapa, um trabalho de redução de dimensão acontece, justificando a avaliação do mapeamento no âmbito da preservação de topografia. Dessa forma, é fato que não existe preservação topográfica perfeita, mas sim aproximações boas e ruins.

4.3.1 Quantização

Um mapeamento oriundo da aplicação de uma RNA-AO sobre um conjunto de dados tem como resultado uma representação resumida de um conjunto de dados, feita por meio de vetores. Essa é a propriedade de quantização vetorial do mapa. Para FAUSETT (1994), quantização vetorial é a tarefa de formar agrupamentos dos dados (vetores de entrada) com o objetivo de comprimi-los sem perder informações importantes. O conjunto de vetores quantizadores que representam esses dados, comumente chamada de "codebook", pode ser usado para interpretá-los.

De forma mais específica, pode-se dizer que quantizar o espaço significa dividi-lo em regiões distintas e, para cada região, definir um vetor de reconstrução (veja Figura 4.6). Novos pontos deste espaço quantizado poderão ser identificados de acordo com a sua proximidade (segundo

alguma métrica) ao vetor de reconstrução de cada região.

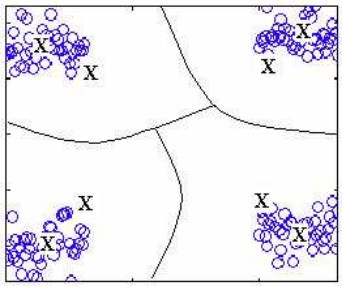


Fig. 4.6: Exemplo da quantização do espaço de um conjunto de dados (círculos) e os vetores de reconstrução (**X**) de cada região (delimitadas pelas linhas curvas).

Para avaliar a capacidade de quantização de uma RNA-AO utiliza-se, geralmente, a medida do Erro de Quantização. Essa medida está baseada no cálculo das médias das distâncias entre os dados e os vetores que melhor os representam, isto é, entre os dados e seus BMUs (veja Figura 4.7). O erro de quantização é dado por 4.4:

$$E_q = \frac{1}{N}\sum_{i=1}^{N} \|x_i - w_{bi}\| \tag{4.4}$$

em que x é um vetor do conjunto sob análise; w_{bi} é o BMU para x_i; N é o número de dados no conjunto de dados. Quanto menor for o erro obtido mais próximo dos dados estão os seus BMUs, isto é, os dados estão melhor representados.

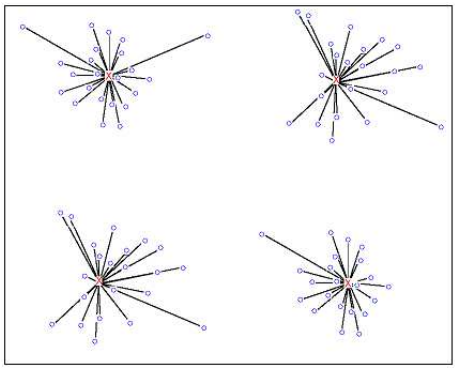


Fig. 4.7: Cálculo do erro de quantização. Os dados estão representados por círculos e ligados por retas (que representam as distâncias) aos seus BMUs (representados pelo **X**).

4.3.2 Topografia

A topografia de um mapa neural pode assumir diferentes formações e a determinação de uma medida (ou um número) que a quantifique tem sido um problema de forte exploração, fato revelado pelas diversas formulações propostas para este fim.

De uma forma simplista, mas não menos significativa que um modelo formal, pode-se definir a topografia de um mapa neural como um meio de expressão das similaridades existentes entre as características do objeto mapeado, por meio de pontos localizados proximamente no mapa. Entretanto, como o conceito de similaridade pode ter diversas facetas em diferentes níveis de análise, o conceito de topografia pode assumir diferentes interpretações.

A primeira interpretação diz respeito à topologia dos elementos envolvidos no mapeamento. Nesta faceta, as relações de vizinhança devem ser preservadas no mapeamento, dando-lhe um caráter de continuidade em relação ao objeto de entrada e a sua correspondência no espaço de saída. Na Figura 4.8(b) é ilustrada essa interpretação (note que as relações entre vizinhos está trocada). Nela, e nas demais figuras do conjunto, os mapas são formados por quatro unidades de entrada e quatro unidades de saída, o mapeamento é ilustrado por meio de flechas que unem cada unidade de entrada à sua correspondente unidade de saída. A Figura 4.8(a) ilustra o mapeamento perfeito.

A necessidade de manutenção das relações de distância entre pares de pontos do objeto envolvido no mapeamento, caracteriza a faceta métrica das relações topográficas (Figuras 4.8(b), (c) e (d)) e, mesmo que de uma maneira mais sutil, podem contribuir como uma forma de manutenção de relações de similaridade.

E por fim, baseando-se ainda nas relações métricas, estabelece-se a terceira faceta que preza pela manutenção do *ranking* das distâncias entre os pontos. Esse conceito está ilustrado nas Figuras 4.8(b) e (c).

A qualidade topográfica de um mapa neural pode ser medida de várias formas (veja BAUER; HERRMANN; VILLMANN (1999) para um estudo comparativo) e cada uma destas formas prioriza um ou mais aspectos de qualidade topográfica. De modo geral, as formas de medição consideram que se a dimensão do espaço de saída é muito baixo, o mapa tenta realizar auto contorções para preencher adequadamente o espaço de maior dimensão ocupado pelo conjunto

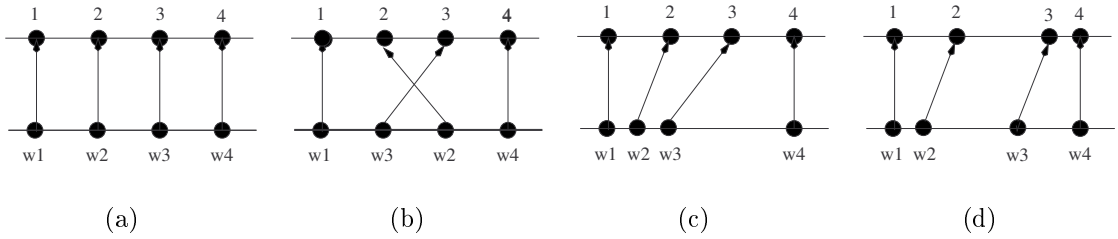


Fig. 4.8: Conceitos de topografia (BAUER; HERRMANN; VILLMANN (1999)): (a) mapeamento perfeito; (b) distorção topológica, métrica e de *ranking* ; (c) distorção métrica e de *ranking*; (c) distorção métrica. Em (b) a violação topológica acontece em pequena escala, em que o par de vizinhos mais próximos de w_1 e w_3 é mapeado para 1-3, que não são vizinhos mais próximos. Além disso, as relações de distâncias estão invertidas: $d_V(w_1, w_2) > d_V(w_1, w_3)$, mas $d_A(1,2) < d_A(1,3)$, onde V é o espaço de entrada (dos dados) e A é o espaço de saída (onde se dá a organização topológica dos neurônios). Em (c) o *ranking* de vizinhos no espaço de entrada difere do *ranking* no espaço de saída - os primeiros dois vizinhos da unidade de saída 3 são as unidades 2 e 4, enquanto que os vizinhos mais próximos da unidade de entrada correspondente, w_3, são w_1 e w_2. Além disso, a distorção métrica é observada porque $d_V(w_3, w_4) > d_V(w_3, w_1)$ enquanto $d_A(3,4) < d_A(3,1)$. Em (d) a distorção métrica é indicada por $d_A(1,2) > d_A(3,4)$ e $d_V(w_1, w_2) < d_V(w_3, w_4)$.

de dados.

A avaliação da qualidade dessas medidas pode estar intimamente ligada ao objetivo para o qual o mapa neural está sendo utilizado, no entanto, BAUER; HERRMANN; VILLMANN (1999) destaca alguns aspectos que devem ser considerados:

- Unicidade: somente quando aplicada a um mapeamento com preservação de topologia perfeita, a avaliação poderia resultar em um valor indicativo de tal efeito;
- Ordenaçao Sistemática: quando aplicada a uma série de mapas com uma mudança sistemática de topologia, como o aumento monotônico da dimensão do mapa para um número constante de neurônios, o valor da medida deveria mudar sistematicamente, expressando qual dos mapas analisados representa melhor os dados;
- Reproducibilidade: para mapas similares a medida deveria apresentar resultados qualitativamente similares;
- Invariância: a medida de topologia deveria depender somente da topologia do mapa, sem alterar-se em virtude de transformações de simetria ou escalonamentos.

A avaliação da preservação topográfica de um mapeamento é equivalente à avaliação da discrepância entre as relações de vizinhança, como um tipo de relação de similaridade verificada em ambos os lados de um mapeamento. De acordo com GOODHILL; SEJNOWSKI (1996) três níveis de classificação de manutenção de relações de similaridade são definidos:

- Nível de manutenção total de relações de similaridade: cada par de pontos no espaço de saída possui valores de similaridade exatamente iguais àqueles, do espaço de entrada, que representam. Esse nível de manutenção é dificilmente alcançado quando existe redução de dimensão no mapeamento estudado;

- Nível de manutenção correlata de relações de similaridade: os valores de similaridade entre os pares de pontos no espaço de saída devem estar correlacionados aos valores de similaridade obtidos entre os pontos no espaço de entrada;

- Nível de manutenção ordenada de relações de similaridade: a comparação entre o conjunto de valores das similaridades entre os pontos no espaço de entrada com o conjunto de valores das similaridades entre os pontos no espaço de saída deve verificar a mesma relação de ordenação nos dois conjuntos.

Neste trabalho está-se particularmente interessado no último desses níveis de classificação. As medidas descritas a seguir enfatizam esse quesito. Para uma revisão sobre outros tipos de medidas (medidas baseadas em relações de similaridade entre pontos, baseados em medição de caminhos mínimos ou baseadas em critérios contínuos) veja GOODHILL; SEJNOWSKI (1996).

Erro Topológico

O Erro Topológico quantifica, proporcionalmente, o número de dados de entrada para os quais o primeiro e o segundo neurônios que melhor os representam não são adjacentes no espaço de saída (veja Figura 4.9). Formalmente,

$$\xi = \frac{1}{n}\sum_{k=1}^{n} u(x_k), \text{ em que } u(x_k) = \begin{cases} 1, \text{ se o primeiro e o segundo neurônios} \\ \text{vencedores forem adjacentes;} \\ 0 \text{ , caso contrário} \end{cases} \quad (4.5)$$

em que n é o número total de dados de entrada e x é um dado de entrada. Essa medida se aproxima de 0 quanto menor for o número de dados para os quais o primeiro e o segundo BMUs não são adjacentes, entretanto, ela não é capaz de fornecer nenhum tipo de informação sobre qual tipo de desvio topológico ocorreu no mapeamento.

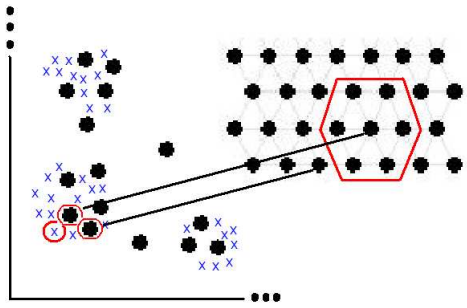


Fig. 4.9: Erro Topológico. Os neurônios circulados representam o primeiro e segundo BMUs do dado **x** circulado. Os dois neurônios destacados não são vizinhos diretos no espaço de saída da RNA-AO.

Produto Topográfico

Esta medida tem como principal objetivo quantificar a preservação de relações de vizinhança em mapeamentos realizados sobre espaços de mesma ou diferentes dimensões. O Produto Topográfico P está formalmente descrito em 4.6. Esse processo de medição de qualidade topográfica relaciona, para cada neurônio, a seqüência de vizinhos no espaço de entrada com a seqüência de vizinhos no espaço de saída. Uma discussão detalhada sobre esta medida é apresentada em BAUER; PAWELZIK (1992) e algumas considerações estão reproduzidas aqui.

$$P = \frac{1}{n^2 - n}\sum_{j=1}^{N}\sum_{k=1}^{N-1} log(\prod_{l=1}^{k} \frac{d_V(w_j, w_{n_l^A(j)})}{d_V(w_j, w_{n_l^V(j)})}\frac{d_A(j, n_l^A(j))}{d_A(j, n_l^V(j))})^{\frac{1}{2k}} \quad (4.6)$$

em que n é o número de neurônios no mapeamento, j é um contador para os neurônios e os representa no espaço de saída, k e j são contadores dos vizinhos de um neurônio, d_V é a distância vetorial entre dois neurônios, d_A é a distância matricial entre dois neurônios, w é a representação do neurônio no espaço de entrada, $w_{n_l^A(j)}$ é a representação, no espaço de entrada, do l-ésimo vizinho mais próximo do neurônio j no espaço de saída, $w_{n_l^V(j)}$ é a representação, no espaço de entrada, do l-ésimo vizinho mais próximo do neurônio j no espaço de entrada, $n_l^A(j)$ é a representação, no espaço de saída, do l-ésimo vizinho mais próximo do neurônio j no espaço de saída e $n_l^V(j)$ é a representação, no espaço de saída, do l-ésimo vizinho mais próximo do neurônio j no espaço de entrada. P é um indicador de contorção de espaços. Um valor maior que 1 para P mostra a contração do espaço de entrada dentro do espaço de saída. Um valor menor do que 1 indica o efeito contrário.

Nessa formulação é utilizada a notação (d_V) para representar as relações de vizinhança no espaço vetorial (a métrica de distância vetorial usada, na aplicação desta tese, é a Distância Euclidiana) e a notação (d_A) para a representação de vizinhança no espaço matricial (onde a métrica de distância matricial escolhida nesta tese é a Distância de Manhattan). Na Figura 4.10 estão mostrados, graficamente, os espaços matricial e vetorial de uma RNA-AO 2-dimensional.

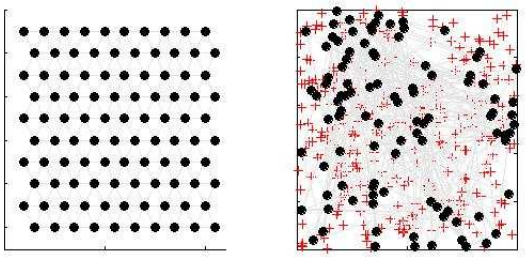


Fig. 4.10: Espaços de um mapeamento RNA-AO: matricial (de saída); vetorial (de entrada).

As relações de vizinhança no espaço vetorial representam as relações de vizinhança no espaço de entrada V do mapeamento, onde analisa-se as relações entre os neurônios por meio de seus vetores de peso. Representa-se por n_k^V, o $k-$ésimo vizinho mais próximo do neurônio j no espaço de entrada do mapeamento, e esse é encontrado a partir da aplicação de 4.7:

$$n_i^V(j) : d^V(w_j, w_{n_i^V(j)}) = \min_{j' \in W} d^V(w_j, w_{j'}) \tag{4.7}$$

em que $i = 1$ identifica o neurônio mais próximo de j, $i = 2$ identifica o segundo neurônio mais próximo de j e assim por diante[4]; w é o vetor de pesos que representa um neurônio; d_V é a métrica para o cálculo da distância entre dois vetores e W é o conjunto de neurônios no mapeamento.

A análise dos neurônios em relação à sua disposição matricial, de acordo com a dimensão e topologia de vizinhança escolhida para o mapeamento, mostra as relações de vizinhança matricial e é feita no espaço de saída do mapeamento (A). Representa-se por n_k^A, o $k-$ésimo vizinho mais próximo do neurônio j no espaço de saída do mapeamento, o qual é encontrado a partir da aplicação de 4.8:

$$n_i^A(j) : d^A(j, n_i^A(j)) = \min_{j' \in A} d^A(j, j') \tag{4.8}$$

em que d^A é uma métrica para o cálculo da distância entre dois pontos de uma matriz. Note que para este caso não se trabalha com os vetores que representam os neurônios, e sim, com o posicionamento dos neurônios na matriz que expressa suas relações de vizinhança topológica. Na Figura 4.11 é possível observar como atuam as medidas descritas acima.

Os termos envolvidos no produtório da fórmula 4.6 são formados por medições de distâncias entre os neurônios e seus vizinhos. Cada um dos termos relaciona os neurônios de uma forma diferente, obtendo medidas que são exploradas de forma a estabelecer um tipo de raio. Segue uma descrição de cada um dos termos.

1. $d_V(w_j, w_{n_l^A(j)})$: Calcula a distância vetorial entre um neurônio j e o seu $l-$ésimo vizinho mais próximo no espaço matricial, ambos representados por seus vetores de pesos. Ou seja, está-se medindo a distância existente, no espaço de entrada do mapeamento, entre um determinado neurônio e aquele que é seu vizinho topológico (veja a Figura 4.12(a)), buscando por distorções de topologia;

[4] i pode assumir valores até o número de neurônios -1, já que no espaço vetorial não existe nenhum ordenamento pré-definido que especifique os possíveis vizinhos de um neurônio.

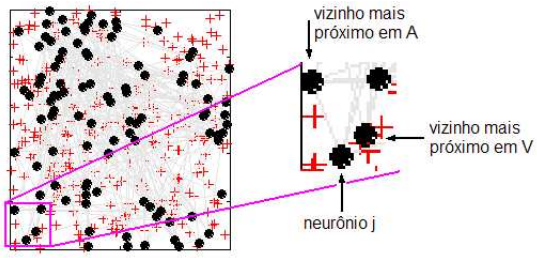
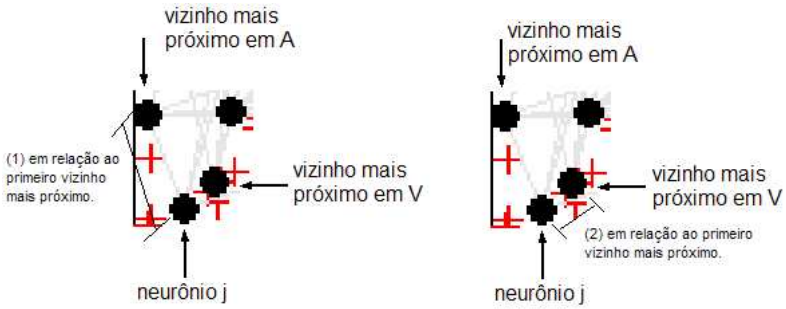


Fig. 4.11: Medidas de distâncias no cálculo do Produto Topográfico. No quadro maior tem-se um mapa RNA-AO plotado sobre os dados de treinamento. Note que o espaço de saída (matricial) está todo contorcido e as relações de vizinhança vetorial não correspondem às relações de vizinhança matricial. O quadro mais à direita é uma ampliação de parte do quadro maior. Note que os vizinhos mais próximos, de um mesmo neurônio, nos dois espaços (matricial - A; vetorial - B) são diferentes.



(a) (b)

Fig. 4.12: Cálculo da distância vetorial entre um neurônio e (a) seu vizinho mais próximo matricialmente e (b) seu vizinho mais próximo vetorialmente.

2. $d_V(w_j, w_{n_l^V(j)})$: Calcula a distância vetorial entre um neurônio j e o seu $l-$*ésimo* vizinho mais próximo no espaço vetorial, ambos representados por seus vetores de pesos. Esse cálculo é idêntico àquele realizado para ordenar os vizinhos mais próximos (veja Figura 4.12(b)).

3. $d_A(j, n_l^A(j))$: Calcula a distância matricial entre um neurônio j e o seu $l-$*ésimo* vizinho mais próximo no espaço matricial, ambos representados pelo índice do neurônio na matriz que determina as relações de vizinhança topológica;

4. $d_A(j, n_l^V(j))$: Calcula a distância matricial entre um neurônio j e o seu $l-$*ésimo* vizinho

mais próximo no espaço vetorial, ambos representados pelo índice do neurônio na matriz que determina as relações de vizinhança topológica. Com esse cálculo está-se avaliando o quanto o vizinho no espaço matricial está distante no espaço vetorial, tentando descobrir violações de topologia. Na Figura 4.13 está representada esta medição.

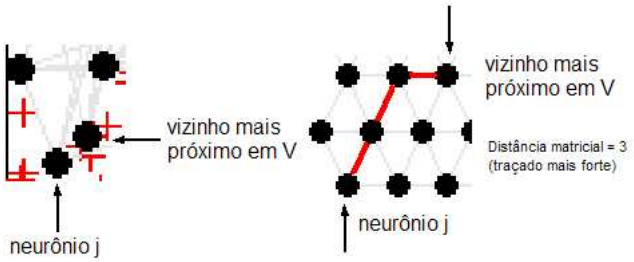


Fig. 4.13: Cálculo da distância matricial entre um neurônio e seu vizinho mais próximo vetorialmente.

As relações estabelecidas a partir dos termos descritos acima são da forma:

$$R_V(j,l) = \frac{d_V(w_j, w_{n_l^A(j)})}{d_V(w_j, w_{n_l^V(j)})} \tag{4.9}$$

$$R_A(j,l) = \frac{d_A(j, n_l^A(j))}{d_A(j, n_l^V(j))} \tag{4.10}$$

Essas relações produzirão o valor 1 sempre que as distâncias matricial ou vetorial entre o vizinho mais próximo de ordem l do neurônio j, no espaço de entrada do mapeamento, forem as mesmas que aquelas observadas entre o seu vizinho mais próximo, de mesma ordem, no espaço de saída do mapeamento, ou seja, $n_l^V(j) = n_l^A(j)$ ou $w_{n_l^A(j)} = w_{n_l^V(j)}$. Isto significa que nenhuma distorção de topologia aconteceu no mapeamento no que tange ao neurônio j e o seu $l-$ésimo vizinho (em uma comparação entre o posicionamento do dois neurônios nos dois espaços do mapeamento).

Analisando esses raios individualmente, existem situações em que se pode cair em um cálculo que não reflete a real situação de distorção de topologia. Analise a situação da Figura 4.14 onde um mapeamento 1-dimensional é mostrado. Nela o $k-$ésimo vizinho mais próximo do neurônio j nos dois espaços do mapeamento não são os mesmos, entretanto, a distância vetorial entre j

e seus dois vizinhos é a mesma e $R_V(j,l) = 1$. Portanto esses raios devem ser combinados, já que para esta mesma situação $R_A(j,l) \neq 1$.

Fig. 4.14: Observação da distância vetorial entre um neurônio e seus vizinhos. A distância vetorial observada entre o neurônio j e seus dois vizinhos "diferentes", nos dois espaços do mapeamento, é a mesma.

Se algum valor diferente de 1 for obtido no cálculo de $R_V(j,l)$ ou $R_A(j,l)$ é indício que existe um desvio topológico observado localmente, junto ao neurônio j. Como observado na Figura 4.14, a linha que compõe o espaço de saída do mapeamento está contorcida e $R_A(j,l) \neq 1$. Note que a sensibilidade desta medida é grande, dado que pequenos desvios métricos serão revelados pela medida quando esta é vista apenas localmente. Analisando com um pouco mais de detalhes o significado dos valores obtidos nas relações $R_V(j,l)$ e $R_V(j,l)$ chega-se às seguintes constatações (a Figura 4.15 é composta por exemplos gráficos para cada um desses casos):

- Um valor maior que 1 será obtido para $R_V(j,l)$ se a distância vetorial entre o neurônio j e o seu $l-$*ésimo* vizinho mais próximo no espaço matricial (de saída) for **maior** que a distância vetorial entre ele e seu $l-$*ésimo* vizinho mais próximo no espaço vetorial (de entrada). Para o caso contrário obtém-se um valor menor que 1.

- Para $R_A(j,l)$, um valor maior que 1 é observado sempre que a distância matricial entre o neurônio j e o seu $l-$*ésimo* vizinho mais próximo no espaço matricial (de saída) for **maior** que a distância vetorial entre ele e o seu $l-$*ésimo* vizinho mais próximo no espaço vetorial (de entrada). Para o caso contrário obtém-se um valor menor que 1.

A obtenção de um valor maior que 1 nesses raios mostra que as distâncias entre as unidades no espaço de entrada são menores do que no espaço de saída, sob o mesmo ângulo de observação

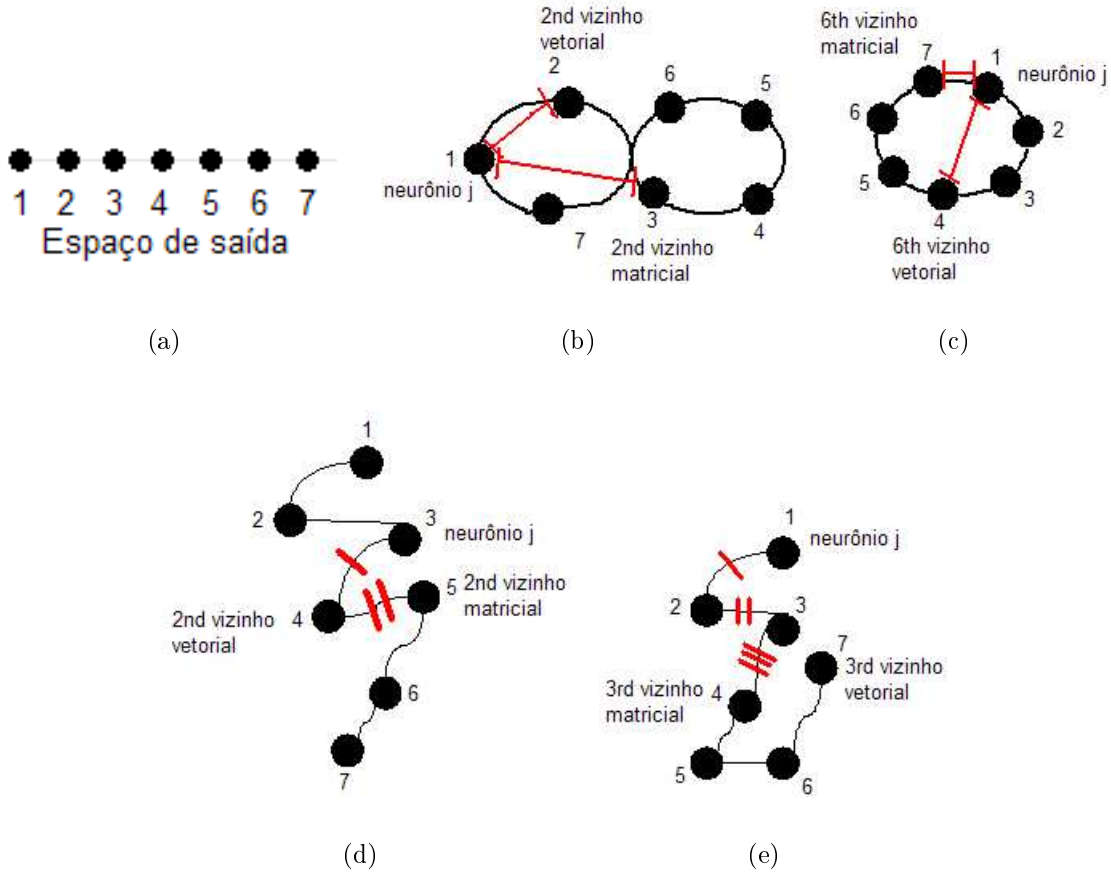


Fig. 4.15: (a) Espaço de saída; (b),(c),(d) e (e) Constatações sobre a configuração do espaço de saída a partir da análise das relações 4.9 e 4.10.

(ou distância vetorial ou distância matricial). Desta forma, percebe-se a contração do espaço de entrada dentro do espaço de saída. Um valor menor do que 1 indica o efeito contrário.

Os termos discutidos acima são calculados para todos os neurônios e os valores obtidos estão relacionados dentro de um produtório. A razão para este tipo de relação é a necessidade de anular efeitos de distorções topológicas muito pequenas, com efeito local, como mostra o exemplo de BAUER; HERRMANN; VILLMANN (1999). Considere os vizinhos mais próximos do nó 3 no espaço de entrada e no espaço de saída (veja Figura 4.16), como segue:

Note que os vizinhos mais próximos coincidem, mas os segundos vizinhos mais próximos não coincidem. Destas medidas segue que:

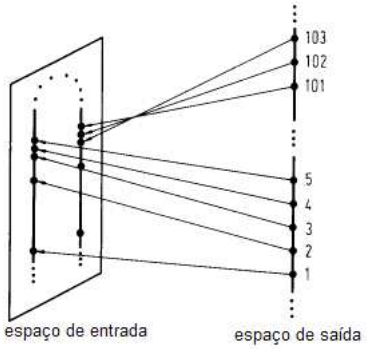


Fig. 4.16: Exemplo de efeitos de distorções topológicas (BAUER; HERRMANN; VILLMANN (1999)).

$$R_V(3,2) = \frac{d^V(w_3, w_2)}{d^V(w_3, w_5)} > 1 \tag{4.11}$$

$$R_A(3,2) = \frac{d^A(3,2)}{d^A(3,5)} < 1 \tag{4.12}$$

Apesar desses resultados que indicam desvios de topologia, os neurônios no espaço de entrada formam uma linha da mesma forma como os neurônios estão dispostos no espaço de saída, ou seja, as relações de vizinhança para os nós 2, 3, 4 e 5 estão preservadas. No produtório (veja 4.13 e 4.14), efeitos como esses podem ser cancelados como mostra o cálculo com os três vizinhos mais próximos do neurônio 3 (veja 4.15 e 4.16), onde o segundo termo do produtório é cancelado pelo terceiro, fazendo com que apenas o primeiro seja considerado na medida e esta acuse a preservação de topologia.

$$P_V(j,k) = (\prod_{l=1}^{k} R_V(j,l))^{\frac{1}{k}} \tag{4.13}$$

$$P_A(j,k) = (\prod_{l=1}^{k} R_A(j,l))^{\frac{1}{k}} \tag{4.14}$$

$$P_V(3,3) = (\frac{d^V(w_3, w_4)}{d^V(w_3, w_4)} \frac{d^V(w_3, w_2)}{d^V(w_3, w_5)} \frac{d^V(w_3, w_5)}{d^V(w_3, w_2)})^{\frac{1}{3}} = 1 \tag{4.15}$$

$$P_A(3,3) = (\frac{d^A(3,4)}{d^A(3,4)}\frac{d^A(3,2)}{d^A(3,5)}\frac{d^A(3,5)}{d^A(3,2)})^{\frac{1}{3}} = 1 \tag{4.16}$$

Ainda, seguindo o exemplo de BAUER; HERRMANN; VILLMANN (1999), observa-se como desvios de topologia mais expressivos são atestados pela medida em questão, como mostra o cálculo para os quatro vizinhos mais próximos do neurônio 3 (4.17 e 4.18).

$$P_V(3,4) = (\frac{d^V(w_3,w_4)}{d^V(w_3,w_4)}\frac{d^V(w_3,w_2)}{d^V(w_3,w_5)}\frac{d^V(w_3,w_5)}{d^V(w_3,w_2)}\frac{d^V(w_3,w_1)}{d^V(w_3,w_{102})})^{\frac{1}{4}} = (\frac{d^V(w_3,w_1)}{d^V(w_3,w_{102})})^{\frac{1}{4}} \approx 1 \tag{4.17}$$

$$P_A(3,4) = (\frac{d^A(3,4)}{d^A(3,4)}\frac{d^A(3,2)}{d^A(3,5)}\frac{d^A(3,5)}{d^A(3,2)}\frac{d^A(3,1)}{d^A(3,102)})^{\frac{1}{4}} = (\frac{d^A(3,1)}{d^A(3,102)})^{\frac{1}{4}} \ll 1 \tag{4.18}$$

Ou seja, foi encontrado um pequeno desvio de 1 para R_V e um desvio mais expressivo para R_A.

A partir do exemplo acima é possível perceber que ao utilizar um valor de k suficiente, os valores finais dos produtos R_V e R_A permanecerão próximos a 1, mostrando um cômputo da médias de contribuições locais. Para os casos onde não existem contribuições locais desta natureza, um resultado de medição de distorções topológicas pode ser obtido combinando os raios numa relação multiplicativa, proporcionando um tipo de ponderação entre as medidas obtidas sob os dois ângulos de observação, conseguindo:

$$R(j,k) = (\prod_{l=1}^{k} R_V(j,l)R_A(j,l))^{\frac{1}{2k}} \tag{4.19}$$

Dessa forma, chega-se à constatação sobre o tipo de distorção topológica encontrado. Valores maiores que 1 para $R(j,k)$ são obtidos quando as maiores distâncias são observadas no espaço de saída, ou seja, o espaço de saída se contrai em um fator maior que o espaço de entrada, mostrando que sua dimensão é mais alta. O caso inverso, quando as distâncias observadas no espaço de saída são menores que no espaço de entrada, levam a um valor menor do que 1 e mostram que o espaço de saída está realizando mais contrações que o espaço de entrada e possui uma dimensão menor.

Finalmente, a fim de analisar todo o mapeamento, os raios descritos acima são calculados para todos os neurônios do mapa e em relação a todas as ordens de vizinhança dele. Os somatórios executam esse procedimento, sendo que o primeiro (somatório em j) diz respeito aos neurônios e o segundo (somatório em k) diz respeito aos seus vizinhos. A função logaritmo é utilizada para facilitar as análises dos desvios de 1 e uma normalização em relação ao número de neurônios na rede é aplicada.

Coeficiente de Spearman ρ

Essa medida analisa a preservação de topologia qualificando-a como uma medida estatística para o grau de correlação entre os *rankings* de ordenação das unidades. Proposta por BEZDEK; PAL (1995), é baseada na idéia de que a preservação topológica métrica refere-se a uma correspondência de ordenação de todas as distâncias entre pares nos espaços de entrada e nos espaços de saída. Assim, essa medida avalia se as posições relativas de todos os vizinhos de um ponto são mantidas no mapeamento.

Segundo o coeficiente de Spearman, um mapeamento é uma transformação com preservação topológica métrica se para todo w_r, sempre que w'_r for o $k-$ésimo vizinho mais próximo em V, o r' será o $k-$ésimo vizinho mais próximo de r em A.

Para calcular este coeficiente (4.20), as distâncias vetoriais entre quaisquer dois vetores de referência em V são concatenadas formando o vetor c_v de comprimento $T = N(N-1)/2$, onde N é o número de unidades na camada de saída do mapa. Da mesma forma, as distâncias em A formam o vetor c_A. Então c_V e c_A são ordenados formando os vetores b_V e b_A. Assim, $b_V(1)$ indica o índice do menor valor de c_V, e idem para b_A.

$$\rho(b_V, b_A) = 1 - \frac{1}{T^3 - T}(6 \sum_{k=1}^{T}(b_A(k) - b_V(k))^2) \neq 1 \tag{4.20}$$

Para ordenações perfeitas, o coeficiente será igual a 1, ou seja, $c_V = c_A$ e, ordenações aleatórias resultarão em um coeficiente próximo de 0. Se o coeficiente for igual a -1, a ordenação do mapeamento é o inverso da ordenação existente no espaço de entrada. Segundo BAUER; HERRMANN; VILLMANN (1999) essa medida é instável, de baixa reprodutibilidade e seus resultados deveriam ser tomados como uma média de várias simulações.

Medida Z

A Medida Z, ou medida de "Zrehen", verifica se para todo par de entradas r e r', a esfera de raio $\|\frac{w_r - w_{r'}}{2}\|$ e centro $\frac{w_r - w_{r'}}{2}$, onde w são vetores de pesos para os BMUs das entradas dadas, não contém qualquer outro vetor de peso. Desta forma, a medida verifica se pares de neurônios vizinhos estão localmente organizados, ou seja, se a linha reta que une seus respectivos BMUs não possui pontos que estejam mais próximos de vetores de pesos de outros neurônios.

Vetores de pesos $w_{r''}$ que violam essa condição são chamados "intrusos" e a soma de todos os intrusos de todos os pares de neurônios vizinhos formam a Medida Z. Os mapas para os quais a Medida $Z = 0$ são considerados mapas com preservação topológica.

Medida C

Tal medida foi estabelecida por Goodhill *et al.* como uma função de custo que assume seu valor mínimo se o *ranking* de distâncias vizinhas coincide em V e A, para todos os pares de unidades de entrada e suas correspondentes unidades de saída. Neste custo funcional diferentes distâncias métricas podem ser utilizadas e ele é formalmente descrito por 4.21:

$$C(\Omega) = \frac{1}{2} \sum_{r} \sum_{r' \neq r} d_V(w_r, w_{r'}) d_A(r, r') \tag{4.21}$$

em que r e r' são dois dados de entrada, w e w' são os vetores de peso dos BMUs para os dados de entrada, d_V e d_A se referem a, respectivamente, distância vetorial no espaço dos dados e distância matricial no espaço de saída da rede neural.

Segundo BAUER; HERRMANN; VILLMANN (1999), a motivação principal para a utilização dessa medida é comparar diferentes configurações de mapas dentro de uma mesma topologia. Por exemplo: diferentes projeções de um espaço de entrada $N \times N$ (quadrado), para um mapa de N^2 neurônios organizados linearmente.

Função Topográfica

Nesta abordagem as relações de vizinhança entre vetores de peso, no espaço de entrada e no espaço de saída, são representadas em grafos. Distâncias entre quaisquer dois vetores de

peso são computadas como distâncias no grafo. Uma comparação dos dois grafos, o do espaço de entrada e o do espaço de saída, resultam em um grau de desvio topológico.

Essa medida analisa o mapeamento com base no espaço de entrada real, ou seja, utiliza as medidas de distâncias existentes dentro do conjunto de dados analisado. Segundo BAUER; VILLMANN (1995) essa característica evita que mapas com contorções estranhas, originadas na tentativa de aproximação de dados com curvaturas também estranhas, sejam considerados como detentores de violações de topologia.

Nos cálculos envolvidos na Função Topográfica, utiliza-se uma estrutura tipo "grafo" D_V correspondente a um caminho $um-para-um$ entre células (Ω_r)[5] de um diagrama de Voronoi, concebido sobre o espaço de entrada.

Nesse processo de avaliação de topologia, as distâncias entre w_r e $w_{r'}$ correspondem ao caminho mínimo $d^{D_V}(w_r, w_{r'})$ entre w_r e $w_{r'}$. Para efeitos de análise topológica, w_r e $w_{r'}$ são considerados adjacentes somente se as células de Voronoi correspondentes são adjacentes.

Na análise dos nós r e r', em relação ao espaço de entrada A, é considerada uma "vizinhança forte" em A, a qual é baseada na distância Euclidiana e no *lattice* do espaço. Além disso, uma vizinhança fraca em A é estabelecida para amenizar os efeitos de *lattice* sobre a topologia. Um valor limite k para as distâncias entre r e r' é estabelecido para determinar o limite entre vizinhança forte e a então denominada vizinhança fraca. Para mais detalhes sobre esta medida veja VILLMANN et al. (1997).

Segundo tal medida, o mapeamento preserva topologia se:

- no mapeamento $V \rightarrow A$, se dois vetores estão conectados no grafo D_V são projeções de dados cuja relação de vizinhança em A é fraca;
- no mapeamento $A \rightarrow V$, se dois vetores possuem sua projeções conectadas no grafo D_V, são referentes a dados cuja relação de vizinhança em A é forte.

As equações 4.22, 4.23 e 4.24 definem a Função Topográfica, em que

$$f_r(k) \stackrel{def}{=} \#\left\{r' \mid \|r - r'\|_{max} > k; d^{D_V}(r, r') = 1\right\} \tag{4.22}$$

[5]Em que r especifica a célula que contém o dado sob análise.

$$f_r(-k) \stackrel{def}{=} \#\{r' \mid \|r - r'\| = 1; d^{D_V}(r, r') > k\} \tag{4.23}$$

sendo que $k = 1, ..., N-1$.

$$\Phi(k) \stackrel{def}{=} \begin{cases} \frac{1}{N}\sum_{r' \in A} f_{r'}(k) & k > 0 \\ \Phi(1) + \Phi(-1)2 & k = 0 \\ \frac{1}{N}\sum_{r' \in A} f_{r'}(k) & k < 0 \end{cases} \tag{4.24}$$

A interpretação de $\Phi(k)$ se dá, segundo BAUER; VILLMANN (1995) da seguinte forma: se $\Phi(k) \equiv 0$, e em particular, $\Phi(0) = 0$, ambos os ($V \to A$ e $A \to V$) mapeamentos preservam topologia; para altos $k^+ > 0$ com $\Phi(k) \neq 0$, a dimensão do espaço de saída não é suficiente para representar o espaço de saída; para baixos $k^- < 0$ com $\Phi(k) \neq 0$ a dimensão do espaço de saída é maior que a dimensão do espaço de entrada. Valores pequenos de k indicam conflitos de topologia locais, enquanto que valores maiores indicam conflitos globais.

4.4 Considerações Finais

Após os estudos e interpretações apresentadas neste capítulo, conclui-se que ele, ao dar continuidade aos passos programados, dedicou-se a fornecer um resumo sobre as principais questões envolvidas no projeto, treinamento e análise de uma RNA-AO. Visto que esta tese se dedica a estudar especificamente questões referentes à capacidade quantizadora e de preservação de informação topológica inerentes ao mapeamento produzido por uma rede neural deste tipo, as medidas de avaliação destes quesitos tiveram destaque e, aquelas aplicadas nesta tese foram discutidas com um nível de detalhamento maior. Objetivou-se com isso dar suporte ao leitor para um entendimento mais preciso sobre o estudo aqui desenvolvido.

A principal contribuição desta tese, o projeto de um novo tipo de dimensão fractal - "dimensão fractal-fuzzy significativa" - será detalhada no próximo capítulo e então, destacar-se-á

a capacidade de quantização e de preservação topológica como os principais objetivos de uma RNA-AO, o dilema envolvido no alcance destes objetivos e como a medida aqui proposta pode contribuir para amenizar os efeitos deste dilema.

Capítulo 5

Dimensão Fractal Fuzzy Significativa

> A palavra "descoberta" carrega consigo uma aura de mistério porque é normalmente anexada a um processo mental que não deixa nenhum traço gravado de seus passos intermediários. Ela é um termo apropriado para o processo de geração de heurísticas, uma vez que traçar novamente os passos evocados nesse processo é, usualmente, uma tarefa difícil! Scheduling Computer and Manufacturing Process - *J. Blazewicz.*

O desenvolvimento de uma heurística consiste em criar, de forma indutiva, uma solução para um problema baseando-se em regras estabelecidas a partir do raciocínio de senso comum ou a partir de experiências. Em uma grande parcela de problemas, soluções baseadas em processos heurísticos são obtidas em um tempo de processamento menor que aquelas soluções obtidas sobre processos exatos, e tais soluções são muitas vezes próximas da ótima. Portanto, dado a sua própria natureza, as heurísticas nem sempre garantem o resultado ótimo, porém, a relação custo-benefício entre uma solução obtida rapidamente e a proximidade de uma boa solução justifica o uso do processo heurístico.

Este capítulo é constituído, em sua maior parte, pela descrição da concepção da principal contribuição desta tese, a Dimensão Fractal Fuzzy Significativa (DFFS), uma medida obtida a partir de um processo heurístico aplicado ao processo de cálculo da dimensão Box-Counting.

No intuito de expor a concepção desta idéia optou-se por, num primeiro momento, contextualizar o leitor sobre como o processo de cálculo da dimensão Box-Counting (o algoritmo Box-Counting) é influenciado pelo tipo de conjunto de dados submetido a ele, e em seguida introduzir a motivação para a concepção da nova medida, bem como, descrever o processo heurístico de seu cálculo. Ainda neste capítulo, é apresentada a modelagem de um sistema que automatiza esse processo heurístico e duas aplicações deste sistema: uma para validação do processo heurístico de análise da curva Box-Counting e outra para a verificação da eficiência do uso da medida DFFS como solução ao problema central desta tese.

5.1 Curva Box-Counting e seu relacionamento com um conjunto de dados.

A execução do algoritmo Box-Counting produz um gráfico do tipo *log/log*. As curvas geradas em gráficos desta natureza possuem um comportamento crescente, comumente se aproximando a uma assíntota em seus últimos pontos, podendo sofrer variações nas inclinações dos segmentos que a compõe. Na Figura 5.1 são apresentadas algumas formas que esta curva pode assumir no contexto das execuções do algoritmo Box-Counting.

Os valores que determinam cada ponto dessas curvas são obtidos em cada iteração do algoritmo Box-Counting, ou seja, a cada imposição da grade de hipercubos sobre o conjunto de dados. Desta forma, cada ponto da curva é uma forma de representação da distribuição dos pontos do conjunto, sob a ótica de observação (ou análise) proporcionada pela grade de hipercubos.

No eixo x desse gráfico representa-se a variável livre no processo: o tamanho dos lados dos hipercubos. Essa variável é plotada como o logaritmo de seu inverso, representando uma medida de precisão. No eixo y representa-se a variável dependente no processo: o número de hipercubos ocupados pelos pontos do conjunto de dados, também plotado como o seu logaritmo.

Do ponto de vista proporcionado pelo processo Box-Counting existe um aumento gradual do nível de detalhamento das observações do conjunto de dados. A diminuição do tamanho dos

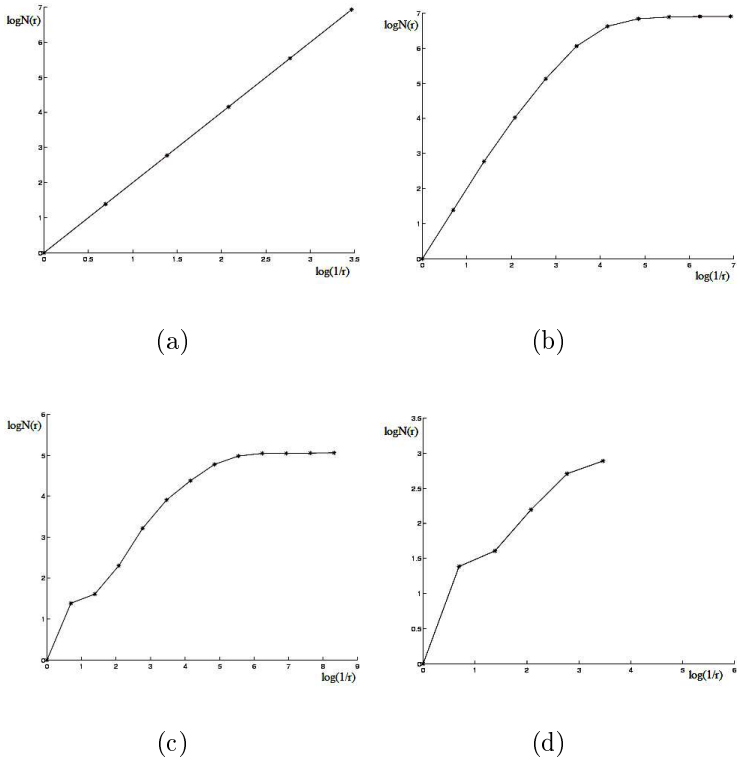


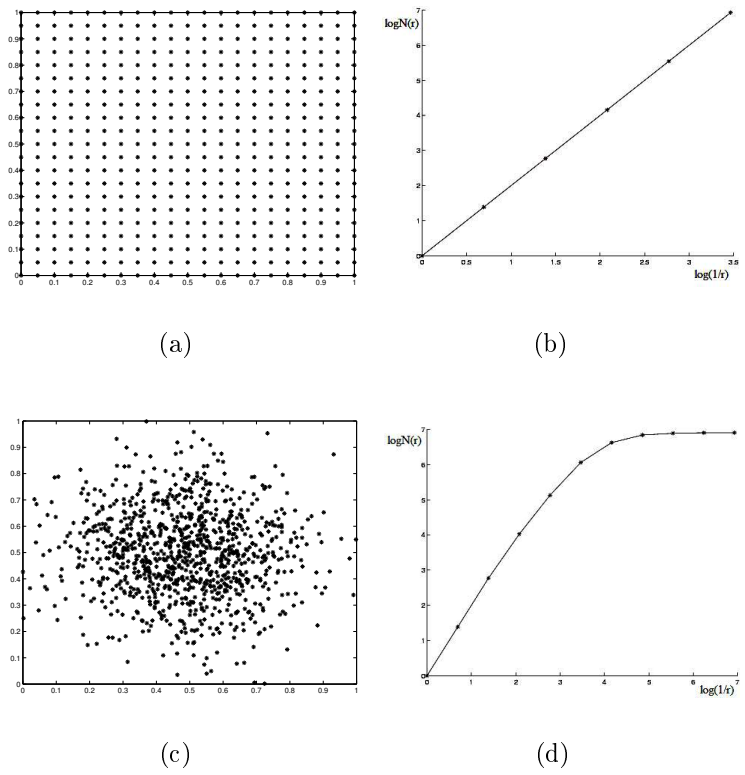
Fig. 5.1: Variações da curva resultante do algoritmo Box-Counting.

hipercubos[1] realiza uma diminuição na granularidade da percepção dos dados no seu espaço, proporcionando uma análise cada vez mais precisa do conjunto de dados. Nesse contexto, com a informação sobre o número de hipercubos ocupados a cada iteração, é possível a observação de algumas características da distribuição dos pontos no espaço, pois a curva resultante do processo iterativo de análise sofre variações de menor ou maior intensidade, conforme o número de hipercubos ocupados sofre alterações leves ou fortes.

Para conjuntos de dados formados por distribuições diferentes, observa-se formações diferentes na curva Box-Counting. Para conjuntos de dados que possuem pontos uniformemente distribuídos no espaço, como mostrado na Figura 5.2(a), o comportamento da curva tende a ser como o de uma reta (Figura 5.2(b)). Variações no comportamento da curva indicam que existem um ou mais pontos que não obedecem à distribuição uniforme. Este fato pode ser verificado na Figura 5.2(d) onde está representada a curva resultante da aplicação do algoritmo Box-Counting

[1]Seguindo a orientação da literatura de Teoria de Fractais, esta diminuição se deu, para todos os exemplos deste trabalho, na ordem de $\frac{x}{2}$, com $x_0 = 1$.

sobre um conjunto de dados formado por uma distribuição normal (Figura 5.2(c)).



(a) (b)

(c) (d)

Fig. 5.2: Curvas resultantes da aplicação do algoritmo Box-Counting em conjuntos com diferentes distribuições de dados: (a) e (b) Dados distribuídos uniformemente; (c) e (d) Dados distribuídos normalmente.

Na Figura 5.3 observa-se um exemplo didático, no qual o comportamento das curvas geradas varia de acordo com as mudanças nas distribuições analisadas. Originalmente o conjunto de dados está distribuído de tal forma que a curva gerada tem um comportamento considerado "padrão" para o contexto deste trabalho (observe a Figura 5.3(a)).

No segundo caso (Figura 5.3(b)) foram inseridos três novos pontos, com localizações relativamente distantes dos demais, fazendo com que os pontos originais apareçam agrupados no canto inferior esquerdo do gráfico. Nesse caso, a curva resultante sofre uma alteração em sua forma (em relação ao comportamento "padrão"), a qual reflete o grande espaçamento existente entre os pontos originais e os novos pontos. A mudança no comportamento da função foi determinada na terceira iteração do algoritmo Box-Counting. As alterações sofridas pela curva da Figura 5.3(c) são provenientes de alterações de natureza diferente da anterior no conjunto de da-

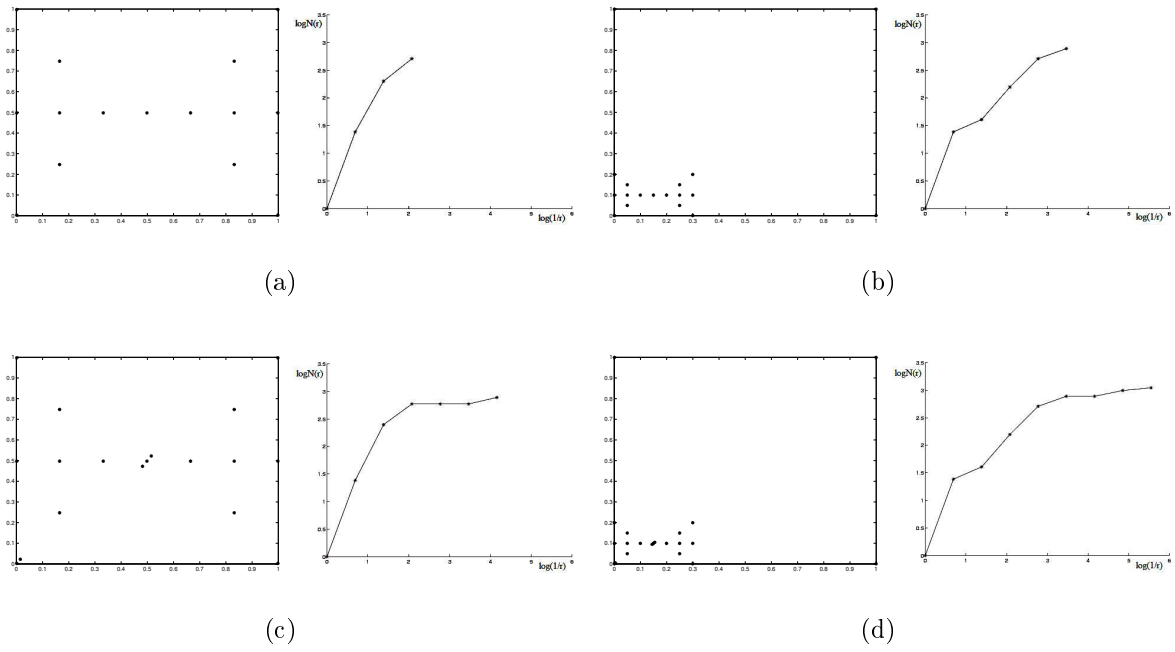


Fig. 5.3: As curvas resultantes da aplicação do algoritmo Box-Counting sobre conjuntos de dados modificados. (a) Conjunto de dados original (KLIR; YUAN (1995)); (b),(c) e (d): Conjunto de dados modificados.

dos original. Nesse exemplo foram inseridos três novos pontos, com localizações bem próximas a pontos já existentes. Dois dos novos pontos estão localizados próximos a um ponto central e o terceiro está localizado próximo ao ponto do canto inferior esquerdo. A curva resultante sofreu alterações a partir da quarta iteração do algoritmo Box-Counting. Na Figura 5.3(d) estão representadas as duas alterações descritas acima. A curva resultante é a combinação das duas anteriores (Figuras 5.3(b) e 5.3(d)).

A partir desses exemplos algumas observações podem ser realizadas:

- a proximidade dos pontos dentro do conjunto de dados influencia no número de iterações do algoritmo Box-Counting e, conseqüentemente, contribui para que a medida de dimensão obtida ao final do processo seja influenciada por estes pontos;

- a distribuição uniforme ou aleatória leva à geração de uma curva com um comportamento próximo a uma reta, aproximando a medida de dimensão Box-Counting à medida de dimensão Topológica do conjunto de dados sob análise;

- a distribuição normal leva à geração de uma curva com comportamento constante ou assintótico em seus últimos pontos;
- um grande distanciamento entre pontos ou conjunto de pontos faz com que a curva apresente um segmento de reta (delimitado por dois pontos consecutivos na curva) com uma inclinação "baixa" em relação ao que se espera em um comportamento "padrão" (veja a curva da Figura 5.3(b) nos segmentos delimitados pelos 2^{o}, 3^{o} e 4^{o} pontos[2]).
- uma grande aproximação entre pontos ou conjunto de pontos faz com que a curva apresente um segmento de reta com uma inclinação "alta" em relação ao que se espera em um comportamento padrão (veja a curva da Figura 5.3(b) nos segmentos delimitados pelos 3^{o}, 4^{o} e 5^{o} pontos[3]).

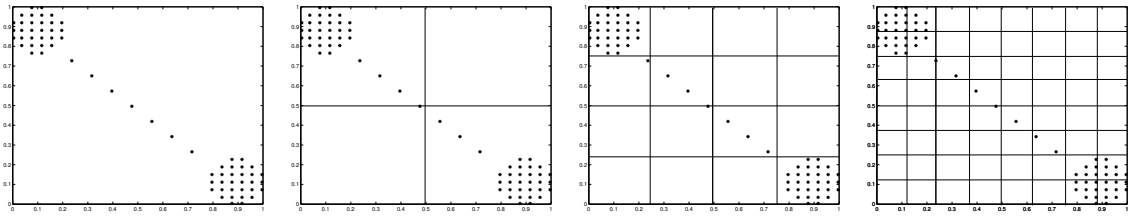
Na Figura 5.4 é ilustrada a imposição da grade de hipercubos sobre um conjunto de dados (YAGER; FILEV (1994)) e é informado o número de hipercubos ocupados a cada iteração. Visualmente é possível perceber que existe um distanciamento entre dois grupos de pontos: o grupo no canto superior esquerdo e o grupo no canto inferior direito. Da terceira para a quarta iteração observa-se uma diferença na ocupação dos hipercubos que insere um comportamento "côncavo" nos segundo e terceiro segmentos de reta da curva resultante (Figura 5.5). A partir da quarta iteração não se percebe, numa análise visual da curva resultante, outro ponto onde o comportamento da curva foge do "padrão".

Baseado nos fatos expostos acima propõe-se a exploração da curva gerada pelo algoritmo Box-Counting para descobrir conhecimento sobre um conjunto de dados, que possa ser utilizado como *conhecimento a priori* por algoritmos e técnicas de análise de dados, e essa proposta será discutida ainda neste capítulo. Outros trabalhos propuseram usar a dimensão Fractal (ou alguma de suas variações como é o caso da dimensão Box-Counting) ou o seu processo de cálculo para auxiliar em tarefas de análise de dados.

Em BARBARá; CHEN (2002) e BARBARá; CHEN (2003) os autores descrevem um algoritmo para descoberta de agrupamentos em conjuntos de dados que utiliza a informação sobre a

[2]Da segunda para a terceira iteração existe pouca variação no número de hipercubos ocupados e da terceira para a quarta iteração a variação é grande.

[3]Da quarta para a quinta iteração existe pouca variação no número de hipercubos ocupados, enquanto que entre as iterações anteriores (terceira para quarta iteração) a variabilidade é grande.



(a) (b) (c) (d)

Fig. 5.4: As quatro primeiras fases do algoritmo Box-Counting. (a) Primeira fase: 1 de 1 hipercubo ocupado (100% de ocupação); (b) Segunda fase: 2 de 4 hipercubos ocupados (50% de ocupação); (c) Terceira fase: 5 de 16 hipercubos ocupados (31, 25% de ocupação); (d) Quarta fase: 12 de 64 hipercubos ocupados (18, 75% de ocupação).

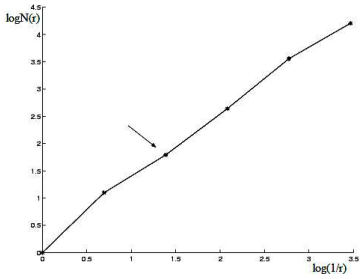


Fig. 5.5: Curva resultante da execução do algoritmo *Box-Counting* sobre o conjunto de dados da Figura 5.4.

dimensão Fractal do conjunto para decidir sobre a formação de agrupamentos. Basicamente, a consideração de que pontos que pertencem a um mesmo agrupamento possuem um grau maior de auto-similaridade do que pontos pertencentes a agrupamentos diferentes forma a base desta aplicação da Teoria de Fractais. O algoritmo proposto por esses autores utiliza a idéia de que pontos de um agrupamento se encontram dispostos de tal maneira que nenhum deles influencia radicalmente a medida da dimensão Fractal do agrupamento a que pertence.

Num primeiro momento, os autores utilizam um algoritmo clássico que determina a primeira estimativa de configuração de agrupamentos no conjunto. Esta inicialização é implementada como um procedimento simples de descoberta de agrupamentos, baseado na escolha randômica de alguns pontos e na concatenação de outros pontos localizados a uma distância mínima pré-determinada (abordagem dos k-vizinhos mais próximos) dessses pontos. Para cada um dos conjuntos de pontos formados na primeira etapa é medida a dimensão Fractal. Em seguida, o

algoritmo, denominado Fractal Clustering, simula a concatenação dos pontos ainda não agrupados, aos grupos já existentes, calculando a nova dimensão Fractal de cada um dos conjuntos. O conjunto onde o ponto causa o menor "impacto fractal", isto é, a mudança mínima entre a dimensão Fractal anterior do grupo e a atual, recebe efetivamente o ponto se o impacto for menor que um limite previamento estabelecido. O resultado de agrupamentos obtido ao término do processo pode diferir totalmente da configuração inicial de grupos, já que o algoritmo é capaz de quebrar ou concatenar conjuntos utilizando operações específicas (veja BARBARá; CHEN (2003)).

Essa técnica, segundo os autores, tem uma boa escalabilidade em relação ao número de pontos e à dimensão Topológica do conjunto de dados, conseguindo alcançar bons resultados mesmo com a existência de ruídos e de agrupamentos de formas arbitrárias.

Em BARBARá; NAZERI (2000), a informação de dimensão Fractal é utilizada para resolver o problema de partição de conjuntos de dados. Esta abordagem segue a linha da abordagem descrita em BARBARá; CHEN (2003). Os autores são motivados pela dificuldade em encontrar regras de associação, em tarefas de mineração de dados, quando o conjunto de dados sob análise apresenta-se inadequadamente particionado por procedimenos arbitrários ou por procedimentos que o particiona em intervalos iguais. A idéia central dessa abordagem é modificar o algoritmo *Apriori* (AGRAWAL; IMIELINSKI; SWAMI (1993) e AGRAWAL; SRIKANT (1994)) de forma que enquanto o algoritmo está computando o suporte de um intervalo para um conjunto de itens de forma categórica (por atributo do item), a dimensão Fractal do conjunto multidimensional seja computada incrementalmente em relação aos itens envolvidos. Assim, as medidas de dimensão Fractal obtidas a cada inserção de um item no conjunto ajudam a decidir sobre os pontos ideais de particionamento. Mais especificamente, uma mudança drástica na medida da dimensão Fractal indica que um elemento que não deve pertencer ao intervalo foi analisado e caracteriza uma boa indicação de uma nova partição. Outros trabalhos dos mesmos autores utilizam a dimensão Fractal na resolução de problemas de análise de dados, veja: BARBARá; WU (2003a), BARBARá; WU (2003b) e BARBARá et al. (2002).

Os autores dos trabalhos TRAINA (2001), TRAINA; TRAINA; FALOUSTOS (1999) e TRAINA et al. (2000) direcionam seus esforços na tentativa de utilizar o conhecimento da

dimensão Fractal, em sua variação Box-Counting, para decidir sobre a seleção de atributos importantes para a representação das informações contidas em um conjunto de dados.

Em TRAINA (2001) e TRAINA; TRAINA; FALOUSTOS (1999) os autores utilizam os recursos da Teoria de Fractais, mais especificamente os conceitos relacionados a dimensão Fractal e auto-similaridade, para estimar a seletividade de consultas por similaridades em Sistemas Gerenciadores de Banco de Dados e para a criação de árvores de indexação "M-trees". No primeiro caso, a medida de dimensão Fractal atua como uma medida de fator de similaridade capaz de estimar o número de tuplas que satisfazem uma dada condição de seleção; no segundo caso, agrupamentos revelados pelas relações de similaridade apontadas pelo cálculo incremental da dimensão Fractal são usados como critérios de divisão de nós em árvores de indexação.

Especificamente em TRAINA et al. (2000), a medida de dimensão Fractal do conjunto de dados original é comparada com a mesma medida calculada sobre o mesmo conjunto de dados com a retirada de um ou mais atributos. A observação de alterações leves (dentro de um limite pré-estabelecido) na dimensão Fractal então obtida, significa que os atributos retirados não afetam o comportamento do conjunto de dados ou das informações que ele descreve (mantendo suas características essenciais). Segundo os autores, a dimensão Fractal de um conjunto de dados não é fortemente afetada por atributos redundantes, e sempre que a contribuição de um atributo é pequena no cômputo da dimensão, ele pode ser removido. Esse estudo é motivado principalmente pela necessidade da obtenção de redução de dimensão sem perda significativa de informação por meio de uma abordagem de eliminação de atributos do conjunto de dados, e apresenta algumas propriedades desejáveis como: a independência de rotação de atributos, ou seja, a técnica proposta não realiza qualquer tipo de movimentação dos dados, o que poderia vir a dificultar a análise direta dos resultados; a capacidade de descobrir dados que possuem correlações não-lineares; a escalabilidade do processo envolvido; a capacidade de estimar quantos atributos devem ser mantidos no conjunto.

Ainda outros trabalhos foram desenvolvidos pelo grupo de pesquisa onde tais autores estão inseridos, relacionando a dimensão Fractal de uma estrutura a: ferramentas para mineração de dados multidimensionais (TRAINA; TRAINA; PAPDIMITRIOU (2001)) e outras aplicações (SOUSA; TRAINA; TRAINA (2005)); geração de dados sintéticos (FERREIRA; BUENO;

TRAINA (2005)); análise de dados "*stream*" (SOUSA; TRAINA; TRAINA (2003)); otimização de consultas em Sistemas Gerenciadores de Banco de Dados (ARANTES; TRAINA; TRAINA (2003)).

5.2 Dimensão Fractal Fuzzy Significativa

Explorar a curva gerada pelo algoritmo Box-Counting, com o intuito de descobrir conhecimento sobre um conjunto de dados, vem ao encontro à hipótese formulada neste trabalho: a ação de encontrar a dimensão ideal para o mapa de saída de uma RNA-AO está diretamente ligada à de encontrar a real dimensão onde as relações entre os dados tomam lugar.

A informação obtida a partir da aplicação de uma RNA-AO está intimamente relacionada às similaridades entre os pontos do conjunto de dados. Estas relações de similaridade obedecem a uma métrica específica, possuem relações de vizinhança que caracterizam uma topografia específica e podem estar organizadas em agrupamentos. O mapeamento topográfico (com ou sem redução de dimensão) realizado pela RNA-AO deve preservar a topologia que está presente no conjunto de dados, ou seja, deve refletir, o quanto for possível, as relações de vizinhança do espaço de entrada no espaço de saída do mapeamento.

Entretanto, a dimensão do espaço de saída de uma RNA-AO deve, idealmente, ser estabelecida com o objetivo de reduzir a dimensão do espaço de entrada, propiciando a construção de um mapeamento topográfico que seja mais suscetível à análises e que preserve ao máximo as características do conjunto de dados. Nesse contexto surgem dois problemas cujas soluções parecem antagônicas:

- Quanto maior for a dimensão do espaço de saída do mapeamento, maior será o esforço necessário para a sua concepção e análise. A alta dimensão impede análises visuais diretas dos vetores (ou neurônios) e de seus relacionamentos. Mesmo com a disponibilidade de técnicas que oferecem meios de análise de dados multi-dimensional, a alta dimensão do espaço analisado exigirá alto esforço computacional;

- Quanto menor for a dimensão do espaço de saída do mapeamento em relação ao espaço dos dados sob análise, maior será a dificuldade de representar, no espaço de saída, as relações

de vizinhança do espaço de entrada. O processo de auto organização da RNA-AO realiza muitas contorções no mapa e incorre, muitas vezes, em representações topográficas falsas.

Observe a Figura 5.6, na qual um conjunto de dados inserido num espaço 2-dimensional é mapeado por duas RNAs-AO. A RNA-AO 1-dimensional realiza um mapeamento mais simples (em termos de esforço computacional e de análise visual direta) que o mapeamento realizado pela RNA-AO 2-dimensional. No entanto, as relações de vizinhança do conjunto de dados estão melhor representadas no mapeamento de dimensão mais alta.

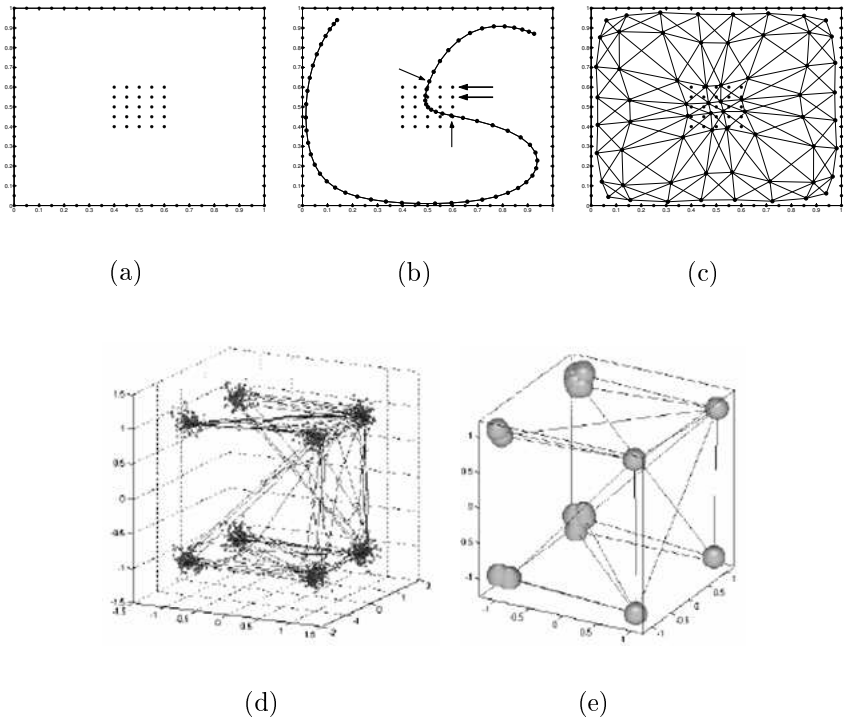


Fig. 5.6: Exemplo de distorção topológica. (a) Conjunto de dados de teste 2-dimensional; (b) Mapeamento 1-dimensional; (c) Mapeamento 2-dimensional; (d) Conjunto de dados de teste 3-dimensional; (e) Mapeamento 2-dimensional.

Na Figura 5.6(b) é mostrado o problema de distorção de topologia. Observe que existem dois pontos vizinhos no conjunto de dados (apontados pelas flechas paralelas), que não tem representantes vizinhos no mapeamento (apontado pelas outras duas flechas), ilustrando a necessidade da contorção do mapeamento para a realização da aproximação. Esse problema alcança maiores proporções conforme o estendemos para espaços de dimensões mais altas.

Nas Figuras 5.6(d) e (e) está ilustrada uma situação onde um mapeamento 2-dimensional não consegue representar todas as relações de vizinhança existentes entre os grupos de pontos localizados em um espaço 3-dimensional. Outros problemas de distorção de topologia são descritos em BAUER; HERRMANN; VILLMANN (1999).

A possível solução para este problema está na minimização de perdas: encontrar uma medida de dimensão que minimize o esforço requerido na interpretação do mapeamento, mas que também permita a maximização da qualidade das informações nele representadas.

Considerando que um objeto, representado por um conjunto de dados, apesar de ser residente em um espaço N-dimensional, pode possuir uma dimensão real (M) menor ou igual à desse espaço, infere-se que suas características topológicas sejam residentes em um espaço M-dimensional. Assim, supôs-se factível usar esta informação para inspecionar a dimensão ideal da camada de saída de uma RNA-AO.

A medida da dimensão Box-Counting de um conjunto de dados é uma forma de análise da sua complexidade, de medição de sua real dimensão, ou ainda uma forma de dimensionar qual a real porção do espaço que é efetivamente ocupado por ele. Logo, a utilização desta medida como um guia heurístico para a automação da decisão sobre o parâmetro de projeto da RNA-AO discutido (a dimensão do espaço de saída), apresenta-se como uma solução promissora ao dilema formulado previamente.

Uma linha de raciocínio similar à discutida aqui é seguida por Speckmann, Raddatz e Rosenstiel. Em SPECKMANN; RADDATZ; ROSENSTIEL (1994a) e SPECKMANN; RADDATZ; ROSENSTIEL (1994b) os autores desenvolvem um método para melhorar o resultado do aprendizado de uma RNA-AO, testando-a e avaliando o seu desenvolvimento qualitativo durante o processo de aprendizado. A dimensão Fractal dos dados é utilizada na otimização desse processo, no sentido de melhorar a sua habilidade de preservação de topologia da RNA-AO. O aprendizado é monitorado por meio do cálculo da dimensão Fractal do espaço de saída do mapeamento, a fim de estabelecer a sua qualidade em cada fase do processo de aprendizado e determinar um critério de parada para ele.

Para ilustrar o seu método, os autores utilizaram dois conjuntos de dados (um homogêneo e outro heterogêneo, sendo este último formado por quatro agrupamentos). A dimensão Fractal

dos conjuntos foi calculada por meio de diferentes processos (resultando em diferentes medidas para a dimensão Fractal). Para o conjunto homogêneo a dimensão Fractal encontrada foi um pouco menor do que 2, apresentando uma pequena variação em relação aos tipos de dimensão Fractal considerados. A análise dos diferentes tipos de dimensão Fractal para o conjunto heterogêneo mostrou que sua dimensão fractal não chegava a 3. Assim, uma RNA-AO com espaço de saída 2-dimensional foi utilizada para o aprendizado nos dois casos por considerar-se que ela poderia convergir para um mapeamento sem defeitos topológicos.

Os autores identificaram a convergência das medidas de dimensão Fractal do mapa para a medida de dimensão Fractal do conjunto de dados em estágios avançados do treinamento. Entretanto, em experimentos com conjuntos de dados reais e de dimensão Fractal mais alta, os autores perceberam que a aproximação da dimensão Fractal do mapeamento à do conjunto de dados era obtida apenas mediante alguma perda na preservação de topologia.

Outra constatação dos autores foi que mesmo mapas com diferentes dimensões topológicas para o espaço de saída, quando aplicados a um mesmo conjunto de dados, convergem para uma configuração tal que suas dimensões Fractais se aproximam a uma mesma medida. Entretanto, as configurações onde estes mapas encontram uma boa preservação de topologia acontece em fases anteriores ao alcance de tais medidas de dimensões Fractais. A insistência no treinamento, a partir desse ponto, inicia um processo de distorção topológica que converge para resultados ruins de preservação de topologia, sendo então necessário monitorar os dois aspectos: a dimensão do espaço de saída do mapa e a evolução das características topológicas durante o treinamento.

Assim, os autores passaram a monitorar os dois aspectos citados, variando a largura da função de adaptação nos momentos anteriores àquele onde o mapa inicia a distorção topológica. Desta forma foram obtidos melhores resultados de preservação topológica e de aproximação de medidas de dimensão Fractal, independentemente da dimensão do espaço de saída.

Isto posto, nesta tese advoga-se que o uso da medida de dimensão Fractal em sua forma original pode não ser adequado para alguns contextos. Dado que o resultado final de um mapeamento topográfico e quantizador realizado por uma RNA-AO é a descoberta de agrupamentos e, visto que agrupamentos são formados por dados que possuem relações de similaridade e que

estas são atestadas pela proximidade dos dados no seu espaço de distribuição (JAIN; DUBES (1988)), supôs-se também que a medida de dimensão Fractal obtida pelo processo Box-Counting não corresponde, necessariamente, à medida da real dimensão do conjunto de dados quando visto como um conjunto de agrupamentos. Além disso, como se está interessado em mapear as relações de vizinhança topológica existentes entre os agrupamentos, a análise destes como um único ponto não prejudicaria o processo, ao contrário, o tornaria mais simples já que não interessa aqui as relações intra-grupos.

Neste sentido, para analisar a topologia do conjunto de dados representado na Figura 5.3 dever-se-ia desconsiderar a influência exercida pela proximidade dos três pontos centrais sobre a formação da curva Box-Counting (geração dos pontos finais da curva) e conseqüentemente sobre a medida de dimensão resultante. Parar o processo de análise dos dados na iteração do algoritmo que isola estes três pontos dentro de um hipercubo faria com que a curva fosse "truncada" e a medida de dimensão obtida estaria refletindo uma análise menos específica dos dados. Ou seja, está-se supondo que, para fins de análises de agrupamentos, poder-se-ia desconsiderar a influência sofrida pela medida por conta da proximidade dos dados integrantes de grupos, tomando-os como um único ponto.

A questão então seria determinar em que momento da execução do algoritmo Box-Counting a proximidade dos pontos passa a influenciar "negativamente" a análise dos dados, ou seja, em que ponto do processo a análise proporcionada pela grade de hipercubos já seria suficiente para identificar as dimensões onde estão as relações topológicas entre os grupos. O trabalho de encontrar este ponto foi realizado por meio de um estudo do comportamento da curva Box-Counting obtida pela execução completa do algoritmo homônimo. Percebeu-se que a existência de agrupamentos atestada por tipos específicos de comportamento dessa curva seria um indicativo de que a execução até aquela iteração seria suficiente para atestar a, então chamada, Dimensão Fractal Fuzzy Significativa ("significativa" para fins de análise de agrupamentos) do conjunto de dados.

Os estudos sobre a relação do comportamento da distribuição dos dados no conjunto com a forma da curva Box-Counting proporcionou a criação do novo critério de parada para o algoritmo Box-Counting. Na realidade, a análise desta curva nos fornece informações sobre o

tipo de distribuição do conjunto de dados observada a cada iteração do algoritmo e, quando a distribuição observada indica a dificuldade em separar os pontos em novos hipercubos estar-se-ia no momento de parar o processo. Seguiu-se então para o estabelecimento de uma abordagem para a realização dessa análise, que proporcionasse uma interpretação sistemática da curva Box-Counting.

5.2.1 Análise da curva Box-Counting

Partindo das observações exploradas na seção 5.1, chegou-se à conclusão da necessidade de, durante o processo de execução do algoritmo Box-Counting, ou seja, durante a construção da curva Box-Counting, armazenar informações sobre os segmentos de reta que a compõem, no contexto de sua formação, para que se possa proceder sua análise. De modo mais específico, cada iteração do processo acrescenta um ponto no gráfico log/log e a ligação desse com o ponto referente a iteração posterior forma o segmento ao qual se está fazendo referência. A análise desses segmentos permite interpretar o conjunto de dados sob os vários níveis de observação que as diferentes granularidades de hipercubos proporcionam, gerando uma seqüência de conclusões sobre a distribuição dos dados.

As informações pertinentes a este contexto são a diferença de inclinação, em relação ao eixo x, entre os dois segmentos de reta adjacentes a cada ponto (variável: "dif-seg") e o valor da inclinação do segundo segmento deste par (também em relação ao eixo x) (variável: "inc-2-seg").

Para definir o cálculo dessas duas variáveis, considere uma curva formada por segmentos de reta $seg(p_1, p_2)$, onde p_1 e p_2 são os pontos que delimitam o segmento, que unem no gráfico log/log, os p_i pontos obtidos na execução das I iterações do algoritmo Box-Counting. Considere também que αseg denota a inclinação deste segmento.

As variáveis "dif-seg" e "inc-2-seg" são vetores, com $J = I - 2$ posições, valoradas de acordo com 5.1 e 5.2, respectivamente:

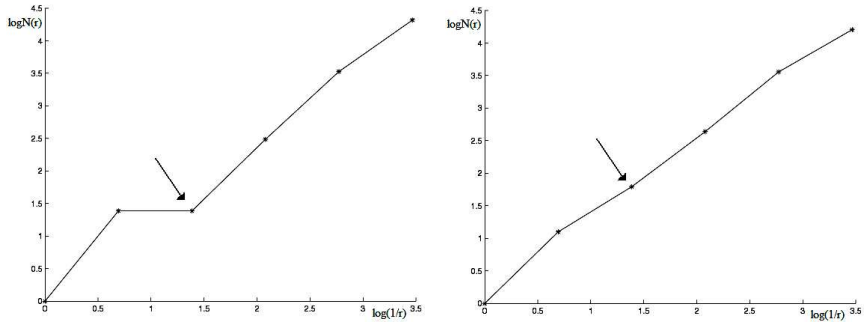
$$\text{dif-seg } [j] = \alpha seg\ (p_i, p_{i+1}) - \alpha seg\ (p_{i+1}, p_{i+2}) \tag{5.1}$$

$$\text{inc-2-seg } [j] = \alpha seg\ (p_{i+1}, p_{i+2}) \tag{5.2}$$

em que i inicia em 1 e j varia de 1 até J. Note que a célula j desses vetores refletem uma análise referente à análise do ponto p_{i+1} da curva Box-Counting.

A variável "dif-seg" tem o papel de verificar, dentro da análise heurística da curva, como se dá a variação de ocupação dos hipercubos de uma iteração do algoritmo em relação às iterações anterior e posterior. Essa variável pode assumir um valor dentro de três intervalos: < 0, 0 ou > 0, ou seja, uma diferença de inclinação "negativa", "neutra" ou "positiva", respectivamente[4].

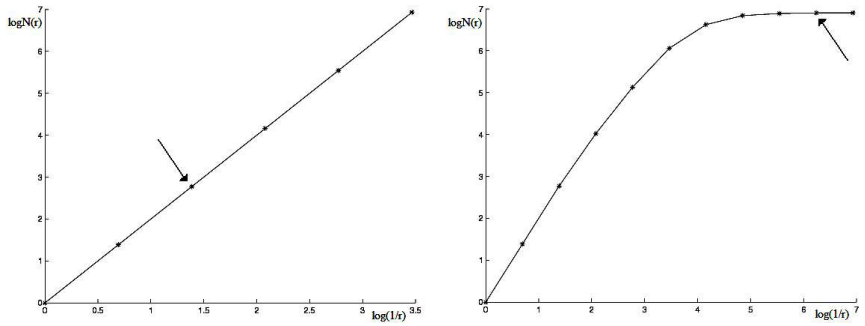
Para entender melhor o papel da variável "dif-seg", observe a Figura 5.7, onde estão destacados dois pontos (ou iterações) em que a referida variável assume um valor negativo. Note que a inclinação do primeiro segmento adjacente ao ponto em destaque é menor que a inclinação do segundo segmento e tal fato revela um aumento, "em algum grau", no número de hipercubos ocupados[5] na iteração que determina a inclinação do segundo segmento sob análise. Essa variação revela a existência de um distanciamento entre pontos ou grupos de pontos. Quando o valor dessa variável se mantém próximo a 0, como pode ser observado na Figura 5.8, significa que o número de hipercubos ocupados se manteve na mesma proporção nas duas iterações. Finalmente, como mostrado na Figura 5.9, o valor "positivo" assumido por essa variável indica que houve um aumento no número de hipercubos existentes mas a taxa de ocupação não acompanhou o aumento. A partir desta última constatação infere-se que existem pontos tão próximos um do outro que seria necessário mais de uma iteração do processo para separá-los.



(a) (b)

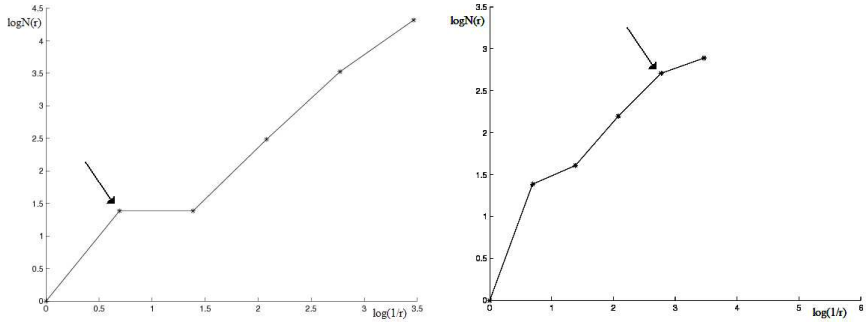
Fig. 5.7: Caso de variação negativa para a variável "dif-seg".

[4]A modelagem desses intervalos será discutida mais a frente, neste capítulo.
[5]Número de hipercubos ocupados em relação ao ao número de hipercubos existentes na iteração.



(a) (b)

Fig. 5.8: Caso de variação neutra para a variável "dif-seg".

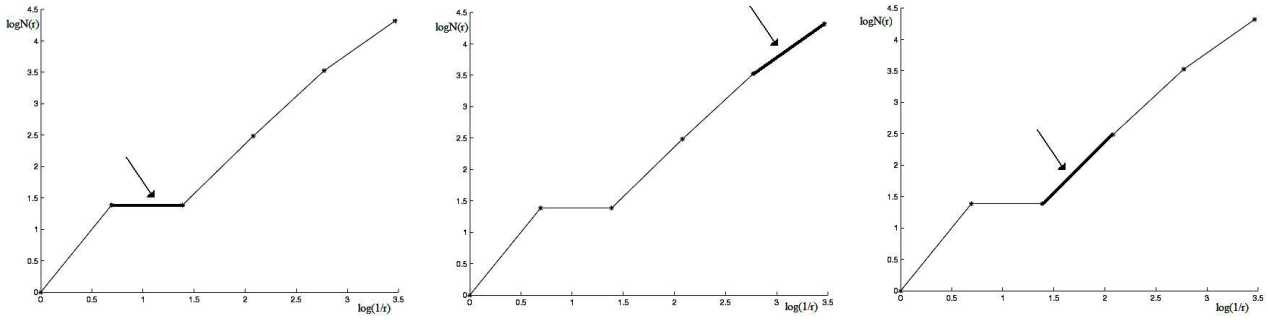


(a) (b)

Fig. 5.9: Caso de variação positiva para a variável "dif-seg".

Contudo, se essas informações indicam ou não algum comportamento específico da distribuição dos dados, só se saberá com a análise da outra variável envolvida no processo. A variável - "inc-2-seg" - tem o papel de analisar se a inclinação do segundo segmento do par em questão é forte o suficiente para definir a inferência de um tipo específico de distribuição no conjunto de dados. Tal variável pode assumir um valor de inclinação considerado "alto", "médio" ou "baixo". Observe na Figura 5.10 três exemplos onde ocorrem: (a) inclinação "baixa", (b) inclinação "média" e (c) inclinação "alta".

A análise conjunta de como se comportam essas duas variáveis permite inferir (para cada granularidade de grade de hipercubos) se os dados estão dispostos em uma de três possibilidades: distribuição uniforme (onde os dados estariam, na situação mais uniforme, ocupando

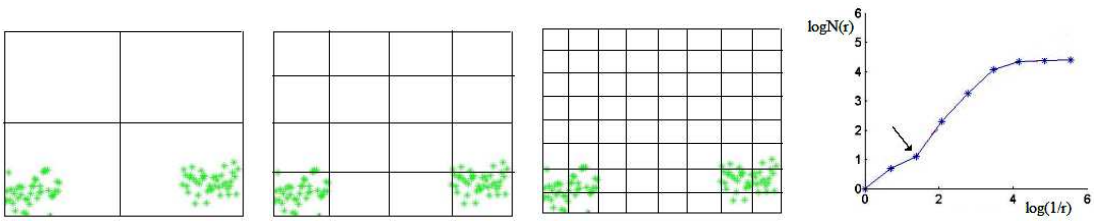


(a) (b) (c)

Fig. 5.10: Caso de variação positiva para a variável "inc-2-seg".

todos os hipercubos da grade em qualquer granularidade); distribuição normal (onde os dados estariam concentrados em uma região); distribuição com agrupamentos (onde os dados estariam concentrados em regiões diferentes) e neste caso uma variação na intensidade da concentração poderia apenas indicar uma aproximação de alguns poucos pontos. Essa análise é feita a partir das seguintes observações:

1. agrupamentos são observados quando a taxa de ocupação de hipercubos aumenta "muito" de uma iteração para outra. O aumento é constatado por $dif - seg = negativa$ porém, a intensidade do aumento é constatada pela $inc - 2 - seg = alta$. Veja a Figura 5.11;

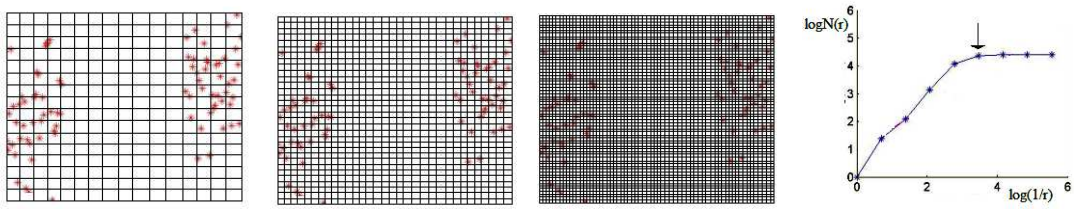


(a) (b) (c) (d)

Fig. 5.11: Iterações do algoritmo Box-Counting, e gráfico resultante, que constata a existência de agrupamentos. (a), (b) e (c) são as iterações do algoritmo e (d) o gráfico resultante com destaque para os pontos referentes às iterações destacadas nos quadros. A iteração (a) é representada pelo segundo ponto na curva; a iteração (b) é representada pelo terceiro ponto na curva; a iteração (c) é representada pelo quarto ponto na curva.

2. a aproximação de alguns poucos pontos é revelada pela ocorrência de uma diferença de

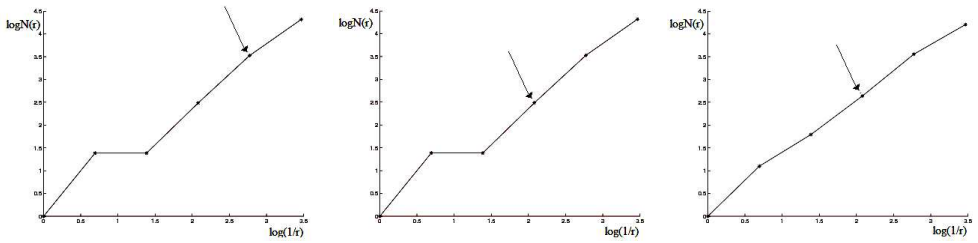
inclinação "positiva" acompanhada de uma inclinação "baixa" para o segundo segmento do par analisado, ou seja, a evolução do aumento de hipercubos ocupados é muito pequena (o processo precisa iterar mais para separar os pontos). Veja a Figura 5.12;



(a) (b) (c) (d)

Fig. 5.12: Iterações do algoritmo Box-Counting, e gráfico resultante, que constata a existência de aproximação entre pontos: (a), (b) e (c) são as iterações do algoritmo e (d) o gráfico resultante com destaque para os pontos referentes às iterações destacadas nos quadros. A iteração (a) é representada pelo quinto ponto na curva; a iteração (b) é representada pelo sexto ponto na curva; a iteração (c) é representada pelo sétimo ponto na curva.

3. as combinações que revelam que a disposição dos dois segmentos se aproxima de uma reta (crescente) revelam um comportamento de ocupação de hipercubos uniforme. São elas: (a)$dif - seg = positiva$ e $inc - 2 - seg = média$ ou (b)$alta$ e (c)$dif - seg = neutra$ e $inc - 2 - seg = alta$. Essas situações são mostradas nos gráficos da Figura 5.13;

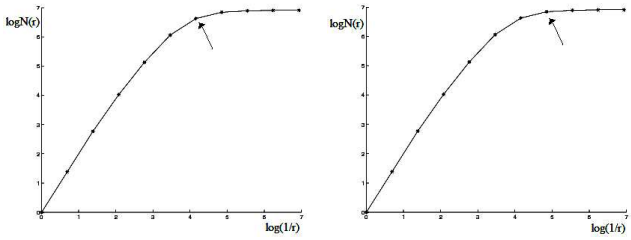


(a) (b) (c)

Fig. 5.13: Situações que ilustram a ocupação "uniforme" de hipercubos.

4. as combinações que revelam um pequeno aumento na ocupação de hipercubos ou um comportamento constante na ocupação desses caracteriza uma situação semelhante à encontrada na análise de conjuntos compostos por uma distribuição normal. São elas

(a)$dif - seg = negativa$ e $inc - 2 - seg = baixa$ ou $media$ e (b)$dif - seg = neutra$ e $inc - 2 - seg = baixa$ ou $media$. Veja alguns exemplos na Figura 5.14.



(a) (b)

Fig. 5.14: Situações que ilustram a ocupação "normal" de hipercubos.

Portanto tem-se que a análise da curva Box-Counting produz como resultado um vetor de "conclusões" com $J = I - 2$ posições que, a partir das observações realizadas sobre o relacionamento das variáveis "dif-seg" e "inc-2-seg", são valorados com as seguintes possibilidades: distribuição uniforme (**U**), distribuição normal (**N**), distribuição com agrupamentos (**A**) e aproximação de alguns poucos pontos (**P**). Assim, a variável "conclusões" é definida por 5.3

$$\text{conclusões } [j] \in \{U, N, A, P\}, \text{ com } j \text{ iniciando em 1.} \tag{5.3}$$

em que para cada j, uma regra do tipo "dif-seg $[j]$ AND inc-2-seg $[j] \rightarrow \{U, N, A, P\}$" é disparada. Neste contexto, as variáveis "dif-seg" e "inc-2-seg" se tornam premissas de regras lógicas conectadas pelo conector lógico AND. A especificação dessas regras é de acordo com as observações discutidas nesta seção e será descrita mais à frente neste capítulo. Note que a posição j do vetor "conclusões" reflete a observação realizada sobre o ponto p_{i+1} da curva Box-Counting (de acordo com as definições das variáveis envolvidas nas premissas das regras).

5.2.2 O ponto de corte e o cálculo final da Dimensão Fractal Fuzzy Significativa

Como já comentado neste texto, a DFFS é uma medida resultante da inserção de um novo critério de parada ao processo de cálculo da medida de dimensão Box-Counting. Assim como o

processo de análise da curva Box-Counting é baseado em experiências e observações, a decisão sobre o critério de parada também o é.

O critério original para encerramento da execução do algoritmo Box-Counting é a separação de todos os pontos do conjunto de dados sob análise em hipercubos distintos. Assim, quanto mais próximos os pontos estão uns dos outros, mais iterações do processo ocorrerão. Quanto mais iterações ocorrerem na análise de um conjunto de dados, mais pontos serão acrescentados à curva Box-Counting e menor será a inclinação da reta que aproxima a curva (resultando em uma dimensão Box-Counting menor).

A questão destacada neste trabalho é que a dimensão fractal seria capaz de medir a real porção do espaço ocupada pelo conjunto de dados e que é nesta porção que as relações de vizinhança dos dados devem tomar lugar. Logo, a dimensão fractal estaria indicando a dimensão onde um mapeamento que pretente analisar essas relações de vizinhança topológica deveria atuar. Juntamente a essa idéia está o fato de que, quando o mapeamento tem o objetivo de analisar grupos e suas relações, dever-se-ia considerar as relações de vizinhança entre grupos, sem a necessidade de estudar o que acontece entre os pontos desses grupos.

Neste contexto insere-se a idéia de usar a análise da curva Box-Counting, descrita na seção anterior, para inferir o ponto em que as relações intra-grupos começam a ser consideradas pelo algoritmo Box-Counting e aí então encerrar o seu processamento. A dimensão final calculada a partir da curva obtida nesse processo "truncado" é a Dimensão Fractal Fuzzy Significativa (DFFS).

O corte da curva Box-Counting, ou interrupção do processo do algoritmo Box-Counting, é realizado no ponto seguinte àquele gerado pela iteração em que a granularidade de observação dos dados proporcionou uma das três conclusões, nesta ordem de prioridade:

1. distribuição com agrupamentos;
2. distribuição com pontos próximos;
3. distribuição normal.

A razão para se realizar o corte no ponto seguinte, como mencionado, vem do desejo de analisar a curva em relação às observações que geraram a conclusão sobre a distribuição. Este

ponto, j_{corte}, é definido em (5.4).

$$j_{corte} = j + 2 \begin{cases} \text{se conclusões } [j] = A, \text{ senão} \\ \text{se conclusões } [j] = P, \text{ senão} \\ \text{se conclusões } [j] = N. \end{cases} \tag{5.4}$$

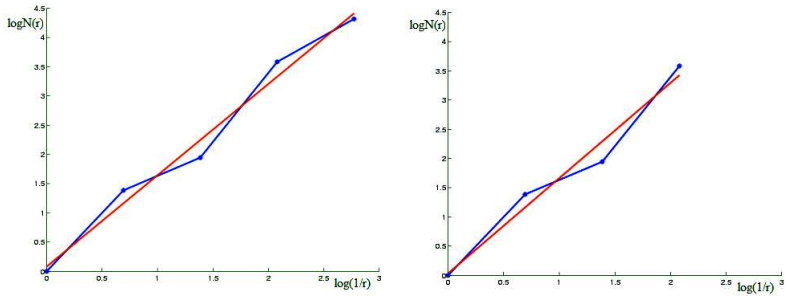
Assim, de forma similar ao teorema 2.6 define-se DFFS como em 5.5,

$$\text{DFFS } = \frac{Nn(\Lambda)}{(\frac{1}{2^n})} \ /n \text{ varia de 1 até } j_{corte}. \tag{5.5}$$

em que Λ é o atrator sob análise (ou seja, conjunto de dados sob análise), $Nn(\Lambda)$ é o número de hipercubos com lado de comprimento $(\frac{1}{2^n})$.

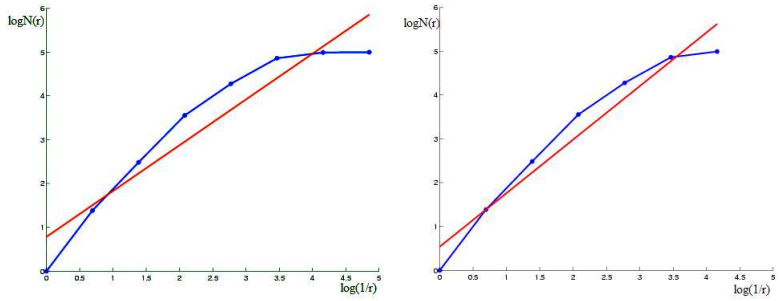
Essa ordem de prioridade foi estabelecida porque se a conclusão de agrupamentos foi obtida, em alguma granularidade de observação, é porque os grupos estão evidentes no conjunto sob essa iteração. Com um pouco menos de evidência descobre-se agrupamentos mais discretos originando a conclusão "distribuição com pontos aproximados". Por fim, quando nenhuma dessas duas conclusões forem obtidas, o início da aproximação da curva Box-Counting a uma função constante ou assintótica é considerado suficiente para análise do conjunto, dado que a partir desse ponto ocorreria a influência de grupos de pontos muito próximos na distribuição (como o grupo central de pontos de uma distribuição normal).

Para os casos onde só se observa o comportamento de distribuição uniforme, não se realiza qualquer corte na curva, dado que a análise da curva está mostrando que não existem grupos. Neste caso, a dimensão obtida como real porção do espaço ocupada pelos dados é muito próxima à dimensão do próprio espaço de distribuição dos dados. Assim, entende-se que as relações de vizinhança topológica estão distribuídas por todas as dimensões do espaço. Nas Figura 5.15, 5.16 e 5.17 são mostrados as três situações de corte discutidas acima.



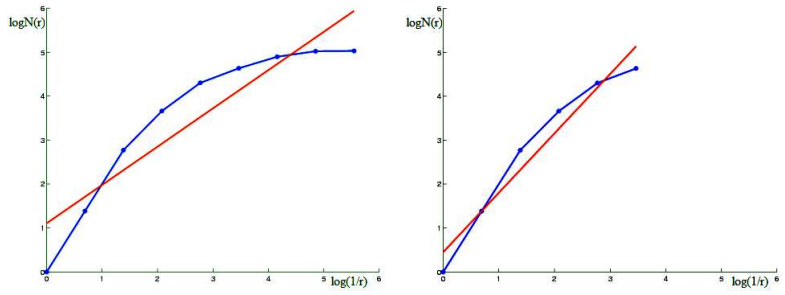
(a) (b)

Fig. 5.15: Situação que ilustram o corte da curva Box-Counting sob descoberta de agrupaentos (3ª iteração): (a) curva Box-Counting original e (b) curva Box-Counting truncada.



(a) (b)

Fig. 5.16: Situação que ilustra o corte da curva Box-Counting sob descoberta de pontos aproximados (7ª iteração): (a) curva Box-Counting original e (b) curva Box-Counting truncada.



(a) (b)

Fig. 5.17: Situação que ilustra o corte da curva Box-Counting sob descoberta de distribuição normal (5ª iteração): (a) curva Box-Counting original e (b) curva Box-Counting truncada.

5.3 Implementação para o cálculo da Dimensão Fractal Fuzzy Significativa

Como explicitado na seção anterior, as variáveis envolvidas no processo de cálculo da DFFS podem assumir valores que são classificados dentro de intervalos. Para automatizar o processo optou-se por dimensionar esses intervalos e os relacionar por meio de regras em um sistema de inferência.

A primeira tentativa de modelagem foi realizada dentro das possibilidades da lógica clássica. Para determinar os intervalos realizou-se um trabalho de inspeção humana, guiada por meio de um procedimento supervisionado. A supervisão à qual se refere aqui trata de um trabalho de calibragem dos limites dos intervalos, o qual foi realizado por meio da observação de curvas Box-Counting referentes a conjuntos de dados de distribuição conhecida. Três dos conjuntos utilizados neste procedimento e as respectivas curvas Box-Counting são mostrados na Figura 5.18. Note que esses conjuntos possuem as características que se deseja inferir na análise da curva com um grau de evidência alto.

Contudo, ao aplicar-se o modelo concebido ao problema de Tendência a Agrupamentos[6] não obteve-se resultados suficientemente acurados que justificassem o seu uso na resolução do problema principal desta tese. Assim, decidiu-se usar o Raciocínio Aproximado Fuzzy para "relaxar" a modelagem dos intervalos e regras de inferência, como uma tentativa de aumentar a acuidade dos resultados. Novamente utilizou-se um procedimento supervisionado, com os mesmos conjuntos de dados utilizados na modelagem anterior, para determinar os parâmetros das funções características das variáveis fuzzy em substituição aos intervalos *crisp*. O modelo fuzzy foi validado também por meio de sua aplicação no problema de Tendência a Agrupamentos e mostrou-se adequado.

A fim de tentar refinar o modelo fuzzy e de explorar mais possibilidades de parametrização, foi realizada uma tentativa de calibragem de parâmetros com o uso de Algoritmos Genéticos. A melhora obtida se deu apenas em termos dos pesos de duas das regras de inferência, contribuindo para algumas melhoras no resultado de validação do modelo. A seguir é apresentada

[6]Esse problema foi utilizado como forma de validar a verossimilhança dos modelos concebidos para determinação da DFFS.

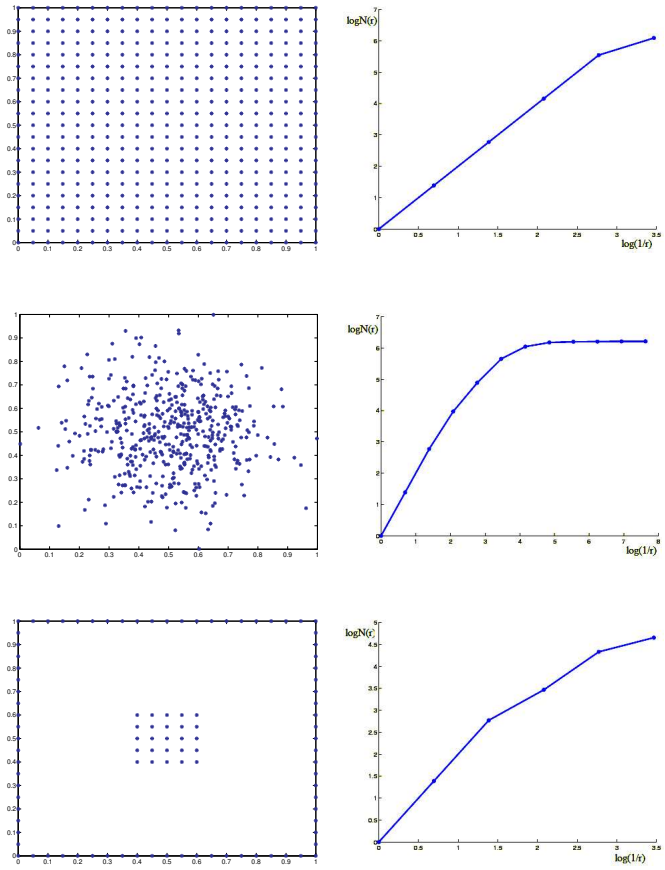


Fig. 5.18: Exemplos de conjuntos de dados com suas respectivas curvas Box-Counting, utilizados no processo de modelagem supervisionada dos limites dos intervalos para instanciação das variáveis "dif-seg" e "inc-2-seg".

a modelagem genética utilizada.

- O modelo do cromossomo: O cromossomo foi construído de modo a representar todos os parâmetros do sistema fuzzy. Ele foi dividido em quatro partes: uma para representar cada uma das três variáveis linguísticas e uma para representar as regras fuzzy (veja a Figura 5.19). Os parâmetros das funções de mapeamento representados no cromossomo são o tipo da função[7] e os parâmetros que a definem[8]; os parâmetros das regras representados no cromossomo são as premissas da regra (termos lingüísticos, conforme a

[7]Foram utilizados os tipos de funções de pertinência disponível da Toolbox Fuzzy do software Matlab®.

[8]O cromossomo tem uma estrutura estática, portanto, se o tipo de função representada é definido por menos de quatro parâmetros, as células excedentes no cromossomo devem ser anuladas para não atuarem na definição da forma da função.

modelagem fuzzy descrita na seqüência, das variáveis 1 - "dif-seg" - e 2 - "inc-2-seg"), a conclusão da regra (termo lingüístico da variável 3 - "conclusoes"), o conector lógico das premissas (AND, OR) e o peso da regra (um valor > 0 e ≤ 1). Veja um exemplo de um cromossomo na Figura 5.20[9].

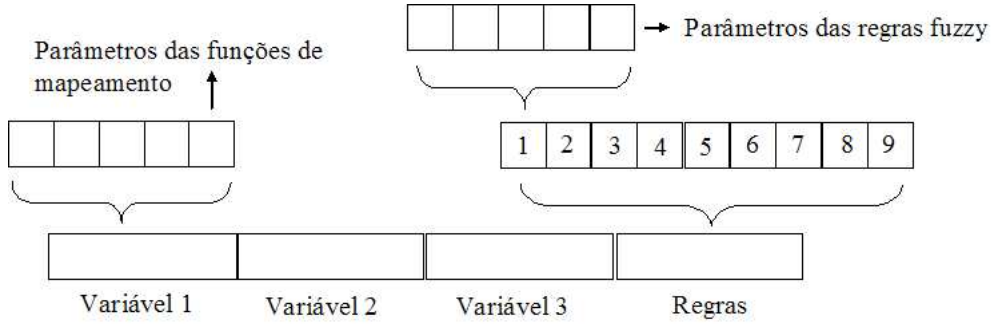


Fig. 5.19: Visão geral de um cromossomo.

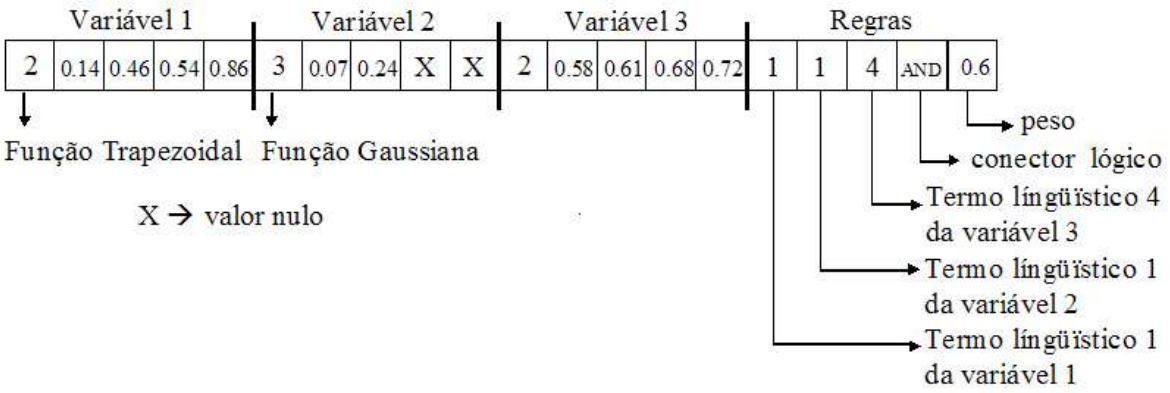


Fig. 5.20: Exemplo de um cromossomo.

- Operadores genéticos: Os operadores genéticos utilizados foram os clássicos *crossover* e mutação simples. O crossover de um ponto foi aplicado em posições que permitiram inserir diversidade de "prole" sem no entanto criar cromossomos infactíveis. A mutação foi aplicada em qualquer gene, porém, com o cuidado de não inserir mutações que infrigiam as regras de formação dos cromossomos e de suas partes constituintes.

[9]Por questões de estética visual optou-se por representar apenas uma regra ao invés de nove regras.

- Função fitness: A função fitness foi estabelecida como sendo o resultado da execução do sistema fuzzy, com os métodos citados mais à frente neste texto, derivado do cromossomo, sobre o problema de Tendência a Agrupamentos em relação a conjuntos de dados de distribuição conhecida. Conjuntos esses que também foram usados na calibragem manual de parâmetros.

- Demais parâmetros do algoritmo: as probabilidades de crossover e mutação, assim como número de gerações, tamanho da população e uso do elistismo foram combinados sob diferentes variações. Na modelagem final manteve-se: 20 cromossomos na população, 40 gerações, uso de elitismo, 5% de probabilidade de mutação, realização de 10 crossovers por geração.

Enfim, os esforços para a modelagem do sistema de inferência fuzzy resultou na seguinte estrutura: variável linguística de entrada "diferença de inclinação", com os seguintes termos lingüísticos: negativa, neutra e positiva (veja Tabela 5.1); variável linguística de entrada: "inclinação do segundo segmento", com os seguintes termos linguísticos: baixa, média e alta (veja Tabela 5.1); variável de saída "conclusão" com quatro termos linguísticos: uniforme, normal, agrupamentos e pontos próximos (veja Tabela 5.2); seis regras de inferência (veja Tabela 5.3); métodos utilizados: "max" para implementação da disjunção e agregação; "min" para implementação da conjunção e da implicação; "centróide" para defuzificação.

Tab. 5.1: Parâmetros das variáveis de entrada do sistema de inferência fuzzy.

Nome	Domínio	Tipo de Função Característica	Termo Linguístico	Parâmetros da Função Característica
dif-inc	[-3 3]	Z-curve	Negativa	[-2.500 1e-006]
		Trapezoidal	Neutra	[-0.077 0.051 0.3900 0.62800]
		S-curve	Positiva	[0.0769 2.500]
inc-2-seg	[0 3]	Z-curve	Baixa	[0.2000 0.800]
		Gaussiana	Média	[0.0765 0.808]
		S-curve	Alta	[0.8000 1.500]

Tab. 5.2: Parâmetros da variável de saída do sistema de inferência fuzzy.

Nome	Domínio	Tipo de Função Característica	Termo Linguístico	Parâmetros da Função Característica
Conclusão	[0 1]	Trapezoidal	Pnts	[0.1900 0.2840]
		Trapezoidal	Norm	[0.1920 0.2520 0.2990 0.3570]
		Z-curve	Agrup	[0.3270 0.3700 0.4210 0.4810]
		S-curve	Unif	[0.3420 0.5570]

Tab. 5.3: Regras de inferência fuzzy.

Regra	Premissa 1	Premissa 2	Conector	Conclusão	Peso
1	Negativa	Baixa	And	Norm	1
2	Negativa	Média	And	Norm	1
3	Negativa	Alta	And	Agrup	1
4	Neutra	Baixa	And	Norm	1
5	Neutra	Média	And	Norm	1
6	Neutra	Alta	And	Unif	1
7	Positiva	Baixa	And	Pnts	1
8	Positiva	Média	And	Unif	0.5
9	Positiva	Alta	And	Unif	0.6

Na Figura 5.21 é mostrado como se dá a integração do cálculo da dimensão Box-Counting com o sistema de inferência fuzzy e com o procedimento de corte da curva BC. Esta integração permite o cálculo da DFFS.

O sistema descrito por essa arquitetura é composto por três módulos de processamento de informações e um conjunto de regras heurísticas para tomada da decisão final. O primeiro módulo, alimentado pelos conjuntos de dados, é o que executa o algoritmo Box-Counting, salvando as informações sobre as inclinações e diferenças de inclinações dos segmentos de reta que compõem a curva resultante. Essas informações são mapeadas pelos conjuntos fuzzy no segundo módulo e o mapeamento alimenta o processo de inferência fuzzy. Como resultado deste processo é obtida uma série de conclusões sobre as observações feitas a respeito da distribuição dos dados em cada iteração do processo Box-Counting. Sobre estas conclusões é executado um

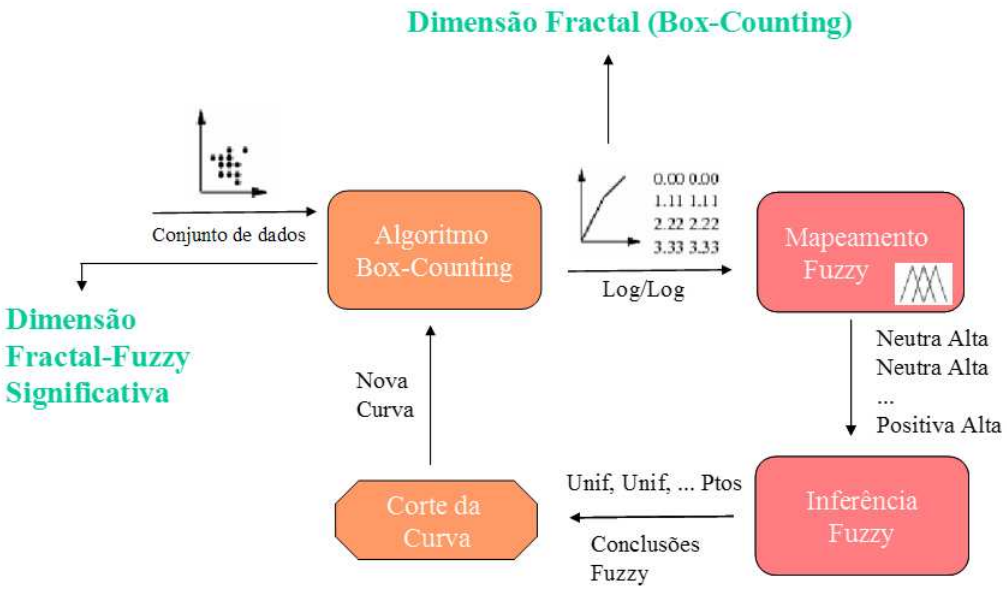


Fig. 5.21: Arquitetura do sistema de cálculo da Dimensão Fractal Fuzzy Significativa.

processo de decisão sobre o ponto ideal em que o algoritmo Box-Counting deve ser interrompido. Esta decisão tem a intenção de analisar a curva Box-Counting somente nos pontos que dizem respeito às interações que não sofrem a influência das relações intra-grupos.

5.4 Validação e aplicação do modelo heurístico

A análise da curva Box-Counting, sob a visão da abordagem Fractal Fuzzy, foi aplicada ao problema de Tendência a Agrupamentos e os resultados obtidos foram comparados àqueles fornecidos pelo uso da abordagem clássica de resolução desse problema (Abordagem de Hopkins). O intuito dessa comparação foi servir como uma forma de validação do processo usado para o cálculo da DFFS. O problema, a maneira como se utilizou a abordagem Fractal Fuzzy sobre ele e os resultados obtidos são detalhados no Capitulo 6. Na Figura 5.22 é mostrada a arquitetura do sistema implementado para resolução do problema de Tendência a Agrupamentos.

Os três primeiros processos dessa arquitetura são iguais ao da arquitetura apresentada para o cálculo da DFFS. O diferencial está na substituição do processo de decisão do corte da curva,

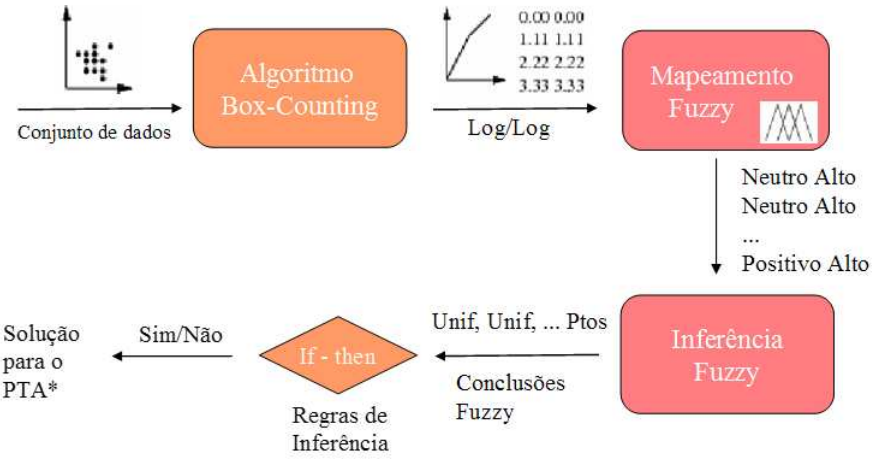


Fig. 5.22: Arquitetura do sistema de resolução do problema de análise da Tendência a Agrupamentos com uso da abordagem Fractal Fuzzy.

por um conjunto de regras heurísticas que analisam as conclusões e decidem se existe ou não tendência a agrupamentos no conjunto de dados sob análise.

Como aplicação direta do modelo como ele foi concebido (como na Figura 5.21) sobre o problema desta tese, configura-se o uso da DFFS para inferir a dimensão adequada do espaço de saída de uma RNA-AO. A análise do desempenho dessa abordagem bem como a forma como ela foi avaliada estão descritas no Capítulo 7.

5.5 Considerações Finais

A exploração da hipótese formulada nesta tese culminou, de forma natural, na elaboração de uma heurística. A Dimensão Fractal Fuzzy Significativa foi criada com o intuito de atender às necessidades específicas da tarefa de encontrar a dimensão ideal para o mapa de saída de uma RNA-AO, no entanto, resultou em uma nova medida de dimensão fractal e em um processo de análise de dados com potencial para ser aplicado em diferentes contextos.

Apesar da primeira arquitetura de sistema apresentada neste capítulo ter sido utilizada para validar a parametrização da decisão que leva à medida DFFS, ela representa uma aplicação válida para o processo de análise da curva Box-Counting. Já a segunda arquitetura de sistema consiste na resolução do problema formulado nesta tese. Além dessas aplicações, outras podem

ser submetidas ao recurso criado aqui e essa possibilidade será discutida nas conclusões deste trabalho. Os próximos capítulos apresentam os resultados obtidos com as avaliações de cada uma dessas arquiteturas.

Capítulo 6

Tendência a Agrupamentos

"...Dê-me a matéria e com ela eu construirei um mundo." (René Descartes em GLEISER (2001)), ... mesmo sem uma formulação matemática precisa, o modelo proposto por Descartes foi o primeiro a descrever o cosmo em termos dinâmicos, baseado em relações causais. *Marcelo Gleiser* em O fim da Terra e do Céu: o apocalipse na ciência e na religião.

Neste capítulo é apresentado o problema de tendência a agrupamentos como parte de uma metodologia para desenvolvimento de trabalhos de análise de dados. Uma abordagem clássica (de Hopkins) para resolução deste problema é apresentada e aplicada mediante uma série de conjuntos de dados. A esse mesmo problema e para os mesmos conjuntos a abordagem Fractal Fuzzy é aplicada. Uma comparação entre o desempenho das duas formas de resolução foi realizada com o intuito de verificar se os parâmetros do sistema Fractal Fuzzy estavam adequadamente calibrados para o seu uso em análise de dados e os resultados estão comentados neste capítulo.

6.1 Introdução

Em JAIN; DUBES (1988) é proposta uma metodologia genérica para guiar o procedimento de análise de dados exploratória cujo principal objetivo é a análise de agrupamentos. A proposição prevê que diferentes problemas de análise de dados apresentam características diferentes que podem inutilizar ou invalidar um ou mais passos dessa metodologia. Contudo, a análise dos passos propostos pode ser útil para fazer com que o processo de análise de dados a ser desenvolvido tenha menores possibilidades de incorrer em erros ou em ações desnecessárias.

Segundo o mesmo autor a metodologia é concebida como um processo inserido em um "*looping*" em que é possível obter novas idéias e colocá-las em operação no decorrer do processo. Como resultado final, uma ferramenta de análise pode ter sido construída ou conclusões informais, obtidas. Os passos estão mostrados na Figura 6.1 e são brevemente descritos na seqüência.

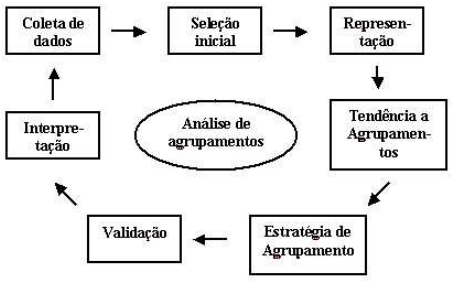


Fig. 6.1: Metodologia para análise de agrupamentos (JAIN; DUBES (1988)).

- Coleta de dados: os dados devem ser coletados e armazenados cuidadosamente observando-se as necessidades da área de aplicação. A quantidade e o tipo dos dados influenciam fortemente as estratégias de análise deles, o que justifica algumas iterações no "*looping*" da metodologia antes que se escolha o procedimento para realização desta coleta e registro;

- Seleção inicial: os conjuntos de dados sob análise usualmente requerem algum procedimento de normalização ou padronização. Ainda nessa fase a metodologia sugere que algumas análises de mais alto nível sejam realizadas (análises de gráficos, histogramas, etc), a fim de eliminar características menos significativas;

- Representação: os dados devem ser formatados de tal modo que facilite a aplicação da técnica de análise escolhida (veja JAIN; DUBES (1988) para estudo de técnicas de formatação de dados). Espera-se como resultado desta fase um tipo de matriz de padrões ou de matriz de proximidade, conforme for mais adequado à área de aplicação ou à técnica de análise;

- Tendência a agrupamentos: nesta fase pretende-se justificar a análise de agrupamentos no conjunto de dados em questão. Isso se deve ao fato que os dados podem não apresentar qualquer formação de grupos e, nesse caso, a aplicação de uma técnica de análise pode criar/ver grupos onde eles não existem realmente, além de gastar tempo do analista. Além disso, técnicas de análise de tendência a agrupamentos podem, de maneira rápida, fornecer algumas informações sobre a natureza dos dados que podem ser de grande valia em outros passos da metodologia;

- Estratégia de agrupamento: diferentes formas de análise de agrupamentos estão disponíveis ao analista de dados. A estratégia a ser aplicada deve ser escolhida entre vários tipos existentes e deve levar em consideração uma série de características, tais como: número de dados existentes, número de características que os descrevem e quais as exigências sobre a apresentação dos resultados.

- Validação: este passo pretende avaliar a qualidade da análise realizada. Índices externos (JAIN; DUBES (1988)) de validação de agrupamentos comparam os resultados da análise àquilo que o analista "gostaria" de ver. Índices internos (JAIN; DUBES (1988)) avaliam o mérito dos resultados em relação a objetivos básicos;

- Interpretação: nesta fase os resultados da análise de agrupamentos é estudada com o fim de descobrir o que os grupos encontrados significam no contexto da área de aplicação.

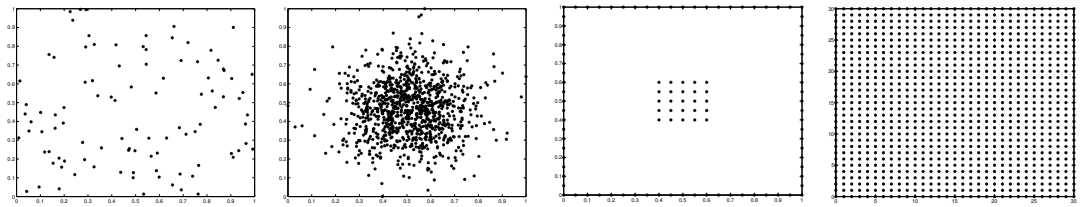
6.2 Tendência a Agrupamentos

O termo "tendência a agrupamentos" (do inglês "*clustering tendency*") refere-se ao problema de decidir se um conjunto de dados possui ou não predisposição a dividir-se em grupos naturais

sem no entanto indentificá-los (JAIN; DUBES (1988)). O objetivo da aplicação de algoritmos para esse fim é prevenir o analista da aplicação inapropriada de técnicas de descoberta de agrupamentos.

O teste de tendência a agrupamentos pode ser encarado como um problema de verificar se os dados de um conjunto estão distribuídos aleatoriamente no espaço (sem presença de grupos). Se assim estiverem, considera-se que eles não apresentam agrupamentos. Por exemplo, um teste de tendência a agrupamentos que analisa um arranjo espacial (um conjunto de dados ou uma matriz de padrões com n linhas e d colunas) em relação à proximidade dos pontos, medida por meio de uma métrica como a distância Euclidiana, pode resultar e três conclusões (JAIN; DUBES (1988)):

- os pontos estão distribuídos aleatoriamente (Figura 6.2(a));
- os pontos estão agrupados (em um ou mais grupos) exibindo um tipo de atração mútua (Figura 6.2(b) e Figura 6.2(c));
- os pontos estão uniformemente espaçados exibindo um tipo de repulsão mútua (Figura 6.2(d));



(a) (b) (c) (d)

Fig. 6.2: Tipos de distribuição: (a) aleatoriedade (gerado com BUENO; FERREIRA; JUNIOR (2004)); (b) e (c) atração mútua; (d) repulsão mútua.

Os dados que estão distribuídos aleatoriamente no espaço e aqueles que apresentam repulsão mútua não revelam tendência a agrupamentos e, portanto, não devem ser alvos de técnicas de agrupamento.

Dependendo da abordagem utilizada para realizar o teste, outros tipos de informação sobre a natureza dos dados podem ser descobertas. JAIN; DUBES (1988) descreve em detalhes

algumas dessas abordagens. Nesta tese aplicou-se a clássica abordagem de Hopkins para servir como comparativo à aplicação da abordagem Fractal Fuzzy como meio de testar a tendência a agrupamentos e com isso verificar a validade da abordagem proposta.

6.2.1 Abordagem de Hopkins

A abordagem de Hopkins fornece um indicador numérico sobre se existe ou não tendência a agrupamentos no conjunto de dados sob análise. Essa abordagem analisa se os pontos do conjunto de dados estão distribuídos de tal forma que contradiga a asserção de que eles estão distribuídos uniformemente. A análise se dá por meio da comparação de uma distribuição uniforme[1] com a distribuição dos pontos do conjunto de dados sob análise. Se não existe similaridade entre os dois conjuntos comparados, então a tendência a agrupamentos está certificada. Se existe, ela não está certificada.

O índice de Hopkins pode ser aplicado a conjuntos multi-dimensionais e é determinado por 6.1, em que: y_j é o conjunto de m-pontos uniformemente distribuídos no espaço d-dimensional do conjunto de dados sob análise; x_i é o conjunto de n-pontos escolhidos aleatoriamente do conjunto de dados sob análise (x_i é uma amostra do conjunto) com $m \leq n$; U é o conjunto de distâncias mínimas entre cada ponto de y_j e seu vizinho mais próximo em x_i; e W é o conjunto de distâncias entre m-pontos, aleatoriamente escolhidos em x_i e seu vizinho mais próximo. A Figura 6.3 ilustra essas medições.

$$H_{hpq} = \frac{\sum_{j=1}^{m} u_j}{\sum_{j=1}^{m} u_j + \sum_{j=1}^{m} w_j} \tag{6.1}$$

em que u é um elemento do conjunto U e w é um elemento do conjunto W.

Esse processo compara as distâncias mínimas existentes entre os pontos do conjunto de dados original com as distâncias mínimas entre esses pontos e aqueles do conjunto gerado artificialmente para o teste. Trata-se de um processo que deve ser repetido várias vezes, dada a sua natureza aleatória.

Quanto mais próximo de 1 está o valor obtido para o índice[2], maior é a confiança que se tem

[1] Distribuições aleatórias também podem ser utilizadas nesse processo, conforme JAIN; DUBES (1988).
[2] Os conjuntos comparados no teste não são similares.

Fig. 6.3: Distâncias utilizadas no cálculo do índice de Hopkins (adaptado de JAIN; DUBES (1988)).

na existência de tendências a agrupamentos. E quanto mais próximo de zero[3], menor é essa tendência. Em LAWSON; JURS (1990), os autores consideram que a tendência a agrupamentos é verificada se o índice assume um valor igual ou superior a $0,75$. Alternativamente, pode-se analisar este índice utilizando uma política de intervalos: um valor num intervalo ao redor de $0,5$ indicaria uma distribuição normal, valores acima desse intervalo indicariam tendência a agrupamentos enquanto valores abaixo indicariam a ausência de tal tendência.

A complexidade computacional da abordagem de Hopkins é $n * m * d$ em que: n é o número de pontos, do conjunto de dados sob análise, escolhidos para alimentar o processo de cálculo do índice, m é o número de pontos da distribuição gerada para alimentar o processo e d é a dimensionalidade do espaço onde o conjunto de dados está localizado. Assim, o limite superior da função de complexidade para este processo é $\theta(n^2 * d)$ (quando $n = m$).

6.2.2 Abordagem Fractal Fuzzy

A abordagem descrita no Capítulo 5 foi adaptada para atuar no problema de análise de tendência a agrupamentos. A partir desta adaptação, denominada abordagem Fractal Fuzzy, a tendência a agrupamentos é verificada com relação à tendência a:

- **agrupamento**: significa uma resposta afirmativa ao problema tratado, ou seja, existência de tendência a agrupamentos;
- **normalidade**: significa uma resposta negativa ao problema tratado, porém com a indicação da existência de um agrupamento de pontos que se destaca na distribuição, como

[3] Os conjuntos comparados no teste são similares.

numa distribuição normal;

- **uniformidade**: significa uma resposta negativa ao problema tratado, ou seja, os dados estão uniformemente distribuídos no espaço (de forma aleatória ou não).

Como discutido previamente, a conclusão sobre as características do conjunto de dados é identificada por meio da análise da forma da curva Box-Counting. A ocorrência de "anomalias" na curva mostra a existência de grupos no conjunto de dados. A identificação de que a curva resultante se aproxima de uma reta resulta na conclusão da existência de uma uniformidade na ocupação do espaço. A aproximação da curva resultante a uma curva "padrão" leva à conclusão de que os dados analisados possuem um distribuição similar a uma distribuição normal.

De forma mais específica, dado a "seqüência de conclusões" obtidas da aplicação da abordagem Fractal Fuzzy ao conjunto de dados, a conclusão final sobre o problema de tendência a agrupamentos é obtida mediante as seguintes regras, apresentadas aqui em uma ordem de importância[4]:

- Se aparece, em qualquer ponto da "seqüência de conclusões", a conclusão **agrupamentos**, a alteração na curva Box-Counting foi tal que as taxas de ocupações dos hipercubos foram suficientemente diferentes, entre duas iterações seguidas, para revelar que existem pontos formando agrupamentos;

- Se aparece a conclusão **pontos próximos** logo no início da "seqüência de conclusões" (situação comum em conjuntos de dados localizados em altas dimensões), infere-se que os dados estão agrupados, visto a queda na taxa de ocupações de hipercubos;

- Se a conclusão **pontos próximos** aparece dividindo sub-seqüências de conclusões **uniformidade** e **normalidade**, assume-se uma das conclusões das duas próximas regras com uma ressalva que caracteriza uma situação de limite de decisão;

- Se a maiora das conclusões na seqüência é formada com conclusões **uniformidade** e estas são adjacentes, a curva está se aproximando de uma reta e a distribuição dos dados se aproxima de uma distribuição uniforme, sem agrupamentos;

[4]Ou seja, se a regra 1 se aplica, as demais não precisam ser analisadas, senão, se a regra 2 se aplica, as demais não precisam ser analisadas e assim por diante.

- Se a maioria das conclusões na seqüência é formada com conclusões **normalidade** e estas são adjacentes, a curva tem um comportamento normal àquela obtida na aplicação dessa abordagem a um conjunto formado por uma distribuição normal;

Além disso, a resposta vem acompanhada de um *grau de pertinência* que representa a força com que os dados podem ser classificados em uma das categorias citadas.

6.3 Testes e Resultados

O problema de constatação de tendência a agrupamentos foi resolvido utilizando as duas abordagens descritas acima. Os testes foram realizados com os conjuntos de dados numéricos apresentados na Tabela 6.1. Todos os conjuntos de dados foram normalizados dentro do intervalo [0,1].

Os conjuntos 24 a 35 foram gerados sob uma distribuição normal, com parâmetros $\mu = 0$ e $\sigma = 1$, com variação do número de pontos e da dimensão do espaço onde estão localizados. As distribuições dos demais conjuntos de geração própria podem ser obtidas no CD-ROM anexo a esta tese. Alguns conjuntos possuem duas opções de agrupamentos e isso se deve ao fato de tais conjuntos admitirem interpretações diferentes em relação à formação dos grupos existentes neles. O Apêndice A traz os gráficos dos conjuntos de dados localizados em espaços 2-dimensionais e 3-dimensionais. Os conjuntos 36 a 42 são provenientes de NEWMAN et al. (1998) e representam informações de situações reais.

O índice de Hopkins foi calculado para cada um dos conjuntos de dados sob quatro parametrizações diferentes:

1. utilizando 10% dos pontos do conjunto de dados, $n = m$ e executando o algoritmo uma vez;

2. utilizando 10% dos pontos do conjunto de dados, $n = m$ e executando o algoritmo 10 vezes;

3. utilizando 100% dos pontos do conjunto de dados, $n = m$ e executando o algoritmo 1 vez;

Tab. 6.1: Descrição dos conjuntos de dados analisados.

N^o	Nome	Dimensão	Tamanho	Fonte	Distribuição
1	Dist. Regular	2	441	Sintético	Uniforme
2	Quadrados	2	121	Sintético	Com grupos
3	Quadrado Central	2	105	Sintético	Com grupos
4	Quadrados Dispersos	2	75	Sintético	Com grupos
5	Borboleta	2	15	KLIR; YUAN (1995)	Com grupos
6	2 Grupos Distintos	2	75	DUDA; HART; STORK (2000)	Com grupos
7	Circulo	2	76	Sintético	Com grupos
8	3 Quadrados Distintos	2	75	YAGER; FILEV (1994)	Com grupos
9	2 ou 3 Grupos	2	67	YAGER; FILEV (1994)	Com grupos
10	Misto1	2	127	Sintético	Com grupos
11	Cruz-Linha	2	148	LIU; XIE (1995)	Com grupos
12	Grupos Pequenos	2	39	LIU; XIE (1995)	Com grupos
13	Misto2	2	111	LIU; XIE (1995)	Com grupos
14	Espiral	2	190	CASTRO; VON ZUBEN (2001)	Com grupos
15	4 Grupos Definidos	2	164	ROUBENS (1982)	Com grupos
16	4 Grupos Dispersos	2	164	ROUBENS (1982)	Com grupos
17	4 Grupos Sobrepostos	2	168	ROUBENS (1982)	Com grupos
18	Misto3	2	321	Sintético	Com grupos
19	Misto4	2	31	Sintético	Com grupos
20	2 Grupos Distantes	2	312	KAYMAK; SETNES (2000)	Com grupos
21	Iris2d	2	150	Adap. NEWMAN et al. (1998)	Com grupos
22	Iris3d	3	150	Adap. NEWMAN et al. (1998)	Com grupos
23	Iris4d	4	150	NEWMAN et al. (1998)	Com grupos
24	Dist-norm-100-1	1	100	Sintético	Normal
25	Dist-norm-100-2	2	100	Sintético	Normal
26	Dist-norm-100-3	3	100	Sintético	Normal
27	Dist-norm-100-4	4	100	Sintético	Normal
28	Dist-norm-100-5	5	100	Sintético	Normal
29	Dist-norm-3000-1	1	3000	Sintético	Normal
30	Dist-norm-3000-2	2	3000	Sintético	Normal
31	Dist-norm-3000-3	3	3000	Sintético	Normal
32	Dist-norm-3000-4	4	3000	Sintético	Normal
33	Dist-norm-3000-5	5	3000	Sintético	Normal
34	Dist-norm-5000-1	1	5000	Sintético	Normal
35	Dist-norm-5000-3	3	5000	Sintético	Normal
36	Abalone	8	4177	NEWMAN et al. (1998)	Com grupos
37	Dermato	34	366	NEWMAN et al. (1998)	Com grupos
38	Auto-mpg	7	398	NEWMAN et al. (1998)	Com grupos
39	Va	13	200	NEWMAN et al. (1998)	Com grupos
40	Cleveland	13	303	NEWMAN et al. (1998)	Com grupos
41	Hungria	13	294	NEWMAN et al. (1998)	Com grupos
42	Switzerland	13	123	NEWMAN et al. (1998)	Com grupos
43	Aleatório	2	153	Sintético	Aleatória

4. utilizando 50% dos pontos do conjunto de dados, $n = m$ e executando o algortimo 5 vezes;

Na Tabela 6.2 são apresentados os índices obtidos em cada uma das configurações de teste. Para as configurações em que mais de uma execução foi realizada, obteve-se valores diferentes para o índice de Hopkins a cada execução (devido a aleatoriedade da escolha dos pontos amostrados do conjunto de dados e da distribuição uniforme usada como comparador). Nesses casos, estão listados na tabela os valores mínimo e máximo de índices obtidos, o intervalo obtido entre o valor mínimo e máximo e o valor do índice de Hopkins como a média dos valores obtidos nas execuções.

As variações nos índices resultantes das execuções da abordagem de Hopkins com diferentes configurações de parâmetros não causou diferenças de respostas à questão de tendência a agrupamentos na maioria dos conjuntos de dados.

Quando a decisão foi limitada pelo índice $0,5$, em apenas dois conjuntos (12 e 24), houve divergência em uma das configurações. Para o caso em que a decisão é limitada pelo índice $0,75$, como nas aplicações que vêem o índice de forma mais restritiva, houve divergência apenas no conjunto 43.

Quanto às variações de índices obtidos por diferentes execuções do processo para uma mesma configuração de parâmetros, existiu maior divergência de decisões na configuração 2 que na 4. Para a configuração 2, com índice limitante $0,5$, a divergência ocorreu em 6 conjuntos de dados e com o índice limitante em $0,75$ a divergência ocorreu em 12 conjuntos. Para a configuração 4 a divergência se deu em apenas 1 conjunto para o limitante $0,5$ e em 3 conjuntos para o limitante $0,75$.

Utilizando a política de intervalos para analisar as respostas da aplicação da abordagem de Hopkins ao problema em questão, percebeu-se que existe divergência de respostas para 2 conjuntos de dados. A verificação das respostas dos índices mínimos e máximos das configurações 2 e 4 mostrou uma divergência em 14 e 4 conjuntos, respectivamente, em relação à política de intervalos.

Os baixos índices de divergência para a configuração 4 contribuíram para que ela fosse escolhida para a realização das comparações com os resultados da aplicação da abordagem

Tab. 6.2: Conjuntos de dados analisados quanto à tendência a agrupamentos, de acordo com o índice de Hopkins.

N^o	1 iteração amostra 10% H_{hpk}	10 iterações amostra 10% H_{hpk} H_{min} H_{max} $H_{intervalo}$	1 iteração amostra 100% H_{hpk}	5 iterações amostra 50% H_{hpk} H_{min} H_{max} $H_{intervalo}$
1	0.2789	0.2856 0.2712 0.3017 0.0304	0.2859	0.2864 0.2798 0.2904 0.0105
2	0.7609	0.8167 0.7406 0.8689 0.1282	0.8194	0.8153 0.8013 0.8315 0.0302
3	0.6536	0.7054 0.6322 0.7491 0.1169	0.7023	0.7214 0.7122 0.7395 0.0273
4	0.8662	0.8297 0.7973 0.8569 0.0596	0.8236	0.8242 0.8110 0.8329 0.0219
5	0.2003	0.3876 0.1517 0.6347 0.4830	0.3927	0.4120 0.3490 0.4822 0.1332
6	0.7184	0.6120 0.4158 0.7530 0.3372	0.6615	0.6620 0.6438 0.6869 0.0431
7	0.7584	0.8163 0.7635 0.8782 0.1147	0.8386	0.8230 0.7985 0.8671 0.0686
8	0.8726	0.8641 0.8339 0.8864 0.0525	0.8637	0.8671 0.8557 0.8783 0.0226
9	0.8545	0.8643 0.7564 0.9169 0.1605	0.8414	0.8882 0.8681 0.9060 0.0379
10	0.8754	0.8683 0.7624 0.9143 0.1519	0.8775	0.8566 0.6320 0.9353 0.3033
11	0.8021	0.8643 0.8189 0.9058 0.0869	0.8685	0.8517 0.6046 0.9424 0.3378
12	0.6374	0.4767 0.2235 0.6736 0.4501	0.5349	0.5600 0.5355 0.5899 0.0544
13	0.9051	0.8701 0.6946 0.9431 0.2485	0.8963	0.9017 0.8844 0.9117 0.0273
14	0.6112	0.6987 0.6140 0.8032 0.1892	0.6851	0.7149 0.6964 0.7448 0.0484
15	0.9319	0.9151 0.8762 0.9373 0.0611	0.9157	0.9195 0.9143 0.9248 0.0105
16	0.6109	0.6664 0.5678 0.7559 0.1881	0.6946	0.6889 0.6447 0.7162 0.0715
17	0.6395	0.6459 0.5711 0.7168 0.1457	0.6632	0.6497 0.6357 0.6796 0.0439
18	0.9450	0.9328 0.9100 0.9530 0.0430	0.9340	0.9338 0.9236 0.9375 0.0139
19	0.9463	0.8668 0.6540 0.9505 0.2965	0.9082	0.9047 0.8888 0.9242 0.0354
20	0.9857	0.9766 0.9715 0.9808 0.0092	0.9784	0.9771 0.9755 0.9799 0.0044
21	0.8995	0.8686 0.8396 0.9041 0.0645	0.8808	0.8771 0.8606 0.8939 0.0333
22	0.8668	0.8931 0.8565 0.9132 0.0567	0.8754	0.8938 0.8882 0.8979 0.0097
23	0.8711	0.8300 0.7991 0.8584 0.0593	0.8405	0.8391 0.8190 0.8443 0.0303
24	0.7136	0.4643 0.2077 0.7395 0.5318	0.3339	0.3806 0.2635 0.5595 0.2960
25	0.5170	0.6423 0.4751 0.8079 0.3328	0.6539	0.6194 0.5664 0.6476 0.0812
26	0.5101	0.6273 0.5429 0.7449 0.2020	0.6233	0.6344 0.6189 0.6584 0.0395
27	0.5354	0.5718 0.4907 0.6880 0.1973	0.5545	0.5593 0.5459 0.5756 0.0296
28	0.6350	0.6141 0.5615 0.6599 0.0984	0.6033	0.5961 0.5859 0.6113 0.0254
29	0.9620	0.9033 0.7190 0.9801 0.2610	0.9019	0.9086 0.8689 0.9621 0.0931
30	0.9362	0.9377 0.9178 0.9513 0.0334	0.9245	0.9302 0.9278 0.9343 0.0065
31	0.8449	0.8512 0.8409 0.8591 0.0182	0.8503	0.8505 0.8458 0.8563 0.0105
32	0.7854	0.7907 0.7686 0.8121 0.0435	0.7904	0.7872 0.7847 0.7913 0.0066
33	0.7815	0.7774 0.7686 0.7849 0.0163	0.7775	0.7678 0.7656 0.7700 0.0044
34	0.9706	0.9694 0.9495 0.9890 0.0395	0.9764	0.9790 0.9729 0.9867 0.0137
35	0.8518	0.8573 0.8516 0.8655 0.0139	0.8597	0.8585 0.8559 0.8611 0.0052
36	0.9601	0.9582 0.9553 0.9606 0.0054	0.9586	0.9580 0.9574 0.9589 0.0015
37	0.7374	0.7339 0.7245 0.7563 0.0318	0.7333	0.7343 0.7284 0.7389 0.0105
38	0.8352	0.8403 0.8315 0.8596 0.0282	0.8516	0.8482 0.8456 0.8523 0.0068
39	0.7256	0.7065 0.6865 0.7253 0.0427	0.7177	0.7205 0.7035 0.7384 0.0349
40	0.7014	0.7051 0.6646 0.7382 0.0735	0.7025	0.7020 0.6937 0.7097 0.0160
41	0.8238	0.8537 0.8234 0.8799 0.0565	0.8514	0.8535 0.8479 0.8612 0.0132
42	0.7441	0.7377 0.7219 0.7525 0.0306	0.7270	0.7281 0.7192 0.7382 0.0190
43	0.5680	0.7469 0.6081 0.8624 0.2545	0.7194	0.7540 0.7125 0.7853 0.0728

Fractal Fuzzy. Como resposta à análise de tendência a agrupamentos a partir da abordagem de Hopkins, no contexto deste trabalho, decidiu-se por usar o índice valorado em $0,5$ como limitante, estabelecendo um relaxamento sobre a restrição aplicada por LAWSON; JURS (1990): conjuntos de dados com índice $\geq 0,5$ apresentam tendência a agrupamentos.

Os resultados da aplicação da abordagem Fractal Fuzzy estão listados na Tabela 6.3. Nela estão representadas as seguintes informações: a seqüência de conclusões resultante da aplicação da abordagem, na forma de variáveis linguísticas (**U** - uniformidade, **A** - agrupamentos, **N** - normalidade e **P** - pontos próximos); a tendência resultante (de acordo com a discussão da Seção 6.2.2); o grau de pertinência ao conjunto *fuzzy* que representa a tendência resultante.

A partir da abordagem Fractal Fuzzy classifica-se os conjuntos que apresentam distribuição uniforme ou aleatória como conjuntos que não apresentam tendência a agrupamentos e aqueles que apresentam a distribuição normal, como conjuntos que apresentam tendência a agrupamento mas que não necessitam ser submetidos a algoritmos de descoberta de agrupamentos (por apresentar um único grupo).

Na Tabela 6.4 estão relacionados os resultados obtidos, em relação à tendência a agrupamentos, por meio da abordagem de Hopkins (com a quarta configuração de parâmetros e limitante no índice $0,5$) e por meio da abordagem Fractal Fuzzy, comparando-as com as classificações conhecidas (resultado esperado) para cada conjunto analisado: distribuição uniforme, distribuição normal, distribuição aleatória, distribuição com grupos e distribuição sem grupos (ou com um único grupo). As respostas acompanhadas de $*$ indicam que a abordagem Fractal Fuzzy tomou a decisão em uma situação de fronteira.

Em relação à resposta para tendência a agrupamentos, pode-se então considerar duas situações: na primeira os conjuntos que apresentam características de uma distribuição normal não deveriam ser submetidos a um algoritmo de descoberta de grupos, e então, a tendência a agrupamentos teria sido atestada corretamente para 29 $(1-4; 6; 8-24; 36-42)$ conjuntos de dados pela abordagem de Hopkins e 32 $(1-4; 7-10; 15-22; 24-33; 37-40; 42-43)$ para a abordagem Fractal Fuzzy; na segunda, se esses conjuntos forem considerados como suscetíveis a aplicação de um algoritmo de descoberta de grupos, então o desempenho da abordagem de Hopkins sobe para 39 $(1-4; 6; 7-23; 25-42)$ conjuntos e da abordagem Fractal Fuzzy sobe

Tab. 6.3: Conjuntos de dados analisados, quanto à tendência a agrupamentos, pela Abordagem Fractal Fuzzy. U - uniformidade; N - normalidade; P - pontos próximos; A - Agrupamentos.

Nº	Resultado	Tendência	Grau de Pertinência
1	U U U P	Regularidade	1.0
2	U U A P	Agrupamento	1.0
3	U U A P	Agrupamento	0.3
4	P A A P	Agrupamento	1.0
5	U N	Regularidade	1.0
6	U U N	Regularidade	1.0
7	U U N N P	Normalidade	1.0
8	N A A P	Agrupamento	1.0
9	A A A N	Agrupamento	0.5
10	U U U A N N N	Agrupamento	0.7
11	U U U N N N	Normalidade	0.7
12	U U P	Regularidade	1.0
13	U U U N N N N	Normalidade	0.2
14	U U U N N N	Normalidade	0.9
15	U A U U P N N	Agrupamento	1.0
16	U U U A P N N	Agrupamento	0.3
17	U U U A P N N N	Agrupamento	1.0
18	U A U U N P	Agrupamento	1.0
19	A P A U A	Agrupamento	1.0
20	A A U U U U N N N N N N	Agrupamento	1.0
21	A U U A N	Agrupamento	1.0
22	U A A N N	Agrupamento	1.0
23	U U P N N	Normalidade Agrupamento	1.0
24	U U U U U U N N N N N N N	Normalidade	0.8
25	U U U P N N N	Normalidade Agrupamento	1.0
26	U U N N	Normalidade	1.0
27	U P N	Normalidade Agrupamento	1.0
28	U P N	Normalidade Agrupamento	1.0
29	U U U U U U U U U U U U N N N N N N N N N N N	Normalidade	0.6
30	U U U U U U P N N N N N N	Normalidade Agrupamento	1.0

Tab. 6.4: Análise e comparação dos testes realizados.

N^o	Resultado Hopkins	Índice de Hopkins	Resultado Fractal Fuzzy	Grau de Pert.	Resultado Esperado
1	Uniformidade	0.2864	Uniformidade	1.0	Distribuição Uniforme
2	Agrupamento	0.8194	Agrupamento	1.0	Com Grupos
3	Agrupamento	0.7214	Agrupamento	0.3	Com Grupos
4	Agrupamento	0.8236	Agrupamento	1.0	Com Grupos
5	Uniformidade	0.4120	Uniformidade	1.0	Com Grupos
6	Agrupamento	0.6620	Uniformidade	1.0	Com Grupos
7	Agrupamento	0.8230	Normalidade	1.0	Sem Grupos
8	Agrupamento	0.8671	Agrupamento	1.0	Com Grupos
9	Agrupamento	0.8882	Agrupamento	0.5	Com Grupos
10	Agrupamento	0.8566	Agrupamento	0.7	Com Grupos
11	Agrupamento	0.8517	Normalidade	0.7	Com Grupos
12	Agrupamento	0.5600	Uniformidade	1.0	Com Grupos
13	Agrupamento	0.9017	Normalidade	0.2	Com Grupos
14	Agrupamento	0.7149	Normalidade	0.9	Com Grupos
15	Agrupamento	0.9157	Agrupamento	1.0	Com Grupos
16	Agrupamento	0.6889	Agrupamento	0.3	Com Grupos
17	Agrupamento	0.6497	Agrupamento	1.0	Com Grupos
18	Agrupamento	0.9338	Agrupamento	1.0	Com Grupos
19	Agrupamento	0.9082	Agrupamento	1.0	Com Grupos
20	Agrupamento	0.9771	Agrupamento	1.0	Com Grupos
21	Agrupamento	0.8771	Agrupamento	1.0	Com Grupos
22	Agrupamento	0.8938	Agrupamento	1.0	Com Grupos
23	Agrupamento	0.8391	Normalidade*	1.0	Com Grupos
24	Uniformidade	0.3806	Normalidade	0.8	Distribuição Normal
25	Agrupamento	0.6194	Normalidade*	1.0	Distribuição Normal
26	Agrupamento	0.6233	Normalidade	1.0	Distribuição Normal
27	Agrupamento	0.5545	Normalidade*	1.0	Distribuição Normal
28	Agrupamento	0.6033	Normalidade*	1.0	Distribuição Normal
29	Agrupamento	0.9019	Normalidade	0.6	Distribuição Normal
30	Agrupamento	0.9245	Normalidade*	1.0	Distribuição Normal
31	Agrupamento	0.8503	Normalidade*	1.0	Distribuição Normal
32	Agrupamento	0.7872	Normalidade*	1.0	Distribuição Normal
33	Agrupamento	0.7678	Normalidade*	1.0	Distribuição Normal
34	Agrupamento	0.9790	Agrupamento	1.0	Distribuição Normal
35	Agrupamento	0.8585	Uniformidade*	1.0	Distribuição Normal
36	Agrupamento	0.9580	Normalidade*	0.5	Com Grupos
37	Agrupamento	0.7333	Agrupamento	1.0	Com Grupos
38	Agrupamento	0.8482	Agrupamento	1.0	Com Grupos
39	Agrupamento	0.7205	Agrupamento	1.0	Com Grupos
40	Agrupamento	0.7020	Agrupamento	1.0	Com Grupos
41	Agrupamento	0.8535	Normalidade*	1.0	Com Grupos
42	Agrupamento	0.7281	Agrupamento	1.0	Com Grupos
43	Agrupamento	0.7540	Normalidade	0.9	Distribuição Aleatória

para 37 $(1-4; 8-11; 13-34; 36-42)$ conjuntos de dados corretamente analisados.

Os erros da abordagem de Hopkins ocorreram, com maior evidência, nos conjuntos "Borboleta" e "Círculo" que possuem poucos pontos, e no conjunto "Aleatório" que, como pode ser observado no Apêndice A, possui alguns poucos pontos mais próximos entre si do que de outros, podendo caracterizar, em algum nível de análise, a formação de grupos.

Os erros da abordagem Fractal Fuzzy ocorreram principalmente nos conjuntos reais com poucos pontos: "Borboleta", "2 Grupos Distintos", os quais podem ser interpretados como uniformemente distribuídos (veja o Apêndice A); e em um conjunto com distribuição normal, porém, a decisão ficou no limite entre Uniformidade e Normalidade.

6.4 Considerações Finais

Neste capítulo discutiu-se a capacidade de resolução de um problema de análise de dados com a aplicação da abordagem Fractal Fuzzy. Concluiu-se que essa abordagem apresenta resultados que a qualificam como tão boa quanto a clássica abordagem de Hopkins para a resolução do problema de tendência a agrupamentos, tanto em termos de alcance de resultados, quanto em termos de complexidade computacional. A complexidade da abordagem Fractal Fuzzy é determinada pela complexidade do algoritmo Box-Counting, para o qual existem implementações bastante eficientes (Veja Capítulo 2).

Assim, mostrou-se que a compatibilidade da abordagem Fractal Fuzzy a uma abordagem clássica de resolução de um problema de análise de dados justifica a parametrização do sistema Fractal Fuzzy como foi estabelecida. O sistema que implementa essa abordagem foi parametrizado, com ajuda da aplicação do Algoritmo Genético, apenas com a supervisão realizada com alguns poucos conjuntos de dados e, quando aplicado a um conjunto maior, teve um desempenho adequado, como atestado na comparação da resolução do problema de Tendência a Agrupamentos. A partir da verificação desta parametrização o sistema está pronto (adequadamente parametrizado) para ser aplicado ao problema de determinação da dimensão dos mapas de saída de uma RNA-AO.

Capítulo 7

Dimensão ideal para uma Rede Neural Artificial Auto Organizável

O objetivo neste capítulo é mostrar que a medida DFFS tem potencial para atuar no apoio à tomada de decisão sobre a dimensão ideal para o espaço de saída de uma RNA-AO. A fim de proporcionar o entendimento dos procedimentos de análise, optou-se por descrever neste capítulo os conjuntos de dados utilizados (informações complementares sobre os conjuntos de dados reais estão listadas no Apêndice B), as RNAs-AO treinadas e as avaliações de suas qualidades topográfica e quantizadora, os critérios utilizados para escolher, dentre as RNAs-AO, aquela que melhor atendeu às expectativas de mapeamento topográfico e quantizador, os resultados referentes às comparações das abordagens de inferência da dimensão ideal do mapa (DFFS, dimensão Box-Counting e Análise de Componentes Principais) e, finalmente, a avaliação de desempenho dessas abordagens. Para efeito de complementação, o formato de algumas tabelas geradas para organizar informações utilizadas para análise de resultados são apresentadas no Apêndice C.

7.1 Conjuntos de dados de teste

Os conjuntos de dados utilizados neste estudo estão divididos em duas classes: a primeira é formada por conjuntos gerados segundo uma distruibuição conhecida e, portanto, os grupos[1] formados podem ser classificados segundo essas funções de distribuição; a segunda é formada por conjuntos de dados reais, com classificação obtida junto à fonte onde a base de dados está disponibilizada.

O objetivo do uso dos conjuntos de dados sintéticos foi analisar o desempenho da abordagem Fractal Fuzzy sob condições controladas, de forma a facilitar o entendimento das vantagens e limitações da mesma. Esse grupo de conjuntos de teste foi projetado com variabilidade em: número de dimensões do espaço de residência, número de pontos, número de funções geradoras de "grupos", forma dessas funções, sobreposição de distribuições, densidade do conjunto, predominância de classe e entropia.

Para alguns desses conjuntos de dados, gerados randomicamente sob uma ou mais distribuições de probabilidade, optou-se por gerar duas instâncias sob as mesmas condições para testar se a abordagem Fractal Fuzzy sofre influência do fator "randomicidade", já que os mapeamentos obtidos para cada uma destas instâncias podem ser diferentes.

Os conjuntos de dados reais foram testados com o objetivo de verificar se, para dados de diferentes contextos, o desempenho da abordagem seria adequado ou não, e conseqüentemente, estimular ou não a sua utilização para problemas do mundo real. Segundo PRECHELT (1994), até a data de escrita do artigo em questão, a grande maioria das pesquisas da área de redes neurais artificiais era testada apenas com conjuntos de dados sintéticos e raramente com mais do que dois ou três. Havia várias justificativas para isso, entre elas, a dificuldade em gerar dados a partir de contextos reais e a capacidade reduzida de tempo de processamento do *hardware* envolvido. Entretanto, atualmente existem inúmeros repositórios de bases de dados reais e tempo de processamento não deve mais ser apresentado como um empecilho para a realização de testes.

Na escolha desses conjuntos procurou-se satisfazer quesitos de variabilidade similares aos dos conjuntos sintéticos: número da dimensão do espaço de residência, número de pontos, número

[1] Ou " clusters".

de "grupos", densidade do conjunto, predominância de classe e entropia. Além disso, buscou-se na literatura por bases de dados que façam parte de conjuntos para *benchmark* de algoritmos de classificação e/ou agrupamento. Entre as diversas bases disponíveis optou-se por utilizar os conjuntos de dados de NEWMAN et al. (1998) e PRECHELT (1994), por apresentarem as características satisfatórias para esse trabalho.

Na Tabela 7.1 encontra-se a descrição dos conjuntos de dados sintéticos. Os conjuntos de dados reais e as respectivas classificações dos "grupos" neles presentes estão descritos, resumidamente, na Tabela 7.2. Nessas tabelas, a primeira coluna refere-se a um indexador útil para identificação do conjunto de teste durante as discussões dos resultados. As demais colunas apresentam atributos descritivos dos conjuntos, com destaque para a última coluna que lista o número de pontos pertencentes a cada grupo, na ordem em que eles aparecem no arquivo de dados dos conjuntos.

Os testes realizados com o conjunto 01s tem o objetivo de analisar o comportamento da abordagem para a situação em que os pontos estão distribuídos por todo o espaço disponível, com a concentração inerente ao tipo de função de distribuição utilizada. Os conjuntos de 02s até 05s ilustram a situação em que duas ou mais funções de distribuição são utilizadas, caracterizando dois "grupos" diferentes dentro do conjunto.

Dois conjuntos (06s e 07s) foram elaborados com o intuito de analisar o comportamento da abordagem quanto ao fator de existência ou não de sobreposição de grupos. Os conjuntos 08s e 09s foram criados com funções de distribuição de probabilidades diferentes nos eixos que compõem o espaço. Por fim, o conjunto 10s foi criado para verificar o comportamento da abordagem em conjuntos de dados localizados dentro do espaço com um grande número de dimensões.

Os conjuntos da Tabela 7.2 podem ser divididos em dois tipos: conjuntos que caracterizam problemas de classificação (01r, 02r, 03r, 04r, 05r, 06r, 07r), e os que caracterizam problemas de aproximação de funções (08r, 09r, 10r, 11r, 12r, 13r e 14r).

Tab. 7.1: Conjuntos de dados sintéticos.

Id.	Nome	Dimensão de Imersão	Nº de pontos	Fonte	Função	Parâmetros	Nº de grupos	Nº de pontos por grupo
01s	Norm1	5	3000	Própria	Gauss	$\mu = 0$, $\sigma = 1$	1	3000
02s	Norm2	5	2000	Própria	Gauss	$\mu = 0$, $\sigma = 100$	2	1000
								1000
03s	Norm3	5	10000	Própria	Gauss	$\mu = 0$, $\sigma = 100$	2	5000
					Gauss	$\mu = 300$, $\sigma = 100$		5000
04s	Norm4	2	2000	Própria	Gauss	$\mu = 0$, $\sigma = 100$	2	1000
					Gauss	$\mu = 300$, $\sigma = 100$		1000
05s	Norm5	5	10000	Própria	Gauss	$\mu = 0$, $\sigma = 100$	2	7000
					Gauss	$\mu = 300$, $\sigma = 100$		3000
06s	Norm6	10	10000	Própria	Gauss	$\mu = 0$, $\sigma = 40$	5	2000
					Gauss	$\mu = 300$, $\sigma = 80$		2000
					Gauss	$\mu = 700$, $\sigma = 60$		2000
					Gauss	$\mu = 1000$, $\sigma = 50$		2000
					Gauss	$\mu = 1500$, $\sigma = 70$		2000
07s	Norm7	10	10000	Própria	Gauss	$\mu = 0$, $\sigma = 100$	5	2000
					Gauss	$\mu = 300$, $\sigma = 100$		2000
					Gauss	$\mu = 700$, $\sigma = 70$		2000
					Gauss	$\mu = 1000$, $\sigma = 80$		2000
					Gauss	$\mu = 1200$, $\sigma = 80$		2000
08s	Norm8	5	2000	Própria	Gauss	$\mu = 0$, $\sigma = 200\ 3$[2]	2	1000
					Gauss	$\mu = 50$, $\sigma = 500\ 2$		1000
					Gauss	$\mu = 500$, $\sigma = 50\ 3$		
					Gauss	$\mu = 500$, $\sigma = 200\ 2$		
09s	Norm9	2	2000	Própria	Gauss	$\mu = 0$, $\sigma = 100$	2	1000
					Gauss	$\mu = 300$, $\sigma = 100$		1000
10s	Norm10	35	20000	Própria	Gauss	$\mu = 0$, $\sigma = 100$	2	10000
					Gauss	$\mu = 400$, $\sigma = 100$		10000

A representação dos valores dos atributos dos conjuntos de dados foi realizada da seguinte forma: para atributos valorados com números reais, os valores foram normalizados para o intervalo $[0, 1]$; os valorados com números inteiros foram tratados como os números reais; atributos ordinais (com m valores diferentes) foram tratados como valores pseudo-reais, aplicando sobre eles o mesmo procedimento de normalização já citado; atributos nominais foram representados com a codificação *"1-of-m"*; finalmente, os atributos com valores desconhecidos foram substituí-

[2]Número de eixos em que estes parâmetros regem a distribuição.

Tab. 7.2: Conjuntos de dados reais.

Id.	Nome	Dimensão de Imersão	Nº de pontos	Fonte	Nº de grupos	Nº de pontos por grupo
01r	Íris	4	150		3	50-50-50
02r	Dermatology	34	366		6	112-61-72-49-52-20
03r	Glass	9	214		7	70-17-76-0-13-9-29
04r	Ionosphere	34	351		2	
05r	Internet	1558	3279		2	2821-458
06r	Letter	16	16000		26	789-766-736-805-768-775-773-734-755-747-739-761-792-783-753-803-783-758-748-796-813-764-752-787-786-734
07r	Zôo	16	101		7	41-20-5-13-4-8-10
08r	Abalone	8	4177		28	1-1-15-57-115-259-391-568-689-634-487-267-203-126-103-67-58-42-32-26-14-6-9-2-1-1-2-1
09r	Auto-mpg	7	398		3	249-70-79
10r	Heart Deseases - Va	12	200		4	51-56-41-42-10
11r	Heart Deseases - Cleveland	12	303		5	164-55-36-35-13
12r	Heart Deseases - Hungria	12	294		5	188-37-26-28-15
13r	Heart Deseases - Switzerland	12	123		5	8-48-32-30-5
14r	Flare	10	323		3	

dos por valores nulos dentro do contexto dos atributos, conforme indicação da documentação da base de dados.

Nas Tabelas 7.3 e 7.4 é apresentado um resumo dos conjuntos de dados e sua relação com as principais características que podem influenciar na tarefa de análise de dados (adaptado de ZHENG (1993)). As informaçãoes contidas nas tabelas foram obtidas a partir da documentação dos conjuntos de dados e estão divididas da seguinte forma:

- a dimensão de imersão é considerada "pequena" se o número de atributos que descreve o conjunto de dados é menor que 10, "média" se este número está entre 10 e 30 e "grande" se o número é maior que 30;

- o tamanho do conjunto é considerado "pequeno" quando o conjunto de dados possui menos

de 210 exemplos, "médio" se este número está entre 210 e 3170 e "grande" se o conjunto possui mais de 3170 exemplos;

- o número de classes é "pequeno" para uma única classe ou duas, "médio" para 3 a 10 classes e "grande" para mais de 10 classes;
- a densidade é " baixa" para valores menores que $1,00 * 10^{-18}$ (veja 7.1 para o cálculo da densidade), "média" para valores entre $1,00 * 10^{-18}$ e $6,00 * 10^{-7}$ e "alta" para valores maiores que $6,00 * 10^{-7}$;
- a predominância de classe é "baixa" se ela representar menos de 40% dos dados no conjunto, "média" se representar algo entre 40% e 75% e "alta" se representar mais do que 40% dos dados;
- a entropia é considerada "baixa" se for menor que $0,80$ bits, "média" se estiver entre $0,80$ e $1,58$ bits e "alta" se for maior que $1,58$ bits (veja 7.3 para o cáculo da entropia).

Tab. 7.3: Conjuntos de dados sintéticos X características.

Id.	Dimensão de Imersão	Tamanho do Conjunto	Número de Classes	Densidade	Predominância de Classe	Entropia
01s	Pequena	Médio	Pequeno	Média	Alta	Baixa
02s	Pequena	Médio	Pequeno	Média	Média	Média
03s	Pequena	Grande	Pequeno	Média	Média	Média
04s	Pequena	Médio	Pequeno	Alta	Média	Média
05s	Pequena	Grande	Pequeno	Média	Média	Média
06s	Média	Grande	Médio	Baixa	Baixa	Alta
07s	Média	Grande	Médio	Baixa	Baixa	Alta
08s	Pequena	Médio	Pequeno	Média	Média	Média
09s	Pequena	Médio	Pequeno	Média	Média	Média
10s	Grande	Grande	Pequeno	Baixa	Média	Média

Segundo ZHENG (1993), um algoritmo de aprendizado pode aprender melhor a partir de um grande número de padrões de treinamento do que a partir de poucos exemplos. Entretanto, considerando diferentes domínios e diferentes descrições de espaço, é difícil dizer que um conjunto de dados com mais de N exemplos é grande e que um outro conjunto com menos de N

Tab. 7.4: Conjuntos de dados reais X características.

Id.	Dimensão de Imersão	Tamanho do Conjunto	Número de Classes	Densidade	Predominância de Classe	Entropia
01r	Pequena	Pequeno	Médio	Alta	Baixa	Média
02r	Grande	Médio	Médio	Baixa	Baixa	Alta
03r	Pequena	Médio	Médio	Média	Baixa	Alta
04r	Grande	Médio	Médio	Baixa	Média	Média
05r	Grande	Grande	Pequeno	Baixa	Alta	Baixa
06r	Média	Grande	Grande	Média	Baixa	Alta
07r	Média	Pequeno	Médio	Alta	Média	Baixa
08r	Pequena	Grande	Grande	Média	Baixa	Alta
09r	Pequena	Médio	Grande	Média	Baixa	Alta
10r	Média	Pequeno	Médio	Média	Baixa	Alta
11r	Média	Médio	Médio	Média	Média	Média
12r	Média	Médio	Médio	Média	Média	Média
13r	Média	Pequeno	Médio	Média	Baixa	Média
14r	Média	Grande	Médio	Alta	Alta	Média

exemplos é pequeno. Assim, para ajudar nessa análise define-se a **densidade** de um conjunto como:

$$\text{Densidade} = \frac{\text{Número de exemplos}}{\text{Tamanho do espaço de descrição}} \tag{7.1}$$

$$\text{Tamanho do espaço de descrição} = \prod_{i=1}^{n} N_i \tag{7.2}$$

onde n é o número de atributos e N_i é, para o i-ésimo atributo, o número de valores diferentes (se o atributo é nominal ou binário) ou o número de valores diferentes no conjunto de dados (se o atributo é contínuo).

A **predominância de classe** é a freqüência relativa da classe mais comum e reflete a dificuldade de um problema de classificação em algum sentido. **Entropia** relaciona a predominância de classe e a distribuição de classes do conjunto de dados e pode ser definida como:

$$\text{Entropia} = -\sum_{i=1}^{N} P(C_i) * log_2 P(C_i) \text{bits} \tag{7.3}$$

onde N é o número de classes e PC_i é a probabilidade da classe C_i

A entropia de um conjunto de variáveis também pode ser vista como um fator de incerteza deste conjunto. Assim, entende-se a entropia do conjunto de dados como sendo uma medida que quantifica o quão "misturados" estão os dados do conjunto.

Na Tabela 7.5 é apresentado um resumo da distribuição, em quantidade, dos conjuntos em relação aos intervalos estabelecidos para cada característica.

Tab. 7.5: Conjuntos de dados X características.

Dimensão			Tamanho			Classes		
Pequena	Média	Grande	Pequeno	Médio	Grande	Pequeno	Médio	Grande
11	9	4	4	11	9	9	12	3
Densidade			Predominância			Entropia		
Baixa	Média	Alta	Baixa	Média	Alta	Baixa	Média	Alta
6	14	4	10	11	3	3	13	8

7.1.1 Análise dimensional dos conjuntos de dados

Seis abordagens para inferência da dimensão adequada para o mapa de saída da RNA-AO foram analisadas: duas baseadas na Dimensão Fractal Fuzzy Significativa, duas baseadas na Dimensão Box-Counting e duas baseadas em Análise de Componentes Principais (PCA). Tendo em vista que as abordagens baseadas em dimensão fractal, em geral, inferem dimensões fracionárias, cada uma delas foi aproximada para um número inteiro de duas formas diferentes: aproximando-a para o maior inteiro ou arredondando-a.

A determinação da dimensão segundo a PCA foi feita com duas ordens de restrição para a escolha do número de componentes que seriam, supostamente, necessárias para melhor representar os dados e que, conseqüentemente, determinariam a dimensão adequada para a atuação da RNA-AO. As ordens de restrição representam a razão máxima tolerada entre os autovalores da matriz de covariância dos dados.

Cada uma das abordagens descritas foi aplicada a todos os conjuntos de dados, obtendo para cada qual seis sugestões sobre a dimensão ideal para o espaço de saída da RNA-AO.

Na Tabela 7.6 são indicadas as inferências obtidas[3]. Estes dados foram comparados com os resultados das análises feitas sobre as RNA-AOs, aplicadas aos mesmos conjuntos de dados, e são apresentados mais à frente neste capítulo. Note que o relaxamento das restrições aplicadas sobre o PCA levam a resultados menos significativos.

Tab. 7.6: Inferências sobre as dimensões ideais para o mapa de saída da RNA-AO.

Conjunto de dados	FF 1	FF 2	BC 1	BC 2	PCA (0.5)	PCA (0.9)
01s	3	3	2	2	5	5
02s	3	3	2	2	1	2
03s	4	3	2	2	1	1
04s	2	2	1	1	1	2
05s	3	2	2	2	1	2
06s	4	3	3	2	1	1
07s	4	3	3	2	1	1
08s	3	3	2	2	1	1
09s	2	2	1	1	1	1
10s	5	5	5	5	1	1
01r	2	2	2	1	1	2
02r	5	4	5	4	2	6
03r	3	2	1	1	1	5
04r	5	4	2	1	1	7
05r	6	6	1	1	2	23
06r	5	4	4	3	2	9
07r	6	6	2	2	2	6
08r	3	2	2	1	1	2
09r	3	3	2	2	1	3
10r	4	4	2	1	1	5
11r	5	4	2	1	3	7
12r	3	3	2	1	2	5
13r	4	3	2	2	1	6
14r	3	2	2	2	2	6

[3]Legenda para a abreviação utilizada nesta e em outras tabelas: FF 1 - dimensão Fractal Fuzzy Significativa aproximada para o maior inteiro; FF 2 - dimensão Fractal Fuzzy Significativa arredondada; BC 1 - dimensão Box-Counting aproximada para o maior inteiro; BC 2 - dimensão Box-Counting arredondada; PCA (0,5) - Análise de Componentes Principais com razão máxima de $0,5$ entre os autovalores da matriz de covariância; PCA (0,9) - Análise de Componentes Principais com razão máxima de $0,9$ entre os autovalores da matriz de covariância.

7.2 Projeto e treinamento das Redes Neurais Artificiais Auto Organizáveis

Juntamente com uma boa definição dos conjuntos de dados que serão usados para a análise de uma abordagem, é necessário fornecer informações que possibilitem a reprodução e a comparação dos resultados obtidos.

A fim de verificar a hipótese formulada no Capítulo 5, referente ao relacionamento entre a DFFS e otimalidade da escolha da dimensão do espaço de saída de uma RNA-AO, os conjuntos de dados descritos na seção anterior foram utilizados para treinar um conjunto de RNAs-AO.

Tal conjunto de RNAs-AO é composto por subconjuntos de mapas caracterizados pelo seu número de neurônios e pela dimensão topológica do seu espaço de saída. A máxima dimensão de espaço de saída utilizada foi 5. Esse número foi escolhido inicialmente de forma empírica, e após a realização das análises, constatou-se que, para a maioria dos casos, mapas de dimensões mais altas não seriam necessários.

O treinamento desses conjuntos de RNAs-AO variou quanto ao método de inicialização de pesos (linear ou randômica), sendo que para a randômica um mesmo mapa foi treinado três vezes, e quanto ao número de épocas (10, 100, 1000 e 10.000)[4].

O método de treinamento utilizado foi "em lote" e os *lattices* para o mapeamento foram o retangular e o diamante. A taxa de aprendizado foi inicializada com $0,5$ e decrementada linearmente até o valor mínimo de $0,05$. A função de vizinhança utilizada foi uma Gaussiana com variância inicializada segundo 7.4 e finalizando em 1 por meio de uma redução linear[5]

$$R_{init} = \begin{cases} max(1, max(msize)/2), & \text{se o método de inicialização de pesos for randômico.} \\ max(1, max(msize)/8), & \text{se o método de inicialização de pesos for linear.} \end{cases} \tag{7.4}$$

sendo que $msize$ é um vetor onde cada posição representa uma dimensão do espaço de saída da RNA-AO e o valor em cada posição (célula) do vetor é o número de neurônios existentes

[4] Considerou-se esses números de épocas devido à experiências citadas na literatura (FAUSETT (1994), SPECKMANN; RADDATZ; ROSENSTIEL (1994a) e ZUCHINI (2003)).

[5] Configuração padrão fornecida pela ferramenta utilizada para implementação das RNAs.

naquela dimensão e $max(msize)$ é a quantidade de neurônios presentes na dimensão do espaço de saída que contém o maior número de neurônios. Esses dois últimos parâmetros são sugeridos pela ferramenta utilizada para implementação, a SOM Toolbox (VESANTO et al. (2000)). Segundo PRECHELT (1994) o uso de formulações padrões é estimulado para simplificar a apresentação e compreensão dos resultados, reduzindo a variabilidade dos parâmetros de "*benchmark*" e facilitando futuros trabalhos de reprodução e/ou comparação.

Na Tabela 7.7 estão listadas as configurações dos grupos de RNAs-AO utilizadas para a análise dos conjuntos de dados. Para cada conjunto de dados, identificado na primeira coluna da tabela, aplicou-se RNAs-AO organizadas por número de neurônios, os quais estão listados na segunda coluna. Cada grupo com um determinado número de neurônios foi treinado sob 4 configurações diferentes de número de épocas (10, 100, 1000 e 10000 épocas), com inicialização de pesos linear e randômica, sendo que nesse último caso, 3 treinamentos foram realizados, constituindo 48 ou 64 sub-grupos de RNAs-AO, dependendo da variação do número de neurônios (veja a terceira coluna da tabela). Por fim, cada sub-grupo foi composto por 4 ou 5 RNAs-AO (dependendo do número de neurônios e da dimensão do espaço dos dados[6]), com variação na dimensão do seu espaço de saída. O número de RNAs-AO treinadas para cada conjunto de dados está listado na última coluna.

A escolha sobre os números de neurônios empregados foi baseado no número de instâncias do conjunto de dados. Procurou-se restringir o número de neurônios ao limite superior de metade do número de instâncias no conjunto de dados, não ultrapassando a quantidade de 256.[7] Assim justifica-se o uso de uma RNA-AO tanto como redutor de dimensão como um quantizador de conjunto.

As distribuições dos neurônios através das dimensões que caracterizam o espaço de saída das RNAs-AO foram realizadas com o intuito de distribuir os neurônios tão igualmente quanto fosse possível, visto que qualquer tipo de conhecimento referente a influência que cada dimensão do espaço exerce sobre o comportamento dos dados não é utilizado no processo de concepção ou treinamento das RNAs-AO. Na Tabela 7.8 estão disponíveis as configurações obtidas.

[6] A dimensão do espaço dos dados influencia na inicialização linear.

[7] Um número maior de neurônios leva a uma execução muito onerosa do algoritmo de avaliação que implementa o "Produto Topográfico". 256 neurônios foi considerado como o limite máximo para a realização desta análise.

Tab. 7.7: Configurações de grupos de RNAs-AO.

Id.	Variação no número de neurônios	Número de sub-grupos	Número de RNAs-AO treinadas - por conjunto de teste
01s a 03s	16,32,64,256	64	304
04s	16,32,64,256	64	260
05s a 08s	16,32,64,256	64	304
09s	16,32,64,256	64	260
10s	16,32,64,256	64	304
01r	16,32,64	48	216
02r	16,32,64,192	64	304
03r	16,32,64,96	64	304
04r	16,32,64,192	64	304
05r	16,32,64,256	64	304
06r	32,64,256	48	240
07r	16,32,64	48	224
08r	16,32,64,256	64	304
09r	16,32,64,192	64	304
10r	16,32,64,96	64	304
11r	16,32,64,144	64	304
12r	16,32,64,144	64	304
13r	16,32,64	48	224
14r	16,32,64,256	64	304

Tab. 7.8: Configurações de topologia de vizinhança (dimensão X número de neurônios).

Número de neurônios	Dimensão 1	Dimensão 2	Dimensão 3	Dimensão 4	Dimensão 5
16	[16,1]	[4,4]	[4,2,2]	[2,2,2,2]	X
32	[32,1]	[4,8]	[4,4,2]	[4,2,2,2]	[2,2,2,2,2]
64	[64,1]	[8,8]	[4,4,4]	[4,4,2,2]	[4,2,2,2,2]
96	[96,1]	[12,8]	[6,4,4]	[4,3,4,2]	[4,3,2,2,2]
144	[144,1]	[12,12]	[12,4,3]	[4,3,4,3]	[4,3,2,2,3]
192	[192,1]	[16,12]	[8,8,3]	[8,4,3,2]	[4,4,3,2,2]
256	[256,1]	[16,16]	[16,4,4]	[4,4,4,4]	[2,2,4,4,4]

A metodologia de distribuição de neurônios pelas dimensões do espaço de saída, garante que,

ao menos pelas condições iniciais e na inicialização linear, os mapeamentos gerados representarão, de forma real, a dimensão desejada. Uma configuração com distribuição não uniforme, pode levar a um mapeamento com características similares a um mapeamento de menor dimensão. Por exemplo, o mapeamento realizado em uma configuração [50,2] pode ter resultados semelhantes ao mapeamento realizado em uma configuração [100,1], representando uma estrutura 1-dimensional. Ao passo que, uma configuração [10,10] tem maiores chances de, após o mapeamento, representar uma estrutura 2-dimensional.

7.2.1 Avaliação das Redes Neurais Artificiais Auto Organizáveis

Cada um dos mapeamentos foi avaliado sob quatro medidas de qualidade: Produto Topográfico, Erro Topológico com vizinhança em diamante, Erro Topológico com vizinhança retangular e Erro de Quantização. A Tabela 7.9 representa o formato geral de um sub-conjunto de RNA com as medidas de avalição de qualidade.

Tab. 7.9: Formato das informações de qualidade referentes a um sub-grupo de RNA-AO.

	Conjunto de dados	XX		
Parâmetros configuráveis: inicialização: *linear*	número de neurônios: 144	número de épocas: 100		
Dimensão do Espaço de saída	P	ξ^d	ξ^r	E_q
1-dimensional	-0.0340	0.0000	0.5055	0.0000
2-dimensional	-0.0162	0.0231	0.4277	0.0297
3-dimensional	0.0003	0.0825	0.4978	0.0066
4-dimensional	0.0039	0.0462	0.5380	0.0000
5-dimensional	0.0134	0.0099	0.6108	0.0000

As informações sobre qualidade, obtidas sob tais medidas, foram analisadas mediante diferentes critérios com o intuito de descobrir em qual das 4 ou 5 configurações de dimensão do espaço de saída obteve-se os melhores resultados em relação à qualidade quantizadora e de preservação de relações topológicas. A razão para criação de vários critérios de análise das medições de qualidade do mapamento é que as medidas podem ser vistas sob diferentes aspec-

tos, e esses foram então utilizados na forma de "critérios de avaliação", os quais estão descritos e justificados a seguir.

- **Critério 1 ($P+$): menor Produto Topográfico não negativo.** Medidas negativas para o Produto Topográfico são obtidas quando as maiores distâncias são observadas no espaço de entrada da RNA-AO, ou seja, o espaço de saída está contorcido revelando que sua dimensão é baixa para aproximar o conjunto de dados. Medidas positivas apontam a situação contrária e, sob este aspecto, o mapeamento com medidas positivas para o Produto Topográfico é o mais adequado.

- **Critério 2 ($\mid P \mid$): menor Produto Topográfico em módulo.** Quanto mais próxima de zero estiver essa medida menos contorções foram observadas no mapeamento, em qualquer um de seus espaços. Independentemente de qual dos espaço se contraia, na RNA-AO em que a contração é menor é que se encontra o melhor mapeamento topográfico.

- **Critério 3 (ξ_2): menor Erro Topológico com precisão de duas casas decimais.** Segundo o Erro Topológico, quanto menores forem as distorções, menor será o número obtido na medição. Este critério foi dividido em duas versões: a) considerando a vizinhança em diamente (ξ_2^d); b) considerando a vizinhança retangular (ξ_2^r)[8].

- **Critério 4 ($2\xi_2$): segundo menor Erro Topológico com precisão de duas casas decimais.** Medidas de erro topológico muito pequenas (próximas a zero) indicam que quase não houve contorções no mapeamento. Este fato pode indicar uma rigidez no mapa, principalmente se acompanhado de um Erro de Quantização alto (fenômeno do sub-ajuste). Assim, optou-se por também realizar as análises sob o segundo menor valor obtido para essa medida. As vizinhanças em diamante ($2\xi_2^d$) e retangular ($2\xi_2^r$) foram consideradas nestes critério.

- **Critério 5 (ξ_4): menor Erro Topológico com precisão de quatro casas decimais.** Idem ao critério 3, porém com maior precisão.

[8]Nesta estrutura de vizinhança o número de vizinhos de um neurônio aumenta consideravelmente com o aumento da dimensão do mapa.

- **Critério 6 ($2\xi_4$): segundo menor Erro Topológico com precisão de quatro casas decimais.** Idem ao critério 4, porém com maior precisão.

- **Critério 7 ($< E_q$): menor Erro de Quantização.** O mapeamento com menor Erro de Quantização é aquele que melhor aproxima os seus neurônios aos dados do conjunto sob análise. O fenômeno de sobre-ajuste é comum aparecer em mapeamentos de Erro de Quantização muito baixos, entretanto, apenas quando o número de neurônios é muito próximo ou maior que o número de dados. Esse não é o caso de nenhum dos testes deste trabalho. Para tal critério, as variações referentes à precisão em casas decimais não apresentou resultados diferentes.

- **Critério 8 ($\mid P \mid \succ \mu E_q$): Produto Topográfico $\succ$ Erro de Quantização.** Critério criado como uma variação da heurística de avaliação utilizada em ZUCHINI (2003). Dentre os três menores valores obtidos para o Produto Topográfico, foi escolhido o mapeamento com Erro de Quantização mais próximo da média do Erro de Quantização dos três mapeamentos envolvidos. Para esse critério, as variações referentes à precisão em casas decimais não apresentou resultados diferentes.

- **Critério 9 ($\xi^d \succ \mu E_q$): Erro Topológico $\succ$ Erro de Quantização.** Critério criado como uma variação da heurística de avaliação utilizada em ZUCHINI (2003). Dentre os três menores valores obtidos para o Erro Topológico foi escolhido o mapeamento com o Erro de Quantização mais próximo da média do Erro de Quantização dos três mapeamentos envolvidos. Apenas a vizinhança em diamante foi considerada[9]. Para esse critério, as variações referentes à precisão em casas decimais não apresentou resultados diferentes.

- **Critério 10 ($< E_q \succ P$): Erro de Quantização $\succ$ Produto Topográfico.** Nesse critério, dentre os três menores valores obtidos para o Erro de Quantização, foi escolhido como melhor mapeamento aquele que apresentou o melhor resultado segundo o produto topográfico. A obtenção do melhor resultado foi dividida em duas versões: a) menor valor não negativo para o Produto Topográfico ($< E_q \succ P+$); b) menor valor em módulo para

[9] A vizinhança retangular não foi considerada porque eleva exponencialmente o número de neurônios vizinhos quando a dimensão do espaço de saída aumenta.

o Produto Topográfico ($< E_q \succ \mid P \mid$).

- **Critério 11 ($< E_q \succ \xi$): Erro de Quantização $\succ$ Erro Topológico.** Nesse critério, dentre os três menores valores obtidos para o Erro de Quantização, foi escolhido como melhor mapeamento aquele que apresentou: a) o menor valor para o Erro Topológico ($< E_q \succ \xi^d$); b) o segundo menor valor para o Erro Topológico. Apenas a vizinhança em diamante foi considerada ($< E_q \succ 2\xi^d$).

- **Critério 12 ($\mid P \mid \succ \pm E_q$): Produto Topográfico $\succ$ Erro de Quantização Intermediário.** Este critério é mais uma variação da heurística de avaliação de RNAs-AO utilizada por ZUCHINI (2003). Nesse critério, dentre os três menores valores obtidos para o Produto Topográfico, foi escolhido como melhor mapeamento aquele que apresentou um valor intermediário para o Erro de Quantização.

- **Critério 13 ($\xi^d \succ \pm E_q$): Erro Topológico $\succ$ Erro de Quantização Intermediário.** Esse critério é a heurística de avaliação de RNAs-AO utilizada em ZUCHINI (2003). Nesse critério, dentre os três menores valores obtidos para o Erro Topológico, foi escolhido como melhor mapeamento aquele que apresentou um valor intermediário para o Erro de Quantização. Apenas a vizinhança em diamante foi considerada.

Dado que diversos critérios de avaliação foram empregados e que cada um oferece uma interpretação diferente das análises de qualidade feitas sobre os mapeamentos, conclusões diferentes sobre a dimensão ideal para os espaços de saída foram obtidas (como pode ser observado nas discussões da Seção 7.3).

Neste contexto faz-se necessário estabelecer um "*ranking*" referente à qualidade destes critérios. Para isso, tomou-se o critério $< E_q$ como base, já que ele é o que melhor reflete o desempenho da RNA-AO em relação à quantização do espaço (ou descoberta de grupos). Os resultados obtidos sob esse critério foram comparados aos resultados obtidos sob o ponto de vista dos demais. Na Tabela 7.10 está o resultado obtido nestas comparações e o "*ranking*" criado está representado na ordem de apresentação dos critérios na tabela (primeira coluna). Na segunda coluna está o número de conjuntos de dados sintéticos para os quais a dimensão sugerida sob o critério em análise (na linha) coincidiu com a dimensão indicada pelo critério

base, por exemplo: o critério 11b avaliou como mais adequadas, em 8 conjuntos de dados sintéticos, as mesmas dimensões de espaços de saída que o critério base julgou serem as melhores. Na terceira coluna está o resultado referente aos conjuntos de dados reais e na última coluna, o número total de indicações coincidentes.

Tab. 7.10: Ranking de critérios (Análise geral).

Critérios		Conjuntos Sintéticos	Conjuntos Reais	Total
11b	$< E_q \succ 2\xi^d$	8	8	16
10a	$< E_q \succ P+$	8	6	14
4a	$2\xi_2^d$	3	9	12
4b	$2\xi_2^r$	3	8	11
6a	$2\xi_4^d$	3	7	10
1	$P+$	6	3	9
2	$\mid P \mid$	5	4	9
10b	$< E_q \succ \mid P \mid$	3	6	9
6b	$2\xi_4^r$	1	5	6
8	$\mid P \mid \succ \mu E_q$	1	5	6
12	$\mid P \mid \succ \pm E_q$	2	4	6
3b	ξ_2^r	3	2	5
11a	$< E_q \succ \xi^d$	3	2	5
13	$\xi^d \succ \pm E_q$	1	4	5
3a	ξ_2^d	2	2	4
5a	ξ_4^d	2	2	4
5b	ξ_4^r	2	2	4
9	$\xi^d \succ \mu E_q$	0	4	4

A análise mostrada na Tabela 7.10 fornece um indicativo de quais seriam os melhores critérios para realizar a avaliação da dimensão do espaço de saída para uma RNA-AO, a qual pretende quantizar o espaço dos dados satisfatoriamente e reduzir sua dimensão com perda mínima de informação sobre as sua relações topológicas. A mesma análise foi realizada com um sub-conjunto de RNAs-AO escolhido mediante a heurística de Vesanto[10], que estabelece o número ideal de neurônios para compor o mapa, calculado a partir do número de elementos do

[10]Neste trabalho, por simplificação, está-se denominando a heurística apresentada em VESANTO et al. (2000) como heurística de Vesanto.

conjunto de dados sob análise. A relação é fazer o número de neurônios ser cinco vezes a raiz quadrada do número de dados do conjunto. O segundo "*ranking*" está mostrado na Tabela 7.11 e é um pouco diferente do primeiro, mas na escolha dos melhores critérios existe um alto grau de coincidências entre os dois.

Tab. 7.11: Ranking de critérios com conjuntos escolhidos sob a heurística de Vesanto.

Critérios		Conjuntos Sintéticos	Conjuntos Reais	Total
10b	$< E_q \succ \mid P \mid$	6	7	13
11b	$< E_q \succ 2\xi^d$	3	9	12
2	$\mid P \mid$	7	5	12
4b	$2\xi_2^r$	3	8	11
4a	$2\xi_2^d$	4	7	11
10a	$< E_q \succ P+$	7	3	10
6a	$2\xi_4^d$	3	6	9
11a	$< E_q \succ \xi^d$	5	4	9
6b	$2\xi_4^r$	2	6	8
1	$P+$	4	2	6
3a	ξ_2^d	0	2	2
3b	ξ_2^r	0	2	2
5a	ξ_4^d	0	2	2
5b	ξ_4^r	0	2	2
9	$\xi^d \succ \mu E_q$	0	1	1
12	$\mid P \mid \succ \pm E_q$	1	0	1
13	$\xi^d \succ \pm E_q$	0	1	1
8	$\mid P \mid \succ \mu E_q$	0	0	0

Nesse cenário, serão considerados como melhores critérios para medir a adequação das abordagens estudadas, os seguintes:

- Análise geral: $P+$, $\mid P \mid$, $2\xi_2^d$, $2\xi_2^r$, $2\xi_4^d$, $< E_q \succ P+$, $< E_q \succ \mid P \mid$ e $< E_q \succ 2\xi^d$.
- Análise combinada à heurística de Vesanto: $\mid P \mid$, $2\xi_2^d$, $2\xi_2^r$, $2\xi_4^d$, $< E_q \succ P+$, $< E_q \succ \mid P \mid$, $< E_q \succ \xi^d$ e $< E_q \succ 2\xi^d$.

Nas Figuras 7.1 e 7.2 pode ser observado um resumo sobre as características que são avaliadas considerando cada um dos critérios selecionados como melhores. Observe que dos critérios

que utilizam a medida Produto Topográfico, aqueles que a consideram como principal fator de avaliação foram considerados melhores. Na Figura 7.1 estão apontados por flechas aqueles que utilizam o Produto Topográfico não negativo e circulados aqueles que usam o menor Produto Topográfico em módulo. Já dos critérios que utilizam o Erro Topológico como principal fator de análise, os melhores são, em sua maioria, aqueles que consideram o segundo menor valor (circulados na Figura 7.2). Na mesma figura, o critério apontado por uma flecha é o único que utiliza o menor valor de Erro Topológico.

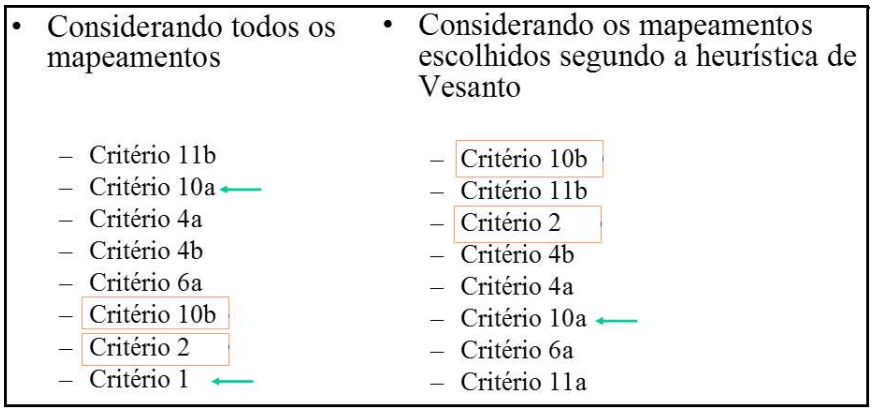


Fig. 7.1: Características dos melhores critérios que utilizam o Produto Topográfico como principal fator de avaliação de uma RNA-AO. Os critérios apontados por flechas utilizam o Produto Topográfico não negativo. Os critérios circulados usam o menor Produto Topográfico em módulo.

7.3 Análise das inferências sobre a dimensão das Redes Neurais Artificiais Auto Organizáveis

Como apresentado na Seção 7.1.1, cada abordagem de inferência da dimensão ideal para o espaço de saída da RNA-AO foi aplicada aos conjuntos de dados escolhidos para a realização dos testes da hipótese desta tese. E, como apresentado na Seção 7.2.1, para cada conjunto de dados todos os mapeamentos sobre eles realizados foram analisados segundo as medidas de qualidade e o resultado desta análise sofreu a avaliação sob critérios específicos. Também, como mostrado na Tabela 7.6, cada conjunto de dados foi analisado pelas seis abordagens afim de saber qual

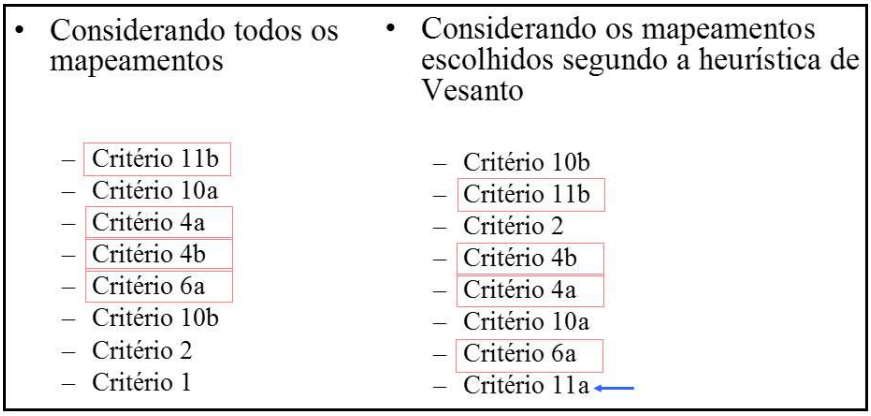


Fig. 7.2: Características dos melhores critérios que utilizam o Erro Topológico como principal fator de avaliação de uma RNA-AO. O critério apontado por uma flecha utiliza o menor valor de Erro Topológico. Os critérios circulados utilizam o segundo menor valor de Erro Topológico.

a sugestão destas em relação à dimensão ideal para o espaço de saída da RNA-AO. Essas duas análises (de qualidade de mapeamento e de inferência de dimensão) foram confrontadas para verificar as coincidências, como mostrado suscintamente na Tabela 7.12 (o número nas células são as dimensões sugeridas e em negrito estão as coincidências). O número de coincidências atesta o desempenho da abordagem e é discutido a seguir.

Tab. 7.12: Exemplo de comparação do resultado das abordagens com o resultado apontado pelo critério 1. Legenda: FF 1 - dimensão Fractal Fuzzy Significativa aproximada para o maior inteiro; FF 2 - dimensão Fractal Fuzzy Significativa arredondada; BC 1 - dimensão Box-Counting aproximada para o maior inteiro; BC 2 - dimensão Box-Counting arredondada; PCA (0,5) - Análise de Componentes Principais com razão máxima de $0,5$ entre os autovalores da matriz de covariância; PCA (0,9) - Análise de Componentes Principais com razão máxima de $0,9$ entre os autovalores da matriz de covariância.

Conjunto	FF 1	FF 2	BC 1	BC 2	PCA (0.5)	PCA (0.9)	Critério 1
Norm3a	4	**3**	2	2	1	1	3
Norm4a	**2**	**2**	1	1	1	**2**	2
Abalone	3	**2**	**2**	1	1	**2**	2
Auto-mpg	**3**	**3**	2	2	1	**3**	3

De forma mais específica, o exemplo na primeira linha da Tabela 7.12 revela que o critério de avaliação 1 ($P+$) inferiu que, das RNAs-AO treinadas sobre o conjunto de dados "Norm3a", as que alcançaram melhores resultados foram, em sua maioria, aquelas que tinham seus espaços de

saída organizados topologicamente dentro de um espaço 3-dimensional. Esse exemplo mostra também que a DFFS obtida para esse conjunto de dados, depois de arredondada, indica o espaço 3-dimensional como sendo onde, provavelmente, estão localizadas as relações de vizinhança dos "grupos" existentes em tal conjunto. Ou seja, das seis abordagens aplicadas, apenas a $FF2$ inferiu o mesmo espaço dimensional que o critério 1, para o conjunto de dados "Norm3". Comparações dessa natureza foram feitas para os 24 conjuntos de dados, em relação aos 19 critérios e sob as 6 abordagens. Contudo, mais do que uma simples comparação quantitativa, a análise do desempenho das abordagens foi realizada com dois objetivos:

- Atestar a abordagem de inferência que melhor se adequa, no sentido de inferir a dimensão adequada para um maior número de casos (análise de desempenho geral) e em relação a quais critérios de avaliação;
- Analisar em quais situações, ou para quais tipos de conjuntos de dados, a aplicação de cada abordagem de inferência é mais apropriada (análise de desempenho por características de conjuntos de dados) e sob quais critérios de avaliação.

7.3.1 Análise de desempenho geral

Duas formas para análise do desempenho geral das abordagens são tratadas aqui. Ei-las:

- A primeira discute a relação existente entre cada uma das seis abordagens e cada um dos dezenove critérios. A discussão é embasada na análise de "quais" critérios apontam como dimensão ideal aquela inferida por uma abordagem específica (desempenho das abordagens por critério de avaliação).
- A segunda analisa o "número" de coincidências entre as inferências realizadas pelas abordagens e as constatações obtidas sob a aplicação dos critérios (desempenho das abordagens num âmbito geral).

Desempenho das Abordagens por Critério de Avaliação

Nesta primeira forma de análise foi possível observar que cada abordagem se destaca para diferentes critérios de avaliação dos mapeamentos. Os destaques ocorrem da forma listada a

seguir, com os melhores critérios acompanhados de um *. A letra "V" junto da identificação do critério indica que ele foi aplicado apenas para os mapeamentos que obedeciam à heurística de Vesanto. O intuito desse procedimento, aplicado em grande parte das análises desta tese, é verificar o potencial de união das duas heurísticas (Vesanto e abordagem FF). Para clarear a interpretação das informações desta listagem, considere o primeiro caso: numa análise onde todos os conjuntos de dados foram considerados, as inferências obtidas pela abordagem Fractal Fuzzy (maior inteiro) estão de acordo, principalmente, com os critérios $\mid P \mid \succ \mu E_q$V, $\xi^d \succ \mu E_q$V, $\mid P \mid \succ \pm E_q$, $\xi^d \succ \pm E_q$V.

1. Considerando todos os conjuntos de dados:

   - Fractal Fuzzy (maior inteiro): $\mid P \mid \succ \mu E_q$V, $\xi^d \succ \mu E_q$V, $\mid P \mid \succ \pm E_q$, $\xi^d \succ \pm E_q$V
   - Fractal Fuzzy (arredondado): $P+$*, $P+$V*, $2\xi_2^d$*, $2\xi_2^d$V*, $2\xi_2^r$V*, $2\xi_4^d$, $2\xi_4^d$V, $\mid P \mid \succ \mu E_q$, $< E_q \succ P+$*, $< E_q \succ P+$V*, $\mid P \mid \succ \pm E_q$V, $\xi^d \succ \pm E_q$V
   - Box-Counting (maior inteiro): $2\xi_2^d$*, $2\xi_4^r$, $2\xi_4^r$V, $< E_q$*, $< E_q$V*, $\mid P \mid \succ \mu E_q$, $\xi^d \succ \mu E_q$, $< E_q \succ P+$*, $< E_q \succ \mid P \mid$V*, $< E_q \succ 2\xi^d$*, $< E_q \succ 2\xi^d$V*, $\xi^d \succ \pm E_q$
   - Box-Counting (arredondado): $\mid P \mid$V*, $< E_q \succ \xi^d$V*, $\xi^d \succ \pm E_q$, $\xi^d \succ \pm E_q$V
   - PCA (0,5): ξ_2^d, ξ_2^dV, ξ_2^r, ξ_2^rV, $2\xi_2^r$*, ξ_4^d, ξ_4^dV, ξ_4^r, ξ_4^rV, $< E_q \succ \mid P \mid$*, $< E_q \succ \xi^d$*
   - PCA (0,9): $P+$V*, $\mid P \mid$, $< E_q$*, $< E_q \succ P+$*, $< E_q \succ \mid P \mid$*

2. Considerando os conjuntos sintéticos:

   - Fractal Fuzzy (maior inteiro): $< E_q \succ \xi^d$V*, $\mid P \mid \succ \pm E_q$
   - Fractal Fuzzy (arredondado): $P+$*, $2\xi_2^d$*, $2\xi_2^d$V*, $2\xi_2^r$*, $2\xi_2^r$V*, $2\xi_4^d$, $2\xi_4^d$V, $2\xi_4^r$V, $\mid P \mid \succ \mu E_q$V, $\xi^d \succ \mu E_q$V, $< E_q \succ \xi^d$V*, $< E_q \succ 2\xi^d$V*, $\mid P \mid \succ \pm E_q$V, $\xi^d \succ \pm E_q$, $\xi^d \succ \pm E_q$V
   - Box-Counting (maior inteiro): $2\xi_4^r$, $< E_q$*, $\xi^d \succ \mu E_q$, $< E_q \succ P+$V*, $< E_q \succ \xi^d$V*, $< E_q \succ 2\xi^d$*
   - Box-Counting (arredondado): $\mid P \mid$V*, $< E_q$*, $< E_q$V*, $\mid P \mid \succ \mu E_q$, $< E_q \succ P+$V*, $< E_q \succ \mid P \mid$V*, $< E_q \succ 2\xi^d$*
   - PCA (0,5): ξ_2^d, ξ_2^dV, ξ_2^r, ξ_2^rV, ξ_4^d, ξ_4^dV, ξ_4^r, ξ_4^rV, $< E_q \succ \xi^d$*

- PCA (0,9): P+V*, $\mid P \mid$*, $< E_q$*, $< E_q \succ P$+*, $< E_q \succ \mid P \mid$*

3. Considerando os conjuntos reais:

   - Fractal Fuzzy (maior inteiro): $\mid P \mid \succ \mu E_q$V, $\xi^d \succ \mu E_q$V, $\mid P \mid \succ \pm E_q$, $\mid P \mid \succ \pm E_q$V
   - Fractal Fuzzy (arredondado): P+*, P+V*, $2\xi_2^d$*, $2\xi_2^d$V*, $2\xi_4^d$, $2\xi_4^d$V, $< E_q \succ P$+*, $< E_q \succ P$+V*, $\mid P \mid \succ \pm E_q$V
   - Box-Counting (maior inteiro): $\mid P \mid$*, $\mid P \mid$V*, $2\xi_2^d$*, $2\xi_2^d$V*, $2\xi_2^r$V*, $2\xi_4^r$V, $< E_q$V*, $\mid P \mid \succ \mu E_q$, $\xi^d \succ \mu E_q$, $< E_q \succ P$+*, $< E_q \succ \mid P \mid$*, $< E_q \succ \mid P \mid$V*, $< E_q \succ 2\xi^d$*, $< E_q \succ 2\xi^d$V*, $\xi^d \succ \pm E_q$
   - Box-Counting (arredondado): $\mid P \mid$*, $\mid P \mid$V*, ξ_2^d, ξ_2^dV, ξ_2^r, ξ_2^rV, ξ_4^d, ξ_4^dV, ξ_4^r, ξ_4^rV, $\mid P \mid \succ \mu E_q$V, $\xi^d \succ \mu E_q$V, $< E_q \succ \xi^d$*, $< E_q \succ \xi^d$V*, $\xi^d \succ \pm E_q$, $\xi^d \succ \pm E_q$V
   - PCA (0,5): $2\xi_2^r$*, $2\xi_4^r$, $< E_q$*, $\mid P \mid \succ \mu E_q$V, $\xi^d \succ \mu E_q$V, $< E_q \succ \xi^d$V*
   - PCA (0,9): P+V*, $\mid P \mid$*, $< E_q \succ P$+*

Essa análise mostra que três abordagens conseguem boas inferências em comparação a vários critérios: Fractal Fuzzy (arredondado), Box-Counting (maior inteiro) e PCA (0,5). Essa última tem o seu destaque pautado, principalmente, em critérios que têm em comum o uso do menor Erro Topológico como parâmetro para atestar o bom mapeamento. Contudo, na realização das análises dos mapeamentos percebeu-se que os menores Erros Topológicos eram, geralmente, iguais ou muito próximos a zero e obtidos, para a grande maioria dos conjuntos de dados, nas RNAs-AO com espaços de saída 1-dimensionais. Tal fato levou à suposição da ocorrência do fenômeno de sub-ajuste sobre esses mapeamentos e da influência desse fenômeno na medida de qualidade[11]. Além disso, o estabelecimento do "ranking" de critérios mostrou que as conclusões sobre a boa capacidade quantizadora são melhores nos mapeamentos que obtiveram os segundos menores Erros Topológicos.

Os desempenhos das abordagens na análise envolvendo todos os conjuntos mostram que os critérios para os quais a abordagem Fractal Fuzzy (arredondado) apresenta bons resultados estão, em sua maioria, relacionados a critérios de avaliação de preservação de topologia. Isso mostra que existe uma tendência da abordagem Fractal Fuzzy em indicar a dimensão onde

[11] Como constatado também nas análises realizadas em ZUCHINI (2003).

as relações topológicas são preservadas, principalmente, sob a avaliação da medida Produto Topográfico (observe na listagem acima a presença dos critérios $P+$, $P+\mathrm{V}$, $\mid P \mid\succ \mu E_q$, $< E_q \succ P+$, $< E_q \succ P+\mathrm{V}$ e $\mid P \mid\succ \pm E_q\mathrm{V}$). A abordagem Box-Counting apresenta o diferencial de conseguir bom desempenho para os critérios $< E_q$ e $< E_q\mathrm{V}$ que analisam apenas a qualidade de quantização. Observa-se ainda que a abordagem Fractal Fuzzy (arredondado) tem um bom desempenho para os conjuntos sintéticos. Para os conjuntos reais o maior destaque fica para a abordagem Box-Counting (maior inteiro).

A abordagem Fractal Fuzzy (arredondado) apresenta um desempenho diferenciado das demais também quando se trata do uso da heurística de Vesanto. Isso mostra que as redes com o número de neurônios indicado pela heurística de Vesanto, combinado ao uso da dimensão de espaço de saída sugerida pela abordagem Fractal Fuzzy (arredondado) têm boas avaliações.

Desempenho das Abordagens num Âmbito Geral

A Tabela 7.13 contém os resultados obtidos na segunda forma de análise e estudados sob diferentes situações (observe as especificações de situações presente na primeira coluna da tabela. A porcentagem exibida em cada célula diz respeito ao número de critérios segundo os quais a abordagem em análise é a que infere a dimensão adequada para o maior número de conjuntos de dados. Por exemplo, das abordagens analisadas, considerando a situação de análise que usa todos os critérios e os conjuntos de dados sintéticos, a abordagem Fractal Fuzzy (arredondado) e a abordagem PCA (0,5) foram as que inferiram a dimensão correta mais vezes em relação a cinco critérios de avaliação ($26,31\%$ dos critérios). Já a abordagem Fractal Fuzzy (maior inteiro) teve o pior desempenho nesta mesma situação, obtendo boas inferências para a maior parte dos conjuntos de dados apenas com relação a um critério ($5,26\%$).

Os resultados apresentados na Tabela 7.13 foram originados de uma série de constatações sobre o desempenho das abordagens, em relação aos conjuntos de dados sob comparação com cada um dos critérios. Por exemplo, dizer que a abordagem Fractal Fuzzy (arredondado) realizou uma boa inferência sob comparação com $26,31\%$ dos critérios significa dizer que esta abordagem conseguiu fazer boas inferências para mais conjuntos de testes sintéticos que as demais abordagens, acertando a inferência para 4, 5 ou 6 conjuntos de dados (dos 10 existentes)

Tab. 7.13: Análise geral realizada sob diferentes aspectos, combinando: todos os critérios (19), os melhores critérios (9), todos os conjuntos (24), apenas os conjuntos sintéticos (10), apenas os conjuntos reais (14) e a heurística de Vesanto. TC - todos os critérios, MC - melhores critérios, V - uso da heurística de Vesanto, CS - conjuntos sintéticos, CR - conjuntos reais e CSR - todos os conjuntos. Legenda: FF 1 - dimensão Fractal Fuzzy Significativa aproximada para o maior inteiro; FF 2 - dimensão Fractal Fuzzy Significativa arredondada; BC 1 - dimensão Box-Counting aproximada para o maior inteiro; BC 2 - dimensão Box-Counting arredondada; PCA (0,5) - Análise de Componentes Principais com razão máxima de $0,5$ entre os autovalores da matriz de covariância; PCA (0,9) - Análise de Componentes Principais com razão máxima de $0,9$ entre os autovalores da matriz de covariância.

Situações de análise	FF 1	FF 2	BC 1	BC 2	PCA (0.5)	PCA (0.9)
TC X CS	5,26	**26,31**	21,05	15,78	**26,38**	21,05
TC X CR	5,26	21,05	**42,10**	36,84	15,78	0
TC X CSR	5,26	26,31	**42,10**	5,26	36,84	10,52
MC X CS	0	**44,44**	22,22	22,22	0	**44,44**
MC X CR	0	44,44	**55,55**	11,11	22,22	0
MC X CSR	0	**44,44**	**44,44**	0	22,22	11,11
TC X CS X V	5,26	**52,63**	10,52	21,05	21,05	5,26
TC X CR X V	15,78	26,31	36,84	**47,36**	15,78	0
TC X CSR X V	10,52	**31,57**	21,05	10,52	21,05	5,26
MC X CS X V	11,11	**55,55**	22,22	44,44	0	0
MC X CR X V	0	33,33	**66,66**	22,22	11,11	0
MC X CSR X V	0	**44,44**	**44,44**	22,22	0	0

mediante 5 dos 19 critérios utilizados para comparação. Isto também significa que para a maior parte dos sub-grupos de RNAs-AO treinadas para esses conjuntos de dados (totalizando, em média, 304 RNAs-AO por conjunto de dados), aquelas com espaço de saída de dimensão igual à dimensão Fractal Fuzzy (arredondado) foram as que alcançaram melhores resultados de quantização e/ou preservação de topologia sob a interpretação destes 5 critérios.

Esta forma de análise chamou a atenção para o bom desempenho das abordagens Fractal Fuzzy (arredondado) e Box-Counting (maior inteiro). A abordagem Fractal Fuzzy (arredondado) mostrou um bom desempenho para os conjuntos sintéticos em relação a todas as situações de comparações: todos os critérios, melhores critérios e também quando se leva em consideração a heurística de Vesanto. Quando são analisados os conjuntos reais, a abordagem Box-Counting apresenta um desempenho melhor. Contudo, na análise geral, a abordagem Box-Counting só

se destaca sobre a abordagem Fractal Fuzzy para o caso de análises que consideram todos os conjuntos de dados sob todos os critérios. Nota-se a partir desta análise que a utilização de PCA não foi considerada uma boa heurística para a estimativa em questão.

7.3.2 Análise de desempenho por características

Foram realizadas análises específicas, em relação a cada característica escolhida como descritora dos conjuntos de dados, para identificar aquelas que quando presentes no conjunto estimulariam a aplicação de uma ou outra abordagem.

Por exemplo, para os conjuntos com poucos dados (4 conjuntos) e em relação à avaliação do critério $P+$, a abordagem Fractal Fuzzy (arredondado) e PCA (0,9) conseguiram as inferências corretas em 2 conjuntos, as abordagens Box Counting (maior inteiro) e Fractal Fuzzy (maior inteiro) obtiveram sucesso em apenas um dos conjuntos e por fim, as abordagens Box-Counting (arredondado) e PCA (0,5) não obtiveram qualquer sucesso de inferência. Se o critério em questão fosse o único parâmetro de avaliação, concluiria-se que as duas abordagens que obtiveram sucesso para 2 conjuntos seriam as mais adequadas para aplicação mediante a características "poucos dados". No entanto, é interessante fazer esse estudo mediante vários critérios e obter um panorama do desempenho de cada abordagem em relação ao conjunto de avaliações. Para tal escolheu-se utilizar o conjunto de "melhores critérios" de avaliação (combinados ou não com a heurística de Vesanto).

Como resultado tem-se as informações listadas nas Tabelas 7.14 - análise geral - e 7.15 - análise combinada com a heurística de Vesanto. Nas células dessas tabelas está representado o número de critérios sob os quais a abordagem obteve destaque no número de inferências corretas, em relação a cada característica. Ou seja, considerando a características "número de dados pequenos", a abordagem Box-Counting (maior inteiro) foi a que mais inferiu dimensões corretas em relação a cinco critérios (dos 9 utilizados). As demais (com exceção dos casos em que o número na célula é 0) obtiveram algum destaque no número de inferências corretas, porém em relação a um número menor de critérios. Em negrito estão os melhores desempenhos (veja o Apêndice C para conhecer a forma de tabulação de dados que permitiu essas conclusões).

Analisando as Tabelas 7.14 e 7.15 é possível constatar que, tanto considerando todos os

Tab. 7.14: Análise por características (melhores critérios).

Situações de análise	FF 1	FF 2	BC 1	BC 2	PCA (0.5)	PCA (0.9)
Número de Dados						
Pequeno	1	2	**5**	2	1	1
Médio	0	**5**	2	0	2	1
Grande	1	3	3	2	3	**5**
Número de Atributos						
Pequeno	0	**5**	2	1	1	1
Médio	0	2	2	1	**4**	0
Grande	0	0	2	0	**8**	3
Número de Classes						
Pequeno	2	**3**	**3**	**3**	1	**3**
Médio	1	3	2	0	**5**	2
Grande	0	1	**5**	4	2	0
Densidade						
Baixa	0	0	3	1	**5**	3
Média	0	4	**5**	2	1	0
Alta	**5**	**5**	2	1	2	2
Entropia						
Baixa	**4**	3	**4**	2	3	2
Média	2	1	**4**	0	2	1
Alta	0	**3**	**3**	0	**3**	2
Predominância de Classes						
Baixa	0	2	**5**	2	3	2
Média	2	**4**	**4**	1	2	2
Alta	**4**	2	2	2	**4**	1

mapeamentos quanto considerando apenas aqueles escolhidos sobre a heurística de Vesanto, as abordagens Fractal Fuzzy (arredondado) e a Box-Counting(maior inteiro) são as que se mostram adequadas para a aplicação na presença da maior parte das características. A abordagem PCA (0.5) também se destaca, mas somente quando a heurística de Vesanto não é considerada. Fazendo então uma análise restrita às abordagens Fractal Fuzzy e Box-Counting, verifica-se que, sob a análise geral dos mapeamentos, a abordagem Box-Counting é a que deve ser aplicada sob a presença da maior parte das características (9 características) e a abordagem Fractal Fuzzy se destaca quando a heurística de Vesanto é utilizada (11 características).

Contudo, para obter uma indicação final para a escolha da abordagem a ser aplicada,

Tab. 7.15: Análise por características (melhores critérios combinados com a heurística de Vesanto).

Situações de análise	FF 1	FF 2	BC 1	BC 2	PCA (0.5)	PCA (0.9)
Número de Dados						
Pequeno	2	**4**	**4**	**4**	2	1
Médio	1	3	**4**	2	0	0
Grande	0	**3**	**3**	**3**	1	1
Número de Atributos						
Pequeno	0	**5**	3	2	0	0
Médio	0	**4**	**4**	3	2	0
Grande	0	2	4	3	**5**	0
Número de Classes						
Pequeno	**5**	4	4	**4**	0	1
Médio	0	4	**5**	3	1	0
Grande	0	2	**7**	2	1	0
Densidade						
Baixa	0	**4**	3	2	3	0
Média	0	**4**	4	1	0	0
Alta	3	**5**	2	2	2	3
Entropia						
Baixa	3	3	2	2	**5**	1
Média	0	**4**	3	2	0	0
Alta	0	**6**	4	3	0	0
Predominância de Classes						
Baixa	0	**4**	2	3	1	0
Média	4	**5**	4	3	0	0
Alta	1	3	2	2	**4**	0

realizou-se uma análise mais restritiva. Sobre os resultados obtidos até então, verificou-se sob quantos, dos "melhores critérios", cada abordagem conseguiu inferir a dimensão correta para mais de 50% dos conjuntos relacionados à característica sob análise, o que indicaria realmente um "bom" desempenho. Isso faria com que, no exemplo citado acima, apesar das abordagens Fractal Fuzzy (arredondado) e PCA (0,9) terem inferido dimensões corretas para 2 conjuntos de dados sob a avaliação do critério $P+$, isso não contribuiria para o cômputo final de seu desempenho em relação à características em questão, visto que o sucesso por elas obtido não ultrapassou metade dos conjuntos que apresentam tal característica. Além disso, para afirmar

que a abordagem é adequada na presença de determinada característica, inclui-se mais duas condições não excludentes:

- a abordagem deve ter o "bom" desempenho em comparação a, pelo menos, 4 critérios de avaliação;
- a abordagem deve apresentar o melhor desempenho dentre todas as abordagens tanto na análise geral quanto na análise com Vesanto[12].

Isto permitiu avaliar as abordagens que seriam "mais" adequadas para a aplicação nos conjuntos de dados com aquela característica. As conclusões obtidas são relacionadas na Tabela 7.16[13]. Nas células dessa tabela encontram-se as abordagens que obtiveram o melhor desempenho para aquela característica e, entre parênteses, está o número de critérios de avaliação sob os quais o melhor desempenho foi atestado. Em negrito estão as indicações de aplicabilidade das abordagens que seguem as duas regras supra-citadas.

7.4 Considerações sobre os Resultados

Os resultados apresentados na Sessão 7.3 mostraram que as três abordagens analisadas, em pelo menos uma de suas duas variações, alcançaram bons resultados, são elas: Fractal Fuzzy (arredondado), Box-Counting (maior inteiro) e PCA (0,5). O destaque para uma ou outra abordagem depende da forma de análise aplicada aos seus resultados e também das características que estão presentes nos conjuntos de dados analisados.

A PCA (0,5) se destaca, principalmente, quando os critérios utilizados para avaliação de seus resultados são aqueles que consideram o menor Erro Topológico como sendo a melhor medida de avaliação de um mapeamento. Quando se analisa, de forma geral, sua adequação a determinados tipos de conjuntos de dados, chega-se à conclusão que esta abordagem é mais adequada para aplicação naqueles que se localizam em espaços de altas dimensões. Os conjuntos

[12]A escolha destas porcentagens e número mínimo de critérios foram feitas empiricamente, com base nas avaliações do conjunto de resultados obtidos na tabulações.

[13]Estas conclusões podem diferir daquelas obtidas sobre as Tabelas 7.14 e 7.15 por considerar um critério mais restritivo onde não basta ser a abordagem que apresenta o melhor desempenho no maior número de vezes, mas sim a que apresenta um desempenho muito bom um maior número de vezes.

Tab. 7.16: Melhores desempenhos por características dos conjuntos de dados.

Características dos conjuntos	Análise Geral	Análise com Vesanto
Número de Dados		
Pequeno	**BC 1 (5)**	**BC 1 (4)**
Médio	FF 2 (2)	FF 2 (2)
Grande	PCA (0.5) (3)	FF 2 (3)
Número de Atributos		
Pequeno	**FF 2 (5)**	**FF 2 (5)**
Médio	BC 1 (2)	BC 2 (2)
Grande	PCA (0.5)(4)	PCA (0.5)(3)
Número de Classes		
Pequeno	FF 1 (2)	FF 1/FF 2 (3)
Médio	BC 1 (2)	BC 1 (2)
Grande	**BC 1 (6)**	**BC 1 (6)**
Densidade		
Baixa	PCA (0.5) (4)	FF 2 (3)
Média	BC 1 (4)	BC 1 (3)
Alta	**FF 2 (6)**	**FF 2 (5)**
Entropia		
Baixa	FF 2 (3)	FF 1 (1)
Média	BC 1 (2)	BC 1 (1)
Alta	FF 2 (2)	FF 2 (4)
Predominância de Classe		
Baixa	FF 2 (3)	FF 2 (4)
Média	BC 1 (2)	BC 1 (1)
Alta	PCA (0.5) (3)	FF 2/PCA(0.5) (4)

"Norm10", Dermatology, Ionosphere e Internet são os que estão localizados em altas dimensões e, a partir da realização da análise específica que verificou o desempenho desta abordagem para estes conjuntos em comparação ao conjunto de critérios de avaliação, constatou-se que dentre estes conjuntos a abordagem PCA (0,5) não mostrou ser a mais adequada apenas para o conjunto Ionosphere.

A fim de validar o desempenho das abordagens para conjuntos de dados com grande número de atributos, decidiu-se por treinar algumas RNAs-AO com espaços de saída organizados topologicamente em mais dimensões. Essas novas RNA-AOs foram organizadas com um número de neurônios maior (1024 neurônios) para que se pudesse organizá-los em espaços topológicos de

6 e 7 dimensões[14]. Dada a grande quantidade de neurônios, a medida de Produto Topológico não foi aplicada por apresentar uma alta complexidade computacional dependente do número de neurônios.

Apurou-se nessas análises que as RNA-AOs organizadas em altas dimensões se destacaram apenas em relaçao à medida de erro topográfico com vizinhança retangular. Caso em que já se espera este resultado dado a "explosão" do número de vizinhos ocasionada pela alta dimensão do espaço de saída. Concluiu-se então que, para este tipo de conjunto de dados, deve-se acatar a decisão da abordagem PCA (0,5) já que as abordagens baseadas em dimensão fractal indicarão, com grande probabilidade, alta dimensão para organização do espaço de saída da RNA-AO, dificultando a análise final do mapeamento sem agregar ganhos significativos.

Os critérios de avaliação de mapeamentos que analisam, com maior ênfase, a capacidade quantizadora mostraram que a abordagem Box-Counting (maior inteiro) se apresenta como uma abordagem adequada para inferência da dimensão ideal para o espaço de saída de uma RNA-AO. Essa adequação se evidencia quando os conjuntos de dados analisados possuem poucos dados e um grande número de classes, e com menos ênfase quando a densidade do conjunto é média, Os conjuntos de dados utilizados para testes neste trabalho só permitiram que a avaliação destas indicações fosse feita se dividida em duas partes:

- na primeira parte avaliou-se os conjuntos "Vá" e "Switzerland" que possuem poucos dados e densidade média.
- na segunda parte avaliou-se os conjuntos "Auto-mpg", "Abalone" e "Letter" que possuem um grande número de classes.

Para os conjuntos "Vá" e "Switzerland" esta abordagem é a que mostra o melhor desempenho. Para o conjunto "Auto-mpg" a abordagem é a que possui o melhor desempenho e para o conjunto "Abalone" esta abordagem tem o mesmo desempenho que a abordagem Fractal-Fuzzy na análise feita com a heurística de Vesanto. Para o conjunto "Letter" esta abordagem não foi adequada como era esperado.

[14]A disposição dos neurônios no espaço de saída foram organizados da seguinte forma: 1-dimensional [1,1024]; 2-dimensional [32,32]; 3-dimensional [16,16,4]; 4-dimensional [8,4,8,4]; 5-dimensional [4,4,4,4,4]; 6-dimensional [4,4,4,4,2,2]; 7-dimensional [4,4,4,2,2,2,2];

Finalmente, o desempenho da abordagem Fractal-Fuzzy foi considerado bom, principalmente, quando se avalia a preservação de topologia via Produto Topográfico ou via o segundo menor Erro Topológico. Essa abordagem também tem um desempenho destacado em relação às outras duas quando a heurística de Vesanto é aplicada como outro fator de automação de decisão sobre o projeto da RNA-AO. Para esta abordagem, a análise em relação às características presentes nos conjuntos de dados revelou que a sua aplicação é indicada principalmente quando o conjunto de dados está localizado em dimensões baixas, com densidade alta. Com menor ênfase, as análises desta tese indicam também a aplicação desta abordagem para conjuntos de dados com predominância de classe baixa e entropia alta.

Dentre os conjuntos de dados testados nesta tese, não existe nenhum que possua todas as características com esses valores. Assim, para avaliação dessa conclusão escolheu-se quatro conjuntos de dados ("Íris", "Glass", "Abalone" e "Auto-mpg") que apresentam três das características e um ("Norm4") que apresenta as duas primeiras (para as quais a indicação desta abordagem mais forte - veja Tabela 7.16. O resultado da avaliação mostrou que a abordagem Fractal Fuzzy (arredondado) realmente infere a dimensão ideal sob comparação com o maior número de critérios (tanto na avaliação geral quanto na avaliação utilizando a heurística de Vesanto) para quatro dos conjuntos selecionados. Para um deles ("Auto-mpg") a abordagem Box-Counting (maior inteiro) obteve melhor resultado na avaliação geral.

A fim de confirmar esses resultados, foi realizado um teste com dois conjuntos de dados adicionais que apresentam algumas das características supra-citadas. Os resultados mostraram que a abordagem Fractal Fuzzy foi capaz de realizar uma boa inferência para a dimensão do espaço de saída da RNA-AO no que tange a busca de preservação de informações topológicas. Os conjuntos utilizados foram Breast Cancer[15] e Chess Databases - "king-rook-vs-king-pawn"[16] (NEWMAN et al. (1998)):

Para o primeiro conjunto a medida DFFS foi 2,65, sugerindo o uso de uma RNA-AO com espaço de saída 3-dimensional (abordagem Fractal Fuzzy arredondado). A medida Box-Counting foi 1,91. Analisando as medidas de qualidade dos mapeamentos percebeu-se que, em relação a

[15]Esse conjunto de dados foi criado pela University Medical Centre, Institute of Oncology, Ljubljana, antiga Yugoslavia. A autora desta tese agradece a M. Zwitter e a M. Soklic.

[16]Base de dados gerada por Michael Bain e Arthur van Hoff, no Turing Institute, Glasgow, UK.

informações de topologia, as RNAs-AO organizadas em espaços 3-dimensionais mostraram um bom desempenho sem exigir altos custos de perda de quantização.

Para o segundo caso, a DFFS não diferiu da dimensão Box-Counting (4,47), sugerindo um espaço de saída organizado na dimensão 4 (abordagem Fractal Fuzzy arredondado). Nesse último caso, o desempenho da abordagem não foi tão promissor. As RNAs-AO que atendiam as condições da inferência das abordagens tiveram medidas de qualidade topográfica muito boas, porém a altos custos de qualidade quantizadora.

As Figuras 7.3 e 7.4 fornecem um resumo das conclusões obtidas. Na primeira figura, o resultado da análise geral é mostrado. Observe que apenas as abordagens Box-Counting 1 (maior inteiro) e Fractal Fuzzy 2 (arredondado) aparecem, porque as demais abordagens não se destacaram como boas inferências para a dimensão do espaço de saída de RNAs-AO. Na segunda figura estão evidenciadas as abordagens que se destacaram como boas formas de inferência para a dimensão do espaço de saída de RNAs-AO mediante a presença de determinadas características nos conjuntos de dados.

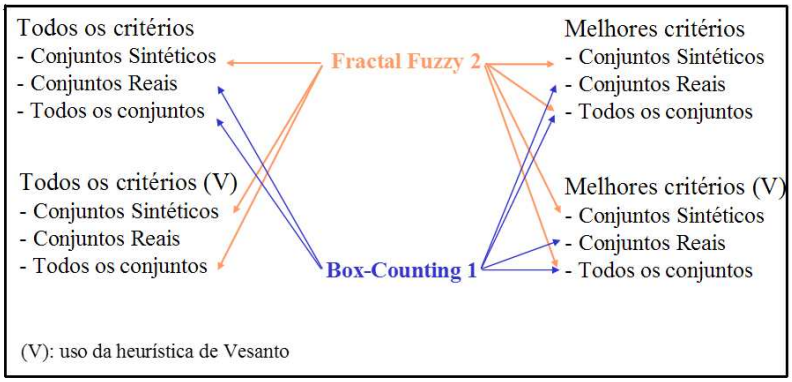


Fig. 7.3: Conclusões sobre a análise geral das abordagens testadas.

7.5 Considerações Finais

Procurou-se neste capítulo explicar o processo de análise de três abordagens para inferência da dimensão ideal para o espaço de saída de uma RNA-AO. Discutiu-se os resultados obtidos bem como averiguou-se que a medida de dimensão fractal DFFS é adequada para resolver,

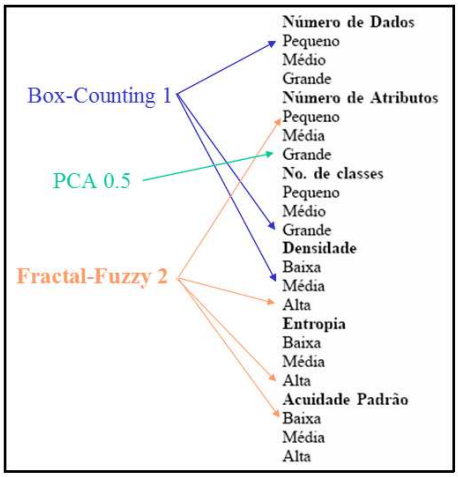


Fig. 7.4: Conclusões referentes ao desempenho das abordagens mediante análise por características.

em situações específicas, o problema estudado nesta tese. Também concluiu-se que as outras abordagens testadas (Box-Counting e PCA) têm seu mérito e seu escopo de aplicabilidade.

Capítulo 8

Conclusão

> "... os livros não são feitos para acreditarmos neles, mas para serem submetidos a investigações. Diante de um livro não devemos nos perguntar o que diz mas o que quer dizer."
> Guilherme para Adso na biblioteca da Abadia. O Nome da Rosa - Umberto Eco.

Esta tese objetivou estudar o problema de preservação de topologia de uma Rede Neural Artificial Auto Organizável, a rede de Kohonen, e apresentou uma abordagem heurística capaz de apoiar o processo de tomada de decisão sobre a determinação de um dos parâmetros livres do projeto desta rede - a dimensão do seu espaço de saída. Mais especificamente, o problema tratado aqui foi a determinação da dimensão mais adequada para que o processo de auto organização fosse iniciado. O estudo e determinação dos demais parâmetros envolvidos no processo e co-responsáveis pelo alcance de bons resultados da RNA-AO foram mantidos fora do escopo desta tese.

O processo de auto organização referido aqui é o que pretende quantizar o espaço dos dados e reduzir sua dimensão de forma a possibilitar uma exploração eficaz. Neste trabalho, refere-se a "dimensão mais adequada" porque procurou-se analisar se a dimensão inferida pela abordagem heurística é aquela que melhor preserva as propriedades topológicas do conjunto de dados e que consegue alcançar uma boa aproximação da densidade de probabilidade do conjunto de dados.

Esta heurística foi projetada, de forma híbrida, com o uso de conceitos da Teoria de Fractais, com ênfase na exploração de uma variação da Dimensão Fractal denominada Dimensão Box-Counting e com uso de conceitos da Teoria de Raciocínio Aproximado Fuzzy. Dentro deste contexto, os conjuntos de dados são tratados como fractais estocásticos, estuda-se sua Dimensão Box-Counting e, a partir de uma análise "fuzzy" do processo de determinação desta medida de dimensão encontra-se uma nova medida, a qual denominou-se Dimensão Fractal Fuzzy Significativa.

O Capítulo 1 forneceu a motivação para o desenvolvimento deste estudo, discutindo aspectos relacionados ao uso do conhecimento intrínseco a um conjunto de dados como suporte ao desenvolvimento de ferramentas para execução de análises sobre eles. Neste sentido, formulou-se a hipótese que a informação de Dimensão Fractal expressa a dimensão na qual residem as relações de vizinhança topológica presentes nos dados, já que ela expressa a real porção do espaço que é efetivamente ocupada pelos dados, e também que o mapeamento realizado por uma RNA-AO poderia atuar nessa dimensão sendo capaz de expressar satisfatoriamente as relações topológicas desses dados.

A Teoria de Fractais e a Teoria de Raciocínio Aproximado Fuzzy foram rapidamente descritas nos Capítulos 2 e 3, dando mais atenção aos aspectos necessários para o entendimento do modo de determinação da Dimensão Fractal Fuzzy Significativa, introduzida nesta tese. O principal objeto de atenção, as RNAs-AO, foram exploradas no Capítulo 4. Além dos principais aspectos relacionados ao funcionamento deste tipo de rede neural artificial, discutiu-se com mais atenção os métodos de avaliação de qualidade de preservação topológica e capacidade quantizadora, já que esses foram usados na análise do desempenho da abordagem proposta.

O uso de conceitos da Teoria de Fractais e da Teoria de Raciocínio Aproximado Fuzzy para determinação da Dimensão Fractal Fuzzy Significativa foi descrito no Capítulo 5, constituindo, no contexto desta tese, a ferramenta de suporte à tarefa de análise de dados que:

1. determina a Dimensão Box-Counting do conjunto de dados, construindo e armazenando informações sobre a curva resultante do processo de cálculo da dimensão em questão, por meio da qual a medida de Dimensão Fractal Fuzzy Significativa é calculada;

2. analisa as informações da curva Box-Counting, por meio da interpretação "fuzzy" das

inclinações dos segmentos de reta que a compõe e da relação entre as inclinações de cada par de segmentos consecutivos, produzindo uma série de conclusões sobre a composição da curva e, conseqüentemente, dada a natureza analítica do processo de construção da curva, produzindo também descobertas sobre a distribuição dos dados em diferentes níveis de observação dos mesmos;

3. conclui um ponto de corte da curva que representa a granularidade adequada de análise de um conjunto de dados quando se visa a tarefa de descoberta de grupos e relações entre grupos;

4. calcula a Dimensão Fractal Fuzzy Significativa a partir da curva resultante do corte no ponto determinado.

Ainda no Capítulo 5 foram apresentadas duas aplicações, referentes à tarefa de análise de dados, para o uso da Dimensão Fractal Fuzzy Significativa ou do processo de análise da curva inerente ao seu cálculo: o problema de Tendência a Agrupamentos, e o problema foco da tese, a determinação da dimensão ideal para o espaço de saída de uma RNA-AO. A modelagem do processo de resolução dos problemas dessas duas aplicações foi descrito ao final do capítulo. O problema de Tendência a Agrupamentos foi utilizado como um método de validação da heurística de análise da curva Box-Counting. Os resultados deste experimento foram apresentados no Capítulo 6. Os resultados dos experimentos com o problema da dimensão ideal foram relatados no Capítulo 7.

8.1 Contribuições

A partir dos esforços dispendidos neste estudo foi possível contribuir tanto para a área de Análise de Dados quanto, especificamente, para a área de Redes Neurais Artificiais Auto Organizáveis. A fim de elucidar essas contribuições, optou-se por organizá-las na lista que segue:

1. Concepção de uma nova medida de dimensão fractal, a Dimensão Fractal Fuzzy Significativa, na forma de uma variação da medida Box-Counting. Tal medida tem o intuito

de mensurar a dimensão fractal de um conjunto de dados considerando os agrupamentos nele existentes.

2. Resolução do problema de decisão sobre a dimensão ideal para o espaço de saída de uma RNA-AO com o uso da DFFS, da dimensão Box-Counting ou de Análise de Componentes Principais, com um estudo que permite escolher uma destas abordagens dependendo do tipo de conjunto de dados que está sob análise e das características qualitativas do mapeamento, as quais se quer priorizar.

3. Concepção de um processo sistemático e heurístico de análise da curva Box-Counting. Esse processo permite inferir conhecimento sobre a distribuição dos dados em um conjunto, sob diferentes níveis de observação, e fornece informações para o cálculo da Dimensão Fractal Fuzzy Significativa.

4. Validação do processo de análise da curva Box-Counting e de cálculo da DFFS por meio da aplicação deste processo na resolução de um clássico problema de Análise de Dados, a análise de Tendência a Agrupamentos.

5. Estudo e discussão de medidas de qualidade para mapeamentos topográficos e, organização, análise e aplicação de um conjunto de critérios de avaliação da qualidade quantizadora e de preservação topográfica de mapeamentos dessa natureza.

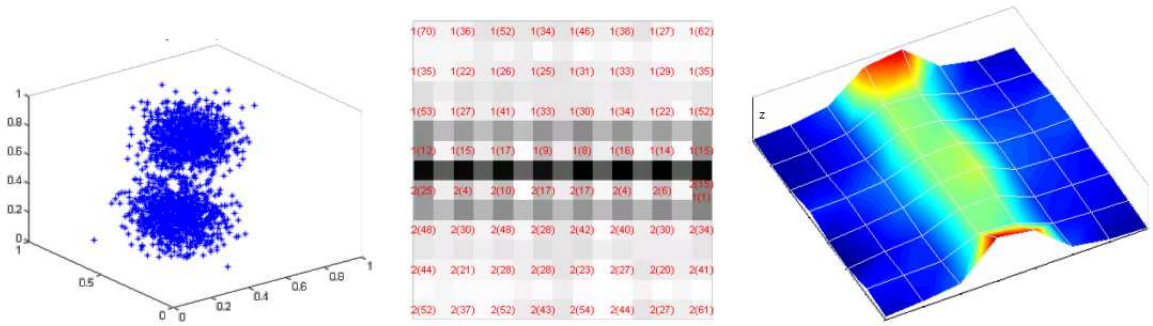
8.2 Trabalhos Futuros

No decorrer do desenvolvimento desta tese notou-se que existem dois caminhos que podem delinear possíveis extensões do assunto aqui tratado: a utilização de diferentes veículos para análise do desempenho da aplicação da DFFS como estimador da dimensão ideal para mapas de saída de RNAs-AO; a utilização da medida DFFS como fonte de conhecimento *a priori* sobre os dados submetidos à análise exploratória, de forma a conceber veículos de otimização, sob diferentes frentes, de algoritmos, modelos e metodologias que se propõem a realizar tarefas de análise de dados.

Sob a perspectiva do primeiro caminho vê-se que a análise empírica de desempenho, como realizada nesta tese, sempre ansiará por um ou vários outros meios de verificação e validação. Assim, sugere-se que este trabalho seja complementado por meio das seguintes propostas de testes e análises:

- Estudo dos agrupamentos formados pelas diferentes RNAs-AO, concebidas sobre diferentes dimensões do espaço de saída, no sentido de verificar, segundo as classificações disponíveis, se os grupos correspondem às reais classes e sob qual dimensão obtém-se os melhores resultados na comparação. Esta análise pode ser suportada pela exploração das Matrizes-U (e Matrizes-U*) (veja Figura 8.1), por meio de algoritmos de discriminação de grupos ou pela análise de informações tais como: número de neurônios que são apontados como BMUs somente para dados que pertencem a uma mesma classe; número de neurônios que são apontados como BMUs para dados que pertencem a diferentes classes (neurônios "confusos"); número de neurônios efetivamente utilizados pela RNA-AO para representação de dados; número de dados que são representados por neurônios "confusos" (classificados erroneamente); número de classes que possuem dados que são classificados erroneamente. Estas informações foram obtidas, para cada uma das RNAs-AO treinadas, durante o desenvolvimento desta tese. A sua exploração é planejada para os próximos passos de extensão desta tese, constituindo uma ação próxima de ser realizada.

- Dado que todas as RNAs-AO treinadas neste trabalho estão armazenadas na forma de seus vetores de peso, é possível aplicar outras medidas de qualidade referentes a capacidade quantizadora ou de preservação de informação topográfica. No Capítulo 4 foram apresentadas algumas medidas que poderiam ser utilizadas, e ainda em KASKI; LAGUS (1996) **apud** VESANTO (1997) é descrita uma integração da análise de capacidade quantizadora e de preservação de informação topográfica em uma única medida, o Erro de Quantização Topológica. Tal medida utiliza como base de cálculo de erro os caminhos mínimos entre os primeiros e os segundos BMUs dentro do espaço de saída de uma RNA-AO. Ainda outras medidas de análise, como a própria medida de dimensão fractal aplicada para analisar a configuração de neurônios do mapa de saída da RNA-AO, podem também ser aplicadas e os resultados comparados àqueles já obtidos, aumentando assim



(a) (b) (c)

Fig. 8.1: (a) Conjunto de dados sob análise; (b) Matriz-U para uma RNA-AO bi-dimensional (8x8): a Matriz-U mostra as distâncias entre unidades vizinhas de forma a mostrar a estrutura de grupos formada pelo mapeamento. Valores altos na Matriz-U significam grandes distâncias entre unidades vizinhas no mapeamento, representando bordas de (entre) agrupamentos (cores escuras). Os grupos são representados por áreas uniformes com baixas distâncias (cores claras). Na figura a Matriz-U é visualizada com "labels" da classe representada por cada unidade do mapa e o número de dados para os quais a unidade é escolhida como BMU. (c) Matriz de distâncias plotada como superfície: a coordenada z indica a distância média entre as unidades vizinhas no mapa. Tons de azul indicam valores de distância média baixos. A progressão das cores até os tons de vermelho indica o aumento nos valores de distância média.

a diversidade de análises dos mapeamentos. Este quesito também faz parte do próximos passos de extensão a serem realizados a curto prazo.

- Exploração, nos moldes do que foi realizado nesta tese, de conjuntos de dados que apresentem outras características além daquelas já utilizadas, por exemplo: tipo de atributo (maioria binária, maioria nominal ou maioria contínua), nível de ruído, freqüência de valores perdidos. Também outras formas de determinação dos intervalos que constituem a variação considerada dentro de cada característica podem ser consideradas. Os resultados obtidos destas novas análises deveriam ser confrontados com aqueles obtidos nesta tese a fim de fortalecer a heurística sobre quais abordagens são mais adequadas e em quais casos.

- Estudar uma formalização para o processo de obtenção da DFFS traria, ainda que concebido para casos específicos e sob condições controladas, uma maior segurança na aplicação desta medida ou na concepção de novas aplicações para ela.

- BRUSKE; SOMMER (1998) apresentaram uma forma de estimar a dimensão intrínseca de um conjunto de dados a partir da análise da dimensão intrínseca da representação topológica deles. Estudar a relação existente entre as RNAs-AO exploradas nesta tese e a dimensão DFFS dos conjuntos de dados representados por elas, a exemplo do trabalho citado, constitue uma forma diferente de validar o processo de obtenção da DFFS.

Para o segundo caminho, as sugestões de trabalhos futuros são delineadas mediante uma abertura de novas possibilidades de aplicação da medida DFFS. Segue algumas delas:

- Mediante os experimentos realizados em (BARBARá; CHEN (2003), BARBARá; NAZERI (2000), BARBARá; WU (2003b), BARBARá et al. (2002), TRAINA et al. (2000) e ARANTES; TRAINA; TRAINA (2003)), sugere-se a substituição da medida de dimensão fractal neles utilizada, pela medida DFFS. Dado que as tarefas abordadas nessas referências são, especificamente, de análise de grupos, espera-se conseguir resultados próximos ou melhores dada a características de concepção da DFFS. Um estudo específico para cada caso será necessário para decidir se a substituição é possível de forma simples ou se adaptações na concepção dos modelos de resolução aplicados podem ser propostas. Para o caso específico dos trabalhos BARBARá; CHEN (2003) e TRAINA et al. (2000) os estudos já estão em andamento.

- Seguindo a linha aplicada por SPECKMANN; RADDATZ; ROSENSTIEL (1994a) e SPECKMANN; RADDATZ; ROSENSTIEL (1994b) sugere-se aplicar a medida DFFS na forma de um controlador do processo de aprendizando de uma RNA-AO. Neste contexto, a disposição dos neurônios da camada de saída de uma RNA-AO seriam tratados como um fractal estocástico, e a sua medida DFFS seria comparada, durante as iterações de treinamento, à medida DFFS dos conjuntos de dados que estão sendo analisados. A aproximação das duas medidas poderia indicar o ponto ideal de parada do treinamento, dado que supostamente, a disposição dos neurônios estaria ocupando a mesma porção do espaço que os dados ocupam. Contudo, seria necessário acompanhar os erros de quantização e de preservação de topologia a fim de averiguar se a disposição dos neurônios é realmente a ideal para representar os dados, como realizado nos trabalhos citados. O

diferencial desta sugestão em relação ao trabalho desenvolvido nessas citações está na medida tomada com dimensão fractal do conjunto de dados e do conjunto de neurônios.

- Um estímulo a mais para o empenho em utilizar a DFFS como apoio à decisões a serem tomadas em projeto e treinamento de uma RNA-AO pode ser encontrada na concepção dos modelos chamados Máquinas de Vetores-Suporte (veja LIMA (2004) e HAYKIN (1999) para explicações e exemplos de aplicações destes modelos). Nessa técnica procura-se controlar a Dimensão Vapnik-Chervonenkis (VAPNIK; CHERVONENKIS (1971)) do modelo durante a fase de treinamento, com o intuito de controlar a complexidade intríseca a ele, tornando-o mais suscetível para a tarefa de minimização de erros em tarefas de análise de dados. Trazendo esta idéia ao contexto da DFFS argumenta-se que, dado que a DFFS é uma medida de análise de dados em nível de grupos (relações inter-grupos), o seu uso para dimensionar um modelo poderia fazê-lo crescer até o nível de complexidade supostamente necessária mediante a complexidade do objeto analisado.

8.3 Publicações

- PERES, Sarajane Marques; ANDRADE NETTO, Marcio Luiz de. Using Fractal Dimension to Fuzzy Pre-Processing of N-Dimensional Datasets. In: XVI ICSE - XVI INTERNATIONAL CONFERENCE ON SYSTEMS ENGINEERING, 2003, Coventry. Proceedings of Sixteenth International Conference on Systems Engineering. Coventry - UK: Burnham, K.J., Haas, O. C. L., v. 2, p. 846-851, 2003.

- PERES, Sarajane Marques; ANDRADE NETTO, Marcio Luiz de. Um Sistema Híbrido para Análise Heurística de Dados Utilizando Teoria de Fractais e Raciocínio Aproximado Fuzzy. Relatório Técnico: Universidade Estadual de Campinas, 2004. .

- PERES, Sarajane Marques; ANDRADE NETTO, Marcio Luiz de. A Fractal Fuzzy Approach to Clustering Tendency Analysis. Lecture Notes in Computer Science, (SBIA'04) Brazil, p. 395-404, 2004.

- PERES, Sarajane Marques; ANDRADE NETTO, Marcio Luiz de. Fractal fuzzy decision making: what is the adequate dimension for self-organizing maps?. In: NAFIPS - NORTH AMERICAN FUZZY INFORMATION PROCESSING SOCIETY, 2004, Banff. Electronic Proceedings for the NAFIPS 2004 conference, 2004.

- PERES, Sarajane Marques; ANDRADE NETTO, Marcio Luiz de. Meaningful Fractal Fuzzy Dimension: a new approach for Data Analysis. TKDE - IEEE Transactions on Knowledge and Data Engineering. Submetido.

Referências Bibliográficas

AGRAWAL, R.; IMIELINSKI, T.; SWAMI, A. N. Mining Association Rules between Sets of Items in Large Databases. In: ACM SIGMOD INTERNATIONAL CONFERENCE ON MANAGEMENT OF DATA, 1993., 1993, Washington, D.C., USA. **Proceedings...** [S.l.: s.n.], 1993. p.207–216.

AGRAWAL, R.; SRIKANT, R. Fast Algorithms for Mining Association Rules. In: INT. CONF. VERY LARGE DATA BASES, VLDB, 20., 1994. **Proceedings...** Morgan Kaufmann, 1994. p.487–499.

ALHOUNIEMI, E.; HIMBERG, J.; PARHANKANGAS, J.; VESANTO, J. **Som Toolbox 2.0**. 2002.

ARABIE, P.; HUBERT, L. J.; DE SOETE, G. **Clustering and Classification**. River Edge, NJ, USA: World Scientific Publishing, 1996.

ARANTES, A. S.; TRAINA, A. J. M.; TRAINA, C. J. The fractal dimension making similarity queries more efficient. In: PROCEEDING OF 2ND WORKSHOP ON FRACTALS AND SELF-SIMILARITY IN DATA MINING: ISSUES AND APPROACHES (IN 9TH ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING), 2003, Washington, DC, USA. **Anais...** [S.l.: s.n.], 2003. p.18–23.

BARBARá, D.; CHEN, P. **Using Self-Similarity to Cluster Large Data Sets**. Virginia, USA: George Manson University, 2002.

BARBARá, D.; CHEN, P. Using Self-Similarity to Cluster Large Data Sets. **Data Minning Knowledge Discovery**, [S.l.], v.7, n.2, p.123–152, 2003.

BARBARá, D.; MENASCE, D.; ALMEIDA, V.; ABRAHAO, B.; RIBEIRO, F. Fractal characterization of web workloads. In: INTERNATIONAL WORLD WIDE WEB CONFERENCE, 11., 2002. **Proceedings...** [S.l.: s.n.], 2002. Finalista para melhor artigo.

BARBARá, D.; NAZERI, Z. **Fractal Mining of Association Rules Over Interval Data**. Acessado em 15/07/2006.

BARBARá, D.; WU, X. Using Fractals to compress real data sets: is it feasible? In: SECOND KDD WORKSHOP ON FRACTALS, POWER LAWS, AND OTHER NEXT GENERATION DATA MINING TOOLS, 2003, Washington, DC, USA. **Proceedings...** [S.l.: s.n.], 2003.

BARBARá, D.; WU, X. Compressing high dimensional datasets by fractals. In: DATA COMPRESSSION CONFERENCE (DCC 2003), 2003., 2003, Snowbird, UT. **Anais...** [S.l.: s.n.], 2003.

BARNSLEY, M. **Fractals Everywhere**. San Diego, California, USA: Academic Press Inc, 1988.

BARTH, A.; BAUMANN, G.; NONNENMACHER, T. F. Measuring Rényi dimensions by a modified box algorithm. **Journal of Physics A**, [S.l.], v.4, n.25, p.381–391, 1992.

BAUER, H. U.; HERRMANN, M.; VILLMANN, T. Neural maps and topographic vector quantization. **Neural Networks**, [S.l.], v.12, n.4-5, p.659–676, 1999.

BAUER, H. U.; PAWELZIK, K. R. Quantifying the Neighborhood Preservation of Self-Organizing Feature Maps. **IEEE Transactions on Neural Networks**, [S.l.], v.3, n.4, p.570–579, July 1992.

BAUER, H. U.; VILLMANN, T. **Growing a Hypercubical Output Space in a Self-Organizing Feature Map**. Berkeley, California, USA: International Computer Science Institute, 1995. (TR-95-030).

BEYER, K.; GOLDSTEIN, J.; RAMAKRISHNAN, R.; SHAFT, U. When Is "Nearest Neighbor" Meaningful? **Lecture Notes in Computer Science**, [S.l.], v.1540, p.217–235, 1999.

BEZDEK, J. C.; PAL, N. R. An Index of Topology Preservation for Feature Extraction. **Pattern Recognition**, [S.l.], n.28, p.381–391, 1995.

BLOCK, A.; BLOH, W. von; SCHELLNHUBER, H. J. Efficiente box-counting determination of generalize fractal dimensions. **Physical Review A**, [S.l.], v.4, n.42, p.1869–1874, 1990.

BOREL, E. **Probabilite et certitude**. [S.l.]: Press Université de France, 1950.

BRUSKE, J.; SOMMER, G. Intrinsic Dimensionality Estimation With Optimally Topology Preserving Maps. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, [S.l.], v.20, n.5, p.572–575, 1998.

BUENO, R.; FERREIRA, M. R. P.; JUNIOR, C. T. **DBGen Manual da Ferramenta**. São Carlos, SP: Instituto de Ciências Matemáticas e de Computação, USP-São Carlos, 2004. (246).

ABBASS, H. A.; SARKER, R. A.; NEWTON, C. S. (Ed.). **Data Mining**: a heuristic approach. USA: Idea Group Publishing, 2001. p.231–259.

CHACHERE, G. A fast O(N) and memory-efficient algorithm for box counting. In: ABSTRACTS OF THE SIAM CONFERENCE ON APPLICATIONS OF DYNAMICAL SYSTEMS, 1992, Snowbird, Utah. **Anais...** [S.l.: s.n.], 1992.

COSTA, J. A. **Classificação Automática e Análise de Dados por Rede Neurais Auto-Organizáveis.** 1999. Tese (Doutorado em Ciência da Computação) — Universidade Estadual de Campinas, Campinas, São Paulo, Brasil.

COSTA, J. A.; NETTO, M. L. A. Clustering of complex shaped data sets via Kohonen maps an mathematical morphology. In: DATA MINING ADN KNOWLEDGE DESCOVERY: THEORY, TOOLS, AND TECHNOLOGY III, 2001, Orlando, Florida, USA. **Anais...** [S.l.: s.n.], 2001. (Proceedings of the SPIE).

DUDA, R. O.; HART, P. E.; STORK, D. G. **Pattern Classification**. 2nd.ed. [S.l.]: John Wiley Sons, Inc., 2000.

ERWIN, E.; OBERMAYER, K.; SCHULTEN, K. Self-organizing maps: stationay states, metastability and convergence rate. **Biological Cybernetics**, [S.l.], n.67, p.35–45, 1992.

ERWIN, E.; OBERMAYER, K.; SCHULTEN, K. Self-organizing maps: ordering, convergence properties and energy funcions. **Biologiucal Cybernetics**, [S.l.], n.67, p.47–55, 1992.

FALCONER, K. **Fractal Geometry**: mathematical foundations and applications. England: John Wiley Sons Ltd., 1990.

FAUSETT, L. **Fundamentals of Neural Netoworks**: architectures, algorithms, and applications. New Jersey, USA: Prentice-Hall, Inc., 1994.

FEENY, B. Fast Multifractal Analysis by Recursive Box Covering. **International Journal of Bifurcation and Chaos**, [S.l.], v.10, n.9, p.2277–2287, 2000.

FERREIRA, M.; BUENO, R.; TRAINA, C. J. Dbgen - gerador de dados sintéticos com ditribuição fractal. In: 2005, Uberlandia, MG, Brasil. **Anais...** [S.l.: s.n.], 2005.

GLEISER, M. **O fim da Terra e do Céu**: o apocalipse na ciência e na religiãao. São Paulo, Brasil: Companhia das Letras, 2001.

GOODHILL, G. J.; SEJNOWSKI, T. J. Quantifying neighbourhood preservation in topographic mappings. In: 1996, University of California, San Diego and California Institute of Technology, 6, Pasadena, CA. **Anais...** [S.l.: s.n.], 1996. p.61–82. (Proceedings of the 3rd Joint Symposium on Neural Computation).

GRASSBERGER, P. On efficient box-counting algorithms. **International Journal of Modern Physics C**, [S.l.], v.4, n.37, p.1711–1724, 1993.

GRINSTEIN, G.; TRUTSCHL, M.; CVEK, U. High-Dimensional Visualizations. In: 2001, San Francisco, California. **Anais...** [S.l.: s.n.], 2001. p.1–14. (Proceedings of the 7th Data Mining Conference - KDD).

HAYKIN, S. **Neural Networks**: a comprehensive foundation. 2nd.ed. New Jersey, USA: Prentice Hall Inc, 1999. 842p.

HEBB, D. O. **The Organization of Behavior**: a neuropsychological theory. [S.l.]: New York:Wiley, 1949.

HOU, X.-J.; GILMORE, R.; MINDLIM, G.; SOLARI, H. An efficient algorithm for fast O(N*ln(N)) box counting. **Physics Letter A**, [S.l.], v.(1,2), n.151, p.43–46, 1990.

JAIN, K. J.; DUBES, R. C. **Algorithms for Clustering Data**. New Jersey, USA: Prentice-Hall, Inc., 1988.

KANERVA, P. Associative Neural Memories: theory and implementation. In: 1993, New York, USA. **Anais...** Oxford University Press, 1993. (chapter Sparse Distributed Memory).

KASKI, S.; LAGUS, K. Comparing self-organizing maps. In: ICANN'96, 1996. **Proceedings...** [S.l.: s.n.], 1996. p.809–814.

KAYMAK, U.; SETNES, M. **Extended fuzzy clustering algorithms**. Erasmus Research Institute of Management (ERIM): Faculdade Bedrijfskunde: In: ERIM : Report Series Research in Management, 2000. ERS : 2000 51 LIS.

KLIR, G. J.; YUAN, B. **Fuzzy Sets and Fuzzy Logic**: theory and applications. [S.l.]: Prentice-Hall, 1995.

KONONEN, T. **Self-Organizing Maps**. 3rd.ed. [S.l.]: Springer, 2000. 521p.

KRUGER, A. Implementation of a fast Box-Counting algorithm. **Computer Physics Communications**, [S.l.], v.98, p.224–234, 1996.

LAWSON, R.; JURS, P. J. New index for clustering tendency and its application to chemical problems. **Journal of Chemical Information and Computer Sciences**, [S.l.], v.30, p.36–41, 1990.

LIEBOVITCH, L. S.; TOTH, T. A fast algortihm to determine fractal dimensions by box counting. **Physics Letter A**, [S.l.], v.(8,9), n.141, p.386–390, 1989.

LIMA, C. A. M. **Comitê de Máquinas**: uma abordagem unificada empregando máquinas de vetores-suporte. 2004. Tese (Doutorado em Ciência da Computação) — Universidade Estadual de Campinas, Campinas, São Paulo, Brasil.

LIU, J.; XIE, W. A genetic-based approach to fuzzy clustering. In: PROCEEIDINGS OF FUZZY-IEEE'95., 1995, Yokohama, Japão. **Anais...** [S.l.: s.n.], 1995. v.4.

MALSBURG, C. Von der. An Introduction to Neural and Eletronic Networks. In: 1990, San Diego, CA, USA. **Anais...** Academic Press, 1990. (chapter Network self-organization).

MOLTENO, T. C. A. Fast O(N) box-counting algorithm for estimating dimension. **Physical Review E**, [S.l.], v.5, n.45, p.R3263–R3266, 1993.

NEWMAN, D. J.; HETTICH, S.; BLAKE, C. L.; MERZ, C. J. **UCI Repository of machine learning databases**. 1998.

PEDRYCS, W.; GOMIDE, F. **An Introduction to Fuzzy Sets**: analysis and design. [S.l.]: A Bradford Books, 1998. 465p.

PEITGEN, H. O.; JURGENS, H.; SAUPE, D. **Chaos and Fractals**: new frontiers of science. New York, USA: Springer-Verlag New York Inc., 1992.

PRECHELT, L. **Proben1 - a set of neural network benchmark problems and benchmarking rules**. Karlsruhe, Germany: Fakultat für Informatik, Universitat Karlsruhe, 1994. (21).

RIESENHUBER, M.; BAUER, H. U.; GEISEL, T. Analysing Phase Transition in High-Dimensional Self Organizing Maps. **Biological Cybernetics**, [S.l.], v.75, n.5, p.397–407, 1996. Submitted to Biological Cybernetics, December 14, 1995 and revised version, July 14, 1996.

ROUBENS, M. Fuzzy clustering algorithms and their cluster validity. **European Journal of Operational Research**, [S.l.], v.10, n.3, p.294–301, 1982. Euro IV. Special Issue.

SCHUSTER, H. G. **Deterministic Chaos**. Weinheim,Germany: Physik-Verkag, 1984.

SOUSA, E. P. M. T.; TRAINA, A. J. M.; TRAINA, J. C. Next Generation of Data Mining Applications. In: ACADEMIC PRESS, 2005. **Anais...** Eds. Wiley/IEEE Press, 2005. p.30. (chapter Using Fractals in Data Mining).

SOUSA, E. P. M.; TRAINA, C. J.; TRAINA, A. J. M. Sid: calculating the intrinsic dimension of data streams. In: PROCEEDING OF 2ND WORKSHOP ON FRACTALS AND SELF-SIMILARITY IN DATA MINING: ISSUES AND APPROACHES (IN 9TH ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING), 2003, Washington, DC, USA. **Anais...** [S.l.: s.n.], 2003. p.18–23.

SPECKMANN, H.; RADDATZ, G.; ROSENSTIEL, W. Improvement of learning results of the selforganizing map by calculating fractal dimensions. In: ESANN 1994 - 2ND EUROPEAN SYMPOSIUM ON ARTIFICIAL NEURAL NETWORKS, 1994, Bruxelles, Belgiun. **Proceedings...** [S.l.: s.n.], 1994.

SPECKMANN, H.; RADDATZ, G.; ROSENSTIEL, W. Relations between generalized fractal dimensions and Kohonen's self-organizing map. In: NEURO NIMES, PARIS, 1994. **Anais...** [S.l.: s.n.], 1994.

TASDEMIR, K.; MERéNYIT, E. Considering Topology in the Clustering of Self-Organizing Maps. In: WORKSHOP ON SELF-ORGANIZING MAPS, 5., 2005. **Anais...** [S.l.: s.n.], 2005. n.5, p.439–446.

TRAINA, A.; TRAINA, C. J.; PAPDIMITRIOU, S. Tri-Plots: scalable tools for multidimensional data mining. In: THE SEVENTH ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA-MINING, 2001, San Francisco, CA, USA. **Anais...** [S.l.: s.n.], 2001. p.184–193.

TRAINA, C. J. **Aplicações da Teoria de Fractais em Bancos de Dados**. 2001.

TRAINA, C. J.; TRAINA, A. J. M.; FALOUSTOS, C. **Distance exponent**: a new concept for selectivity estimation in metric trees. Pittsburg, PA, USA: Carnegie Mellon University, 1999.

TRAINA, C. J.; TRAINA, A.; WU, L.; FALOUSTOS, C. Fast feature selection using fractal dimension. In: XV BRAZILIAN DATABASE SYMPOSIUM, 2000, João Pessoa, PA, Brasil. **Anais...** [S.l.: s.n.], 2000. p.158–171.

TURING, A. M. Computing machinery and intelligence. **Mind**, [S.l.], v.59, p.433–460, 1950.

VAPNIK, V.; CHERVONENKIS, A. On the uniform convergence of relative frequencies of events to their probabilities. **Theory of Probability and its Applications**, [S.l.], v.16, n.2, p.264–280, 1971.

VESANTO, J. **Data Mining Techniques Based on the Self-Organizing Map**. 1997. Dissertação (Mestrado em Ciência da Computação) — Helsinki University of Tecnology.

VESANTO, J.; HIMBERG, J.; ALHOUNIEMI, E.; PARHANKANGAS, J. **SOM toolbox for matlab 5**. [S.l.]: Helsink Univeristu of Technology, 2000. (A57).

VILLMANN, T.; DER, R.; HERRMANN, M.; MARTINETZ, T. Topology Preservation in Self-Organizing Feature Maps: exat definition and measurement. **IEEE Transactions on Neural Networks**, [S.l.], v.8, n.2, p.256–266, 1997.

YAGER, R.; FILEV, D. Approximate clustering via mountain method. **IEEE Transactions on Systems, Man and Cybernetics**, [S.l.], v.24, n.8, p.1279 1284, 1994.

ZHENG, Z. **A benchmark for classifier learning**. [S.l.]: Basser Department of Computer Science, University of Sydney, 1993. (TR474).

ZUCHINI, M. H. **Aplicações de Mapas Auto-Organizáveis em Mineração de Dados e Recuperação de Informação**. 2003. Dissertação (Mestrado em Ciência da Computação) — Universidade Estadual de Campinas, Campinas, São Paulo, Brasil.

Apêndice A

Gráficos dos conjuntos de dados

Os gráficos dos conjuntos de dados são nomeados de acordo com os nomes já atribuídos aos conjuntos no decorrer da tese. Veja a Tabela 6.1 para fazer o relacionamento entre o conjunto mostrado aqui e o contexto em que ele é utilizado na tese.

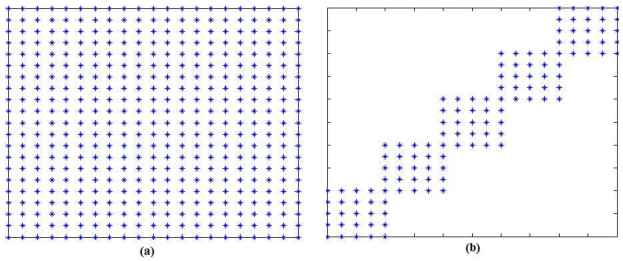


Fig. A.1: Conjunto de dados: (a) "Distribução Regular" (N^o 1) e (b) "Quadrados" (N^o 2).

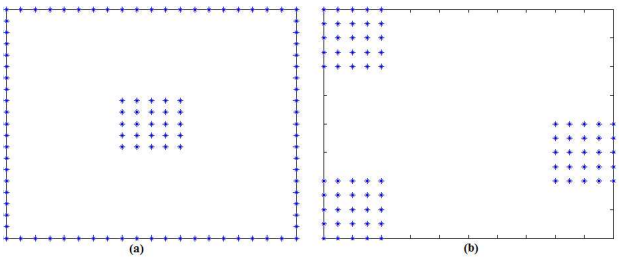


Fig. A.2: Conjunto de dados: (a) "Quadrado Central" (N^o 3) e (b) "Quadrados Dispersos" (N^o 4).

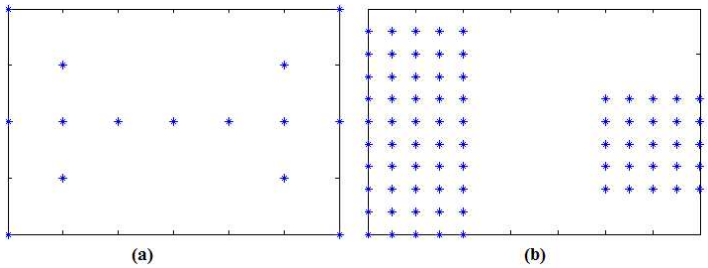


Fig. A.3: Conjunto de dados: (a) "Borboleta" (N^o 5) e (b) "Dois Quadrados" (N^o 6).

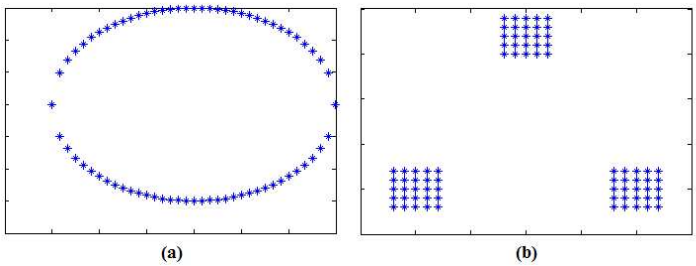


Fig. A.4: Conjunto de dados: (a) "Círculo" (N^o 7) e (b) "Três Quadrados Distintos" (N^o 8).

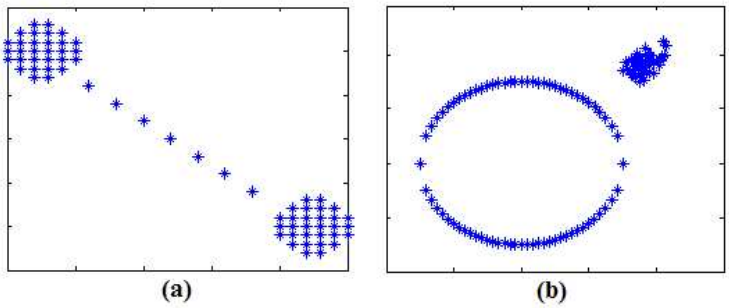


Fig. A.5: Conjunto de dados: (a) "Dois ou três grupos" (N^o 9) e (b) "Misto 1" (N^o 10).

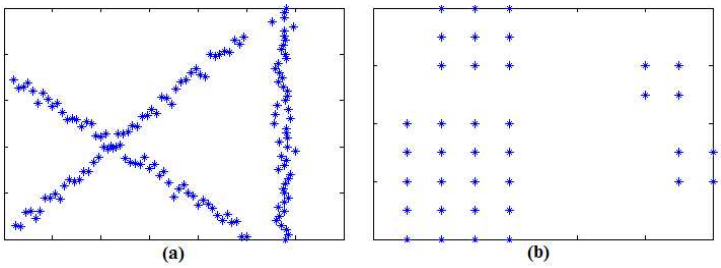


Fig. A.6: Conjunto de dados: (a) "Cruz-Linha" (N^o 11) e (b) "Grupos Pequenos" (N^o 12).

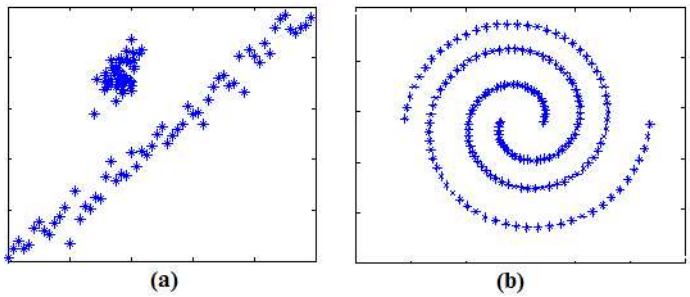


Fig. A.7: Conjunto de dados: (a) "Misto 2" (N^o 13) e (b) "Espiral" (N^o 14).

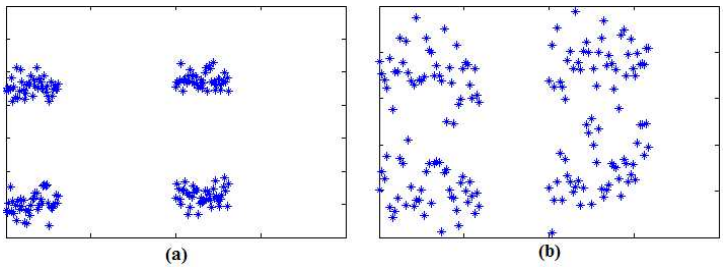


Fig. A.8: Conjunto de dados: (a) "4 Grupos Definidos" (N^o 15) e (b) "4 Grupos Dispersos" (N^o 16).

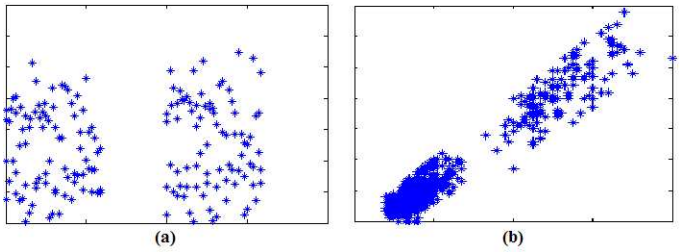


Fig. A.9: Conjunto de dados: (a) "4 Grupos Sobrepostos" (N^o 17) e (b) "Misto 3" (N^o 18).

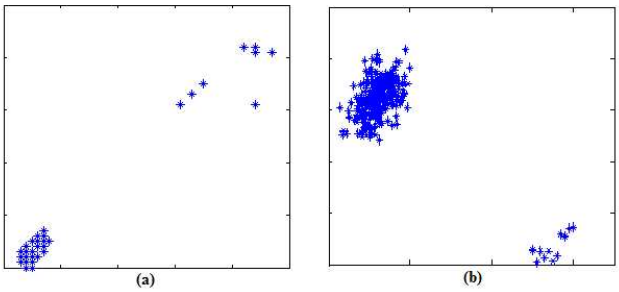


Fig. A.10: Conjunto de dados: (a) "Misto 4" (N^o 19) e (b) "Dois Grupos Distantes" (N^o 20).

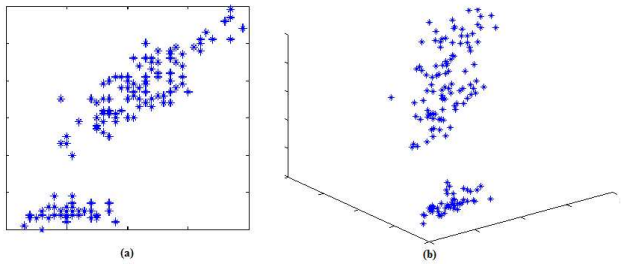


Fig. A.11: Conjunto de dados: (a) "Iris2D" - componentes 1 e 2 - (N^o 21) e (b) "Iris3D" - componentes 1, 2 e 3 - (N^o 22).

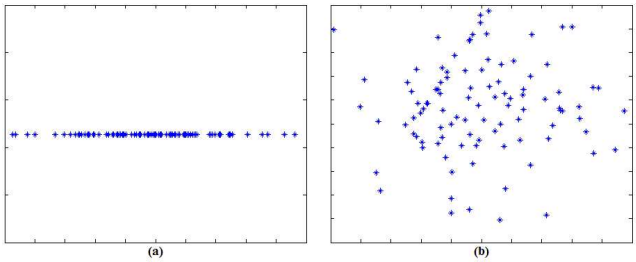


Fig. A.12: Conjunto de dados: (a) "Dist-norm-100-1" (N^o 24) e (b) "Dist-norm-100-2" (N^o 25).

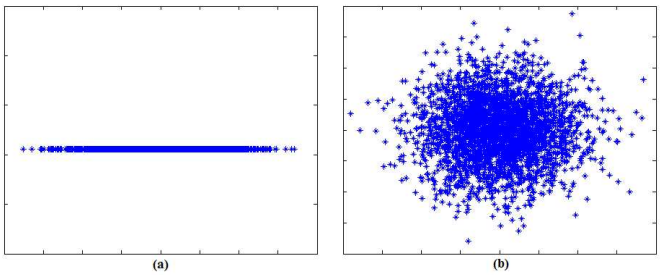


Fig. A.13: Conjunto de dados: (a) "Dist-norm-3000-1" (N^o 29) e (b) "Dist-norm-3000-2" (N^o 30).

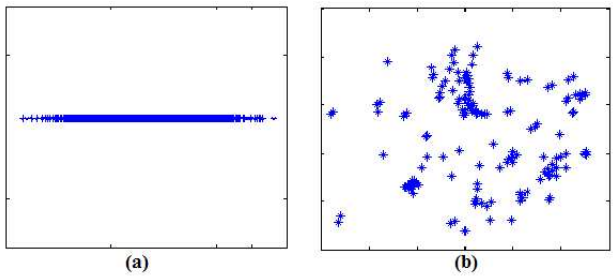


Fig. A.14: Conjunto de dados: (a) "Dist-norm-5000-1" (N^o 34) e (b) "Aleatório" (N^o 43).

Apêndice B

Descrição resumida dos conjuntos de dados reais

O conjunto de dados Íris, criado por R. A. Fisher em julho de 1988, é um dos mais citados em trabalhos referentes à análise de dados. Trata-se de um conjunto contendo 3 classes (Setosa, Versicolor e Virginica) de um tipo de planta - Íris. Das 3, uma é linearmente separável das outras duas. Os 4 atributos que descrevem os 150 dados desse conjunto são numéricos, não existem valores desconhecidos, e tratam do comprimento e largura das folhas das plantas.

O conjunto de dados Dermatology é formado por 33 atributos numéricos (discretos multi-valorados) e 1 nominal. Nele estão armazenadas informações sobre pacientes com 6 diferentes tipos de doenças dermatológicas: psoríase, dermatite, seborréica, *lichen planus*, pitiríase rosa, dermatite crônica e pitiríase vermelha. 8 valores no atributo "idade" são desconhecidos e foram substituídos por 0.

Criado por B. German, em 1987, o conjunto de Glass reúne dados referentes a amostras de vidro de uso doméstico e industrial. Esses são classificados em janelas de edifício (laminado), janelas de veículos (laminado), janelas de edifícios (não laminado), janelas de veículos (não laminado) - que não possui representante na base de dados - vidro de recipientes, louça de mesa, lâmpadas e faróis. Os vetores descritivos desses dados são compostos pelo índice de refração e por 8 elementos químicos que os compõem (nenhum valor desconhecido). Os dados são representados por números contínuos e são originários de trabalhos de investigação

criminológica.

O conjunto Ionosphere é composto por dados descritos por vetores formados por 17 pares de atributos contínuos obtidos a partir de uma função de autocorrelação que processa pulsos (em relação ao tempo de pulso e número de pulsos) de alta freqüência, disparados contra a ionosfera (nenhum valor desconhecido). Os objetos desse conjunto são classificados em dois grupos: aqueles que evidenciam algum tipo de estrutura na ionosfera - os "bons" - e aqueles que representam ruído de fundo - os "ruins".

O conjunto de dados Internet é composto, basicamente, por atributos binários. É uma base de dimensionalidade muito alta, com 1558 atributos, dos quais 3 são contínuos (com um ou mais valores desconhecidos em 28% das instâncias) e os demais são binários. Esse conjunto de atributos descreve características de uma página da Internet, com o objetivo de classificá-la em páginas de propagandas ou não.

O conjunto de dados Zôo, foi criado por Richard Forsyth em 1990 e é constituído por vetores compostos por 1 atributo numérico e 15 atributos binários. É um bom conjunto para testar a habilidade de uma abordagem em tratar dados binários em altas dimensões. O conjunto é constituído por 101 dados (sem valores desconhecidos), referentes às classes de animais: mamíferos, aves, répteis, peixes, anfíbios, insetos e moluscos e crustáceos. Segundo ZUCHINI (2003) trata-se de um conjunto de dados bem comportado, com boa separação entre as classes.

O conjunto Letter (criado por David J. Slater em 1991) é formado por dados residentes em um espaço composto por 16 atributos inteiros (sem valores ausentes), normalizados para o intervalo $[0, 15]$ e que representam 26 letras maiúsculas do alfabeto inglês (cada letra constitui uma classe diferente). Os dados foram obtidos a partir de imagens de 20 tipos diferentes de fonte.

O conjunto Abalone diz respeito à informações sobre um "abalone", as quais permitem a predição da idade desse animal. Os atributos se referem a características físicas do animal e formam um conjunto de 8 atributos, sendo que um deles é nominal (e foi transformado, para este trabalho, em discreto multivalorado) e os outros 7 são contínuos (nenhum valor desconhecido).

O conjunto Auto-mpg é usado para aprendizado com o objetivo de predição. Diz respeito a dados (cilindro, *displacement*, potência - com 6 valores desconhecidos -, peso, aceleração, ano

do modelo e origem) referentes a automóveis e prediz a quantidade de consumo em milhas por galão. São 4 atributos contínuos e 3 atributos discretos multivalorados. Além dos atributos citados, existe ainda nesse conjunto, um atributo simbólico referente ao nome do carro, único para cada instância. Tal atributo foi considerado dispensável neste trabalho por não apresentar informação de caracterização.

Os conjuntos Va, Cleveland, Hungria e Switezerland dizem respeito a informações sobre doenças do coração e foram obtidos respectivamente de: 1.V.A. Medical Center, Long Beach, CA - Robert Detrano; 2. Cleveland Clinic Foundation - Robert Detrano; 3. Hungariam Institute of Cardiology, Budapeste - Andras Janosi; 4. University Hospital, Zurich e Basel, Switzerland - William Steinbrunn e Matthias Pfisterer. Os dados são descritos por vetores de 74 atributos, dos quais 13 são usados para tarefas de análise de dados (9 são discretos multi-valorados, 3 são binários e 1 é contínuo). Essa base também é usada para preparação de ferramentas de predição. Vários valores são desconhecidos e substituídos por -9.0, como indicado na documentação correlata.

Predição de labaredas é a tarefa a ser realizada com os dados do conjunto Flare. Trata-se de supor o número de labaredas solares pequenas, médias e grandes que ocorrerão numa região da superfície solar num período de 24 horas. Os 10 atributos discretos multivalorados descrevem a atividade na superfície solar e a história da região analisada.

Apêndice C

Suporte para análise de dados

Este apêndice tem o objetivo de mostrar um pouco mais sobre o trabalho de análise feito nos testes referentes ao uso das abordagens Fractal-Fuzzy, Box-Counting e Análise de Componentes Principais para decidir sobre a dimensão ideal do espaço de saída de uma RNA-AO. Para tal, estão reproduzidas aqui algumas tabelas que ilustram o processo de geração de informações necessário para suportar os resultados mostrados nesta tese.

A análise de desempenho das abordagens mediante cada critério de avaliação gerou 19 pares de tabelas (um para cada critério de avaliação), uma referente a todos os mapeamentos e outra referente aos mapeamentos escolhido sob a heurísticas de Vesanto, como a Tabela C.1 e a Tabela C.2. Nas células da primeira tabela está representado a dimensão inferida por cada abordagem para cada conjunto de dados. Os casos em que esta inferência coincide com a análise do critério em questão está destacado em negrito. Nas células da segunda tabela está representada a porcentagem de conjuntos de dados com uma determinada característica para a qual a abordagem infere resultados corretos segundo o critério em questão. Em negrito estão os melhores desempenhos para cada característica.

Tab. C.1: Comparação com análises de qualidade feitas com o Produto Topológico (Critério 1)

Id.	FF 1	FF 2	BC 1	BC 2	ACP (0,5)	ACP (0,9)
01s	3	3	2	2	**5**	**5**
02s	3	3	**2**	**2**	1	**2**
03s	4	**3**	2	2	1	1
04s	**2**	**2**	1	1	1	**2**
05s	3	**2**	**2**	**2**	1	**2**
06s	4	3	3	**2**	1	1
07s	4	3	3	**2**	1	1
08s	3	3	**2**	**2**	1	1
09s	**2**	**2**	1	1	1	1
10s	5	5	5	5	1	1
Total	2	4	3	**5**	1	4
01r	**2**	**2**	**2**	1	1	**2**
02r	5	4	5	4	2	6
03r	**3**	2	1	1	1	5
04r	5	4	2	1	1	7
05r	6	6	1	1	2	23
06r	5	**4**	**4**	3	2	9
07r	6	6	2	2	2	6
08r	3	**2**	**2**	1	1	**2**
09r	**3**	**3**	2	2	1	**3**
10r	4	4	2	1	1	5
11r	5	**4**	2	1	3	7
12r	**3**	**3**	2	1	2	5
13r	4	**3**	2	2	1	6
14r	3	2	2	2	2	6
Total	4	7	3	0	0	3
Total Geral	6	11	6	5	1	7

Tab. C.2: Desempenhos, de acordo com o Critério 1, considerando os conjuntos de dados agrupados por características.

Características		FF 1	FF 2	BC 1	BC 2	ACP (0,5)	ACP (0,9)
Número	Pequeno	45	**64**	45	27	9	55
de	Médio	11	**44**	11	22	0	11
Atributos	Grande	0	0	0	0	0	0
Tamanho	Pequeno	25	**50**	25	0	0	50
do	Médio	45	**45**	18	18	9	36
Conjunto	Grande	0	**33**	33	**33**	0	11
Densidade	Baixa	17	0	17	**33**	0	0
do	Média	29	**64**	36	21	7	36
Conjunto	Alta	25	**50**	0	0	0	**50**
Número	Pequeno	22	**44**	33	33	11	**44**
de	Médio	25	**33**	8	17	0	17
Classes	Grande	33	**100**	67	0	0	33
Predominância	Baixa	30	**60**	30	20	0	30
de	Média	27	**45**	27	27	0	27
classe	Alta	0	0	0	0	**33**	33
	Baixa	0	0	0	0	**33**	**33**
Entropia	Média	31	**62**	31	23	0	31
	Alta	25	**38**	25	25	0	13

A organização de dados, similar à estrutura da Tabela C.3, foi necessária para suportar a obtenção de resultados mais concisos e expressivos. Essas informações se referem à totalização dos desempenhos de cada abordagem sob cada um dos critérios de análise utilizados. O desempenho está mostrado em termos do número de conjuntos (sintéticos e reais) para os quais a abordagem conseguiu inferir a dimensionalidade que se mostrou mais adequada segundo cada critério de avaliação. As demais informações similares a essas mostradas no exemplo se referem a: todos os critérios X conjuntos sintéticos; todos os critérios X conjuntos reais; todos os critérios X todos os conjuntos; melhores critérios X conjuntos sintéticos; melhores critérios X conjuntos reais; melhores critérios X todos os conjuntos; os critérios X conjuntos sintéticos X heurística Vesanto; todos os critérios X conjuntos reais X heurística de Vesanto; melhores

critérios X conjuntos sintéticos X heurística de Vesanto; melhores critérios X conjuntos reais X heurística de Vesanto; melhores critérios X todos os conjuntos X heurística de Vesanto. A última linha indica o número de critérios segundo os quais a abordagem obteve sucesso.

Tab. C.3: Desempenho: todos os critérios X todos os conjuntos X heurística de Vesanto.

Critérios	FF 1	FF 2	BC 1	BC 2	ACP (0,5)	ACP (0,9)
Critério 1	5	**8**	6	3	4	**8**
Critério 2	3	6	8	**9**	5	6
Critério 3a	0	0	4	10	**16**	7
Critério 3b	0	0	4	10	**16**	7
Critério 4a	8	**12**	8	4	2	3
Critério 4b	8	**12**	8	5	4	4
Critério 5a	0	0	4	10	**16**	7
Critério 5b	0	0	4	10	**16**	7
Critério 6a	6	**13**	7	3	2	3
Critério 6b	6	7	**10**	5	2	4
Critério 7	3	7	**13**	11	6	3
Critério 8	**7**	6	3	4	4	0
Critério 9	**7**	6	3	4	4	0
Critério 10a	5	**10**	7	5	1	6
Critério 10b	2	3	**9**	8	4	2
Critério 11a	5	6	5	**8**	6	2
Critério 11b	4	8	**11**	7	3	2
Critério 12	7	**9**	2	1	1	1
Critério 13	5	5	3	5	4	1
Total (dentro de 19 critérios)	2	6	4	2	4	1

Um outro conjunto de informações gerado para suportar as análises finais relacionou o número de vezes que a abordagem inferiu um resultado bom, em relação a cada um dos melhores critérios (com ou sem o uso da heurística de Vesanto), quando da análise de conjuntos de dados com determinada característica (18 possibilidades de características com 9 possibilidades de critérios). Por exemplo, a Tabela C.4 ilustra a forma como os dados foram estruturados

para conjuntos com número de dados pequeno (4 conjuntos) em relação aos melhores critérios escolhidos a partir do uso da heurística de Vesanto.

Tab. C.4: Relação abordagens X melhores critérios (com o uso da heurística de Vesanto) para conjuntos com número de dados pequeno

Critérios	FF 1	FF 2	BC 1	BC 2	ACP (0,5)	ACP (0,9)
Critério 2	0	0	2	2	1	0
Critério 4a	2	1	0	0	0	0
Critério 4b	1	1	1	0	1	1
Critério 6a	0	1	0	0	0	0
Critério 10a	1	2	0	1	1	1
Critério 10b	1	2	0	0	1	1
Critério 11a	1	0	0	0	0	0
Critério 11b	0	2	0	2	2	0

Apêndice D

Mídia com informações adicionais

Faz parte desta tese, uma mídia que contém os programas referentes às implementações utilizadas nesta tese, os conjuntos de dados sintéticos, os resultados obtidos a partir destes programas (as redes neurais artificiais e suas análises) e o conjunto de arquivos em formato .pdf e .eps referentes ao texto e às figuras presentes no mesmo. Os conjuntos de dados reais podem ser encontrados na bibliografia refenciada no texto.