

**UNIVERSIDADE FEDERAL DE SANTA CATARINA
PROGRAMA DE PÓS-GRADUAÇÃO EM
ENGENHARIA DE PRODUÇÃO**

**Famílias de T-Normas Diferenciáveis,
Funções de Pertinência Relacionadas e
Aplicações**

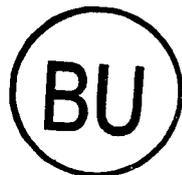
**Tese submetida à Universidade Federal de Santa Catarina para a obtenção do
grau de Doutor em Engenharia.**

Maria Bernadete Zanusso



0.273.775-1

UFSC-BU



Florianópolis

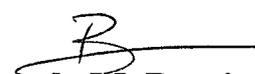
Santa Catarina

Dezembro de 1997

Autora: Maria Bernadete Zanusso

**Famílias de T-Normas Diferenciáveis,
Funções de Pertinência Relacionadas e
Aplicações**

Esta tese foi julgada adequada para a obtenção do título de "Doutor em Engenharia", especialidade Engenharia de Produção e aprovada em sua forma final pelo Programa de Pós-Graduação.


Ricardo M. Barcia, PhD

Coordenador do Programa de Pós-Graduação em Engenharia de Produção

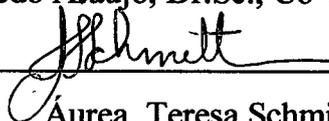
Banca Examinadora:



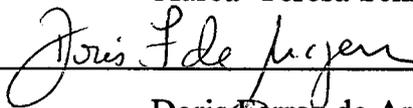
Luiz Fernando Jacintho Maia, Dr.Eng, Orientador



Aldrovando Luís Azeredo Araújo, Dr.Sc., Co-Orientador

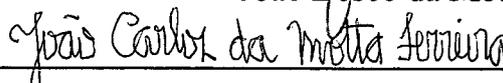


Aurea Teresa Schmitt, Dr.Ent.

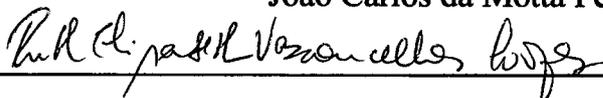


Doris Ferraz de Aragon, Dr.Sc.


João Bosco da Mota Alves, Dr.Eng.



João Carlos da Motta Ferreira, Dr.



Ruth Elisabeth Vasconcellos Lopes, M.Sc.

**Ao Marcelo
e à Janaina**

AGRADECIMENTOS

Desejo agradecer a todos que, direta ou indiretamente, contribuíram com estímulos, trabalho e alegria para que eu pudesse escrever esta tese, desde a minha família, aos colegas do Departamento de Computação e Estatística da Universidade Federal do Mato Grosso do Sul e os amigos da Universidade Federal de Santa Catarina e de Florianópolis.

Gostaria, porém de enfatizar o meu agradecimento a alguns em particular por seus conhecimentos e sua boa vontade em transmití-los a mim.

Em primeiro lugar quero mencionar o dr. Luiz Fernando Jacyntho Maia, meu orientador, por ter encontrado em sua tese o tema para a minha, o qual correspondia ao projeto que tinha idealizado ao sair de Campo Grande. Além disto, deu-me a idéia básica para as funções de caracterização. Não fosse pela confiança que depositou em mim, e por sua paciência, este trabalho certamente não se teria realizado.

Ao dr. Aldrovando Luís Azeredo Araújo, professor do departamento de Matemática, meu coorientador, agradeço pelo entusiasmo, amizade, bom humor, principalmente por ter dedicado seu conhecimento e dias de trabalho, contribuindo consideravelmente para a elevação do nível desta tese.

Agradeço ao dr. Jorge M. Barreto, atualmente professor coordenador do Curso de Pós-Graduação em Ciência da Computação, pela atenção e papos que tivemos sobre reconhecimento de padrões e aprendizado em redes neurais, contribuindo muito para meu amadurecimento neste campo.

Ao dr. Fernando M. de Azevedo, professor chefe do Departamento de Engenharia Biomédica, pelo interesse por nossas t-normas e conversas que tivemos estimulando a proposta do modelo de classificador que propus.

A Lourdes M. Brasil, aluna do doutorado em Engenharia Biomédicas e amiga que, além de apoio afetivo nos momentos difíceis, me estimulou a usar o editor de texto Latex passando-me os primeiros passos e também pelos estudos em conjunto.

Ao dr. José Leomar Todesco e outros professores da Pós-Graduação em Engenharia de Produção, pelo empréstimo de livros, pelos papos sobre modelos de redes neurais e o software MATLAB.

Ao dr. João Carlos da Motta Ferreira, professor do Departamento de Matemática da Universidade Federal do Mato Grosso do Sul, pelo seu interesse no tema da tese e amizade.

Ao dr. José Eduardo Corrente, Professor do Departamento de Estatística da USP-ESALQ, pela amizade que mantivemos desde o tempo do curso de graduação e também por seu estímulo e sugestões da área de Estatística.

Agradeço à dra. Doris F. Aragon, pesquisadora senior do ILTC-RJ na área de inteligência artificial; à dra. Áurea T. Schmitt, entomologista da EPAGRI-EEItajaí; a M.Sc Ruth E. V. Lopes, professora da UFSC-DLV e a dr. João B. M. Alves, pesquisador em robótica, por terem aceitado participar da banca de avaliação da tese.

Finalmente, parabênzo o Programa de Pós Graduação em Engenharia de Produção, junto com seus membros, pela sua política que motiva a ampliação de conhecimentos em todos os ramos de atividade, principalmente quanto aos seus relacionamentos e à CAPES pela bolsa do PICDT.

Sumário

Sumário	iii
Lista de Figuras	iv
Lista de Tabelas	v
Resumo	v
Abstract	vi
1 Introdução	1
1.1 Preâmbulo	1
1.2 Descrição do Problema e Objetivo	3
1.3 Tópicos que Motivaram a Pesquisa	4
1.3.1 Mudança de paradigma	4
1.3.2 Formação de conceitos e modelos	4
1.4 Justificativas	4
1.4.1 Aplicações da teoria dos conjuntos difusos	4
1.4.2 Outras t-normas e t-conormas	5
1.4.3 O contexto de aplicação	6
1.5 Estrutura e Perfil da Dissertação	7
2 REVISÃO: Conjuntos Difusos	9
2.1 Introdução	9
2.2 Um Breve Histórico	9
2.3 Fontes de Imprecisão	10
2.4 Aplicações e Vantagens da Imprecisão	10
2.5 Conjuntos Difusos e Pertinência	11
2.5.1 Qual é a relação de V com X ?	12
2.5.2 Representação	12
2.6 Experiências Empíricas	13
2.7 Operações sobre os Conjuntos Difusos	14
2.7.1 Operadores padrões	14
2.7.2 Outros operadores	15
2.7.3 A dependência do contexto	16
2.8 T-Normas e T-Conormas	16
2.8.1 Caracterização de interseção e união difusa	17
2.8.2 A força da interseção e da união	18
2.9 A Teoria de Protótipos	18

2.10	Incerteza Probabilística e Incerteza Difusa	19
3	REVISÃO: Taxionomia Biológica	20
3.1	Introdução	20
3.2	Definição de Termos	21
3.3	Definição de Espécie	21
3.3.1	Operações com conjuntos ordinários e taxionomia	22
3.3.2	Problemas com a definição de espécie	23
3.3.3	As flores íris	25
3.4	Conjuntos Difusos e Outras Abordagens	25
3.5	Problemas de Identificação	26
4	Famílias de T-Normas e Funções de Pertinência	28
4.1	Introdução	28
4.2	Definições e Caracterizações	30
4.3	Famílias de T-Normas e T-Conormas	33
4.4	Espaço de Funções de Pertinência	37
4.4.1	Diferenciabilidade das t-normas e t-conormas	37
4.4.2	Uma família particular de função de pertinência	38
4.4.3	Relação de ordem e convexidade	44
4.5	Quando a Conjunção não é T-Norma:	51
5	APLICAÇÃO: Caracterização de Conceitos	52
5.1	Introdução	52
5.2	Uma Aplicação no Contexto da Taxionomia Biológica	53
5.3	Sobre o Conjunto de Dados	56
5.4	Função de Pertinência para um Conceito e um Atributo	56
5.5	A Identificação Proposta	59
5.5.1	Um conceito como conjunção de conceitos	59
5.5.2	Um conceito como disjunção de conjunções	61
5.6	Classificador	61
5.6.1	Definição de classificador	62
5.6.2	O erro médio quadrático	62
5.6.3	Algoritmo de minimização	64
5.6.4	Fase de teste	65
5.6.5	Funções de caracterização	65
5.6.6	Alguns gráficos	67
5.6.7	Conclusões das subseções acima	69
5.7	Aproximação de funções	70
6	APLICAÇÃO: Rede Neural	71
6.1	Introdução	71
6.2	A Versão Difusa do PMC	71
6.3	T-Norma e T-Conorma	72
6.4	A medida de erro	72
6.4.1	Derivada das t-normas	73
6.4.2	Derivadas do erro	73

7 Conclusões	75
8 Sugestões	77
Apêndices	77
Referências Bibliográficas	77

Lista de Figuras

4.1	Função de pertinência	29
4.2	Variações de α e de β na família de t-normas	34
4.3	Gerador decrescente	36
4.4	Gerador crescente e seu inverso	37
4.5	Algumas variações de k com p fixo	39
4.6	Variações de k com p fixo	41
4.7	Jovem e velho	41
4.8	velho e muito velho	42
4.9	Relação com máximo e mínimo	43
4.10	Não convexidade da conjunção em \mathcal{F}	44
4.11	Convexidade da conjunção	49
4.12	Convexidade da conjunção	49
4.13	A convexidade da conjunção e não convexidade da disjunção	50
4.14	Relação com o mínimo	51
5.1	Os dados íris de Anderson	57
5.2	Os dados íris de Anderson	58
5.3	Um conceito que caracteriza a classe G	59
5.4	Conceito primário e suas modificações	60
5.5	Classificador (básico) de padrões	62
5.6	O classificador que propomos	63
5.7	Parte que reconhece A	64
5.8	Gráfico de EA com $txa = 0.5$	67
5.9	Gráfico de EB	68
5.10	Gráfico de EC	68
5.11	Gráfico de EA com dois parâmetros variando	69
6.1	Rede neural AND e OR para a classificação e geração de regras	74

Lista de Tabelas

3.1	Fêmeas do gênero <i>Meloidogyne</i> (em micrômetro: μm)	25
3.2	Medidas da Sépala e da Pétala para Flores Íris(em centímetros: cm) .	25
5.1	Resultados Obtidos por Nosso Classificador	65

Resumo

Os operadores conjuntivos e disjuntivos para conjuntos difusos são uma extensão da definição de interseção e união, respectivamente, para os conjuntos ordinários. São muitos os operadores propostos na literatura. O que há de comum entre eles é a não-satisfação da lei do meio excluído. Na tentativa de caracterizar a interseção e a união difusa e também de criar um sistema lógico difuso, vários operadores foram propostos, com base em diferentes axiomas que deveriam ser satisfeitos. Entretanto, as experiências empíricas mostraram que cada operador é adequado para um contexto diferente. Reconhecidamente, a interseção e a união difusa ficam caracterizadas com os quatro axiomas básicos para t-normas e t-conormas. A tese propõe uma família de t-normas e t-conormas diferenciáveis. A diferenciabilidade é uma qualidade que destaca essa família entre as que dependem do máximo ou do mínimo. Além disto, há uma estreita relação com a família de funções de pertinência que dependem de parâmetros que permitem a calibração de conjunções e disjunções delas para representar conceitos em diferentes contextos. Para sistematizar a calibração propõe-se um classificador que tem como saída funções caracterizadoras, ou discriminantes, de conceitos, utilizando conjunções e disjunções de conceitos elementares. Neste contexto vamos entender conceito, categorias ou classes como um conjunto de objetos com determinados valores de atributos. Este é um modelo diferente dos comumente encontrados na literatura, baseados em funções discriminantes lineares ou probabilísticas. Existem poucos pesquisadores em Taxionomia Numérica que utilizam conjuntos difusos, ainda que os taxionomistas freqüentemente encontrem problemas com as formas intermediárias e sua nomenclatura. Na tese, o conjunto de dados íris é usado para treinar o classificador, que no final fornece funções que caracterizam três espécies de flores íris. Na literatura, é comum encontrar redes neuronais classificadoras, ou geradoras de regras, que utilizam somas ponderadas ou superposição de funções sigmóides e neurônios *and* e *or* com as t-normas do máximo e do mínimo. Na tese são apresentadas as derivadas para a implementação do algoritmo da descida do gradiente numa rede neuronal com neurônios *and* e *or* que utilizam as t-normas propostas ao invés das comumente utilizadas na literatura.

Abstract

Conjunctive and disjunctive fuzzy set operators are extensions of the definitions of intersection and union for crisp sets, respectively. There are many such operators whose common characteristic is that they don't satisfy the law of excluded middle. In attempt to qualify the fuzzy intersections and fuzzy unions and to create a fuzzy logic system, several operators were proposed based on different axioms that should be satisfied. In the mean-time, empirical evidences show that each operator is adequate for different contexts. In the literature it is recognized that the characterization of fuzzy intersection and fuzzy union remain well defined with the four basic axioms of t-norms and t-conorms. This dissertation proposes a differentiable family of t-norms and t-conorms, an outstanding quality between them which depends on maximum and minimum. Besides, there is close relation with a family of membership functions that depends on parameters which allow the calibration of conjunctions and disjunctions to represent concepts in different contexts. To put this calibration in a system a classifier is proposed giving as output characterization functions, or discriminants, of concepts, using conjunctions and disjunctions of elementary concepts. This is a different model among the ones found in the literature, based on linear or probabilistic discriminant functions. There are few researchs on Numerical Taxonomy where fuzzy sets are utilized, although taxonomists frequently meet problems of intermediate forms and their nomenclature. In this dissertation, the data *iris* are used to train a classifier, that produces as output functions that characterize three species of iris flowers. In the literature it is frequently found neural networks classifiers, or rule generators, which use ponderate sums or superpositions of a sigmoidal function and neurons *and* and *or* with maximum and minimum t-norms. This dissertation presents the derivatives to the implementations of the gradient descent algorithm in a neuronal network with *and* e *or* neurons using the proposed t-norms. In this context, we are considering concepts, categories or classes as a set of objects with certain values of attributes.

Capítulo 1

Introdução

1.1 Preâmbulo

Pesquisadores como Leach [20] e Piaget [54] concordam que o ambiente físico e social de uma criança é percebido como um contínuo, que não contém *coisas* separadas. Mas por que passam a separá-las quando deixam de ser crianças? Ou, por que os índios, considerados de cultura primitiva, classificam os pássaros pelo canto?(Jensen[3]).

A verdade é que no dia-a-dia temos bastante sucesso em lidar com um ambiente que muda continuamente. Pois, ainda que seja raro observarmos um mesmo objeto exatamente nas mesmas condições, as variações não parecem interferir em nosso reconhecimento desses objetos. Por outro lado, os contextos em que eles aparecem, o tamanho, a forma, a cor e também o material de que são feitos podem mudar e no entanto sempre são reconhecidos corretamente. Vejam por exemplo os funis, freqüentemente os encontramos em lugares diferentes, nas mais variadas formas, material, etc., e no entanto sempre são reconhecidos corretamente. Por trás deste fenômeno existe a habilidade de segmentar o mundo em classes de equivalências. Sem ela, toda instância de cada tipo de objeto, eventos ou situações sobre as quais se possa pensar pareceriam novos toda vez que fossem encontrados. É através do uso dos processos e das representações envolvidas quando abstraímos alguma coisa de fundamental dos objetos, que construímos modelos a partir das observações empíricas. Estes modelos se aprimoram desde os estereótipos que fazemos das pessoas em nossos encontros diários até os modelos complexos da física quântica.

A aprendizagem ou formação de conceitos é a área das ciências cognitivas que trata destes mecanismos. Neste contexto, entende-se o conceito como uma representação mental de uma categoria de objetos do mundo, dada por uma seqüência de conjunções ou disjunção de conjunções. Como, as conjunções e disjunções estão associadas ao *e* e ao *ou* lógicos, estamos num contexto lógico de representação do conhecimento.

Aristóteles foi quem primeiro criou um sistema lógico para representar o conhecimento que extraia de suas observações empíricas dos objetos do mundo. Em seu sistema, um objeto pertencia a uma categoria se tivesse os atributos correspondentes a ela. Se não tivesse um dos atributos, o objeto estaria numa outra categoria. Esta foi a base para o cálculo de predicados em que as sentenças ou são verdadeiras ou são falsas. Contudo, são freqüentes os casos intermediários, em que não é possível determinar com exatidão a que categoria pertence um objeto. Isto

ocorre devido a várias fontes de incerteza.

Körner [63], em 1966, ofereceu uma alternativa para a determinação exata destes problemas, propondo uma lógica de três valores, que permitia uma terceira categoria distinta de *verdadeira* e *falsa* para acomodar os casos limite. Mas isto não resolveu o problema satisfatoriamente, pois exigia que fosse traçada uma linha marcante entre os casos limite e os verdadeiros ou falsos. Assim, seguramente se impunha uma precisão artificial, já que insistir que a uma dada altura um homem deixa de ser um caso limite e se torna claramente alto é a mesma coisa que insistir que a uma dada altura um homem deixa de ser não-alto para tornar-se claramente alto .

Zadeh [65], em 1975, estabeleceu a base formal para sua lógica difusa, que está baseada na teoria dos conjuntos difusos. Enquanto na teoria clássica dos conjuntos um objeto ou é ou não é membro de um dado conjunto, na teoria dos conjuntos difusos a pertinência é uma questão de grau; o grau de pertinência de um objeto a um conjunto difuso varia num intervalo contínuo de 0 a 1, com 0 denotando não-pertinência e 1, pertinência total. Em 1965, Zadeh [65] propôs que o grau de pertinência à interseção de dois conjuntos difusos fosse dado pelo mínimo dos graus de pertinência a cada conjunto envolvido e o grau de pertinência à união de dois difusos fosse dado pelo máximo. Estas operações ficaram conhecidas como *standards*.

Os resultados sobre operações que sejam extensões das operações de interseção e união sobre conjuntos ordinários são inúmeros na literatura. De certo modo elas estão axiomatizadas e caracterizadas com base no trabalho de Schweizer e Sklar [2]. A saber, tais operações podem ser caracterizadas por expressões envolvendo funções com propriedades bastante gerais. Em Klir [36] apresenta-se a t-norma, operação que representa a interseção de conjuntos difusos, com a seguinte caracterização: existe $f: [0, 1] \rightarrow R$, contínua decrescente com $f(1) = 0$ que admite uma pseudo-inversa f^{-1} tal que,

$$i(a, b) = f^{-1}(f(a) + f(b)),$$

em que f são denominadas geradores, pois com elas podemos gerar t-normas à vontade, para os mais variados contextos de aplicações.

Nossa proposta é bastante diferente, pois estaremos trabalhando no intervalo $[-1, 1]$ e usaremos geradores com singularidades. Mais precisamente nossos geradores estão bem definidos em $[-1, 1)$ ou $(-1, 1]$ e são inversíveis.

Como o procedimento natural para estender operações sobre números para funções é estendê-la pontualmente, então para cada t-norma existe um gerador decrescente e vice-versa. Assim, nossos operadores são estendidos pontualmente para funções $\varphi: (0, +\infty) \rightarrow (-1, 1)$. Usando uma estrutura métrica adequada somos capazes de fazê-los operar sobre funções com imagem em $[-1, 1]$, preservando todas as propriedades de t-normas e t-conormas.

Contudo, as t-normas também compartilham da mesma propriedade do mínimo de que não pode existir um objeto que tenha maior grau de pertinência à interseção do que a cada um dos conjuntos envolvidos. Alguns pesquisadores, como Osherson e Smith [23] e Mervis e Rosch [21], argumentaram com exemplos que desestimulam o uso dos operadores de máximo e de mínimo e de qualquer t-norma ou t-conorma para combinar ou compor conceitos associados a conjuntos difusos. Mas seus exemplos também foram revistos e criticados por outros pesquisadores.

Entretanto, na tese propomos uma família de t-normas e t-conormas que variam de acordo com parâmetros que sob determinadas condições produzem t-normas e sob outras produzem resultados que podem se adequar às argumentações de Osherson e Smith.

1.2 Descrição do Problema e Objetivo

A teoria dos conjuntos difusos fornece instrumentos para representar e manipular conceitos lingüísticos vagos e dependentes do contexto. Ela permite expressar e manipular várias relações, funções e equações que envolvem estes conceitos.

A dependência do contexto envolve não somente o significado dos termos lingüísticos, associados às variáveis lingüísticas, mas também o significado das operações de composição sobre eles, as quais permitem criar conceitos complexos. As investigações empíricas demonstram que os conectivos dos termos lingüísticos, tais como a conjunção *e*, a disjunção *ou*, a negação *não* e o condicional *se-então* têm diferentes significados em diferentes contextos.

Na tese propomos três operadores básicos para a combinação de conceitos difusos, a saber, conjunção, disjunção e negação. Na literatura, estas operações são caracterizadas pelos axiomas da definição de t-norma e t-conorma. Assim, o nosso primeiro objetivo na tese é mostrar a validade destes axiomas para os nossos operadores. Além disso, nos estenderemos para a verificação de outras propriedades, com o objetivo de possivelmente modelar um sistema lógico.

Como aplicação, propomos uma agregação, ou composição, de funções de pertinência tipo sigmóide para representar conceitos no contexto da Taxionomia Biológica, com o objetivo de produzir uma função que caracterize o conceito de espécie. Isto vai ser feito utilizando um modelo de classificador que implementa a heurística do gradiente descendente.

Na literatura, encontramos redes neuronais lógicas, que utilizam neurônios *and* e *or* que efetuam operações conjuntivas e disjuntivas para classificar ou gerar regras num sistema especialista híbrido. Assim, propomos um modelo de perceptron de múltiplas camadas classificador, que utiliza os nossos operadores no lugar de somas ponderadas de funções sigmóides. A regra delta, generalizada para o modelo da rede, foi facilmente deduzida, dada a diferenciabilidade de nossos operadores, em contraste com os operadores dependentes do máximo e do mínimo que requerem um extenso trabalho de diferenciação.

Estudos futuros poderão verificar em que contexto do conhecimento estes operadores são mais adequados. Contudo, escolhemos o campo da Taxionomia Biológica para desenvolver nossa aplicação porque, além de ser rico em casos de imprecisão na definição das classes, estimula nossas reflexões sobre a mudança do paradigma da lógica clássica de Aristóteles (em que se baseia o sistema de classificação de Lineu e, ainda hoje, muitas outras ciências) para, talvez, um sistema lógico difuso.

Uma questão que surge é o que aconteceria com a Taxionomia Biológica se se abandonasse o paradigma lineano que cria classes mutuamente exclusivas. Os objetos que pertencem ou não a uma classe (espécie, gênero, etc.) passariam a ter pertinência variando num contínuo de graus de pertinência, ou seja, poderiam ter

0.1 de pertinência à classe A, 0.5 à classe B e 0.9 à classe C. Quais são as implicações desta graduação de pertinência para o estudo dos agrupamentos e hierarquias da taxionomia?

1.3 Tópicos que Motivaram a Pesquisa

1.3.1 Mudança de paradigma

O conceito de *paradigma científico* foi estudado por Thomas Kuhn, em seu importante livro *The Structure of Scientific Revolutions* (Univ. of Chicago Press, 1962). Ele é definido como um conjunto de teorias, padrões, princípios e métodos pressupostos pela comunidade científica em uma dada área da ciência. Considerando que é sobre um sistema lógico que as várias ciências se desenvolvem, porque precisam de um padrão, formado por critérios de validade a fim de obter maior rigor e exatidão, isto não poderia deixar de estimular a reflexão quanto à possibilidade de um novo paradigma para a ciência baseado em um sistema lógico difuso. Entretanto, torna-se difícil imaginar até mesmo o que ocorreria no caso da Taxionomia Biológica, que tem como paradigma a lógica de Aristóteles.

1.3.2 Formação de conceitos e modelos

Conjunções e disjunções estão estreitamente relacionados com os modelos de conceitos. Uhr [64], em suas experiências com cartões em que se desenham figuras para estudar como se realiza o processo de formação de conceito, faz a seguinte suposição: um *conceito* será alguma conjunção particular, tal como *grandes círculos pontilhados*, ou uma disjunção de conjunções, tal como *círculos pontilhados ou quadrados sólidos* que o sujeito deve aprender.

Diferentes noções de conceitos podem ser encontradas em sistemas de representação de conhecimento em inteligência artificial, nas teorias de compreensão da linguagem natural, em teorias da lógica e semântica e em considerações psicológicas da memória semântica. Entretanto, segundo Cohen e Murphy [39], parece haver muita confusão sobre o que trata a teoria de conceitos.

Não é um dos objetivos da tese estudar os modelos de conceitos. Mas este é um tema motivador, considerando que certamente poderemos desenvolver pesquisas de como nossos operadores se aplicam a vários contextos e também fazer comparações de seus desempenhos.

1.4 Justificativas

1.4.1 Aplicações da teoria dos conjuntos difusos

A teoria dos conjuntos difusos foi introduzida por Zadeh [65], [66], [67] para descrever uma situação em que é impossível determinar se certos elementos pertencem ou não a um conjunto particular, devido à presença de ambigüidades nas características dos elementos e na formação de nossos conceitos e julgamentos. Assim, os conjuntos difusos estão estreitamente relacionados a problemas básicos de reconhecimento de padrões e de tomada de decisões.

Muitas são suas aplicações tanto no contexto acadêmico quanto nos comerciais, financeiros, transportes, medicina, etc. Uma delas é a classificação dos aboios do Nordeste, que são os cantos improvisados pelos boiadeiros para conduzir boiadas ou procurarem um animal perdido. Eles não podem ser tocados por um piano, porque este não produz os infinitos microtons entre um dó e um ré, típicos dos aboios. Poderiam ser tocados por outros instrumentos, mas a dificuldade de classificação permaneceria, segundo estudiosos.

São numerosos os campos de aplicação de operadores conjuntivos e disjuntivos quando associados aos termos *and* e *or*, respectivamente. Particularmente podemos dizer, de acordo com Brasil [29], Mitra [104], Pedrycz [6], Fu [76], Keller [43] e Machado [5], que a lógica difusa tem sido usada para a classificação e geração de regras utilizando um sistema especialista conexionista difuso.

No campo da comunicação por meio da máquina, existe o problema do reconhecimento da fala. A aplicação da teoria aparece nos trabalhos de Pal [12] e Mitra [104]. Utilizando métodos de classificação com base em algoritmos difusos, eles procuram responder à seguinte pergunta: Qual é a vogal contida em uma expressão desconhecida em um grande número de palavras faladas, e quem é o seu informante?

A teoria de conjuntos difusos representa uma ferramenta conveniente na modelagem de processos complexos de transporte e tráfego, cujos parâmetros, segundo Teodorović [11], Pappis [82] e Sugeno [85] são caracterizados por incerteza, subjetividade, imprecisão e ambigüidade.

Aplicações na área financeira foram desenvolvidas por Pacheco [93] e Rodrigues [8]. Lapolli [26] desenvolveu aplicações no processamento de imagens digitais. Wee [108] a utiliza na teoria dos autômatos e Tamura [102] e Marinos [86] na teoria da linguagem e do reconhecimento de padrões.

Na medicina, para diagnóstico médico, podemos citar os trabalhos de Azevedo [27] e Barreto [28].

1.4.2 Outras t-normas e t-conormas

Estas funções surgem naturalmente no estudo das desigualdades triangulares generalizadas para espaços métricos estatísticos. Menger [62] foi o primeiro a usar a denominação *norma triangular*, resumindo *t-norma*. Schweizer [2] afirmou a importância de se ter à mão grande repertório de t-normas e foi quem estabeleceu para elas uma caracterização que nos permite construí-las à vontade. Klir [36] faz uma revisão dos operadores de interseção e união difusa colocando as t-normas e t-conormas como funções qualificadas para efetivá-las. Segundo ele, isto ocorre rotineiramente na literatura.

As operações-padrão de interseção, união e complementação de conjuntos difusos foram definidas por Zadeh [65] de modo a se reduzirem à definição usual de interseção e união quando os graus de pertinência estão no conjunto $\{0,1\}$. Ou seja, elas são generalizações das correspondentes operações clássicas sobre conjuntos ordinários. Mas estas não são as únicas generalizações possíveis. Existe, para cada operação, uma classe ampla de funções, qualificadas como generalizações das operações clássicas, em que cada uma é caracterizada por axiomas adequadamente justificados.

Estas classes de operações de complementação, interseção e união di-

fusa permitem muitas alternativas de aplicação e avaliação de suas propriedades dentro de um determinado contexto. Assim, não só a função de pertinência de conjuntos difusos, mas também as operações sobre eles são dependentes do contexto. A capacidade de determinar funções de pertinência apropriadas a operações difusas significativas para o contexto de cada aplicação particular é muito importante para a prática da teoria de conjuntos difusos.

Os operadores de conjunção e disjunção de conjuntos difusos foram descritos inicialmente por Maia [71]. Em sua tese, Maia propõe a utilização destes operadores para a caracterização e reconhecimento de conceitos, na qual considera o conceito como um padrão, ou categoria simples ou complexa, obtido por um processo de abstração de atributos comuns a objetos.

Na tese, daremos maior formalização matemática a estes operadores, mostrando que de fato caracterizam interseção e disjunção de conjuntos difusos, isto é, que são t-normas e t-conormas.

Aplicações da teoria da decisão é um dos campos em que há maior discussão em torno da adequabilidade dos operadores. Temos que avaliar um conjunto de objetos de acordo com algumas de suas propriedades. Isto implica a tomada de alguma decisão, entendendo o termo decisão como o ato de dividir os objetos em dois ou mais grupos com base na avaliação de suas propriedades. Ou eles satisfazem um nível esperado, ou eles ficam abaixo deste nível. Entretanto, a decisão só poderá ser tomada se a avaliação dos objetos com base em suas propriedades puder ser agregada de alguma maneira, isto é, um operador de agregação deve ser construído. Neste sentido, em nossa aplicação propomos uma agregação de conceitos elementares, associados a cada atributo observado num objeto, através da aplicação sucessiva de operadores conjuntivos e disjuntivos.

1.4.3 O contexto de aplicação

Na Taxionomia Biológica, observa-se grande dificuldade em estabelecer o conceito de espécie, ou gênero, diante da alta frequência de indivíduos intermediários, levando a muitas discussões metodológicas e filosóficas entre a fenética e a filogenética.

Sneath e Sokal [96] observam que a tendência clássica em Sistemática Biológica tem sido a classificação hierárquica dos organismos em classes mutuamente exclusivas e aninhadas. Seu sucesso em representar o sistema da natureza tem levado ao estudo dos arranjos taxonômicos biogeográficos e ecológicos, especialmente da ecologia de plantas.

Além da economia de memória, várias motivações encorajam os taxonomistas a preferir este tipo de classificação dos organismos vivos, cada um em sua área de atuação, seja com plantas, animais, seja com músicas, filosofias, técnicas de agrupamento, solos, etc.. Parece uma tendência humana geral destacar a clareza de distinção entre categorias e superenfatizar as lacunas no espectro das variações fenéticas.

Conforme Sneath e Sokal [96], as categorias mutuamente exclusivas são usadas conceitualmente pelos seres humanos, ainda que estejamos constantemente prevenidos contra eventos e indivíduos intermediários. Mesmo assim, sucumbimos a uma tendência *natural*, ou contingencial, e continuamos evitando os conjuntos que

se interceptam, que resultariam em alguns indivíduos que seriam simultaneamente membros de mais de um conjunto. E somos tão obedientes ao sistema de Lineu, que requer categorias mutuamente exclusivas e hierarquicamente ordenadas, que o processo de classificação se tem tornado sinônimo, na crença de muitos biólogos, de um mapeamento da diversidade da natureza ao sistema de Lineu.

Em Sistemática Biológica, o movimento contra a classificação hierárquica tradicional vai além do desenvolvimento de medidas de distorção de matrizes de similaridade ou dissimilaridade quando estas são interpretadas por meio de fenogramas. Tais medidas, como o coeficiente de correlação cofenética

proposto por Sokal e Rohlf [35], e outras medidas de destaque, têm levado à observação de que, freqüentemente, as classificações hierárquicas são representações pobres do verdadeiro relacionamento fenético encontrado na natureza.

Jardine e Sibson [91] mostraram matematicamente que algumas propriedades de classificações que eles consideravam desejáveis podem ser violadas pelas classificações aninhadas, e sugeriram trocá-las por classificações hierárquicas que se sobrepõem .

Sneath [89] faz uma revisão dos últimos 30 anos de Taxionomia Numérica e cita o desenvolvimento de programas de computador para a identificação de espécimes e o trabalho de Bezdek [57], que utilizam conjuntos difusos nessa área. Para Sneath, a lógica difusa está estreitamente relacionada com a Taxionomia Numérica.

Bezdek observa que os conjuntos convencionais impõem restrições matemáticas aos modelos de classificação que freqüentemente dificultam sua capacidade para refletir com precisão a estrutura que existe nos dados [58].

Assim, atualmente parece haver uma tendência geral em direção a uma maior flexibilidade e experimentação nos tipos de sistemas de classificação aceitáveis para os biólogos. Dentro disto, propõem-se os conjuntos difusos, nos quais se considera que os indivíduos (plantas, animais, etc.) têm um contínuo de graus de pertinência a uma classe.

No presente trabalho descrevemos um processo de reconhecimento do conceito de espécie, que envolve a utilização dos operadores conjuntivos e disjuntivos difusos. Através dele pode-se decidir, com algum índice de pertinência, se um determinado espécime pertence ou não a uma espécie ou conceito. Este trabalho foi desenvolvido utilizando o clássico conjunto de dados composto de 150 observações de flores íris, que se subdividem em três espécies.

Particularizamos nossas reflexões e aplicação para a Taxionomia Biológica, porque neste campo, talvez mais que em outros, se evidenciam as conseqüências da utilização da lógica clássica aristotélica como fundamento. Assim, dado o grande número de problemas que surgem, pode ser o campo perfeito para a reflexão sobre a possibilidade de um novo sistema lógico. Observamos, todavia, que podem ser desenvolvidas aplicações dos operadores lógicos que propomos a vários contextos.

1.5 Estrutura e Perfil da Dissertação

A seguir damos uma descrição resumida de cada capítulo, para que se possa ter desde já uma idéia do que será feito na tese.

- **Capítulo 1:** Esta introdução.
- **Capítulo 2:** Apresenta uma pequena revisão dos conjuntos difusos, concentrando-se mais nos tópicos relacionados ao tema da imprecisão, da definição da função de pertinência e da caracterização da interseção e união através de uma t-norma e t-conorma, respectivamente.
- **Capítulo 3:** Apresenta uma revisão do conceito de espécies, mostrando como se definia uma espécie com as operações clássicas sobre conjuntos ordinários e, depois, com o objetivo de dar convencimento do seu caráter difuso, damos alguns exemplos de publicações que o demonstram.
- **Capítulo 4:** Aqui introduzimos nossa família de t-normas e explicitamos sua estreita relação com uma família particular de funções de pertinência. Além disto verificamos algumas de suas propriedades algébricas e analíticas que nos fazem acreditar em seu melhor desempenho em aplicações, quando se comparam com operadores que envolvem o máximo e o mínimo.
- **Capítulo 5:** Desenvolvemos aplicações das t-normas para caracterizar o conceito de espécie, utilizando como base um classificador que determina funções discriminantes, ou de caracterização, que dão o grau de pertinência de uma flor, cujo vetor de atributo entra no classificador, a cada espécie caracterizada.
- **Capítulo 6:** Desenvolve a regra delta, generalizada para o treinamento de uma rede neuronal lógica, melhor dizendo, um perceptron de múltiplas camadas difuso, composto de unidades processadoras que realizam operações *and* e *or* sobre as entradas. Contudo, não fizemos o treinamento da rede.
- **Capítulo 7:** Conclusões
- **Capítulo 8:** Sugestões.
- **Apêndices:** O conjunto de dados *iris* usado no capítulo 5, o programa desenvolvido com o software MATLAB e alguns de seus resultados. Os Apêndices 5 e 6, mostram uma tabela de t-normas e t-conormas, respectivamente, apresentadas em Klir [36]. Na tese não pretendemos fazer comparações entre estas e a que propomos, mas certamente este é um trabalho a ser feito posteriormente.

Capítulo 2

REVISÃO: Conjuntos Difusos

2.1 Introdução

Neste capítulo, a partir da literatura, apresentamos uma resenha dos conceitos fundamentais associados à teoria dos conjuntos difusos, incluindo a descrição de algumas experiências empíricas para a definição da função de pertinência; um histórico da caracterização de t-normas e t-conormas e outros axiomas que fundamentam a teoria. O objetivo é dar não só uma base de definições empregadas nos capítulos seguintes, mas também rever a seqüência básica de procedimentos teóricos e empíricos para a formalização desta teoria, desde a definição da função de pertinência até a caracterização da interseção e união difusa pelos axiomas de t-normas e t-conormas, respectivamente. Na tentativa de explicar por que se consideram tais axiomas, também fazemos uma descrição rápida do significado de cada um, além de uma revisão das considerações feitas quando da escolha dos operadores a serem utilizados nas aplicações. Nisto também passamos pelos conceitos de *força* ou *intensidade* de uma interseção ou união difusa. Antes, porém, introduzimos alguns tópicos mais relacionados com histórico, fontes de imprecisão, suas vantagens e aplicações. No final, fazemos uma breve colocação sobre a teoria de protótipos, e sobre a diferença entre incerteza probabilística e incerteza difusa.

2.2 Um Breve Histórico

Imprecisão e incerteza têm uma história bem estabelecida como conceitos dignos de serem estudados por filósofos, psicólogos, lingüistas e, mais recentemente, por matemáticos e cientistas da computação. Alguns tópicos de interesse para a teoria e as aplicações dos conjuntos difusos foram discutidos por Goguen [56]. Dificuldades resultantes da imprecisão na linguagem foram notadas por cientistas da computação como Winograd [105], bem como por lingüistas como Lakoff [37].

Surgiu um maior interesse neste tema depois que o engenheiro eletrônico Zadeh [65], em 1965, lançou as bases da teoria dos conjuntos difusos. Neste campo, uma parte das pesquisas teóricas e aplicadas foram realizadas por psicólogos [31], [1], [83], por cientistas da computação e engenheiros da área de reconhecimento de padrões [44], [10] e sistemas de controle [24],[25] e [106].

2.3 Fontes de Imprecisão

Ao tratar de problemas de tomada de decisão automatizada, existe uma crescente necessidade de modelar e administrar a incerteza. A incerteza pode ocorrer pela complexidade do problema, pela imprecisão nos cálculos, por ruídos de vários tipos, etc. Uma revisão sobre as fontes de imprecisão pode ser encontrada em Hisdal [18]. Aqui são descritas três das que podem dar origem a um grau de pertinência parcial de um objeto a uma classe. Todas elas estão conectadas com a consciência da possibilidade de erro e com o desejo e a capacidade de levá-los em consideração.

Fonte 1: Antecipação, pelo sujeito, de erros de observações sob condições inexatas. Mesmo quando o sujeito realiza um experimento exato, em que ele mede o valor exato do atributo de cada objeto, ele está ciente do fato de que sob condições não exatas de observação do dia-a-dia, sua estimativa da medida do valor do atributo pode assumir valores que variam de acordo com alguma distribuição de probabilidade ou *função de erro*. Hisdal [19] mostra como esta consciência faz com que apareçam curvas de pertinência em S e na forma de *sino* na teoria dos conjuntos difusos.

Fonte 2: Representação em um universo de dimensão inferior e antecipação, feita pelo sujeito, da distribuição em um universo de dimensão maior. Por exemplo, em um diagnóstico médico, um sujeito (o médico ou um sistema especialista) pode não ter à sua disposição todos os dados exigidos para o diagnóstico não-difuso de uma doença num paciente. Estando ciente desta limitação, ele designa um grau de pertinência parcial para o paciente naquela doença, com base em sua estimativa da frequência de ocorrência dos dados ausentes e da resultante frequência de classificação como um dado conceito (um termo lingüístico), de acordo com a partição não-difusa em um universo de dimensão maior.

Fonte 3: Um sujeito sabe que diferentes pessoas podem ter diferentes limiares de valores de altura, no caso de um experimento com vários sujeitos respondendo a *sim* ou *não*, em que um objeto não é mais *médio*, mas *alto*.

Nesta tese supomos que temos mais imprecisão do tipo 1, na qual nos baseamos para afirmar o aspecto difuso do conceito de espécie.

2.4 Aplicações e Vantagens da Imprecisão

Com os conjuntos difusos, os conceitos como *pequeno*, *grande*, *jovem*, *velho*, *alto* ou *baixo* podem ser transferidos de maneira utilizável para o computador. De fato, os conjuntos difusos e a lógica difusa estão encontrando aplicações cada vez maiores em uma ampla classe de problemas, desde o controle de processos industriais e o reconhecimento de padrões, até a previsão do tempo, diagnóstico médico e planejamento em agricultura. Os conjuntos difusos também fazem parte dos esforços para desenvolver máquinas inteligentes que podem *pensar como pessoas* (Zadeh [69]).

Ainda que os conjuntos difusos sejam vistos como indesejáveis, por alguns pesquisadores, a flexibilidade dos conjuntos difusos dá a eles inúmeras vantagens sobre os conjuntos convencionais. Primeiro, eles evitam a rigidez do raciocínio matemático convencional e da programação para computador. Por exemplo, é a

rigidez do computador que torna difícil ou impossível ler, automaticamente, caracteres escritos a mão e sem muito cuidado. Um humano pode facilmente reconhecer que um dado caractere escrito a é como um outro caractere a , enquanto o computador o considerará diferente se eles não forem similares a um padrão rigidamente estabelecido. Se é difícil para um computador que lê caracteres levar em conta alguma pequena variabilidade mesmo nos caracteres impressos, quanto mais nos escritos a mão.

Os conjuntos difusos simplificam a tarefa de comunicação entre o raciocínio humano, que é inerentemente flexível, e a rígida operação dos computadores digitais. Em particular, no raciocínio de senso comum, os humanos tendem a usar palavras em vez de números para descrever como um sistema se comporta. Com os conjuntos difusos, o programador pode traduzir valores linguísticos, tais como *jovem*, *alto*, *grande* e *baixo*, diretamente em um programa de computador fornecendo-lhe a função de pertinência que os definem. Modificadores tais como *não*, *muito*, *altamente* e *levemente*, etc. também podem ser traduzidos facilmente e sem ambigüidades.

Tais traduções diretas são particularmente importantes para os sistemas especialistas e os controles de processo, nos quais é necessário programar instruções que sejam essencialmente adquiridas pelo senso prático e a experiência, ao invés de regras ou cálculos exatos. Com os conjuntos difusos e as variáveis linguísticas isto é relativamente fácil. A representação de conceitos linguísticos por conjuntos difusos pode ajudar na grande dificuldade da representação da linguagem natural em máquina.

Enquanto muitas aplicações difusas estão ainda em primeiro estágio de desenvolvimento, parece provável que nas próximas décadas a lógica difusa se torne rotineiramente aplicada em muitas áreas da inteligência artificial, nas quais a interação com as pessoas ou a imitação de seu processo de pensamento está envolvida. Isto pode ajudar a criar uma ponte na separação existente entre o pensamento analógico e flexível dos humanos e a estrutura rígida dos computadores atuais (Zadeh [69]).

2.5 Conjuntos Difusos e Pertinência

Turksen [50] especifica cinco símbolos, que considera necessários para uma representação adequada dos conjuntos difusos e descreve seus significados e interrelações.

- (i) O conjunto de elementos $\theta \in \Theta$. Por exemplo, *homem* em *homens*.
- (ii) Uma *variável linguística* V em um conjunto de *variáveis linguísticas*, que é um nome para um atributo dos elementos $\theta \in \Theta$. Por exemplo, *altura de homens* digamos no Sul do Brasil.
- (iii) Um termo linguístico A de uma variável linguística, um adjetivo ou advérbio, em um conjunto de termos linguísticos possíveis para a variável, em um dado contexto. Por exemplo, *homem alto* ou *homem baixo*, associadas com *altura de homens* no sul do Brasil. Na tese também denominamos o termo linguístico por conceito.

- (iv) Um intervalo mensurável $\mathbf{X} \subset [-\infty, +\infty]$ de designações numéricas possíveis, conhecido como *conjunto referencial* para o nome do atributo particular, \mathbf{V} , do conjunto de elementos $\theta \in \Theta$. Por exemplo, $[0, 2]$ metros para *altura de homens*.
- (v) Uma designação numérica subjetiva $\mu_A(\theta)$, conhecida como valor de pertinência gradual ao conjunto, para representar o grau com que um elemento θ pertence ao conjunto de elementos identificados pelo termo lingüístico \mathbf{A} . Esta designação subjetiva é a avaliação (ou julgamento, ou crença), da compatibilidade do atributo \mathbf{V} de um elemento com o termo lingüístico \mathbf{A} , dado por um especialista, um observador ou um taxionomista. Por exemplo, ele é o grau de pertinência designado a um *homem* em um grupo de homens por um observador quando ele/ela usa o termo lingüístico *alto* de acordo com seu/sua compreensão, ou conceitualização, do adjetivo, dentro do contexto da *altura de homens no Sul do Brasil*.

2.5.1 Qual é a relação de \mathbf{V} com \mathbf{X} ?

Vamos considerar as vezes $\mu_A: \Theta \rightarrow [0, 1]$, ou as vezes $\mu_A: \mathbf{X} \rightarrow [0, 1]$, significando que existe, na verdade, uma composição $\mu_A \circ \mathbf{V}$, sendo $\mathbf{V}: \Theta \rightarrow \mathbf{X}$, ou seja

$$\mu_A \circ \mathbf{V}: \Theta \rightarrow \mathbf{X} \rightarrow [0, 1].$$

Mas, vale observar que se estivermos fazendo uma experiência de avaliação subjetiva do grau de pertinência, perguntando a vários sujeitos com que grau θ é \mathbf{A} , na prática, o que se tem é uma função subjetiva

$$\mathbf{V}: \Theta \rightarrow \mathbf{A}.$$

Nos dois casos, não há dúvida quanto à existência da relação direta entre um atributo e seu conjunto referencial, a função auxiliar $\mathbf{V}: \Theta \rightarrow \mathbf{X}$ fica implícita.

2.5.2 Representação

Definição 2.1 (Turksen [50]) [*Função de pertinência*] *Seja um conjunto difuso \mathbf{A} sobre um conjunto referencial \mathbf{X} (universo de discurso) com elementos x . Uma função de pertinência é uma função μ_A ,*

$$\mu_A: \mathbf{X} \rightarrow [0, 1],$$

em que cada elemento $x \in \mathbf{X}$, pertence a \mathbf{A} com um grau $\mu_A(x)$, de modo que $\mu_A(x) = 1$ significa pertinência completa e $\mu_A(x) = 0$ significa não-pertinência ao conjunto difuso \mathbf{A} .

A função de pertinência *mapeia* elementos de um dado conjunto universal \mathbf{X} , que é sempre ordinário, em um número real no intervalo $[0, 1]$. Para conjuntos ordinários o intervalo $[0, 1]$ é igual a $\{0, 1\}$. O intervalo de graus de pertinência mais usado é $[0, 1]$. Na tese usamos o intervalo $[-1, 1]$, onde supomos que

graus negativos indicam não pertinência ao conjunto difuso A , sendo que -1 indica não pertinência total. Os positivos indicam pertinência. O grau de pertinência nulo é associado ao ponto $x \in X$ em que ocorre a transição entre pertencer e não pertencer.

Notação:

$$\mu_A : X \rightarrow [0, 1]$$

2.6 Experiências Empíricas

Uma atividade humana característica é a formulação de um modelo para as observações empíricas. Este modelo vai desde os estereótipos que algumas pessoas fazem em seus encontros pessoais diários até os modelos complexos da física quântica. É uma ferramenta necessária na luta pela sobrevivência. Se não fosse por estes modelos a grande quantidade de *inputs* de nosso sistema (próprio, pessoal) logo nos tornaria incapazes de funcionar (Yager [98]).

O trabalho de Macvicar-Whelan [90]- Ele fez um estudo experimental e teórico da categorização da altura humana e apresentou um modelo teórico para este processo de classificação. Pediu que indivíduos de ambos os sexos, cujas idades variavam entre 6 e 72 anos, classificassem a altura de um grupo de pessoas usando as denominações:

baixíssimo
medianamente baixo
baixo
alto
medianamente alto
altíssimo

Os resultados experimentais confirmaram as afirmações de Zadeh acerca da existência de classificação difusa (a falta de fronteiras claras para as classes). Entretanto, não se encontraram evidências para dar suporte à utilização da curva S para representar o grau de pertinência das alturas pertencentes a uma dada classe. Enquanto se poderia esperar que a curva de pertinência para um atributo fosse do tipo S , observou-se que nenhum sujeito conectou a certeza de verdade com a certeza de falsidade com este tipo de curva. Assim, utilizou-se a aproximação com uma linha reta.

O trabalho de Dombi [52]- Ele usou os dados publicados por Zysno [110]. Neste projeto, 64 sujeitos de 21 a 25 anos de idade, classificaram 52 declarações diferentes sobre idades, com relação a um dos 4 conjuntos difusos:

muito jovem
jovem
velho
muito velho

Dombi usou estes dados e um método de minimização do erro médio quadrado (LMS) com o qual encontrou bom ajuste para o seu modelo de função de pertinência, dado por

$$\mu(x) = \frac{(1 - \nu)^{\lambda-1}(x - a)^\lambda}{(1 - \nu)^{\lambda-1}(x - a)^\lambda + \nu^{\lambda-1}(b - x)^\lambda}$$

Assim, a determinação de qualquer função de pertinência é possível com quatro parâmetros: primeiro, deve ser dado o intervalo (a, b) . Os outros dois parâmetros também têm significado: λ tem influência sobre a transição da função de pertinência e ν poderia ser interpretado como um nível de expectativa. Esta função tem gráfico na forma de **S**.

O trabalho de Zimmermann e Zysno [46]- Apresentam dois modelos para determinar, empiricamente, a função de pertinência. Elas diferem com relação às suas propriedades de mapeamento. Para eles, medir significa designar números a objetos de modo que certas relações entre números reflitam relações análogas entre objetos. Posto de maneira diferente, medição é o mapeamento de relações entre objetos em relações numéricas do mesmo tipo.

Outros trabalhos empíricos chegaram à conclusão de que uma curva **S** era mais adequada para representar uma função de pertinência: Norwich e Turksen [49] para o conceito *alto*. Hisdal [19] explica o processo de arredondamento ou suavização para obter uma curva em **S**, ou na forma de **sino**, quando estuda o conceito *alto*.

2.7 Operações sobre os Conjuntos Difusos

Na teoria dos conjuntos ordinários, as operações sobre conjuntos produzem um outro conjunto. Do mesmo modo que o cálculo proposicional, com suas operações de conjunção, disjunção, negação e condicional, está associado a conjuntos ordinários, a lógica difusa com as mesmas operações está associada aos conjuntos difusos. Assim, por exemplo, as operações sobre conjuntos difusos são relevantes para os sistemas especialistas difusos, porque o seu processo de inferência está baseado no processamento de conectivos lógicos sobre as regras de produção difusas armazenadas em sua base de conhecimento. A seguir damos um exemplo de regra usada em um sistema especialista difuso para diagnóstico e resolução de problemas financeiros, extraídas de Pacheco [93]:

```
IF          LIQUIDITY
.   and     Seasonal is yes
.   and     Interest is HIGH
.   and     Accounts receivable in Days of Sales is VERY HIGH
.   and     Inventory is VERY HIGH
THEN
.           Reduce Inventories is MEDIUM
.   and     Reduce Accounts Receivable is MEDIUM.
```

As operações são úteis não só para os sistemas especialistas, mas também para todos os sistemas de aprendizado que envolvem a caracterização e reconhecimento de conceitos, ou padrões (que também podem ser considerados como sendo conceitos).

2.7.1 Operadores padrões

No primeiro artigo sobre conjuntos difusos, Zadeh [65] introduziu as seguintes definições para as operações de complemento, interseção e união de conjuntos difusos:

$$\mu_{\bar{A}} = 1 - A(x),$$

$$\mu_{A \cap B}(x) = \min[A(x), B(x)],$$

$$\mu_{A \cup B}(x) = \max[A(x), B(x)],$$

para todo $x \in X$. Estas operações são denominadas operações difusas padrões ou *standards*. Além de elas serem as únicas que satisfazem os axiomas de Bellman [78], também são consideradas naturais por serem significativas em relação à informação ordinal, isto é, por serem independentes da escala, conforme verificam Turksen [50] e Yager [99]. Eles discutem esta independência da escala para combinações de conjuntos difusos referentes a uma única variável lingüística e afirmam que neste caso só poderiam ser usados os operadores padrões. De fato, estas operações têm sido usadas com sucesso em aplicações práticas da teoria dos conjuntos difusos (Terano [80], Hirota [61])

Contudo, elas não são a única alternativa para estender as operações clássicas. Muitos operadores têm sido propostos para modelar a união e a interseção de conjuntos difusos. Diferentes definições produzirão modelos de sistemas especialistas difusos diferentes (Kandel [7]). Baldwin [60] faz referência a esta possibilidade de criarmos nosso próprio sistema lógico ao propor um sistema especialista difuso que usa a linguagem FProlog.

A relevância da investigação de operadores sobre conjuntos difusos não está restrita aos fundamentos matemáticos da teoria dos conjuntos difusos (Dubois [41]). Em situações práticas, a escolha de operadores apropriados depende do conhecimento das características do problema, dos operadores alternativos disponíveis e de suas distintas propriedades. Por exemplo, no processo de tomada de decisão, os operadores *max* e *min* não são apropriados se os critérios de decisão são compensáveis, isto é, a adequabilidade de um pode equilibrar a inadequabilidade de outro. Em tais situações, uma escolha melhor seria um operador compensatório, tais como as médias ou médias geométricas definidas por Dubois [40] e Zimmermann [47].

Em sua tese de doutorado, Pacheco [93] diz que a escolha dos operadores de *max* e *min* é devida aos seguintes fatos: primeiro, havia a intenção explícita de satisfazer a estrutura axiomática em que a união e a interseção difusa são descritas (isto é, usando t-normas e t-conormas); segundo, não encontrou razão para adotar uma interseção que *restringisse* mais (isto é, para não usar o maior conjunto difuso produzido pela interseção), ou, em outros termos, a interseção mais fraca, e do mesmo modo para a união mais *relaxada*. Além disso, são as únicas t-norma e t-conorma que satisfazem a propriedade de idempotência, isto é, quando aplicadas a elementos idênticos elas produzem este mesmo elemento. Isto pode ser vantajoso quando os conjuntos envolvidos na conjunção (ou disjunção) incluem repetições. No campo do reconhecimento de padrões e da tomada de decisão com multicritérios, isto pode ocorrer (Yager [81]).

2.7.2 Outros operadores

Existem elementos $\theta \in \Theta$ cujo atributo subjetivo V , que desejamos representar, pode ser composto de duas ou mais componentes subjetivas $V_i, i =$

1, 2, ... cada uma delas afetando o atributo em questão. Um exemplo é *conforto*, composto de várias combinações de *umidade* e *temperatura*. Turksen [50] estuda este caso para os operadores padrões e também para t-normas e t-conormas. Sua conclusão é que se pode usar qualquer *t-norma* ou *t-conorma* para a composição de duas ou mais variáveis lingüísticas componentes a fim de encontrar os valores de pertinência de variáveis lingüísticas compostas.

2.7.3 A dependência do contexto

Além de os conceitos lingüísticos estarem sujeitos a várias fontes de imprecisão, eles também dependem do contexto, que também contribui para a imprecisão.

A dependência do contexto envolve não somente o significado dos termos lingüísticos, mas também o significado das operações sobre os termos lingüísticos (Klir [36]). Pode haver diferentes funções para representar estas operações em diferentes contextos. Assim, não somente as funções de pertinência, mas também as operações sobre os conjuntos difusos dependem do contexto. Vejamos os exemplos a seguir:

1. Sejam $A = \{\text{mamíferos}\}$ e $B = \{\text{carnívoros}\}$. Tem-se
 $\{\text{mamíferos-carnívoros}\} = \{\text{mamíferos}\} \cap \{\text{carnívoros}\}$
2. $\{\text{bom gerente}\} \neq \{\text{bom}\} \cap \{\text{gerente}\}$.

Contudo, na tese não trabalhamos com estas questões. Segundo Klir [36], a construção de funções de pertinência que captem adequadamente os significados dos termos lingüísticos empregados em uma aplicação particular, como também a determinação de significados das operações efetuadas sobre os termos lingüísticos, não são problemas para a teoria dos conjuntos difusos *per se*, mas para a área de aquisição do conhecimento.

2.8 T-Normas e T-Conormas

Atualmente, a caracterização da intersecção e da união difusa utilizando t-normas e t-conormas, respectivamente, é utilizada rotineiramente nas aplicações que envolvem conjuntos difusos. As normas e conormas triangulares surgem nos estudos de Menger [62] sobre as desigualdades triangulares para espaços métricos estatísticos. É crucial saber se uma t-norma, ou t-conorma, acontece em um determinado contexto. Estudos gerais dos procedimentos pelos quais classes destas funções podem ser geradas, foram efetuados por Schweizer e Sklar [2], Ling [45] e outros. Dado que o significado dos conectivos *e* e *ou* dependem do contexto das aplicações, é importante ter à disposição várias t-normas. Assim, é preciso uma caracterização de t-normas que nos permita construí-las à vontade.

A seguir damos algumas definições que serão revistas no capítulo 4 junto com os teoremas de caracterização, mas que por agora nos ajudarão a explicar por que elas são qualificadas para caracterizar a intersecção e a união difusa.

Definição 2.2 (T-Norma) Uma função $T : [-1, 1] \times [-1, 1] \rightarrow [-1, 1]$ que satisfaz as seguintes condições:

- (T1) $T(-1, -1) = -1, T(a, 1) = T(1, a) = a$ (condições limite)
 (T2) $T(a, b) \leq T(c, d)$ quando $a \leq c$ e $b \leq d$ (monotonicidade)
 (T3) $T(a, b) = T(b, a)$ (comutativa)
 (T4) $T(T(a, b), c) = T(a, T(b, c))$ (associativa)

Definição 2.3 (T-Conorma) Uma função $S : [-1, 1] \times [-1, 1] \rightarrow [-1, 1]$ que satisfaz as seguintes condições:

- (S1) $S(1, 1) = 1, S(a, -1) = S(-1, a) = a$ (condições limite)
 (S2) $S(a, b) \leq S(c, d)$ quando $a \leq c$ e $b \leq d$ (monotonicidade)
 (S3) $S(a, b) = S(b, a)$ (comutativa)
 (S4) $S(S(a, b), c) = S(a, S(b, c))$ (associativa)

Em alguns casos exige-se a continuidade e a subidempotência das t-normas. O lema a seguir mostra a relação de toda t-norma e t-conorma com o máximo e o mínimo. Este lema será revisto no capítulo 4.

Lema 2.1 Se T e S são t-norma e t-conorma arquimedianas estritas vale:

$$T(a, b) \leq \inf\{a, b\} \quad (2.1)$$

$$S(a, b) \geq \sup\{a, b\} \quad (2.2)$$

2.8.1 Caracterização de interseção e união difusa

A seguir passamos a fazer uma série de considerações sobre o significado de cada axioma, para que a t-norma seja considerada intuitivamente aceitável como interseções difusas significativas para quaisquer pares de conjuntos difusos.

1. A condição (0.1) para t-norma e t-conorma estabelece que as operações de interseção e união para conjuntos difusos sejam extensões das mesmas operações clássicas sobre os conjuntos ordinários, pois estas operações se reduzem às operações não-difusas quando os graus de pertinência estão restritos ao conjunto $\{0, 1\}$. Também por isto são denominados operadores lógicos Dombi [51]. Quando um argumento de T é 1, expressando pertinência total, as condições-limite e a comutatividade também asseguram, como nossa concepção intuitiva requer, que o grau de pertinência na interseção seja igual ao do outro argumento.
2. A monotonicidade e a comutatividade expressam a exigência natural de que um decréscimo no grau de pertinência no conjunto A ou B não pode produzir um aumento no grau de pertinência da interseção.
3. A comutatividade assegura que a interseção difusa é simétrica, isto é, indiferente quanto à ordem em que os conjuntos a serem combinados são considerados.

4. A associatividade assegura que se pode tomar a interseção para qualquer número de conjuntos em qualquer ordem de agrupamento pareada desejada. Esta operação permite estender a interseção difusa para além de dois conjuntos. O grau de pertinência independe do agrupamento dos conceitos .
5. A continuidade impede que ocorra uma situação em que uma mudança muito pequena no grau de pertinência do conjunto A ou do conjunto B produza uma grande (descontínua) mudança no grau de pertinência de $A \cap B$.
6. A subidempotência expressa que o grau de pertinência ao conjunto difuso $A \cap B$ neste caso especial não pode exceder a . Se se exigisse a idempotência, isto significaria que a conjunção de conjuntos iguais resultaria no mesmo conjunto.
7. A monotonicidade estrita é uma forma mais forte de monotonicidade.

2.8.2 A força da interseção e da união

Pela equação (2.1) e (2.2) sabe-se que a interseção difusa *standard* produz, para quaisquer conjuntos difusos dados, o maior conjunto difuso entre todos aqueles produzidos por todas as possíveis interseções difusas (t -normas). Esta é a interseção mais fraca. Por outro lado, a união difusa *standard* produz o menor conjunto difuso entre os conjuntos difusos produzidos por todas as possíveis uniões difusas (t -conormas). Esta é a união mais forte.

Por exemplo, seja $A_1 = \textit{inteligente}$ e $A_2 = \textit{bonita}$. O grau de pertinência de uma menina ao conceito determinado pela conjunção A_1 e A_2 é maior quando calculado pelo operador mínimo do que por qualquer outra t -norma. Assim, a força do *e* é menor no caso do mínimo. Ter pertinência 0.9 à interseção indica que a medida de satisfação simultânea é maior do que ter pertinência 0.6, mas que a força do *e* é menor.

Note-se que pela declaração lógica A_1 e A_2 requer-se a satisfação simultânea de A_1 e de A_2 . É um fenómeno comum na linguagem falada enfatizar o *e* quando se requer forte satisfação destas duas condições. Assim, pela *força* de um *e* se quer expressar o quão forte se quer esta satisfação simultânea.

2.9 A Teoria de Protótipos

A teoria de protótipos constrói a pertinência na extensão de um conceito como graduada, determinada pela similaridade com o melhor exemplar do conceito. Considera que um conceito é dado como um complexo, consistindo de um protótipo e de uma função que mede o grau com que objetos diferem do protótipo. Então, pergunta-se: Quanto um conceito determina que objetos caem sobre ele e quais não? (Osherson e Smith [23]).

Zadeh [65] deixa claro que segue esta teoria no desenvolvimento de sua teoria de conjuntos difusos, principalmente quando define o núcleo (ou *core*) de um conjunto difuso convexo.

Osherson e Smith [23], usando o exemplo clássico da *maçã-listrada*, discutem os inúmeros problemas que surgem para o significado de conceitos simples

ou complexos quando se usa a teoria de protótipos baseada em conjuntos difusos. Vale dizer que o conceito complexo é obtido pela combinação de conceitos simples.

2.10 Incerteza Probabilística e Incerteza Difusa

Existe uma distinção filosófica entre incerteza probabilística e incerteza difusa. Bezdek [58] usa o seguinte argumento para enfatizar esta diferença: $\mu_{ik} = 0.25$ não indica que a probabilidade seja igual a 0.25 que x_k pertença inteiramente à classe (i), antes de tudo ela estabelece que x_k compartilha cerca de 0.25 das qualidades requeridas para a pertinência inequívoca à classe (i). Assim, depois de observado x_k , a incerteza probabilística quanto à sua identidade é completamente removida, enquanto a incerteza difusa permaneceria a mesma. A incerteza difusa reflete o quão distinto é um indivíduo quando comparado com outros com os quais tenha alguma relação quanto às propriedades ou atributos que caracterizam uma classe.

O grau de pertinência não é uma probabilidade. Basicamente, ele é uma medida da compatibilidade de um objeto com o conceito representado por um conjunto difuso. Por exemplo, 0.3 é a compatibilidade de André com a definição do conjunto difuso dos associados jovens da SBMAC ; 0.3 não é a probabilidade que André venha a ser um sócio jovem da SBMAC.

Capítulo 3

REVISÃO: Taxionomia Biológica

A divisão lógica de um *gênero* não é só de interesse científico. Uma curiosidade é que até os índios Wayampi, dos rios Jari e Amapari, na Amazônia brasileira, estabeleceram uma divisão lógica de alguns representantes da família Tinamidae (inhambus). O *gênero máximo*, por eles chamado Namupewar, é dividido em três espécies pela agregação de *differentiae* baseadas nos tipos de canto dessas aves.

A. A. Jensen(1988)

3.1 Introdução

No sistema clássico de Aristóteles, em que se fundamenta o sistema taxonômico de Lineu, as classes devem ser mutuamente exclusivas. Neste sistema, em particular na taxionomia Biológica, as formas ditas intermediárias, por haver ambigüidades em seus valores de medida dos atributos, seja por ruídos nos instrumentos de medida, ou por serem selecionados em diferentes meios em que variam a temperatura, a acidez do solo, etc., devem pertencer a uma ou outra classe. Problemas em que se discute a pertinência de objetos a categorias e em que muitas vezes não se define precisamente a pertinência são tratados no campo de reconhecimento de padrões, de comunicação de informações, na teoria da decisão e em abstrações relacionadas à formação de categorias e conceitos de uma maneira geral.

A Taxionomia Biológica surge no contexto da tese como o campo primário e um dos mais fecundos em que podemos refletir sobre a formação de classes ou categorias ou a formação de conceitos. Neste contexto, um conceito é representado por um conjunto. Assim, também temos uma base para a reflexão de um possível sistema lógico difuso, que a princípio pode ser definido tomando por base qualquer operador conjuntivo, disjuntivo e de negação que caracterize a intersecção, união e complementar difusos, respectivamente.

Por que a escolha deste contexto? Porque nele, conceito é igual a conjunto e, por ter como fundamento o sistema lógico clássico de Aristóteles, pode facilitar a reflexão para um possível sistema lógico difuso.

Este capítulo faz um pequeno levantamento que apresenta desde a aplicação das operações clássicas sobre conjuntos ordinários, a definição aristotélico-lineana de espécie, até alguns exemplos de publicações em taxionomia que mostram

como aparecem termos lingüísticos e ambigüidades quanto ao intervalo de variação dos caracteres taxonômicos quantitativos contínuos.

Alguns pesquisadores têm maneiras diferentes de abordar o problema quando discutem o aprendizado de conceitos com o fim de os implementar em algoritmos de aprendizado de máquina e também em aspectos cognitivos da formação de conceitos.

Não sabemos ainda em que contexto nossos operadores serão mais bem aplicados, mas o fato é que desenvolvemos uma aplicação no campo da Taxionomia Biológica, que é rico para aplicações dos conjuntos difusos.

3.2 Definição de Termos

Os termos classificação, sistemática e taxionomia são freqüentemente usados para significar uma só coisa. Atualmente, há uma tendência a atribuir significados diferentes para estes termos. As definições a seguir foram extraídas de Simpson [38].

Taxon, cujo plural é taxa, é uma abreviação para grupo taxonômico de qualquer natureza (ou rank).

Sistemática é definida como o estudo científico dos tipos e diversidades de organismos e de qualquer e todo relacionamento entre eles. Esta definição deve ser entendida em seu sentido mais amplo. Ela se refere não somente ao arranjo dos organismos em taxa e sua denominação, mas também com as causas e origem destes arranjos.

Classificação é a ordenação dos organismos em grupos, ou conjuntos, com base em seus relacionamentos. Pode haver confusão sobre o termo relacionamento. Pode significar relacionamento filogenético ou simplesmente semelhança ou similaridade global, quando julgada pelos caracteres dos organismos sem qualquer implicação quanto ao seu relacionamento com os ancestrais. Embora a classificação tenha sido definida como um processo, também tem sido usada para designar o produto final deste processo. Assim, o resultado da classificação é uma classificação. O termo não se restringe a colocar entidades dentro de classes distintas, mas também inclui sua ordenação em um espectro contínuo, ou em outro arranjo. Em Biologia, a classificação deve ser diferenciada de identificação.

Identificação é a alocação de objetos adicionais não identificados, na classe correta, uma vez que a classificação tenha sido estabelecida.

Taxionomia é o estudo teórico da classificação, incluindo suas bases, princípios, procedimentos e regras. Taxionomia também é um processo. Alguns entendem a Taxionomia como sendo a prática diária de identificação de vários tipos de organismos. Mas o que de fato faz com que se forme uma taxionomia é que os objetos podem ser distribuídos em diferentes níveis hierárquicos, relacionados por inclusões de classes.

3.3 Definição de Espécie

As bases lógicas para esta ciência encontram-se no método socrático-platônico da diérese, ou divisão lógica. Ainda que não fosse o interesse central

de Aristóteles, ele foi o fundador da Taxionomia Biológica. O principal problema enfrentado por Aristóteles foi o da definição de gênero. O que é um *gênero* de seres concretos? Por onde iniciar a divisão lógica dos seres vivos? Aristóteles aconselhava selecionar algum grande gênero, *megiston genos*, um grande grupo qualquer de organismos do qual se possui um conhecimento prévio, quase *inato*, como por exemplo, os animais que voam ou os que não voam. Há que empregar a divisão lógica apenas para aqueles caracteres (atributos) que pertencem aos organismos. Por isto, passou longos anos de sua vida no estudo da morfologia, fisiologia, anatomia, embriologia, comportamento, etc. Daí passou à divisão lógica dos grandes gêneros, obedecendo aos seguintes princípios arrolados por Joseph [48]:

1. A divisão deve ser exaustiva; de outra maneira, algo é omitido.
2. As espécies subordinadas a um gênero devem ser mutuamente exclusivas (ou seja, devem ter, duas a duas, diferença simétrica). De outra maneira, haveria uma separação incompleta (uma relação de inclusão entre conjuntos de caracteres), e portanto uma definição falha das espécies.
3. Uma divisão deve ser feita em cada estágio, e tanto quanto possível, em todos os estágios, baseada em um único princípio ou característica.

Aqui, pode-se observar que não há consideração para indivíduos intermediários ou para uma agregação de características.

3.3.1 Operações com conjuntos ordinários e taxionomia

Estes princípios aristotélicos foram corretamente aplicados em Taxionomia Biológica por Carolus Linnaeus, principalmente na décima edição de seu *Systema Naturae Linnaei* [75]. Vejamos alguns exemplos tirados dessa obra, os quais demonstram a maneira como as operações de união, interseção e diferença simétrica são utilizadas.

Exemplo: considere o gênero lineano Myrmecophaga (que inclui os tamanduás). Linnaeus utilizou-se, para caracterizar esse gênero, dos seguintes caracteres:

1. $a = \textit{dentesnulli}$ (dentes ausentes);
2. $b = \textit{linguateresextensilis}$ (língua cilíndrica, extensível);
3. $c = \textit{corpuspilistectum}$ (corpo coberto de pêlos).

Pode-se, então, representar o gênero Myrmecophaga por um conjunto ternário $\{a, b, c\}$. Para obter a divisão lógica do gênero, Linnaeus utilizou dois *fundamenta divisionis*, esgotando todas as possibilidades de cada um deles: o número de dedos na mão (*palmis didactylis*, *palmis tridactylis* e *palmis tetradactylis*) e o número de dedos no pé (*plantis tetradactylis*, *plantis pentadactylis*). Representemos esses caracteres da seguinte maneira:

1. $d = \textit{palmisdidactylis}$;
2. $e = \textit{plantistetradactylis}$;

3. $d' = \textit{palmistridactylis}$;
4. $e' = \textit{plantispentadactylis}$;
5. $d'' = \textit{palmistetradactylis}$.

O universo $U = \{a, b, c, d, d', d'', e, e'\}$ foi chamado por Linnaeus de *deffinitio* (definição) do gênero.

Representam-se aqui, pela mesma letra minúscula, os diversos estados do mesmo *fundamentum divisionis*.

$\{a, b, c\} \cup \{d, e\} = \{a, b, c, d, e\} = \textit{Myrmecophaga didactyla}$.

$\{a, b, c\} \cup \{d', e'\} = \{a, b, c, d', e'\} = \textit{Myrmecophaga tridactyla}$.

$\{a, b, c\} \cup \{d'', e'\} = \{a, b, c, d'', e'\} = \textit{Myrmecophaga tetradactyla}$.

As três espécies desse gênero são válidas: apresentam diferença simétrica, duas a duas. E possuem, todas, caracteres que as distinguem umas das outras.

Se for feita a diferença de uma espécie em relação a cada uma das outras e em seguida se fizer a interseção dessas diferenças, obter-se-á o caráter exclusivo dessa espécie, chamado por Linnaeus de *nomen essentiale* da espécie.

Definição 3.1 *Uma espécie só pode ser formada pela união de conjuntos finitos, não vazios, de caracteres.*

Na prática diária atual da taxionomia isto não é tão simples, como se tivéssemos só caracteres quantitativos obtidos por contagem. Quando se levam em conta medidas de variação contínua, como é comum em taxionomia, entram os aspectos que definem a imprecisão neste contexto.

No capítulo 5 consideramos uma espécie como sendo definida pela conjunção, ou disjunção de conjunções, de funções de pertinência associadas a atributos do objeto que têm um intervalo de variação sujeito a várias fontes de imprecisão. A possibilidade de ajustar os parâmetros do modelo permite uma caracterização que varia de acordo com o contexto.

3.3.2 Problemas com a definição de espécie

O objetivo desta seção é apresentar exemplos de descrições de caracteres taxonômicos, para a identificação de seres vivos, nos quais ocorrem o uso de variáveis lingüísticas como curto, médio ou longo, ou em que há problemas com a divisão de classes, ou imprecisões quanto ao intervalo de variação dos caracteres firmando assim o caráter difuso da definição de espécie sobre o qual nos baseamos em nossas aplicações.

Anderson [17] escreve em seu trabalho de apresentação do conjunto de dados IRIS as seguintes palavras:

Dado que é um fenômeno biológico, o problema da definição de espécies é digno de sérios estudos com um fim próprio, e não com uma simples consequência a ser usada em algum outro campo. É um problema tão fundamental e importante que ele toca muitos campos. Quando se assume o problema, como um problema, e o estuda de diversos pontos de

vista da genética, da taxionomia, da citologia e biometria, observa-se que ele não somente necessita das muitas técnicas existentes, mas que também se deve inventar novas. (Anderson 1936). (Tradução da autora.)

Nesta seção serão citados alguns exemplos que mostram que na prática da taxionomia, no dia-a-dia de um laboratório, não é fácil estabelecer que espécimes pertencem a uma espécie, dentro de um gênero. São muitas as causas de variações nas medidas. Além das variações devidas às diferentes precisões dos instrumentos de medida, existem as de origem geográfica e do hospedeiro da amostra.

Caso 1: Nematóides do gênero *Helicotylenchus*

Fortuner [109] é um grande estudioso deste gênero, que tem cerca de 180 espécies, diferenciadas por seus autores originais por mais de 40 medidas, muitas das quais altamente variáveis. Como consequência desta grande variabilidade, é difícil diferenciar uma espécie de outra. Em um de seus estudos, para a revisão da espécie *H. pseudorobustus*, e uma possível redescricao, utilizou 43 amostras de 1 a 30 espécimes de populações de campo, vindas de 25 países, todas identificadas como *H. pseudorobustus* (Steiner, 1914).

Utilizando técnicas de análise multivariada, Fortuner concluiu que havia três grupos dentro deste: *pseudorobustus* (Steiner, 1914), *microlobus* (Golden, 1956) e *dihystera* (Cobb, 1893). Seus estudos lançaram novas luzes sobre o relacionamento entre *H. pseudorobustus* e *H. microlobus*, em discussão nos últimos 20 anos. Também observou em uma amostra algumas características de *H. pseudorobustus* e algumas características de *H. dihystera*. A conclusão de Fortuner é que ela pode pertencer a uma espécie diferente não identificada.

Entretanto, Fortuner expõe suas dúvidas quanto a se a análise multivariada poderia ser usada para diferenciação específica e identificação no gênero *Helicotylenchus*. Ele sugere, para a identificação prática de espécies, a avaliação de similaridades entre pares de espécies.

Caso 2: Nematóides do gênero *Rotylenchus*

Neste caso vemos como são usados os conceitos *curto*, *médio* e *longo* para descrever um gênero que se divide em dois grupos. De acordo com Castillo [30]:

1. um, com tamanho médio de corpo (em torno de 1mm) e com o lóbulo da glândula do esôfago curto;
2. outro, com corpo mais longo (1.5-2mm) e lóbulo da glândula do esôfago muito mais alongado.

Caso 3: Nematóides do gênero *Meloidogyne*

A história deste gênero começa em 1885. Sua identificação é difícil devido à grande similaridade morfológica. Jepson [103] publica uma *Lattice Key* para identificação de espécies no gênero *Meloidogyne*. Foram dados para cada caráter

quantitativo vários conjuntos de medidas, em micrômetros(μm), referentes a populações diferentes, para tentar cercar a variação a ser encontrada dentro de uma espécie. Uma parte da chave para as fêmeas contém no total informações sobre 6 caracteres de 51 espécies, na qual se destaca ainda mais o aspecto de superposição de intervalos de variação como se pode observar na tabela 3.1.

Tabela 3.1: Fêmeas do gênero *Meloidogyne* (em micrômetro: μm)

Espécie	Compr. do estilete	Comp. do corpo
<i>graminicola</i>	12.6-15.3 10.6-11.2	418-827 445-765
<i>oryzae</i>	14.4-15.3 14.0-18.0	484-718 475-750
<i>ardenensis</i>	17.6-18.9 15-19	526-735 309-520

Fonte: Jepson, S.B., 1987.

3.3.3 As flores íris

Para as flores íris também ocorrem problemas com a definição dos intervalos de variação dos atributos. Anderson [17] fornece os seguintes intervalos em *cm*, descritos na tabela 3.2. Em cada casa, a primeira medida refere-se ao

Tabela 3.2: Medidas da Sétala e da Pétala para Flores Íris (em centímetros: *cm*)

Espécie	Sétala	Pétala
<i>Setosa</i>	4-6	pequenas
"	—	
<i>Versicolor</i>	4-7	2-5
"	—	0.5-2
<i>Virgínica</i>	4-8	3-7
"	—	1-3

Fonte: Anderson, E., 1936

comprimento e a segunda, à largura. Observam-se diferenças nos intervalos da tabela 3.2 e dos publicados por Fisher [92], conforme se pode ver na tabela em anexo no final.

3.4 Conjuntos Difusos e Outras Abordagens

Sneath [89] faz uma revisão dos últimos 30 anos de Taxionomia Numérica

onde afirma: *A lógica difusa está estreitamente relacionada com a Taxonomia Numérica, principalmente com a fenética.*

Em seu artigo, Sneath [89] cita a contribuição de Bezdek ([57], [58], [59]) para a introdução dos conjuntos difusos na área de Taxionomia Numérica. Ele constrói uma extensão difusa do ISODATA, que é um algoritmo para agrupamento, ou *cluster*, criado por Ball e Hall [13] para aplicações em ciências sociais. Com este método, ele analisa a subestrutura do conjunto de dados íris em que é obtida uma solução de agrupamento sem utilizar o conhecimento a priori da classificação das espécies dadas por Anderson [16].

A revisão de alguns artigos da revista *Psychologic Review* mostra como o estudo de modelos de conceitos, em geral, consideram o contexto da Taxionomia Biológica como fundamental.

Medin e Smith [15], em seu artigo sobre formação de conceitos, fazem várias críticas à visão clássica, que diz que todas as instâncias de um conceito compartilham propriedades comuns que são condições necessárias e suficientes para definir o conceito, mas isto é a mesma coisa que a definição de espécie vista anteriormente. Fazem isto dando vários exemplos da área de Taxionomia Biológica e, além disto, iniciam uma grande discussão sobre a teoria de protótipos e terminam por dizer que a teoria de Zadeh [65], com o *max* e *min*, não explica muitos casos que envolvem conceitos complexos.

Mervis e Rosch [21] fazem um estudo sobre a categorização de objetos naturais utilizando exemplos da Taxionomia Biológica e dizem que no paradigma clássico de formação de conceitos os atributos são combinados arbitrariamente para formar itens, mas citam estudos de aspectos antropológicos da linguagem no caso de categorização de animais feitos por Leach [20] : *O ambiente físico e social de uma criança é percebido como um contínuo. Ele não contém qualquer coisa que seja intrinsecamente separada.*

Mervis continua discutindo a indeterminação da representação e da pertinência a categorias e cita os trabalhos empíricos de Labov [107] e outros. Quanto às categorias biológicas, cita o trabalho do taxonomista numérico Sokal [97], que demonstra as conclusões de Rosch e Mervis [83], a saber, que é provável que os membros mais pobres de uma categoria contenham atributos dos agrupamentos (ou *clusters*) correlacionados a atributos de outras categorias.

Em outras publicações da área de inteligência artificial encontram-se os trabalhos de Dietterich e Michalski sobre a formação de conceitos. Abaixo, descrevemos suas considerações a respeito da Taxionomia Biológica:

Uma descrição taxionômica (ou classificação) é uma descrição de uma classe de objetos que subdivide a classe em subclasses. Uma descrição taxonômica é fundamentalmente disjuntiva. A classe global é descrita pela disjunção de descrições de subclasses (Dietterich [100]).

Descrição taxionômica é um tipo de generalização descritiva, ou seja, é um tipo de aprendizado de conceito através de observações feitas numa coleção de objetos (Michalski ([101],[32],[95],[94])).

3.5 Problemas de Identificação

Atualmente, a identificação é feita com base em caracteres filogenéticos

e fenéticos, sendo este último o aspecto a que mais nos referimos nesta tese. Os caracteres fenotípicos são estudados até seus últimos detalhes, incluindo a análise de cromossomos e outras técnicas mais precisas, como, por exemplo, microscópio eletrônico, imagens fotoelétricas e fotografias que podem ser tratadas com técnicas de processamento de imagens. Tudo isto com o objetivo de classificar o espécime observado em uma ou outra classe já existente ou até de criar uma nova classe. Isto leva a muitos problemas com a nomenclatura, surgindo muitas espécies sinônimas e chegando-se até a propor nomenclaturas numéricas (Sneath e Sokal [96]).

A abordagem mais satisfatória para a identificação é usar uma combinação dos caracteres, ou atributos, que mais discriminam, por causa da similaridade morfológica das espécies e da falta de um único atributo que pode diferenciá-las e porque existe variação no grau com que os diferentes atributos podem ser usados para separar as espécies (Jepson [103]).

A observação acima sugere que seria adequada uma calibração, ou equilíbrio, dos parâmetros que definem cada conceito como vamos propor em nossa aplicação. Mas na tese não proponho uma caracterização de espécie num laboratório de taxionomia, pois nele a identificação é feita usando inúmeros caracteres morfológicos e morfométricos, o que nosso classificador atual não estaria em condições de processar.

Capítulo 4

Famílias de T-Normas e Funções de Pertinência

4.1 Introdução

A origem destes operadores está no trabalho de Maia [71]. Ele propôs a utilização da seguinte família de funções de pertinência:

$$\mu_{p,k}(x) = \frac{x^k - p^k}{x^k + p^k} \quad (4.1)$$

com $p > 0$ e $k \neq 0$. Estas funções seriam usadas para modelar conceitos elementares como, por exemplo, *pessoa alta*, *pessoa pesada*, etc. A figura 4.1 ilustra o gráfico correspondente a esta função.

Diferentemente da maioria dos autores nesta área, ele propunha trabalhar no intervalo $[-1, 1]$ e também optou por funções com propriedades assintóticas. Para agregar os conceitos ele propôs operadores que deveriam corresponder essencialmente à conjunção e à disjunção.

Se φ_1 e φ_2 denotam funções de pertinência quaisquer, então definimos

$$C(\varphi_1 : q_1 : w_1; \varphi_2 : q_2 : w_2) = \frac{1 - y}{1 + y} \quad (4.2)$$

onde

$$y = \left(\frac{q_1}{z_1}\right)^{w_1} + \left(\frac{q_2}{z_2}\right)^{w_2} \quad (4.3)$$

com $z_i = \frac{1+\varphi_i}{1-\varphi_i}$, $i = 1, 2$.

$$D(\varphi_1 : q_1, w_1; \varphi_2 : q_2 : w_2) = \frac{y - 1}{y + 1} \quad (4.4)$$

onde

$$y = \left(\frac{z_1}{q_1}\right)^{w_1} + \left(\frac{z_2}{q_2}\right)^{w_2} \quad (4.5)$$

com $z_i = \frac{1+\varphi_i}{1-\varphi_i}$, $i = 1, 2$.

Os operadores C e D , não são operadores conjuntivos e disjuntivos propriamente ditos, no sentido que caracterizam interseção e união de conjuntos

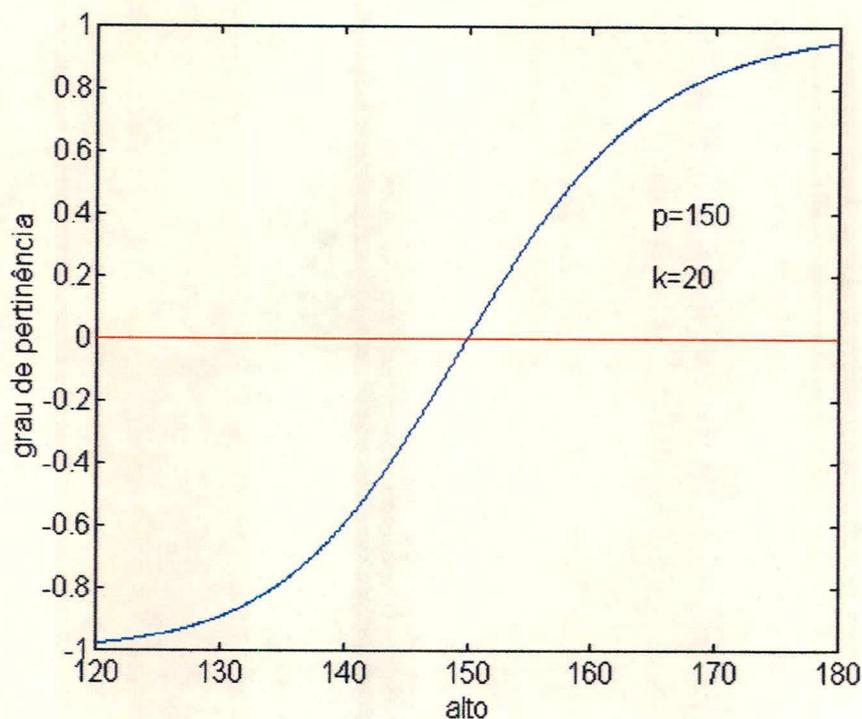


Figura 4.1: Função de pertinência

difusos. De fato, pode-se mostrar que existem φ_1, φ_2 e parâmetros q_1, w_1, q_2, w_2 , tais que existem $0 < x < +\infty$ satisfazendo

$$C(\varphi_1 : q_1 : w_1; \varphi_2 : q_2 : w_2) > \min(\varphi_1(x)\varphi_2(x)) \quad (4.6)$$

e

$$D(\varphi_1 : q_1 : w_1; \varphi_2 : q_2 : w_2) < \max(\varphi_1(x)\varphi_2(x)) \quad (4.7)$$

contrariando uma das propriedades básicas de toda t-norma e t-conorma.

Ao estudarmos estes operadores, observamos que se definissemos as funções

$$Op_1 : (-1, +\infty) \rightarrow [-\infty, 1)$$

$$Op_2 : [-1, 1) \rightarrow [0, +\infty)$$

$$Op_3 : (-1, +\infty) \rightarrow (-1, +\infty)$$

dadas por

$$Op_1(x) = \frac{x-1}{1+x} \quad (4.8)$$

$$Op_2(x) = \frac{1+x}{1-x} \quad (4.9)$$

$$Op_3(x) = \frac{1-x}{1+x}, \quad (4.10)$$

facilmente podíamos ver que

$$C(\varphi_1 : q_1 : w_1; \varphi_2 : q_2 : w_2) = Op_3((q_1 Op_3(\varphi_1))^{w_1} + (q_2 Op_3(\varphi_2))^{w_2}) \quad (4.11)$$

e

$$D(\varphi_1 : q_1 : w_1; \varphi_2 : q_2 : w_2) = Op_1 ((q_1 Op_2(\varphi_1))^{w_1} + (q_2 Op_2(\varphi_2))^{w_2}) \quad (4.12)$$

Se tomarmos $q_1 = q_2 = w_1 = w_2 = 1$, verificamos que os operadores definidos deste modo são t -normas e t -conormas. Assim, para este caso particular estabelecemos as seguintes notações:

$$U_c = C(-, 1, 1; -, 1, 1) \quad (4.13)$$

$$U_d = D(-, 1, 1; -, 1, 1) \quad (4.14)$$

A prova de que U_c e U_d são t -norma e t -conorma, respectivamente, segue com os resultados das seções seguintes, onde, na notação, deixaremos os parâmetros iguais a 1 de lado e usaremos só os valores dos graus de pertinência que serão denotados por a e b , como sendo graus de pertinência, ou φ_1 e φ_2 no caso de duas funções de pertinências. As linhas gerais para esta demonstração também podem ser encontradas em Zanusso [73], [74] e Araújo [70]. A seguir passaremos a usar letras caligráficas para C e D .

4.2 Definições e Caracterizações

Para provarmos que U_c e U_d são t -norma e t -conorma, respectivamente, baseamo-nos nas definições e teoremas dados por Schweizer e Sklar [2], e que passamos para esta seção.

Definição 4.1 (T-Norma) *Uma função $T : [-1, 1] \times [-1, 1] \rightarrow [-1, 1]$ é uma norma triangular (ou t -norma) se*

$$(T1) \quad T(-1, -1) = -1, T(a, 1) = T(1, a) = a \quad (\text{condições limite})$$

$$(T2) \quad T(a, b) \leq T(c, d) \text{ quando } a \leq c \text{ e } b \leq d \quad (\text{monotonicidade})$$

$$(T3) \quad T(a, b) = T(b, a) \quad (\text{comutativa})$$

$$(T4) \quad T(T(a, b), c) = T(a, T(b, c)) \quad (\text{associativa})$$

Se além disto, se T satisfizer

$$(T5) \quad T \text{ é contínua}$$

$$(T6) \quad T(a, a) < a$$

T denomina-se t -norma arquimediana. Finalmente, se a condição (T2) é estrita, quando $a < c$ e $b < d$, T denomina-se t -norma arquimediana estrita.

As t -normas arquimedianas estritas estão completamente caracterizadas pelo teorema seguinte. Antes vamos precisar de uma definição.

Definição 4.2 (Gerador Decrescente) *Uma aplicação*

$$f : (-1, 1] \rightarrow [0, +\infty),$$

contínua e estritamente decrescente, que satisfaz as condições:

$$(f1) \quad f(1) = 0$$

$$(f2) \quad \lim_{x \rightarrow -1} f(x) = +\infty$$

é denominada gerador decrescente.

Teorema 4.1 (Caracterização de T -Norma) *Se T é uma t -norma arquimediana estrita, existe $f : (-1, 1] \rightarrow [0, +\infty)$, um gerador decrescente tal que*

$$T(a, b) = f^{-1}(f(a) + f(b)) \quad (4.15)$$

para $\forall a, b \in (-1, 1]$. Além disso, se g é outro gerador decrescente, existe $\lambda > 0$ tal que $\lambda f = g$.

Para valores de a ou b iguais a -1 , a representação (4.15) admite uma óbvia extensão via limite, coincidindo, neste caso, com os respectivos valores de T . Remetemos o leitor a Schweizer [2].

Definição 4.3 (T-Conorma) *Uma função $S : [-1, 1] \times [-1, 1] \rightarrow [-1, 1]$ é uma t -conorma se*

$$(S1) \quad S(1, 1) = 1, S(a, -1) = S(-1, a) = a \quad (\text{condições limite})$$

$$(S2) \quad S(a, b) \leq S(c, d) \text{ quando } a \leq c \text{ e } b \leq d \quad (\text{monotonicidade})$$

$$(S3) \quad S(a, b) = S(b, a) \quad (\text{comutativa})$$

$$(S4) \quad S(S(a, b), c) = S(a, S(b, c)) \quad (\text{associativa})$$

Se, além disto, S satisfizer:

$$(S5) \quad S \text{ é contínua}$$

$$(S6) \quad S(a, a) > a$$

S denomina-se t -conorma arquimediana. Finalmente, se a condição **S2** é estrita, quando $a < c$ e $b < d$, S denomina-se t -conorma arquimediana estrita.

Definição 4.4 (Gerador Crescente) *Uma aplicação*

$$g : [-1, 1) \rightarrow [0, +\infty),$$

contínua e estritamente crescente, que satisfaz as condições:

$$(g1) \quad g(-1) = 0$$

$$(g2) \quad \lim_{x \rightarrow 1^-} g(x) = +\infty$$

é denominada gerador crescente.

Como no teorema 4.1, vale a seguinte caracterização das t -conormas arquimediana estritas:

$$S(a, b) = g^{-1}(g(a) + g(b)) \quad (4.16)$$

onde g é um gerador crescente. O seguinte lema relaciona geradores crescentes e decrescentes.

Lema 4.1 *Valem as seguintes propriedades:*

1. *Se $f : (-1, 1] \rightarrow [0, +\infty)$ é um gerador decrescente, então $g : [-1, 1) \rightarrow [0, +\infty)$ definida como*

$$g(x) = f(-x) \quad (4.17)$$

é um gerador crescente.

2. *Se $g : [-1, 1) \rightarrow [0, +\infty)$ é um gerador crescente então $f : (-1, 1] \rightarrow [0, +\infty)$ definida via*

$$f(x) = g(-x) \quad (4.18)$$

é um gerador decrescente.

Prova: Provaremos apenas (a), pois a prova de (b) é semelhante. Observa-se que g é obviamente contínua. Além disso, g é estritamente crescente, pois para $\forall x, y, x < y \Rightarrow -y < -x$, como f é decrescente, $f(-x) < f(-y)$, isto é, $g(x) < g(y)$. Também $g(-1) = f(1) = 0$ e

$$\lim_{x \rightarrow 1} g(x) = \lim_{x \rightarrow 1} f(-x) = \lim_{y \rightarrow -1} f(y) = +\infty.$$

c.q.d.

Definição 4.5 (Complemento Difuso) *Uma função c definida em $[-1, 1]$ é um complemento difuso, se satisfaz as seguintes condições:*

$$(C1) \quad c(-1) = 1, \quad c(1) = -1,$$

$$(C2) \quad \forall a, b \in [-1, 1] \text{ com } a \leq b \text{ então } c(a) \geq c(b),$$

$$(C3) \quad c \text{ é involutiva, isto é, } c(c(a)) = a, \forall a \in [-1, 1],$$

$$(C4) \quad c \text{ é contínua.}$$

O lema seguinte permite dar uma interpretação para toda t-norma e t-conorma e sua relação com a t-norma do *min* e a t-conorma do *max*.

Lema 4.2 *Se T e S são t-norma e t-conorma arquimedianas estritas, vale:*

$$T(a, b) \leq \inf\{a, b\} \tag{4.19}$$

$$S(a, b) \geq \sup\{a, b\} \tag{4.20}$$

Prova: Pelo teorema 4.1

$$T(a, b) = f^{-1}(f(a) + f(b))$$

com f estritamente decrescente e positiva. Então

$$f(a) + f(b) \geq f(a)$$

$$f(a) + f(b) \geq f(b).$$

Logo

$$T(a, b) = f^{-1}(f(a) + f(b)) \leq f^{-1}(f(a)) = a. \tag{4.21}$$

Do mesmo modo

$$T(a, b) \leq b \tag{4.22}$$

isto é,

$$T(a, b) \leq \inf\{a, b\}$$

A prova da desigualdade (4.20) é similar e será omitida.

Estamos finalmente prontos para verificar que nossos operadores são t-norma e t-conorma arquimedianas estritas. De maneira original, antes de tudo, generalizaremos, considerando uma família de operadores, que incluem nossos operadores iniciais.

0.273.775-1

Biblioteca Universitária
UFSC

4.3 Famílias de T-Normas e T-Conormas

Provaremos que estas famílias satisfazem as condições para serem t-normas e t-conormas e que, particularizando os parâmetros, fica provado que nossos operadores são t-norma e t-conorma.

Definição 4.6 (Família de t-normas) *Por definição, se $a, b \in (-1, 1]$*

$$U_c^{\alpha, \beta}(a, b) \stackrel{\text{def}}{=} f_{\alpha, \beta}^{-1}(f_{\alpha, \beta}(a) + f_{\alpha, \beta}(b)), \alpha, \beta > 0. \quad (4.23)$$

$$a = 1 \Leftrightarrow U_c^{\alpha, \beta}(a, b) = b$$

$$a = -1 \Leftrightarrow U_c^{\alpha, \beta}(a, b) = -1$$

onde $f_{\alpha, \beta}: (-1, 1] \rightarrow [0, +\infty)$ é definida via

$$f_{\alpha, \beta}(x) = \alpha \left(\frac{1-x}{1+x} \right)^\beta \quad (4.24)$$

Definição 4.7 (Família de t-conormas) *Por definição, se $a, b \in [-1, 1)$*

$$U_d^{\alpha, \beta}(a, b) \stackrel{\text{def}}{=} g_{\alpha, \beta}^{-1}(g_{\alpha, \beta}(a) + g_{\alpha, \beta}(b)) \quad (4.25)$$

$$a = 1 \Leftrightarrow U_d^{\alpha, \beta}(a, b) = 1$$

$$a = -1 \Leftrightarrow U_d^{\alpha, \beta}(a, b) = b$$

onde $g_{\alpha, \beta}: [-1, 1) \rightarrow [0, +\infty)$ é definida via

$$g_{\alpha, \beta}(x) = \alpha \left(\frac{1+x}{1-x} \right)^\beta \quad (4.26)$$

A figura 4.2 mostra que conforme variamos os valores de α e de β damos maior ou menor força para a t-norma.

Lema 4.3 *São satisfeitas as seguintes propriedades:*

(a) $f_{\alpha, \beta}$ é um gerador decrescente;

(b) $g_{\alpha, \beta}$ é um gerador crescente.

(c) Se $c: [-1, 1] \rightarrow [-1, 1]$ é definido por $c(x) = -x$ então

$$f_{\alpha, \beta}(x) = g_{\alpha, \beta}(c(x))$$

$$g_{\alpha, \beta}(x) = f_{\alpha, \beta}(c(x))$$

Prova: Provaremos apenas (a). Observa-se que as condições da definição 4.2 são satisfeitas. De fato:

$$f_{\alpha, \beta}(1) = \alpha \left(\frac{1-1}{2} \right)^\beta = 0$$

$$\lim_{x \rightarrow -1} f_{\alpha, \beta}(x) = +\infty$$

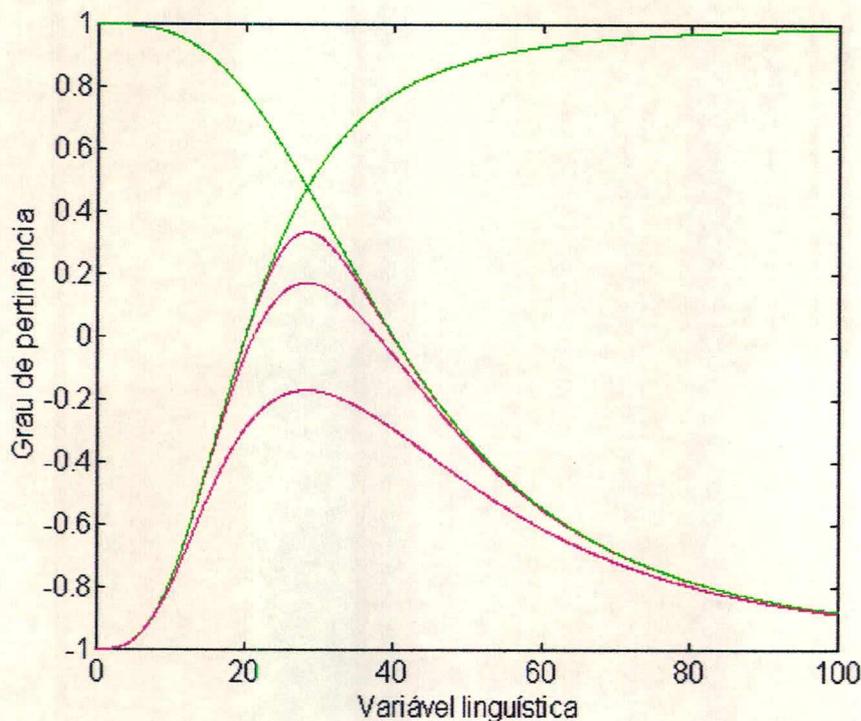


Figura 4.2: Variações de α e de β na família de t-normas

pois $\lim_{x \rightarrow -1} \frac{1-x}{1+x} = +\infty$ e $\alpha, \beta > 0$. Para provarmos que $f_{\alpha, \beta}$ é decrescente, basta estimar sua derivada:

$$f'_{\alpha, \beta}(x) = -\frac{\alpha\beta(1-x)^\beta}{(1+x)^\beta} \left(\frac{1}{1-x} + \frac{1}{1+x} \right)$$

que, obviamente, é estritamente negativa para $\forall x \in (-1, 1)$, provando que $f_{\alpha, \beta}$ é estritamente decrescente.

Portanto, o teorema 4.1 se verifica para a família conjuntiva dada na definição 4.3.

Corolário 4.1 *As famílias $U_c^{\alpha, \beta}$, $U_d^{\alpha, \beta}$ são t-norma e t-conorma arquimedianas estritas, respectivamente.*

Observação: É fácil ver que c , definida em $[-1, 1]$ por $c(x) = -x$, é um complemento difuso.

Definição 4.8 (Tripla Dual) *Dizemos que uma t-norma T e uma t-conorma S são duais relativamente ao complemento c se, e somente se,*

$$c(T(a, b)) = S(c(a), c(b))$$

e

$$c(S(a, b)) = T(c(a), c(b)).$$

A tripla $\langle T, S, c \rangle$ denomina-se tripla dual.

Corolário 4.2 A tripla $\langle U_c^{\alpha,\beta}, U_d^{\alpha,\beta}, c \rangle$, onde $c(x) = -x$, é uma tripla dual.

Prova: Primeiro observemos que valem as seguintes relações:

$$f_{\alpha,\beta}(a) = g_{\alpha,\beta}(-a) \tag{4.27}$$

$$f_{\alpha,\beta}^{-1}(x) = \frac{\alpha^{\frac{1}{\beta}} - x^{\frac{1}{\beta}}}{\alpha^{\frac{1}{\beta}} + x^{\frac{1}{\beta}}} \tag{4.28}$$

$$g_{\alpha,\beta}^{-1}(x) = \frac{x^{\frac{1}{\beta}} - \alpha^{\frac{1}{\beta}}}{x^{\frac{1}{\beta}} + \alpha^{\frac{1}{\beta}}} \tag{4.29}$$

e portanto

$$-f_{\alpha,\beta}^{-1}(x) = g_{\alpha,\beta}^{-1}(x) \tag{4.30}$$

Logo

$$\begin{aligned} c(U_c^{\alpha,\beta}(a, b)) &= -U_c^{\alpha,\beta}(a, b) \\ &= -f_{\alpha,\beta}^{-1}(f_{\alpha,\beta}(a) + f_{\alpha,\beta}(b)) \\ &= g_{\alpha,\beta}^{-1}(f_{\alpha,\beta}(a) + f_{\alpha,\beta}(b)) \\ &= g_{\alpha,\beta}^{-1}(g_{\alpha,\beta}(-a) + g_{\alpha,\beta}(-b)) \\ &= g_{\alpha,\beta}^{-1}(g_{\alpha,\beta}(c(a)) + g_{\alpha,\beta}(c(b))) \\ &= U_d^{\alpha,\beta}(c(a), c(b)) \end{aligned} \tag{4.31}$$

A outra igualdade se prova do mesmo modo.

De maneira semelhante ao lema 4.2, podemos enunciar o seguinte corolário.

Corolário 4.3 Valem as seguintes desigualdades:

$$(a) U_c^{\alpha,\beta}(a, b) \leq \inf\{a, b\}$$

$$(b) U_d^{\alpha,\beta}(a, b) \geq \sup\{a, b\}$$

Observação: Usando as propriedades (T4) e (S4), podemos estender nossos operadores, a saber, fica bem definido

$$U_c^{\alpha,\beta}(x_1, x_2, \dots, x_n) \tag{4.32}$$

e

$$U_d^{\alpha,\beta}(x_1, x_2, \dots, x_n) \tag{4.33}$$

para $\forall (x_1, x_2, \dots, x_n) \in [-1, 1]^n$.

A partir de agora vamos restringir nosso estudo aos operadores $U_c^{1,1}$ e $U_d^{1,1}$, que notaremos por

$$U_c^{1,1} = \mathcal{C} \tag{4.34}$$

$$U_d^{1,1} = \mathcal{D} \tag{4.35}$$

de forma semelhante a 4.13 e a 4.14 mas usando caligráfico.

Usando a notação da introdução, podemos escrever

$$\mathcal{C}(x_1, \dots, x_n) = Op_3(Op_3(x_1) + \dots + Op_3(x_n)) \quad (4.36)$$

$$\mathcal{D}(x_1, \dots, x_n) = Op_1(Op_2(x_1) + \dots + Op_2(x_n)) \quad (4.37)$$

com as óbvias precauções quando algum $x_i = -1$, no caso de \mathcal{C} , e algum $x_i = 1$, no caso de \mathcal{D} . Assumiremos que o leitor está ciente dos valores de \mathcal{C} e \mathcal{D} nestes casos. Para conferir com a notação da definição observe que $f_{1,1} = f_{1,1}^{-1} = Op_3$, $g_{1,1} = Op_2$, e $g_{1,1}^{-1} = Op_1$.

Gráfico dos geradores:

A figura 4.3 mostra o gerador decrescente para a t-norma, fazendo $\alpha = \beta = 1$. E a figura 4.4 mostra o gerador crescente para a t-conorma, fazendo $\alpha = \beta = 1$, junto com seu inverso.

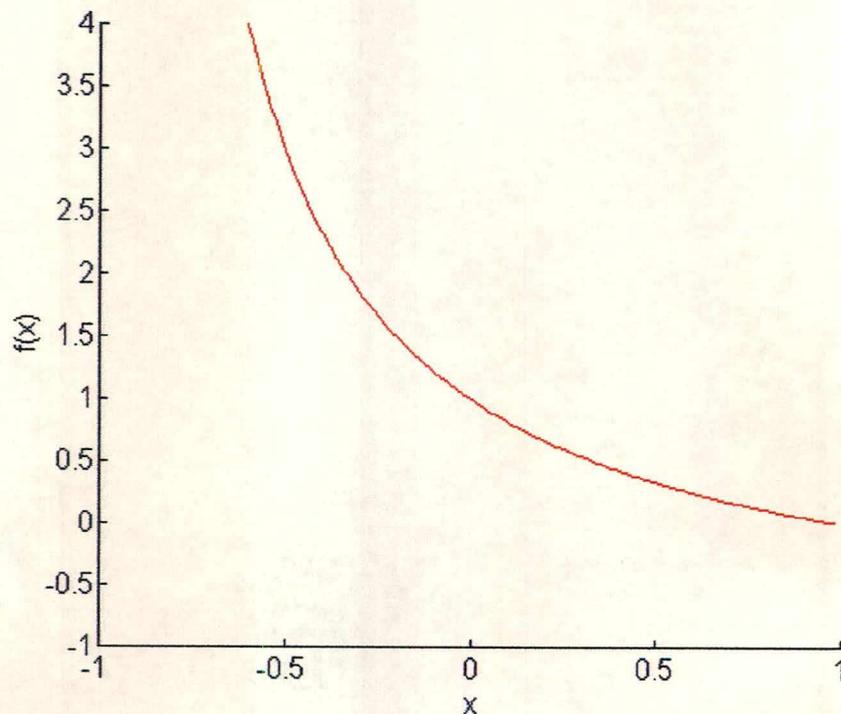


Figura 4.3: Gerador decrescente

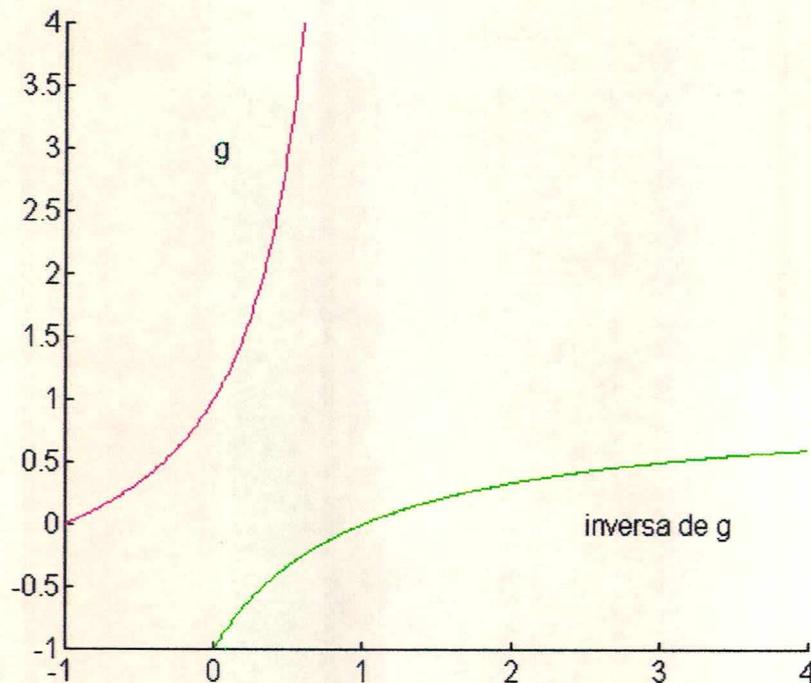


Figura 4.4: Gerador crescente e seu inverso

4.4 Espaço de Funções de Pertinência

Tendo provado as propriedades fundamentais de nossos operadores, podemos agora estabelecer espaços adequados sobre os quais podemos fazer atuar nossos operadores. A idéia é fazê-los atuar de modo pontual sobre um espaço de funções apropriado e provarmos algumas propriedades desta ação. Um tal espaço corresponderá a um espaço de funções de pertinência associadas a conjuntos difusos.

Seja $X \subset \mathbb{R}$ um conjunto fechado. Definimos sobre X o espaço de funções

$$\mathcal{F} = \{ \varphi : X \rightarrow [-1, 1] \mid \varphi \in C^0 \},$$

sobre o qual adotamos a norma do supremo, que denotamos por $\| \cdot \|$,

$$\| \varphi \| = \sup \{ |\varphi(x)| ; x \in X \},$$

que torna $(\mathcal{F}, \| \cdot \|)$ um espaço métrico completo.

4.4.1 Diferenciabilidade das t-normas e t-conormas

Definição 4.9 Dizemos que $\varphi : X \rightarrow [-1, 1]$ é diferenciável se existe $U \supset X$, aberto, e $\bar{\varphi} : U \rightarrow \mathbb{R}$ uma extensão de φ diferenciável.

Teorema 4.2 (Diferenciabilidade) Se $\varphi_1, \dots, \varphi_k \in \mathcal{F}$ e $x \in X$, então

(a) $\mathcal{C}(\varphi_1, \dots, \varphi_k)(x) \stackrel{\text{def}}{=} \mathcal{C}(\varphi_1(x), \dots, \varphi_k(x)) \in \mathcal{F}$

$$(b) \mathcal{D}(\varphi_1, \dots, \varphi_k)(x) \stackrel{\text{def}}{=} \mathcal{D}(\varphi_1(x), \dots, \varphi_k(x)) \in \mathcal{F}$$

(c) Se $\varphi_1, \dots, \varphi_k \in \mathcal{F}$ são diferenciáveis, então

(c1) $\mathcal{C}(\varphi_1(x), \dots, \varphi_k(x))$ é diferenciável.

(c2) $\mathcal{D}(\varphi_1(x), \dots, \varphi_k(x))$ é diferenciável.

Prova: Trivial.

Observe que o teorema garante que nossos operadores preservam a classe de diferenciabilidade das funções envolvidas.

Observação: Vale lembrar que $(\varphi_1(x), \dots, \varphi_k(x)) \in [-1, 1]^k$ e que portanto podemos usar as equações dadas em 4.34 e 4.35.

A partir de agora vamos estudar estes problemas e outros no seguinte contexto: $X = [0, +\infty)$. Logo estaremos considerando o conjunto de funções

$$\mathcal{F}^1 = \{ \varphi : [0, +\infty) \rightarrow [-1, 1] \mid \varphi \in C^1 \}.$$

Um subconjunto importante de \mathcal{F}^1 que vamos considerar é o de funções de pertinência

$$\mathcal{P} = \left\{ \varphi \in \mathcal{F}^1 \mid \lim_{x \rightarrow +\infty} \varphi(x) = \pm 1, \lim_{x \rightarrow 0^+} \varphi(x) = \pm 1 \right\}.$$

Em \mathcal{P} está o subconjunto das funções de pertinência $\mu_{p,k}$ associadas a conceitos elementares, que será denotado por \mathcal{P}_0 e dado por

$$\mathcal{P}_0 = \{ \mu_{p,k}, p > 0, k \neq 0 \}.$$

4.4.2 Uma família particular de função de pertinência

Vamos considerar em nossas aplicações a função de pertinência proposta por Maia [71], dada pela seguinte definição:

Definição 4.10 (Função de pertinência) A função que representa um conceito elementar A sobre um atributo x é

$$\mu_A = \frac{x^k - p^k}{x^k + p^k} \quad (4.38)$$

em que $p(p > 0)$ e $k \neq 0$ são números reais.

A denominação *conceito elementar* refere-se à propriedade monotônica da função e a uma simetria definida pela condição:

$$\frac{x_1}{p} = \frac{p}{x_2} \Leftrightarrow \mu_A(x_1) = -\mu_A(x_2).$$

Esta função efetua uma conversão de uma medida de um atributo físico, que expressa um aspecto da realidade sob observação, em um valor adimensional associado a um conceito ligado a esse atributo. Esta conversão permitirá a

combinação (ou composição) de conceitos simples para formar conceitos mais complexos, sem misturar unidades de medida diferentes, como centímetros e toneladas ou metros e gramas.

Por exemplo, se o objeto θ é uma pessoa, a variável lingüística V altura e o termo lingüístico A , ou conceito A , é *alto*. Um grau de pertinência $\mu_A = 0$ é interpretado como se um sujeito, em um experimento empírico, respondesse que com o valor p de altura a pessoa era *nem alta e nem não alta*, isto é, ela está no limite de concordância ou discordância com o conceito ser pessoa *alta*. Graus de pertinência próximos de 1 significam grande concordância com o conceito, e próximo de -1 , grande discordância. Em outras palavras, isto quer dizer que os valores extremos de concordância são assintóticos. Esta característica parece concordar com a nossa intuição de que sempre é possível conhecer alguém mais alto do que a pessoa mais alta conhecida até então.

Concordamos com Maia [71] quando diz que é mais natural expressar um índice de pertinência por um valor no intervalo $[-1, 1]$ (ou, equivalentemente, entre mais e menos 100 por cento) do que no intervalo $[0, 1]$, que é prática comum na teoria de conjuntos difusos.

A figura 4.5 mostra variações no parâmetro k aplicadas no conceito *alto* da figura 4.1.

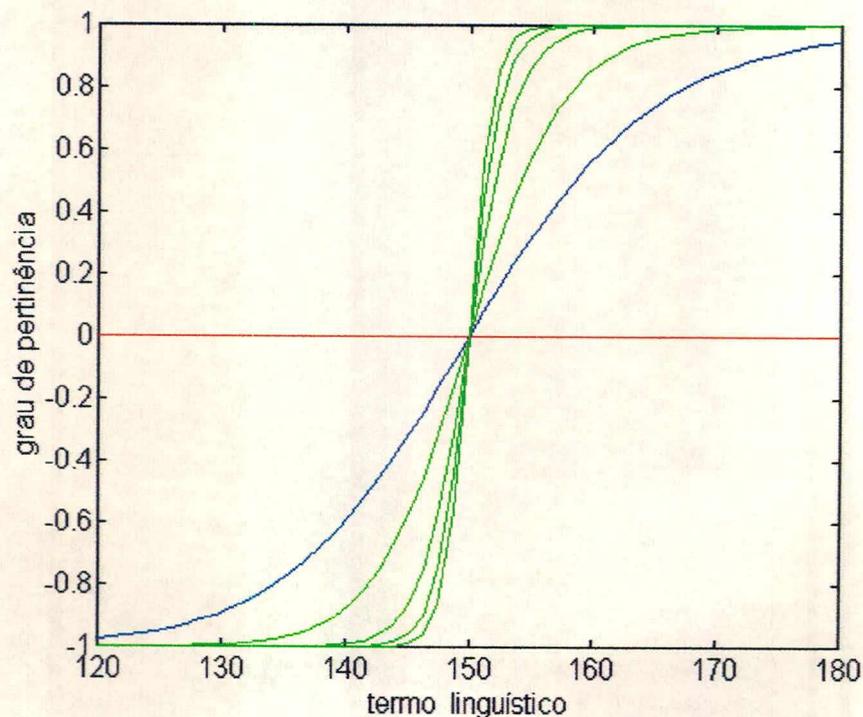


Figura 4.5: Algumas variações de k com p fixo

Observando as figuras deste capítulo é fácil ver que vale o lema a seguir, cujos resultados serão usados nas demonstrações da relação de ordem.

Lema 4.4 São verdadeiras:

(a) $\mu_{p,k} \in C^\infty$.

(b)

$$\lim_{x \rightarrow +\infty} \mu_{p,k}(x) = \begin{cases} 1 & \text{se } k > 0 \\ -1 & \text{se } k < 0 \end{cases}$$

$$\lim_{x \rightarrow 0^+} \mu_{p,k}(x) = \begin{cases} -1 & \text{se } k > 0 \\ 1 & \text{se } k < 0 \end{cases}$$

(c)

$$k > 1 \Rightarrow \begin{cases} \mu_{p,k} & \text{é convexa se } x < p \\ \mu_{p,k} & \text{é côncava se } x > p \end{cases}$$

$$k < -1 \Rightarrow \begin{cases} \mu_{p,k} & \text{é côncava se } x < p \\ \mu_{p,k} & \text{é convexa se } x > p \end{cases}$$

Observação: Algumas considerações particulares devem ser feitas para quando $k = 1$ ou $k = -1$.

Modificadores:

Zadeh [69] define os modificadores como sendo operações que mudam a função de pertinência de um conjunto difuso tornando mais rápida ou mais lenta (*by spreading out*), a transição entre pertinência total e não pertinência total, ou *movendo* a posição da região de transição.

A figura 4.6 mostra como os parâmetros p e k , na equação (4.38), podem efetuar estas modificações de maneira suave e contínua, refletindo-se na caracterização dos conceitos.

Apesar de não termos feito um experimento para verificar a questão dos limiares (*hedges*) linguísticos, podemos variar os parâmetros p e k para obter formas semelhantes às obtidas experimentalmente por Dombi [52] para os

conceitos, ou termos linguísticos, *jovem* e *velho*, na figura 4.7 e *velho* e *muito velho*, na figura 4.8.

A determinação da função de pertinência:

A determinação da função de pertinência, associada a um conceito em um determinado contexto, pela fixação de seus dois parâmetros p e k , pode ser feita de duas maneiras:

- fixando o ponto de transição p e ajustando por tentativas o parâmetro k até obter a caracterização desejada;
- fixando dois pontos distintos no plano $x - z$, de acordo com Maia [71].

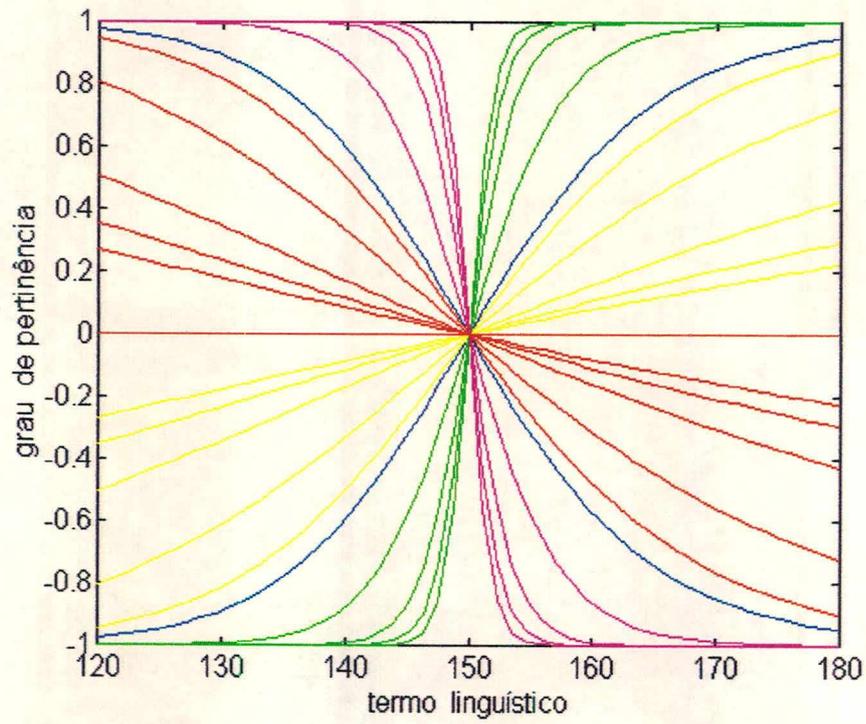


Figura 4.6: Variações de k com p fixo

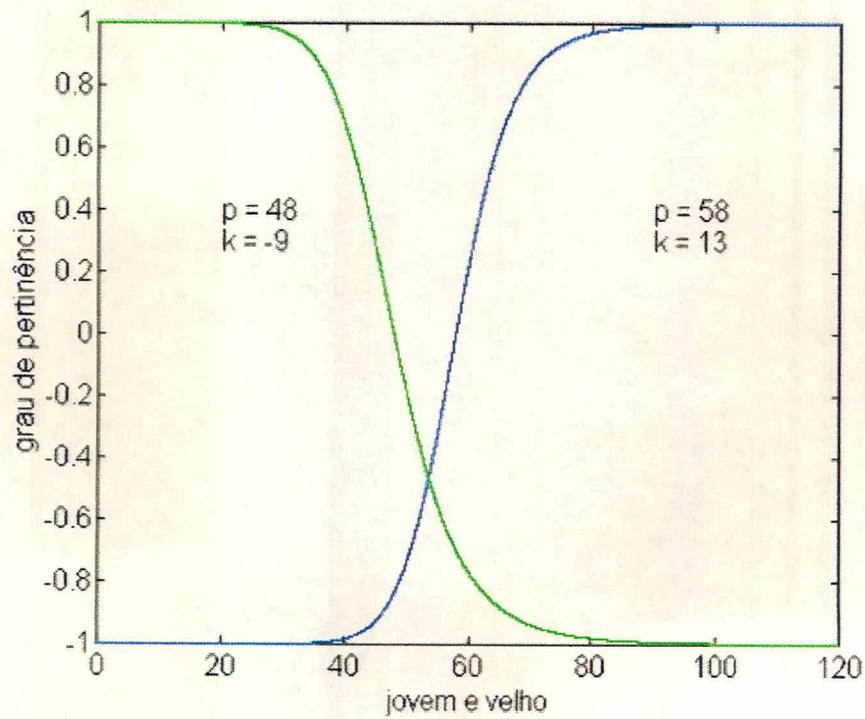


Figura 4.7: Jovem e velho

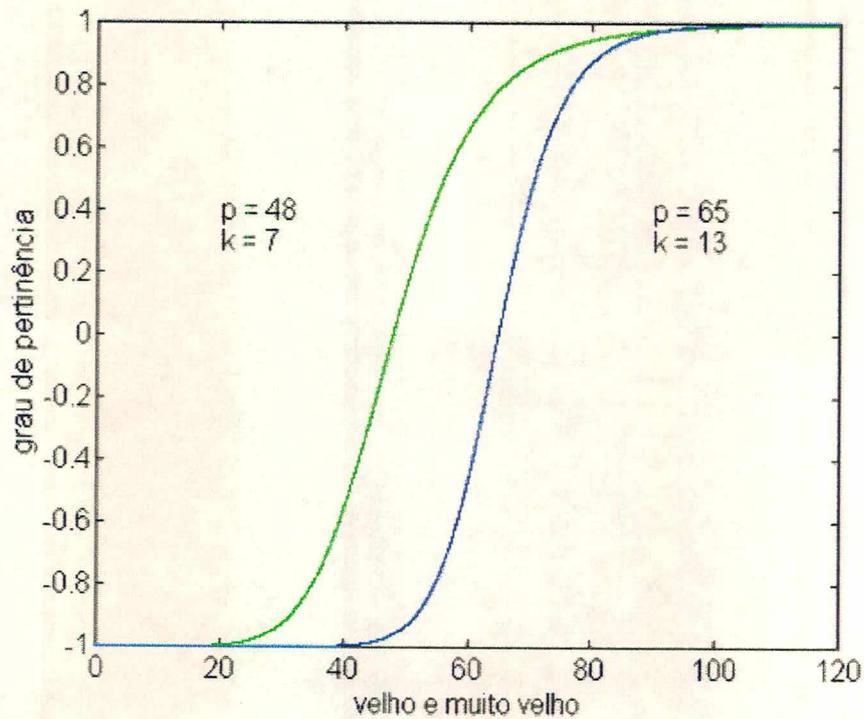


Figura 4.8: velho e muito velho

Relação com os operadores:

Considere os geradores da família de t-normas dada na definição 4.3:

$$f_{\alpha,\beta}(x) = \alpha \left(\frac{1-x}{1+x} \right)^\beta, \alpha, \beta > 0 \tag{4.39}$$

Observe que

$$\alpha \left(\frac{1-x}{1+x} \right)^\beta = y \Rightarrow \left(\frac{1-x}{1+x} \right)^\beta = \frac{y}{\alpha} \Rightarrow \frac{1-x}{1+x} = \left(\frac{y}{\alpha} \right)^{\frac{1}{\beta}}$$

isto é,

$$x = \frac{\alpha^{\frac{1}{\beta}} - y^{\frac{1}{\beta}}}{\alpha^{\frac{1}{\beta}} + y^{\frac{1}{\beta}}}$$

Assim,

$$f_{\alpha,\beta}^{-1}(y) = \frac{\alpha^{\frac{1}{\beta}} - y^{\frac{1}{\beta}}}{\alpha^{\frac{1}{\beta}} + y^{\frac{1}{\beta}}}$$

Vemos que é um elemento de nossa família de funções de pertinência elementares, a menos de multiplicação por 1 ou -1. Em outras palavras, o inverso do gerador decrescente é um elemento da família de conceitos elementares. Isso mostra a estreita relação entre a função de pertinência que estamos usando e o gerador da t-norma $f_{\alpha,\beta}$, conforme a equação (4.39), o que faz com que cheguemos à conclusão

de que existe uma grande adequabilidade de nossos operadores para a função de pertinência escolhida.

Além disso, considere dois elementos particulares das famílias dadas em 4.3, já definidos nas equações 4.23 e 4.24, fazendo $\alpha = \beta = 1$. A saber:

$$\mathcal{C}(x, y) = f^{-1}(f(x) + f(y)) \quad (4.40)$$

$$\mathcal{D}(x, y) = g^{-1}(g(x) + g(y)) \quad (4.41)$$

onde $f(x) = \frac{1-x}{1+x}$ e $g(x) = \frac{1+x}{1-x}$. Dado que

$$g^{-1}(x) = g(x) = \frac{1-x}{1+x} \quad (4.42)$$

$$g^{-1} = \frac{x-1}{x+1} = -f(x) \quad (4.43)$$

Assim, também confirmamos a estreita relação da função de pertinência com o operador disjuntivo.

Relação dos Operadores com o Máximo e o Mínimo:

Na figura 4.9 apresentamos um conjunto de conceitos elementares na cor vermelha, fazemos a conjunção e a disjunção e comparamos com os operadores de máximo e de mínimo, como definidos por Zadeh [65].

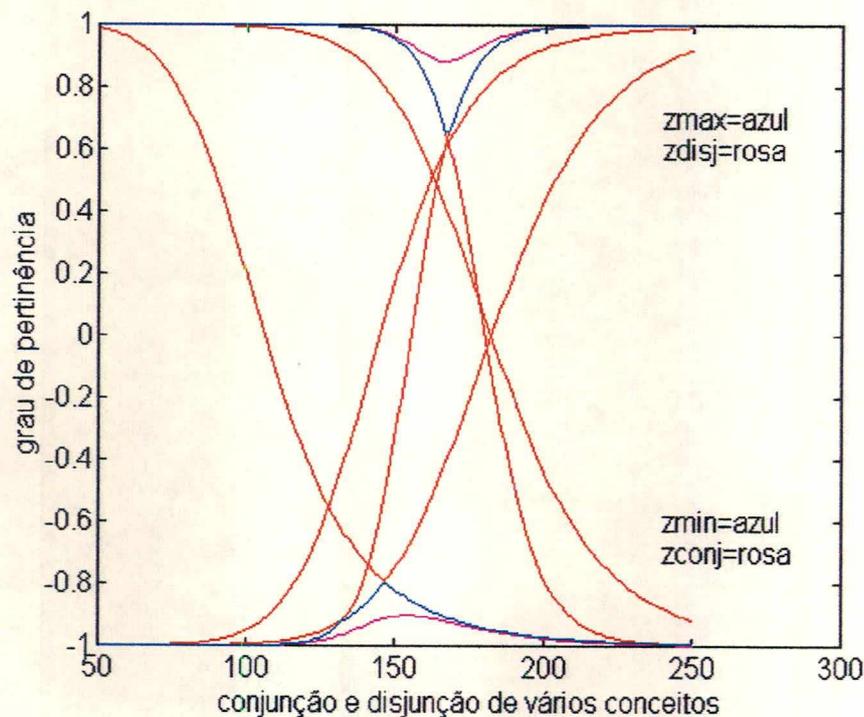


Figura 4.9: Relação com máximo e mínimo

4.4.3 Relação de ordem e convexidade

Uma propriedade importante para aplicações em reconhecimento de padrões e otimização é a de convexidade de um conjunto difuso, introduzida por Zadeh [65].

A figura 4.10 mostra que $\varphi \in \mathcal{F}$ e $\psi \in \mathcal{F}$ convexas, não implica $\mathcal{C}(\varphi, \psi)$ convexa. Considere os termos linguísticos: A_1 = amarelo e A_2 = azul; X = intervalo de valores na unidade de medida da cor. $\mu_A(x)$ = grau com que o objeto que tem x associado pertence ao conjunto dos objetos identificados por azul, ou amarelo.

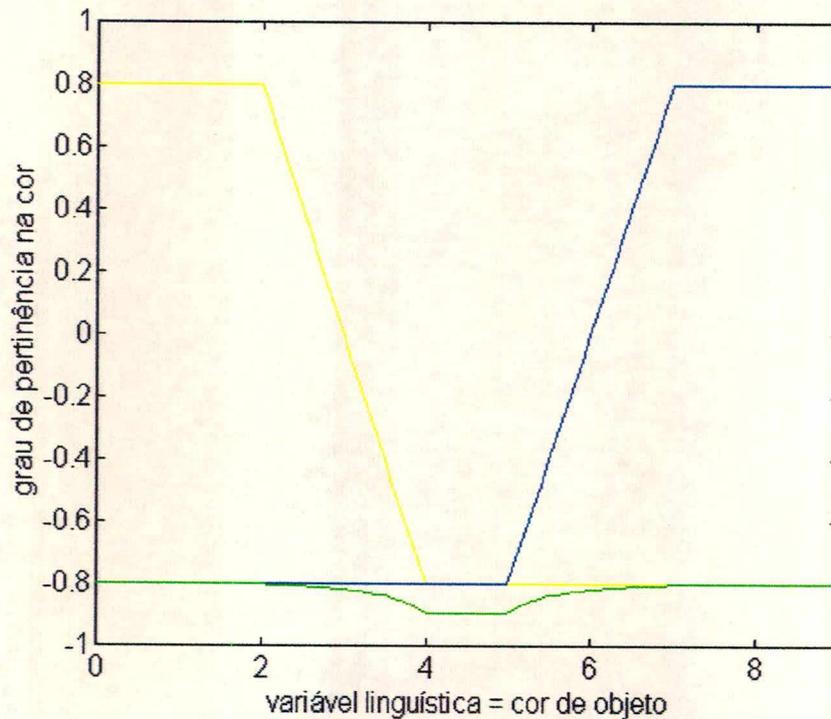


Figura 4.10: Não convexidade da conjunção em \mathcal{F}

A seguir, damos as definições necessárias e mostramos, através de uma relação de ordem caracterizada pelos parâmetros das funções de pertinências, que a conjunção é convexa.

Relação de ordem:

Definição 4.11 (Ordem Parcial) Sobre \mathcal{F}^1 definimos a seguinte ordem parcial

$$\varphi \geq \psi \Leftrightarrow \varphi(x) \geq \psi(x), \forall x \in X.$$

É imediato que \geq é uma relação de ordem sobre \mathcal{F}^1 .

Sobre \mathcal{P}_o , a ordem parcial dada na definição 4.11 admite uma caracterização em relação aos parâmetros p e k . Antes uma observação elementar:

Observação: $\mu_{p,k}$ e $\mu_{q,w}$ são comparáveis $\Leftrightarrow k, w > 0$ ou $k, w < 0$.

Teorema 4.3 (Ordem parcial sobre \mathcal{P}_\circ)

$$\mu_{q,w} \geq \mu_{p,k} \Leftrightarrow \begin{cases} w = k < 0 & e q \geq p \\ w = k > 0 & e q \leq p \end{cases}$$

Prova:(\Leftarrow)

$$\begin{aligned} \mu_{q,w}(x) - \mu_{p,k}(x) &= \frac{x^k - q^k}{x^k + p^k} - \frac{x^k - p^k}{x^k + q^k} \\ &= \frac{x^{2k} - x^k q^k + p^k x^k - p^k q^k - x^{2k} + x^k p^k - x^k q^k + p^k q^k}{(x^k + q^k)(x^k + p^k)} \\ &= \frac{2x^k(p^k - q^k)}{(x^k + q^k)(x^k + p^k)} \\ &\geq 0 \end{aligned}$$

quando $p^k - q^k \geq 0$. Mas isto ocorre pela hipótese, isto é, se $k < 0$ e $q \geq p$ ou se $k > 0$ e $q \leq p$.

(\Rightarrow) Seja

$$\mu_{p,k}(x) - \mu_{q,w}(x) = \frac{2(x^k q^w - x^w p^k)}{(x^k + p^k)(x^w + q^w)}$$

Primeiro será provado que $\mu_{p,k}(x) - \mu_{q,w}(x) = 0$ para algum valor de x se $w \neq k$. Mas isto é equivalente a provar que $x^k q^w - x^w p^k = 0$ com $w = k + \delta, \delta \in \mathbb{R}, \delta \neq 0$. Tem-se que

$$\begin{aligned} x^k q^k q^\delta - x^k x^\delta p^k = 0 &\Leftrightarrow x^k (q^\delta q^k - x^\delta p^k) = 0 \\ &\Leftrightarrow x^\delta p^k = q^\delta q^k \\ &\Leftrightarrow x^\delta = \frac{q^\delta q^k}{p^k} \end{aligned} \tag{4.44}$$

Logo, aplicando \ln em ambos os lados, tem-se $x = \exp\left(\frac{1}{\delta} \ln \frac{q^\delta q^k}{p^k}\right)$.

Neste ponto, as funções ou se tangenciam ou se cruzam. Se elas se tangenciam, neste ponto as derivadas são iguais. Se as derivadas são diferentes, então de fato elas se cruzam. O que contraria a hipótese que $\mu_{q,w}(x) \geq \mu_{p,k}(x)$. Logo $w = k$. As funções só podem ser iguais.

As derivadas das funções são:

$$\mu'_{p,k}(x) = \frac{2kx^{k-1}p^k}{(x^k + p^k)^2}$$

e

$$\mu'_{q,w}(x) = \frac{2wx^{w-1}q^w}{(x^w + q^w)^2}$$

Se $x^\delta = \frac{q^\delta q^k}{p^k}, \delta \neq 0$, tem-se:

$$\mu'_{q,w}(x) = \frac{2(k + \delta)x^{k-1}x^\delta q^k q^\delta}{(x^k x^\delta + q^k q^\delta)^2} = \frac{2(k + \delta)x^{k-1}q^k q^\delta x^\delta}{(x^k x^\delta + q^k q^\delta)^2}$$

Agora,

$$\begin{aligned}
 \frac{\mu'_{p,k}(x)}{\mu'_{q,w}(x)} &= \frac{kp^k}{(x^k + p^k)^2} \frac{(x^k x^\delta + q^\delta q^k)^2}{(k + \delta)q^k q^\delta x^\delta} \\
 &= \frac{(x^k x^\delta + q^k q^\delta)^2}{(x^k + p^k)^2} \frac{k}{k + \delta} \frac{1}{x^{2\delta}} \\
 &= \frac{\frac{(x^k x^\delta + q^k q^\delta)^2}{(p^k)^2}}{\frac{(x^k + p^k)^2}{(p^k)^2}} \frac{k}{k + \delta} \frac{1}{x^{2\delta}} \\
 &= \frac{\left(\frac{x^k x^\delta}{p^k} + \frac{q^k q^\delta}{p^k}\right)^2}{\left(\frac{x^k}{p^k} + 1\right)^2} \frac{k}{k + \delta} \frac{1}{x^{2\delta}} \\
 &= \frac{\left(\frac{x^k x^\delta}{p^k} + x^\delta\right)^2}{\left(\frac{x^k}{p^k} + 1\right)^2} \frac{k}{k + \delta} \frac{1}{x^{2\delta}} \\
 &= \frac{k}{k + \delta} \neq 1
 \end{aligned} \tag{4.45}$$

Logo, as derivadas não são iguais neste ponto, isto é, não pode acontecer $\mu_{q,w}(x) < \mu_{p,k}(x)$. Daí $\mu_{q,w}(x) = \mu_{p,k}(x)$. Mas isto não pode ocorrer com $w \neq k$. Logo $k = w$.

Convexidade:

Definição 4.12 (Convexidade) Um conjunto difuso A é convexo se, e somente se, os conjuntos Γ_α definidos por

$$\Gamma_\alpha = \{x \in X \mid \varphi_A(x) \geq \alpha\} \tag{4.46}$$

são conjuntos convexos de X para todo $\alpha \in (-1, 1]$. $\varphi_A(x)$ é a função de pertinência.

Observação: Observe que há diferença entre esta definição e a definição de função convexa.

A seguinte caracterização de conjuntos convexos está provada em Zadeh [65].

Lema 4.5 A é convexo se, e somente se,

$$\varphi_A(\lambda x + (1 - \lambda)y) \geq \min[\varphi_A(x), \varphi_A(y)] \tag{4.47}$$

para todo $x, y \in X$, $\lambda \in [0, 1]$.

Para os operadores \min e \max de Zadeh vale o seguinte teorema, como se pode ver na figura 4.10 abaixo.

Teorema 4.4 (Zadeh [65]) Se A e B são convexos, então

$$A \cap B (= \min[\varphi_A, \varphi_B]) \tag{4.48}$$

é convexo.

Observação: Na demonstração do teorema abaixo, usamos as notação dada na equação 4.36, em termos de $Op1$, $Op2$ e $Op3$, e a do teorema sobre diferenciabilidade. Usamos g sem estar nos referindo a ao gerador crescente das t -conormas.

Teorema 4.5 *Sejam $\varphi, \psi \in \mathcal{P}_0$.*

$$\varphi(x) = \frac{x^k - p^k}{x^k + p^k} \quad e \quad \psi(x) = \frac{x^w - q^w}{x^w + q^w}$$

com $p, q > 0$, então $\mathcal{C}(\varphi, \psi)$ é convexa.

Prova: Como φ e ψ são funções C^∞ , $\mathcal{C}(\varphi, \psi)$ é C^∞ , logo, para provarmos que \mathcal{C} é convexa, é suficiente provar que \mathcal{C} tem no máximo um máximo local. Além disso, como $Op3$ é monótona decrescente e

$$\mathcal{C}(\varphi, \psi)(x) = Op3(Op3(\varphi(x)) + Op3(\psi(x))) = Op3\left(\left(\frac{p}{x}\right)^k + \left(\frac{q}{x}\right)^w\right) \quad (4.49)$$

é imediato verificar que se existem x, y e z com $x < y < z$, tal que

$$\mathcal{C}(\varphi, \psi)(x) > \mathcal{C}(\varphi, \psi)(y)$$

e

$$\mathcal{C}(\varphi, \psi)(y) < \mathcal{C}(\varphi, \psi)(z)$$

que corresponde à situação que queremos mostrar que não ocorre, então

$$\left(\frac{p}{x}\right)^k + \left(\frac{q}{x}\right)^w < \left(\frac{p}{y}\right)^k + \left(\frac{q}{y}\right)^w$$

$$\left(\frac{p}{y}\right)^k + \left(\frac{q}{y}\right)^w < \left(\frac{p}{z}\right)^k + \left(\frac{q}{z}\right)^w$$

Resumindo, é suficiente mostrar o seguinte resultado: se

$$g(x) = \left(\frac{p}{x}\right)^k + \left(\frac{q}{x}\right)^w, \quad x \in (0, +\infty)$$

admite no máximo um x , tal que $g'(x) = 0$, temos duas situações a considerar:

- (I) Se existe x_0 tal que $\varphi(x_0) = \psi(x_0)$, segue do teorema da relação de ordem que se $k \neq w$. Então

$$\begin{aligned} \frac{x^k - p^k}{x^k + p^k} &= \frac{x^w - q^w}{x^w + q^w} \\ \implies 2 \frac{(x^k q^w - x^w p^k)}{(x^k + p^k)(x^w + q^w)} &= 0 \\ \implies x^k q^w - x^w p^k &= 0 \\ \implies x_0^{k-w} &= \frac{p^k}{q^w} \end{aligned}$$

isto é, o ponto em que $\varphi(x) = \psi(x)$ é este. Derivando g e igualando a zero

$$\begin{aligned}
g'(x) &= \left(\frac{-p}{x^2}\right) + w \left(\frac{q}{x}\right)^{w-1} \left(\frac{-q}{w^2}\right) \\
&= -k \frac{p^k}{x^{k+1}} - w \frac{q^w}{x^{w+1}} \\
&= - \left(\frac{kp^k x^w + wq^w x^k}{xx^k x^w} \right) \\
&= 0 \\
&\implies kp^k x^w = -wq^w x^k \\
&\implies \frac{x^k}{x^w} = -\frac{kp^k}{wq^w} \\
&\implies x^{k-w} = -\frac{k p^k}{w q^w}
\end{aligned}$$

Se k e w têm o mesmo sinal, a equação acima não tem solução em $x > 0$ e portanto g é monótona. Se k e w têm sinais diferentes, a equação acima tem uma única solução em $x > 0$, concluindo a prova neste caso.

(II) $\forall x, x > 0, \varphi(x) \neq \psi(x)$.

Neste caso podemos supor que $\varphi > \psi$. Como φ e ψ são monótonas, é imediato ver que ou ambas são crescentes, ou ambas são decrescentes.

O resultado, segue do lema abaixo.

Lema 4.6 *Se φ e ψ são monótonas crescentes (ou decrescentes), então $\mathcal{C}(\varphi, \psi)$ é convexa.*

Observação Suponhamos que φ e ψ são crescentes, o outro caso é similar. Então, se $x < y$ temos $\varphi(x) \leq \varphi(y)$ e $\psi(x) \leq \psi(y)$. Se $\lambda \in [0, 1]$, temos que

$$\varphi(x) \leq \varphi(\lambda x + (1 - \lambda)y) \leq \varphi(y)$$

$$\psi(x) \leq \psi(\lambda x + (1 - \lambda)y) \leq \psi(y)$$

logo,

$$Op3(\varphi(\lambda x + (1 - \lambda)y) + Op3(\psi(\lambda x + (1 - \lambda)y))) \leq Op3(\varphi(x)) + Op3(\psi(x)),$$

pois $Op3$ é decrescente. Deste modo,

$$\begin{aligned}
&Op3(Op3(\varphi(\lambda x + (1 - \lambda)y)) + Op3(\psi(\lambda x + (1 - \lambda)y))) \\
&\geq Op3(Op3(\varphi(x)) + Op3(\psi(x))) \\
&= \mathcal{C}(\varphi, \psi)(x)
\end{aligned}$$

isto é,

$$\mathcal{C}(\varphi, \psi)(\lambda x + (1 - \lambda)y) \geq \min \{ \mathcal{C}(\varphi, \psi)(x), \mathcal{C}(\varphi, \psi)(y) \}$$

Segue do teorema 4.5 acima que $\mathcal{C}(\varphi, \psi)$ é convexa.

As figuras 4.11, 4.12 e 4.13, ilustram a convexidade do conjunto difuso associado à função conjuntiva (que não é uma função convexa).

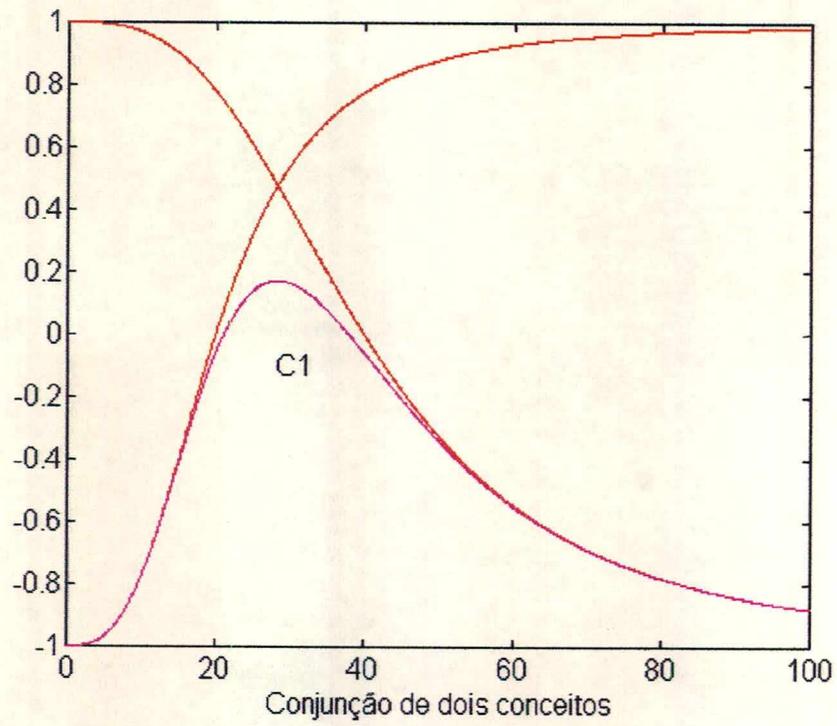


Figura 4.11: Convexidade da conjunção

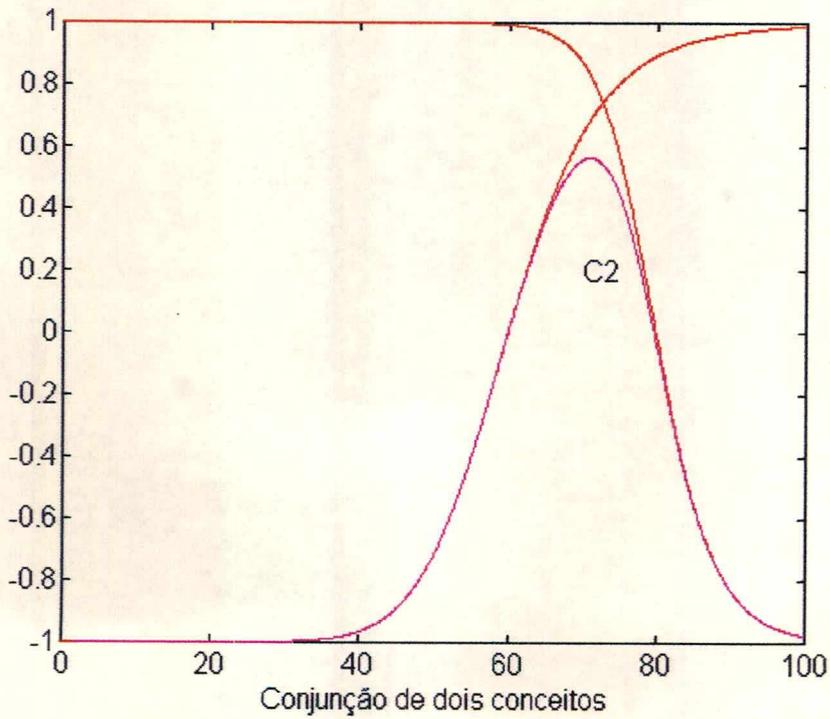


Figura 4.12: Convexidade da conjunção

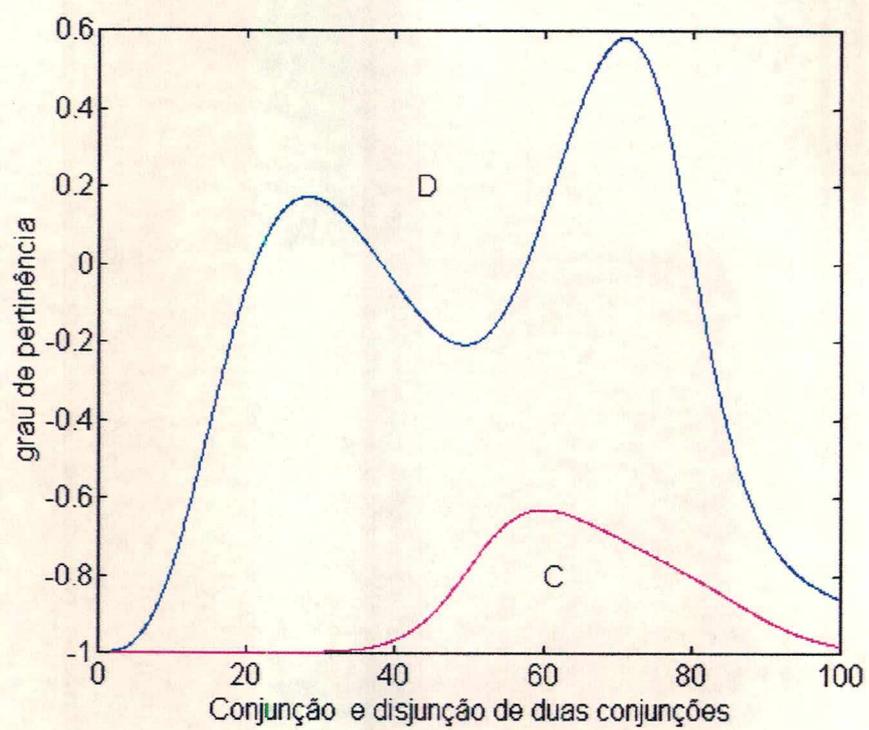


Figura 4.13: A convexidade da conjunção e não convexidade da disjunção

4.5 Quando a Conjunção não é T-Norma:

Conclusão quanto à aplicabilidade de nossos operadores no exemplo da maçã listrada de Osherson e Smith [23]:

Talvez para a caracterização e reconhecimento de quaisquer tipos de conceitos não fosse necessário demonstrar que a função conjuntiva é uma t-norma, caracterizando-se assim, como interseção de conjuntos, pois estas sempre são menores ou iguais ao mínimo, o que não seria, com certeza, nenhuma inovação para o problema da maçã listrada. A esperança é que uma sugestão para este problema seja encontrada quando se variam os parâmetros de forma que se consiga um valor de pertinência do objeto a para o conceito conjuntivo <maçã listrada> maior do que o valor da pertinência ao conceito <maçã>. De fato, existem valores de parâmetros na função conjuntiva que vamos estudar no capítulo 4 que fazem isto acontecer. Zadeh [68] dá algumas referências de como esta discussão pode se desenvolver.

Kamp [9] diz:

Parece que uma versão da teoria de protótipos destinada a analisar graus de pertinência para conceitos em termos da prototypicalidade não é aplicável a todos os conceitos, incluindo muitos que seriam entendidos como prototípicos, de acordo com os vários critérios de Rosch [22].

Podemos ver na figura 4.14 que a conjunção ultrapassou o mínimo.

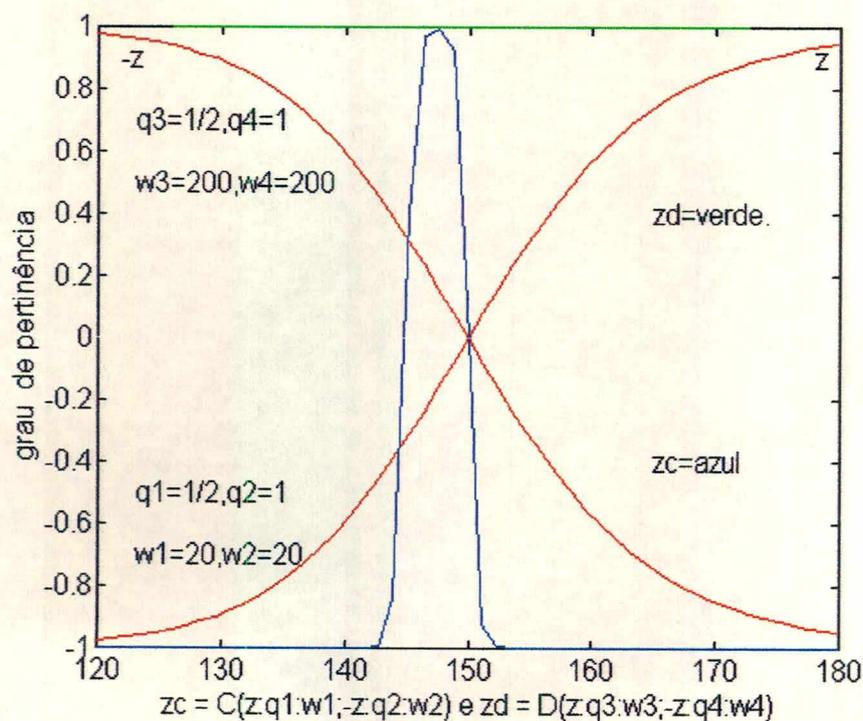


Figura 4.14: Relação com o mínimo

Capítulo 5

APLICAÇÃO: Caracterização de Conceitos

5.1 Introdução

A teoria dos conjuntos difusos tem visto inúmeras e variadas aplicações serem colocadas nos mais diferentes ramos do conhecimento nos últimos anos. Desde que foi proposta por Zadeh [65], sofreu grande desenvolvimento, incluindo especialmente a caracterização completa dos operadores conjuntivos e disjuntivos bem como a formulação de muitas famílias de tais operadores e funções de pertinência com suas extensas propriedades. O problema da escolha das funções de pertinência e dos operadores nas aplicações têm sido tema de vários trabalhos na área e continua sendo até hoje a parte mais polêmica. Uma vez que estamos propondo operadores particulares e funções de pertinência, consideramos de capital importância justificar por que nossas escolhas são bem adequadas no contexto de nossa aplicação.

Uma de nossas aplicações é no reconhecimento de padrões, estreitamente relacionada com a teoria de conceitos. Temos conhecimento da longa discussão a respeito da teoria dos conceitos entre modelos representacionais e modelos de extensão e de todas as críticas ao modelo extensional ao qual nos estamos atendo. No entanto, entrar nesta controvérsia está muito além de nossos objetivos. Estamos, em nosso trabalho, nos restringindo aos modelos de extensão baseados na teoria dos conjuntos difusos. Um conceito é representado por algum conjunto difuso A com sua função de pertinência μ_A . De conceitos elementares para conceito complexos evoluímos por uma seqüência de aplicações de um número finito de conjunções e disjunções. A idéia fundamental por trás desta abordagem, comum a vários ramos novos como da teoria do caos, da geometria fractal, etc. é a seguinte: iterações sucessivas de uma lei simples, partindo-se de elementos bem simples mas suficientemente gerais, pode produzir estruturas extremamente complexas. Nesta direção estamos propondo a seguinte família de funções de pertinência: no intervalo $(0, +\infty)$ consideremos a família, a dois parâmetros, de funções definidas via

$$\mu_{p,k}(x) = \frac{x^k - p^k}{x^k + p^k} \quad (5.1)$$

com $p, k \in R, p > 0, k \neq 0$.

As propriedades destas funções já foram vistas no capítulo anterior, mas cabe ressaltar que elas são descritas com poucos parâmetros que têm influência

direta sobre a transição de *pertinência* para *não pertinência* e sobre a posição da região de transição, provavelmente dando-lhes grande adequabilidade a diversos contextos. Além disto, $\mu_{p,k}$ satisfaz todas as propriedades básicas que a literatura tem exigido para uma função de pertinência.

Também apresentamos os operadores, que podem ser revistos como

$$\mathcal{C}(x, y) = f^{-1}(f(x) + f(y)) \quad (5.2)$$

$$\mathcal{D}(x, y) = g^{-1}(g(x) + g(y)) \quad (5.3)$$

onde $f(x) = \frac{1-x}{1+x}$ e $g(x) = \frac{1+x}{1-x}$. Estes são elementos particulares de uma família de operadores cujos geradores decrescentes são

$$f_{\alpha,\beta}(x) = \alpha \left(\frac{1-x}{1+x} \right)^\beta, \alpha, \beta > 0. \quad (5.4)$$

A propriedade de que a inversa de $f_{\alpha,\beta}$ é um elemento da nossa família 5.1 nos garante a estreita relação entre a função de pertinência e os operadores. Para o sistema de caracterização de conceito, que vamos propor neste capítulo, esta propriedade garante que haverá uma forte conexão entre as *entradas (inputs)* e as *saídas (output)* do sistema.

Como $f_{\alpha,\beta}$ é uma função C^∞ , os operadores são diferenciáveis e, portanto, se φ e ψ são C^∞ então

$$\mathcal{C}(\varphi, \psi) \quad e \quad \mathcal{D}(\varphi, \psi)$$

também o são. Esta propriedade representa uma enorme vantagem operacional. Como as funções de pertinência também são C^∞ , os problemas de aproximação podem ser abordados com as inúmeras técnicas que o cálculo diferencial nos fornece, por exemplo, método de Newton, o método do gradiente descendente, o método do algoritmo genético, etc. Esta vantagem se firma diante dos problemas concretos de classificação onde o número de variáveis é da ordem de dezenas e centenas. O leitor poderá acompanhar na aplicação que faremos adiante à classificação das flores íris.

Voltando aos nossos operadores em 5.2 e 5.3, é importante ressaltar a simplicidade matemática dos geradores bem como de seus inversos. Observemos que

$$f^{-1} = f = \frac{1-x}{1+x} \quad (5.5)$$

$$g^{-1} = \frac{x-1}{x+1} \quad (5.6)$$

5.2 Uma Aplicação no Contexto da Taxionomia Biológica

Decidir se um espécime (ou população) é de uma determinada espécie (que já tem a descrição prototípica), ou é uma nova espécie, depende de atributos quantitativos e qualitativos. Contudo, em nossa aplicação com os dados *íris* vamos considerar só os quantitativos. A variável lingüística em questão é *espécie*, que se define pela composição ou disjunção de vários termos lingüísticos associados a cada

atributo observado. Acrescentamos que cada atributo observado também é uma variável lingüística, portanto *espécie* é uma variável composta.

Alguns exemplos, com partes de descrições taxonômicas, mostram este aspecto de composição (todos os dados em micrômetros):

Espécie A \equiv comprimento do corpo em $[0.56, 0.68]$; comp. do estilete em $[0.1, 0.2]$

Espécie B $\equiv L = 719(s = 49)$; *estilete* = $26(s = 0.4)$; esôfago = $121(s = 1.3)$

Espécie C $\equiv L = 451 \pm 50\mu m$ (349-544); *Oes* = $88 \pm 12\mu m$ (71-11); *EP* = $77.3 \pm 0.9\mu m$ · até 15 variáveis.

Observe que não temos, envolvidos nestas expressões, exatamente termos lingüísticos, mas suponho que um especialista no assunto os teria.

Na tentativa de dar apoio à nossa proposta de função de pertinência, para o caso das variáveis lingüísticas componentes e também para a composta destas componentes, apresentamos, a seguir, algumas observações.

- Os caracteres taxonômicos quantitativos, que envolvem medidas com microscópio, estão sujeitos a muitos ruídos. Além disso, a variabilidade pode vir de fatores como temperatura, umidade, acidez do solo e hospedeiro etc., em que são encontrados os indivíduos.
- O estudo destes caracteres é semelhante ao estudo da altura. Só que a altura todos conhecem sobre ela e podem formar conceitos como *alto*, *médio* e *baixo*, associados a um protótipo. No caso de nematóides, flores e qualquer outro organismo vivo, os taxonomista também usam termos como *curto*, *médio* e *longo*, apesar de haver um esforço cada vez maior para que seja publicada a média do caráter juntamente com o desvio padrão. Ainda assim, há muita divergência de opiniões e publicações aceitas como mais fidedignas ou menos.
- Antecipação, pelo sujeito, de erros de observações sob condições não exatas: mesmo quando o sujeito realiza um experimento exato em que ele mede o valor exato do atributo de cada objeto, ele sabe do fato de que, sob condições de observações não-exatas da vida diária, sua estimativa do valor do atributo pode assumir valores que variam de acordo com alguma distribuição de probabilidade ou *função de erro*. Hisdal [19] mostra como este conhecimento do sujeito dá origem a curvas de pertinência da forma *S* ou de *sino*.
- Por exemplo, valores bem próximos de 1,64 (que deve ser a média de altura da espécie humana do Sul do Brasil) teriam maior pertinência à espécie humana. A pertinência iria diminuindo conforme se distanciasse pela direita, ou pela esquerda, de 1,64. Para a variável lingüística altura, poderiam ser definidos os conceitos (ou termos lingüísticos) das espécies de cachorros, gato, homem, etc.
- Para a altura e outras medidas quantitativas, os estatísticos usam a curva normal para expressar a probabilidade de encontrar, aleatoriamente, um indivíduo dentro de um intervalo. Há maior probabilidade de que a ele esteja associada uma medida próxima da média da espécie, que é atribuída ao protótipo.

- Aqui não estamos falando de função de pertinência para um termo lingüístico, por exemplo, *alto*, da espécie humana ou do cachorro, mas do conjunto difuso determinado pelo conceito *espécie*.
- Ao protótipo é associado grau de pertinência próximo de 1.
- Supomos que o conceito espécie humana, ou espécie cachorro pequinês, para o atributo altura, ou no universo de discurso das alturas, sejam curvas parecidas com a normal, isto é, aproximadamente em forma de *sino*, mas que sejam curvas de pertinência. Ou seja, os graus de pertinência variam em torno do valor igual a 1, ou o mais alto valor, que é associado ao protótipo e, portanto, à média da espécie. Uma argumentação formal que dá apoio a esta suposição pode ser encontrada em Hisdal [19].
- Para representar o conceito de um espécie, segundo um caráter, supomos um conceito difuso para o limite inferior do intervalo de variação de medidas do caráter, com $k > 0$ na expressão do conceito elementar, e um conceito difuso para o limite superior, com $k < 0$. A conjunção destes dois daria a curva que representa a espécie. Na seção a seguir vemos uma representação mais formal para isto.
- O conceito difuso de uma espécie vai ser dado pela conjunção dos conceitos difusos associados a cada atributo envolvido na identificação. Em cada atributo, a função de pertinência do conceito de espécie, tem a forma da conjunção de dois conceitos elementares associados aos atributos. Cada atributo tem um intervalo de variação (*range*) que é o seu *conjunto referencial*.
- Por que supomos que a curva possa ser não-simétrica, para a direita, por exemplo? De fato, como a designação do conceito depende de muitos atributos, pode ocorrer que valores mais à esquerda do intervalo referencial de um atributo V^* dêem menos pertinência para classificá-lo dentro daquela espécie. Outros atributos o determinam (o espécime) como sendo daquela espécie, mas em relação a este atributo V^* ele tem um valor que pode não ser o mais comum dentro da espécie. Ou seja, o espécime que tem este valor do atributo V^* mais à esquerda, no conjunto referencial para este atributo, tem menor grau de pertinência ao conceito daquela espécie no que se refere a este atributo. A mesma observação pode valer para o caso de uma curva não-simétrica para a esquerda. É claro que a curva pode ser simétrica também.
- Não sabemos se a assimetria vai ser mais para a esquerda, ou para a direita. O que vai fazer com que os parâmetros das funções de pertinência, associadas a cada atributo daquela espécie, se definam é a amostra que vamos usar para ajustá-los.
- Conforme Turksen [50], podemos afirmar que o fato de a espécie ser definida como composição de termos de vários atributos favorece a utilização de t-normas ou t-conormas.

5.3 Sobre o Conjunto de Dados

Íris é uma planta da família das iridáceas. Há mais de 1000 espécies, sobretudo nos países temperados, poucas das quais brasileiras. Muitas delas são ornamentais e comuns nos jardins. A conhecida palma-de-santa-rita é uma espécie brasileira.

Anderson [16] observou 150 íris quanto às variáveis: comprimento e largura da sépala (SL e SW) e comprimento e largura da pétala (PL e PW), em cm. Ele mediu 50 plantas de cada subespécie: Íris setosa (SE), Íris versicolor (VC) e Íris virgínica (VG). Assim, o conjunto de dados ÍRIS é formado por 150 linhas com 4 colunas. As subespécies VC e VG foram encontradas e medidas em uma única colônia, enquanto as medidas em SE foram feitas em plantas que cresciam num hábitat diferente.

Anderson fez uma primeira conjectura: que VC era híbrido poliplóide de SE e VG. Esta especulação formou a base para o artigo clássico de Fisher [92], sobre a utilidade da análise discriminante linear, como uma ferramenta para a investigação destes aspectos. Fisher acabou concordando com a teoria da hibridização, sugerindo ainda que VG exerce aproximadamente duas vezes mais influência sobre um paradigmático membro de VC do que SE à medida que a hibridização é considerada (*Even suggesting that VG exerts roughly twice as much influence on a paradigmatic member of VC as does SE as far as hybridization is concerned*).

Este conjunto de dados tornou-se favorito para a análise de agrupamento (*cluster*). Algumas referências são [16],[92],[84], [53],[55],[79],[57], [14],[4],[87]. A figura 5.1 mostra a concordância com a teoria da hibridização de Anderson e Fisher. As características da pétala são as que melhor discriminam.

A figura 5.2 mostra que as características da sépala não discriminam os três grupos.

5.4 Função de Pertinência para um Conceito e um Atributo

Nesta seção, e nas seguintes, vamos utilizar uma notação diferente para a função de pertinência associada a um conceito elementar definido sobre um atributo X , a saber,

$$Z_X = F_o(x : p : k) = \mu_{p,k}(x) \quad (5.7)$$

A seguir introduzimos uma definição para facilitar a notação e a compreensão do problema em estudo. Nós a usaremos no final do capítulo.

Definição 5.1 (Modificação) *Dado um conceito elementar sobre o atributo X , Z_X , qualquer outro conceito elementar definido sobre o mesmo atributo, denotado por $Z'_X = F_o(x : p' : k')$, pode ser obtido efetuando uma modificação sobre Z_X , ou seja,*

$$Z'_X(x) = F_m(Z_X : q : w) = \frac{y^w - q^w}{y^w + q^w} \quad (5.8)$$

onde $y = \frac{1+Z_X}{1-Z_X}$, $w = \frac{k'}{k}$ e $q = \left(\frac{p'}{p}\right)^k$

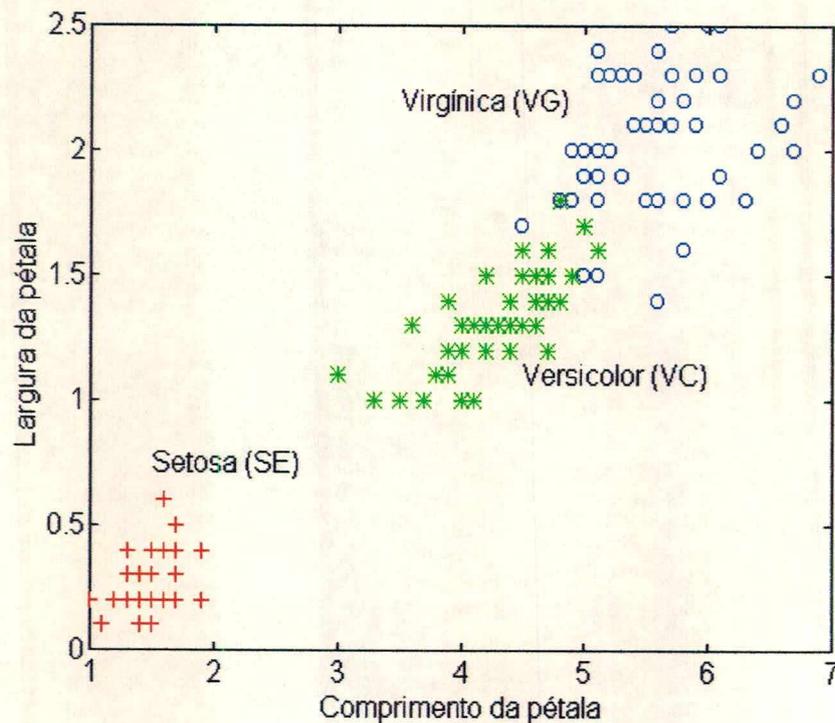


Figura 5.1: Os dados íris de Anderson

Esta definição e alguns gráficos relacionados podem ser encontrados em Maia [71].

Observação: Neste capítulo e no próximo vamos usar a notação U_c para operador conjuntivo e U_d para operador disjuntivo.

Agora estamos prontos para introduzir nossa função de pertinência para um conceito (ou também classe) sobre um atributo que vai compor o conjunto dos atributos importantes para caracterizar uma espécie.

Vamos supor um atributo real X qualquer, cujos valores para uma classe G , sejam conhecidos como variando no intervalo $[I, S]$. Contudo, como é o caso na Taxionomia Biológica, muitos pesquisadores encontram variações nestes limites. Estas podem ocorrer por diferenças nos instrumentos de medição, na natureza química do solo, na temperatura, etc., dificultando a identificação de espécimes numa ou outra classe, o que dá origem a muitas controvérsias entre os taxonomistas. Por isto dizemos que G é uma classe difusa, cujos limites não são bem definidos.

Quando os valores de X se aproximam destes limites, a pertinência ao conceito difuso Z_G , que caracteriza a classe G , se aproxima de 0. Conforme se distancia destes limites, saindo de G , a pertinência a Z_G vai tendendo a -1 , assintoticamente. A figura 5.3 ilustra o gráfico de um possível Z_G , para $I = 3$ e $S = 13$.

Também podemos dizer que existe uma região difusa para os limites I e S . A pertinência dos valores dos atributos, ao conceito que caracteriza a região difusa I , denotado por Z_{XI} , é positiva à direita de I e negativa à esquerda de I . Do mesmo modo define-se o conceito Z_{XS} com pertinência positiva à esquerda de S e negativa à direita.

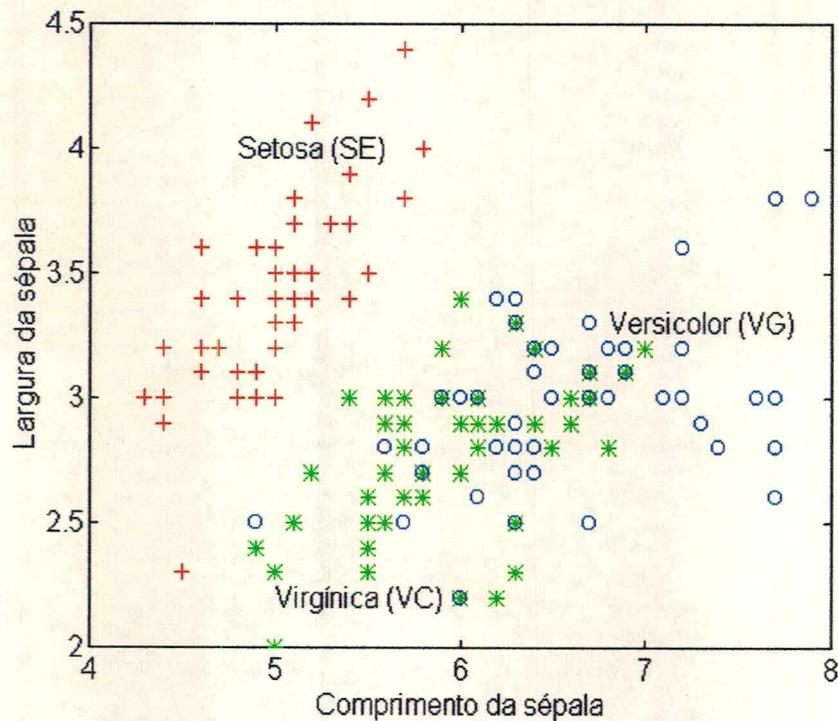


Figura 5.2: Os dados íris de Anderson

Para um atributo X , pode-se considerar um *conceito primário*, elementar, caracterizado por Z_X , como na equação 5.7. Podemos considerar Z_{XI} e Z_{XS} modificações de Z_X , conforme definição 5.1 e a figura 5.4. Assim, temos:

$$Z_X = Fo(x : p : k) \tag{5.9}$$

$$Z_{XI} = Fm(Z_X : q_i : w_i) \tag{5.10}$$

$$Z_{XS} = Fm(Z_X : q_s : w_s) \tag{5.11}$$

Dá para verificar que para todo conceito da forma da figura 5.3 é possível encontrar modificações do *conceito primário* Z_X , de modo que

$$Z_G = U_c(Z_{XI} : 1 : 1, Z_{XS} : 1 : 1) = U_c(Z_X : q_i : w_i, Z_X : q_s : w_s) \tag{5.12}$$

com w_s negativo. A figura 5.4 ilustra o caso em que $I = 3$ e $S = 13$, com os conceitos

$$Z_X = Fo(x : 8 : 2) \tag{5.13}$$

$$Z_{XI} = Fm(Z_X : q_i : w_i) = Fm(Z_X : \frac{9}{64} : 2) = Fo(x : 3 : 4) \tag{5.14}$$

$$Z_{XS} = Fm(Z_X : q_s : w_s) = Fm(Z_X : \frac{169}{64} : \frac{-5}{2}) = Fo(x : 13 : -5) \tag{5.15}$$

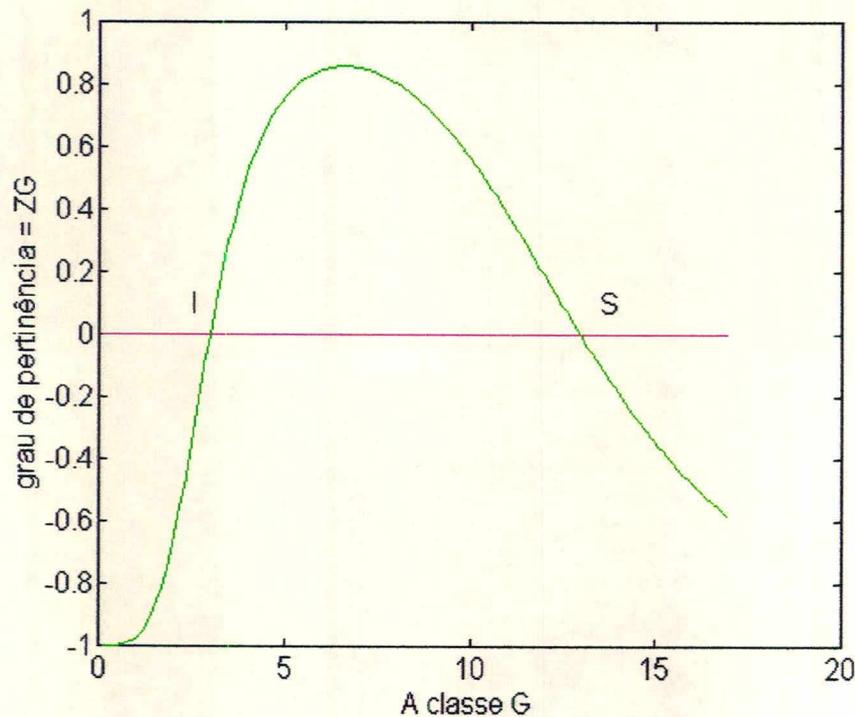


Figura 5.3: Um conceito que caracteriza a classe G

5.5 A Identificação Proposta

Hipótese: As funções obtidas depois do ajuste dos parâmetros devem servir para identificar uma nova observação, não classificada, à classe certa, ou seja, determinada por Anderson.

Observação 5.1 • *Para simplificar, as denominações dadas por Anderson foram modificadas fazendo subespécie SE como classe A, VC como classe B e VG como classe C.*

- *Pretendemos verificar um procedimento de classificação que utilize as funções propostas ou, também, a capacidade classificatória das funções propostas.*

Para isto a cada conceito das classes A, B e C será associada uma função de caracterização, ou de reconhecimento, que depende destes 4 atributos e de 16 parâmetros para A, 32 para B e 32 para C. Cada função é determinada, ou só pela conjunção, ou pela disjunção de conjunções dos conceitos elementares associados a cada atributo. Portanto, representam um conceito complexo.

5.5.1 Um conceito como conjunção de conceitos

A função caracterizadora ou de reconhecimento da classe A, ou do conceito A, é denotada por RA , e é dada por:

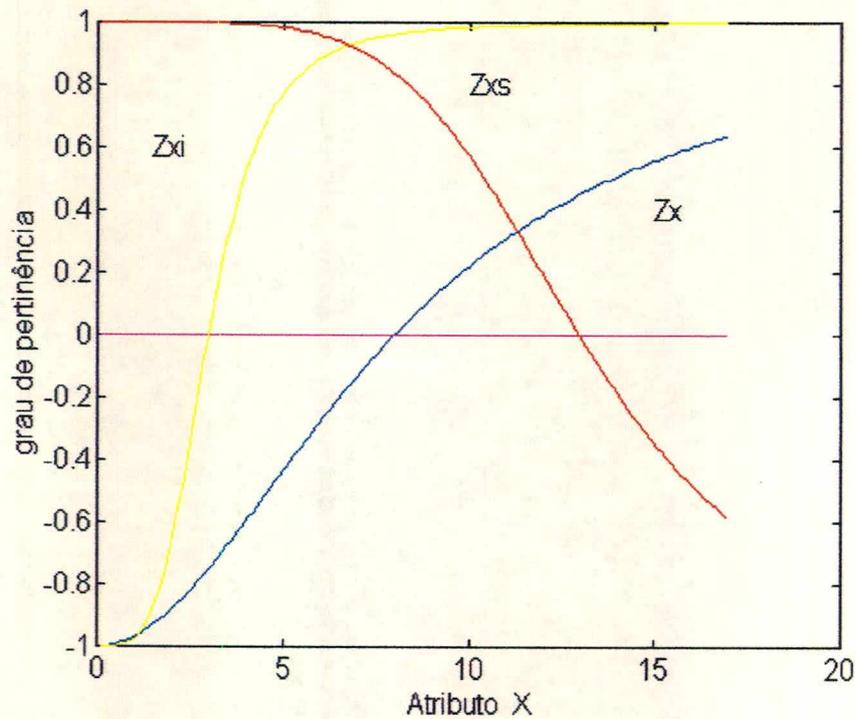


Figura 5.4: Conceito primário e suas modificações

$$\begin{aligned}
 RA = U_c(Z_{SLI} : 1 : 1, Z_{SLS} : 1 : 1, \\
 Z_{SWI} : 1 : 1, Z_{SWS} : 1 : 1, \\
 Z_{PLI} : 1 : 1, Z_{PLS} : 1 : 1, \\
 Z_{PWI} : 1 : 1, Z_{PWS} : 1 : 1)
 \end{aligned}
 \tag{5.16}$$

onde o *I* refere-se a inferior e *S* a superior. Ou seja, temos um conceito inferior e um superior para cada atributo que define a espécie.

É fácil provar que *RA* também pode ser escrito de outra maneira:

$$\begin{aligned}
 RA = C(Z_{SL} : q_{i1} : w_{i1}, Z_{SL} : q_{s1} : w_{s1}, \\
 Z_{SW} : q_{i2} : w_{i2}, Z_{SW} : q_{s2} : w_{s2}, \\
 Z_{PL} : q_{i3} : w_{i3}, Z_{PL} : q_{s3} : w_{s3}, \\
 Z_{PW} : q_{i4} : w_{i4}, Z_{PW} : q_{s4} : w_{s4})
 \end{aligned}
 \tag{5.17}$$

Como exemplo, para o atributo *SL*, *Z_{SLI}* é uma modificação de *Z_{SL}* com parâmetros *q_{i1}* e *w_{i1}* e *Z_{SLS}* com parâmetros *q_{s1}* e *w_{s1}*. Vale o mesmo para todos os atributos *SL*, *SW*, *PL* e *PW*.

O cálculo de *RA* será feito, considerando as equações da introdução do capítulo 4.

Supõe-se que a classe *A*, caracterizada por *RA*, é uma região difusa no espaço *R⁴* e, portanto, caracterizada por um conceito obtido pela combinação de conceitos associados aos 4 atributos. As variações de *RA* em relação a cada

dimensão do espaço devem refletir a influência dessa dimensão na caracterização do conceito da espécie *A*.

Em nosso caso, para o conjunto de dados *íris*, temos as classes *A*, *B* e *C*. Estas serão caracterizadas de maneira semelhante. Contudo, como era evidente a mistura das classes *B* e *C*, supusemos que podia haver uma partição de *B* em duas subclasses *B*₁ e *B*₂, de forma que $B = B_1 \cup B_2$, sendo *B*₁, ou mesmo *B*₂, composto daquelas flores de *B* que estavam mais misturadas com *C*. Fizemos o mesmo para *C*, considerando que $C = C_1 \cup C_2$.

5.5.2 Um conceito como disjunção de conjunções

Assim, a função caracterizadora de *RB* é obtida fazendo a disjunção dos conceitos complexos *RB*₁ e *RB*₂. Da mesma forma obtém-se a função caracterizadora de *RC* utilizando *RC*₁ e *RC*₂. Ou seja,

$$RB = U_d(RB_1 : 1 : 1, RB_2 : 1 : 1) \quad (5.18)$$

e

$$RC = U_d(RC_1 : 1 : 1, RC_2 : 1 : 1) \quad (5.19)$$

que são calculadas utilizando as equações da introdução do capítulo 4.

Os conceitos *RB*₁, *RB*₂, *RC*₁ e *RC*₂ são obtidos por fórmulas semelhantes à dada por *RA*, nas equações 5.16 ou 5.17, isto é, por conjunções que dependem cada uma de 16 parâmetros.

Na verdade, esta divisão se confirmou como mais efetiva em um trabalho de ajuste manual de parâmetros feito antes de desenvolvermos o algoritmo de minimização do erro, que será descrito nas seções seguintes.

5.6 Classificador

O classificador vai ser treinado com esta amostra. Ele vai representar o padrão da espécie. Se o classificador é para três espécies, suas conexões convergirão para três ramos finais, um para cada espécie. Cada ramo vai representar o padrão de uma espécie, com seus valores de parâmetro nas conexões. Também podemos dizer que cada ramo representa o ajuste de uma função, a de reconhecimento (ou discriminatória) de cada espécie.

Quando os dados de um espécime (na forma de um vetor) forem colocados na entrada do classificador, eles irão atravessar os ramos que já têm os valores de parâmetro nas conexões determinados na fase de treino (ou de ajuste), e vai sair (no final) com maior pertinência a um dos ramos, significando que foi ao padrão associado a este ramo que ele foi mais semelhante. Se se pegar uma amostra de um contexto (ou região geográfica) diferente, treinamos o classificador para este contexto (os parâmetros das funções de pertinência vão estar associados a este contexto).

Portanto, a falta de um experimento empírico subjetivo, a priori, para determinar qual a função de pertinência, fica justificada por supormos a princípio que pode ser qualquer uma, isto é, qualquer parâmetro, e que o que faz determinar os parâmetros é o treinamento do classificador com uma amostra já classificada para aquele contexto por um especialista. No caso o taxonomista Anderson [16] foi o especialista.

5.6.1 Definição de classificador

Existem muitas maneiras diferentes de representar classificadores de padrões. Uma forma canônica é em termos de um conjunto de *funções discriminantes* $g_i(\mathbf{x}), i = 1, \dots, c$. O classificador designa um vetor de atributos \mathbf{x} para a classe w_i se

$$g_i(\mathbf{x}) > g_j(\mathbf{x}) \quad \forall j \neq i \tag{5.20}$$

Um classificador é visto como uma máquina que calcula c funções discriminantes e seleciona a categoria correspondente ao maior discriminante (Duda e Hart [88]). A figura 5.5, mostra o percurso da informação em um classificador básico.

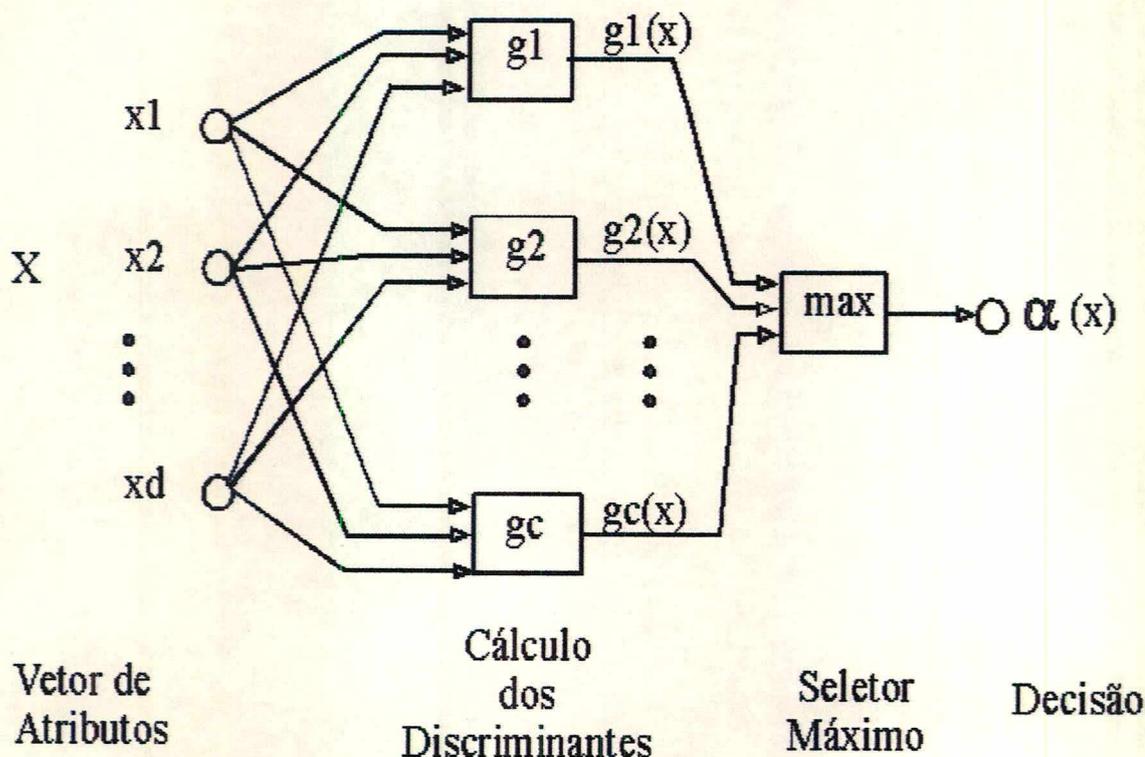


Figura 5.5: Classificador (básico) de padrões

A figura 5.6 ilustra o classificador que estamos desenvolvendo, com mais detalhes sobre as conexões.

A figura 5.7 ilustra a parte do classificador que reconhece a classe A.

5.6.2 O erro médio quadrático

Para cada flor observada, tem-se um vetor da forma $FLOR = (SL, SW, PL, PW)$, o qual já foi classificado por Anderson [16]. Com esta classificação, para cada flor tem-se o grau de pertinência desejado para cada classe A, B e C, denotado por TA, TB e TC , respectivamente. Por exemplo, quando a flor for da classe A, tem-se $TA = 1, TB = -1$ e $TC = -1$. A seguir realizam-se os seguintes passos para cada vetor:

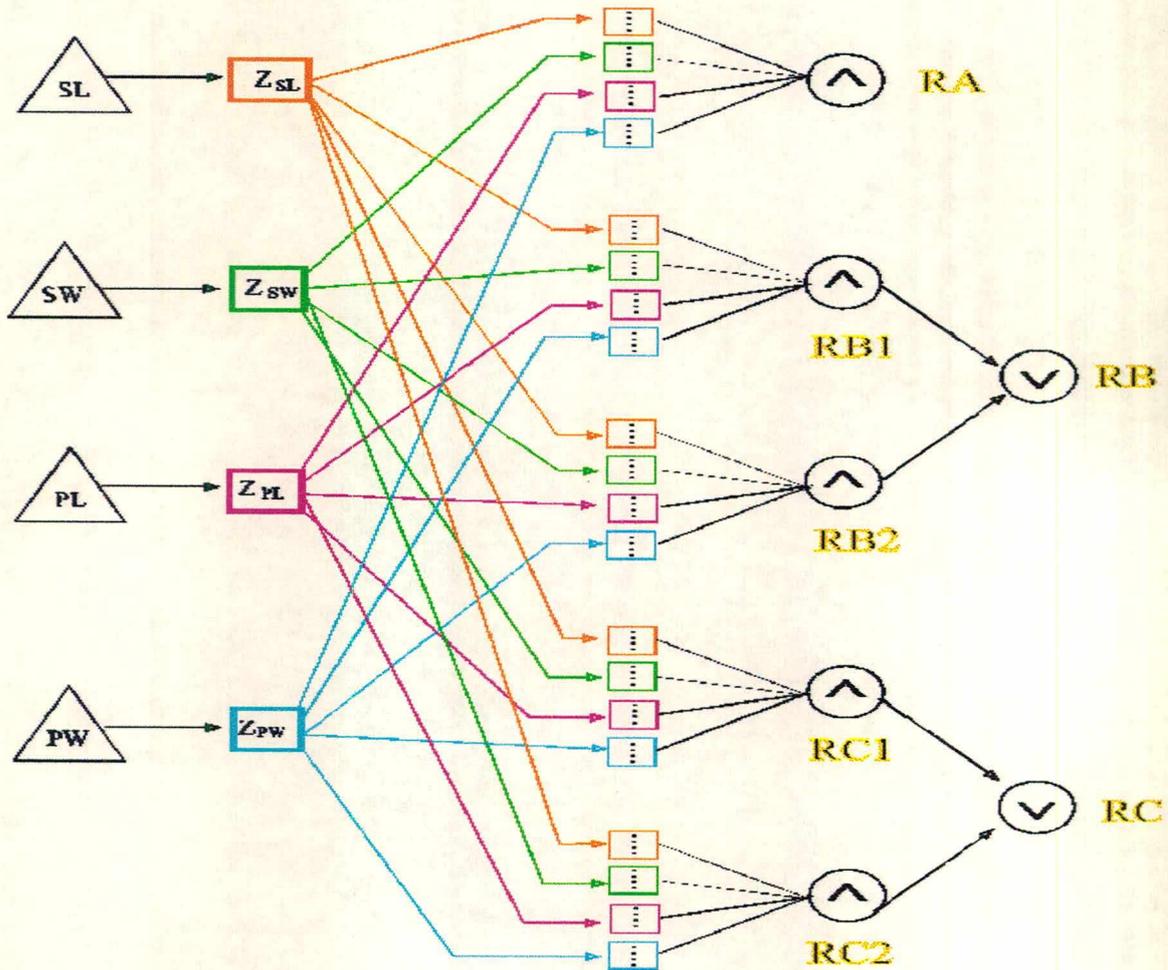


Figura 5.6: O classificador que propomos

1. Calcula-se o grau de pertinência de cada flor às classes A, B e C , utilizando-se as fórmulas para RA, RB e RC , respectivamente.
2. Calculam-se, para cada flor observada, as diferenças:

$$EAo(Wa, Qa) = \frac{1}{2}(RA - TA)^2 \quad (5.21)$$

$$EBo(Wb, Qb) = \frac{1}{2}(RB - TB)^2 \quad (5.22)$$

$$ECo(Wc, Qc) = \frac{1}{2}(RC - TC)^2 \quad (5.23)$$

em que Wa e Qa são vetores com 8 coordenadas cada um. E $Wb = [Wb1, Wb2]$, $Qb = [Qb1, Qb2]$, $Wc = [Wc1, Wc2]$ e $Qc = [Qc1, Qc2]$, sendo cada um com 16 coordenadas.

3. Calcula-se o erro em cada classe para cada vez que passa o conjunto de treino do classificador, ou seja, para cada época, sendo que o conjunto de treino tem N observações e cada época é enotada por n :

$$EA_n(Wa, Qa) = \left(\frac{1}{N}\right) \sum_1^N EAo \quad (5.24)$$

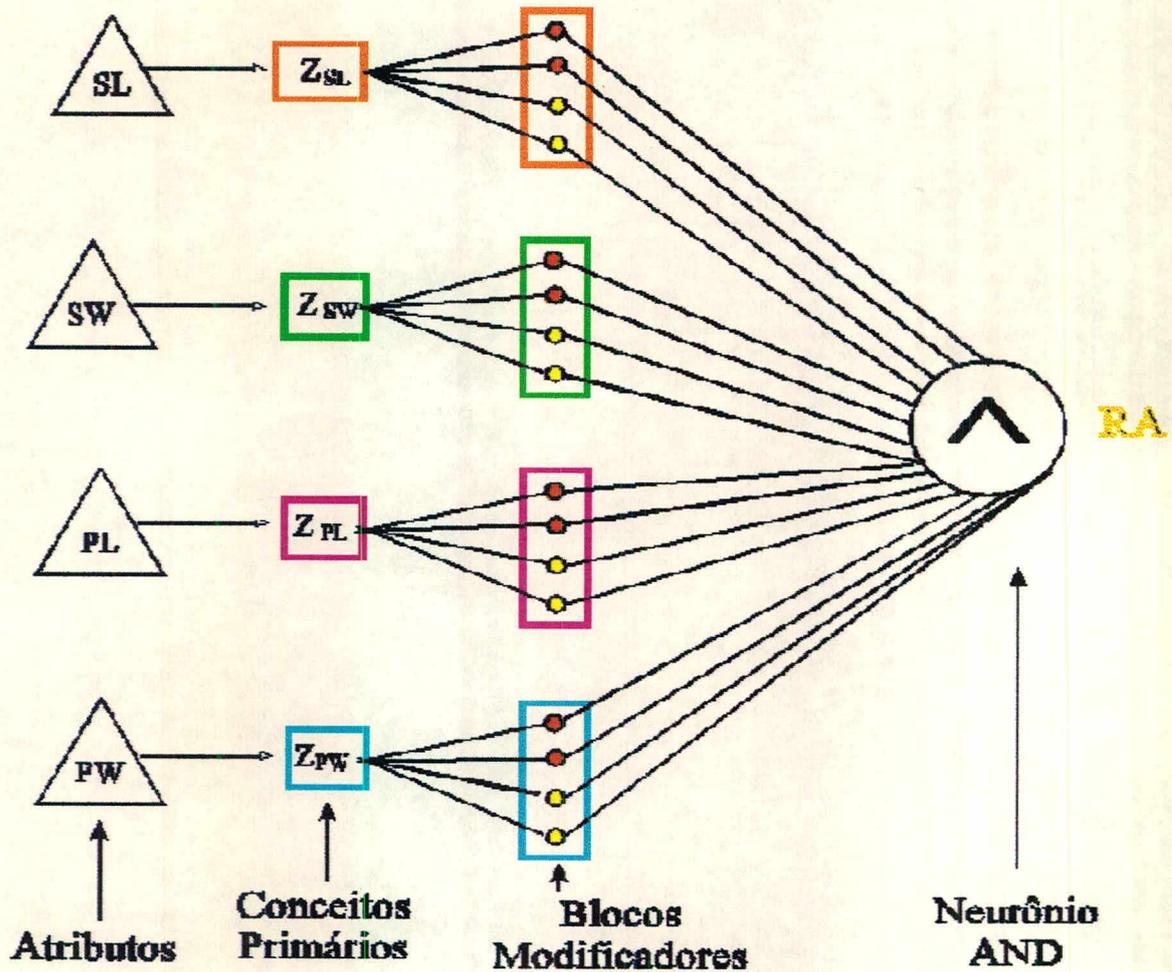


Figura 5.7: Parte que reconhece A

$$EB_n(Wb, Qb) = \left(\frac{1}{N}\right) \sum_1^N EBo \quad (5.25)$$

$$EC_n(Wc, Qc) = \left(\frac{1}{N}\right) \sum_1^N ECo \quad (5.26)$$

4. Calcula-se o erro médio ao final de cada época:

$$Emédio(W, Q) = EA_n + EB_n + EC_n \quad (5.27)$$

onde W e Q são vetores de parâmetros com 40 coordenadas cada um.

5.6.3 Algoritmo de minimização

O objetivo é minimizar o erro médio de classificação. Depois de uma determinada época, se ele for menor do que um *erro objetivo* então consideramos que o classificador está treinado e que já encontramos os parâmetros das funções caracterizadoras. Mas se isto não ocorrer, vamos utilizar o algoritmo da descida do gradiente (gradient descent) para atingir um mínimo. As variáveis independentes, ou parâmetros livres, a serem usadas na minimização são os q 's e os w 's.

Segundo este algoritmo, para encontrar o ponto de mínimo da função de erro, vamos caminhar na direção indicada pelas coordenadas do gradiente da função naquele ponto em que estamos. Se for na primeira época, isto é, $n = 1$, este ponto será o ponto inicial escolhido. Fazemos isto até atingir o erro objetivo, ou atingir um certo número de épocas especificado. Com este algoritmo pode ocorrer que não atingimos o ótimo global, mas sim um ótimo local. Apesar de haver inúmeros procedimentos para atingir o ótimo global de uma função, na tese optamos por ficar com os parâmetros que obtivemos depois de 3000 épocas.

Os Apêndices 2 e 3 mostram o código do algoritmo para o classificador mostrado na figura —, utilizando a linguagem do software MATLAB.

5.6.4 Fase de teste

Para o caso *iris* dividimos aleatoriamente o conjunto de dados em 70% para a fase de treino e 30% para a fase de teste, a qual não podia conter flores do conjunto de treino. Com as observações do conjunto de treinamento foram ajustados os parâmetros das funções de caracterização RA , RB e RC . Para verificar se estas funções caracterizavam bem os conceitos de cada espécie, ou classe, verificamos a taxa de erro de classificação, com base na classificação de Anderson. Com isto podemos avaliar a capacidade de generalização, ou de previsão, do classificador, ou das funções de caracterização.

Ao final da fase de teste vamos produzir a tabela 5.1.

Tabela 5.1: Resultados Obtidos por Nosso Classificador .

Classes	Corretos	Porcentagem
A	15	100
B	14	93,33
C	14	93,33

Se os resultados do teste são favoráveis, então ficamos com as funções de caracterização obtidas pela substituição dos parâmetros nas fórmulas para RA , RB e RC .

5.6.5 Funções de caracterização

Tendo-se ajustado os parâmetros e substituindo-os nas funções RA , RB e RC , conclui-se o processo de definição das funções de caracterização, ou de reconhecimento, das classes A , B e C . Estas funções podem ser usadas para calcular o grau de pertinência de uma flor a cada classe. A decisão quanto a qual classe a flor vai ser identificada depende de um critério estabelecido, que em nosso caso será pelo maior grau de pertinência. Contudo, pode não haver esta decisão *crisp*, se nos contentarmos com ficar cômicos de que a flor pode ter graus de pertinência diferentes para as três classe.

A função de caracterização da classe A é dada pela equação 5.17 e usando as equações do capítulo 4.

Observação: Os valores dos parâmetros que estão nas equações abaixo são obtidos pelo classificador depois de 3000 épocas. Valores dos parâmetros até 8000 épocas são mostrados no apêndice 4.

$$RA = Op3((0,001 \cdot Op3(Z_{SL}))^{1,25} + (3,22 \cdot Op3(Z_{SL}))^{-3,50} + (0,001 \cdot Op3(Z_{SW}))^{1,36} + (6,60 \cdot Op3(Z_{SW}))^{-4,54} + (1,79 \cdot Op3(Z_{PL}))^{-2,61} + (1,72 \cdot Op3(Z_{PL}))^{-2,67} + (1,33 \cdot Op3(Z_{PW}))^{-2,19} + (0,30 \cdot Op3(Z_{PW}))^{-3,94}) \quad (5.28)$$

A função de caracterização da classe B é dada por:

$$RB = Op1(Op2(RB1) + Op2(RB2)) \quad (5.29)$$

com

$$RB1 = Op3((0,001 \cdot Op3(Z_{SL}))^{1,26} + (3,89 \cdot Op3(Z_{SL}))^{-3,54} + (0,001 \cdot Op3(Z_{SW}))^{1,13} + (3,21 \cdot Op3(Z_{SW}))^{-3,29} + (0,0001 \cdot Op3(Z_{PL}))^{1,34} + (2,44 \cdot Op3(Z_{PL}))^{-3,77} + (0,31 \cdot Op3(Z_{PW}))^{3,65} + (1,75 \cdot Op3(Z_{PW}))^{-7,66}) \quad (5.30)$$

e

$$RB2 = Op3((0,001 \cdot Op3(Z_{SL}))^{1,25} + (3,89 \cdot Op3(Z_{SL}))^{-3,54} + (0,001 \cdot Op3(Z_{SW}))^{1,13} + (3,21 \cdot Op3(Z_{SW}))^{-3,29} + (0,0001 \cdot Op3(Z_{PL}))^{1,34} + (2,45 \cdot Op3(Z_{PL}))^{-3,77} + (0,3130 \cdot Op3(Z_{PW}))^{3,65} + (1,75 \cdot Op3(Z_{PW}))^{-7,66}) \quad (5.31)$$

E a função de caracterização da classe C é dada por:

$$RC = Op1(Op2(RC1) + Op2(RC2)) \quad (5.32)$$

com

$$RC1 = Op3((0,0004 \cdot Op3(Z_{SL}))^{1,44} + (5,4966 \cdot Op3(Z_{SL}))^{-3,46} + (0,001 \cdot Op3(Z_{SW}))^{1,22} + (4,0884 \cdot Op3(Z_{SW}))^{-3,31} + (0,000 \cdot Op3(Z_{PL}))^{1,78} + (0,0003 \cdot Op3(Z_{PL}))^{-1,68} + (2,0181 \cdot Op3(Z_{PW}))^{7,29} + (0,0309 \cdot Op3(Z_{PW}))^{-2,31}) \quad (5.33)$$

e

$$RC2 = Op3(((0,0004 \cdot Op3(Z_{SL}))^{1,44} + (5,49 \cdot Op3(Z_{SL}))^{-3,47} + (0,001 \cdot Op3(Z_{SW}))^{1,22} + (4,09 \cdot Op3(Z_{SW}))^{-3,32} + (0,000 \cdot Op3(Z_{PL}))^{1,78} + (0,0003 \cdot Op3(Z_{PL}))^{-1,68} + (2,0181 \cdot Op3(Z_{PW}))^{7,29} + (0,0309 \cdot Op3(Z_{PW}))^{2,31}) \quad (5.34)$$

Com estas funções, pode-se calcular o grau de pertinência de cada flor observada, a cada classe. Estas também são chamadas funções discriminantes.

5.6.6 Alguns gráficos

As figuras a seguir mostram os gráficos do erros obtidos ao final de 3000 épocas de treinamento com taxa de aprendizado igual a 0.5.

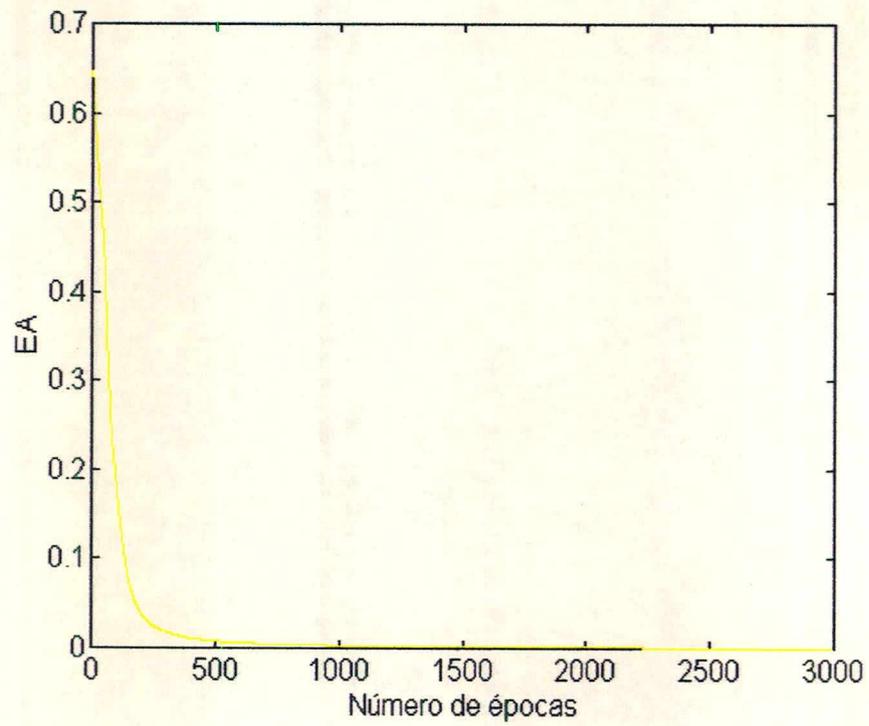


Figura 5.8: Gráfico de EA com txa = 0.5

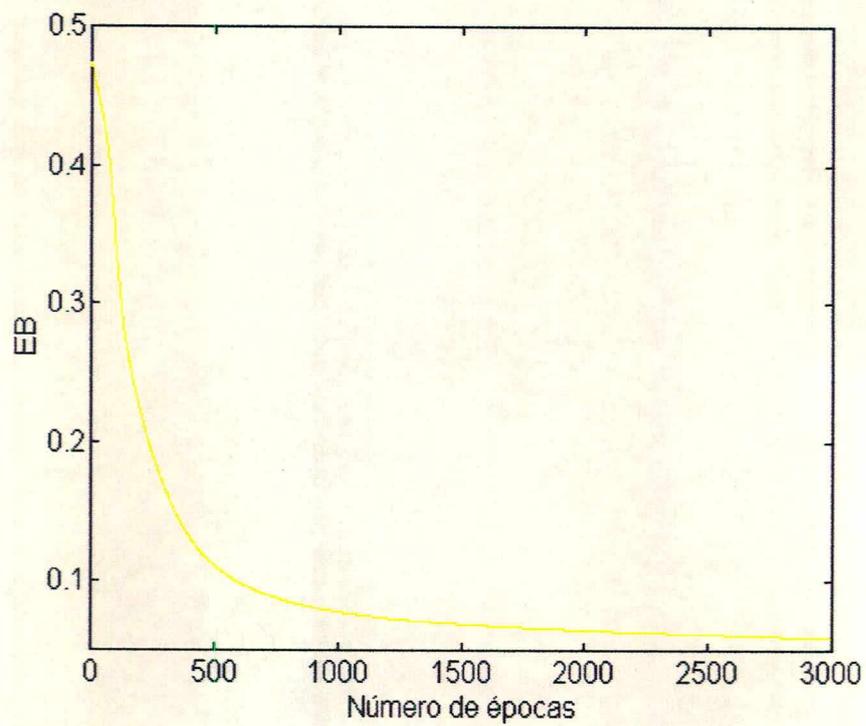


Figura 5.9: Gráfico de EB

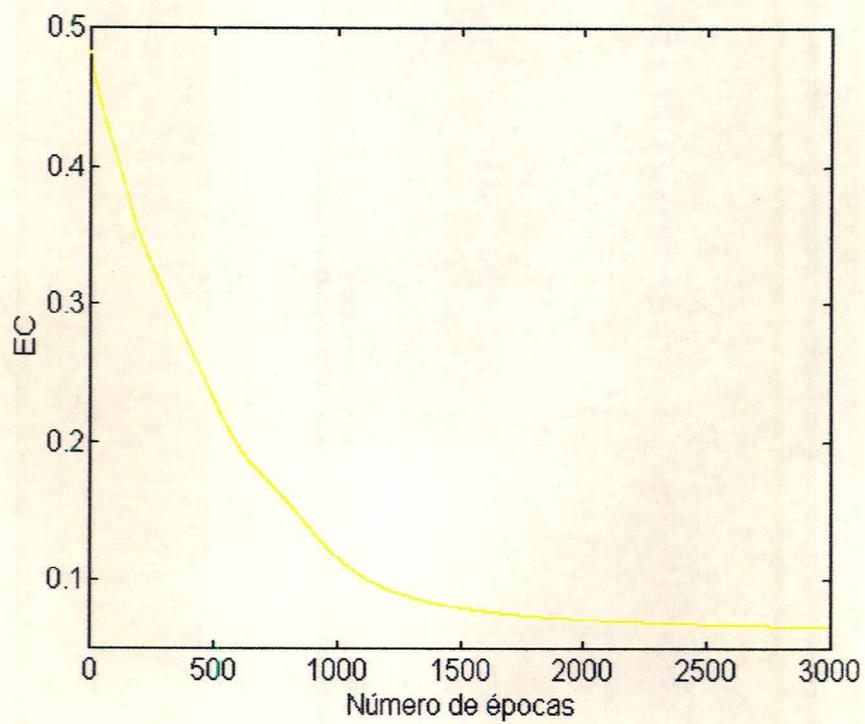


Figura 5.10: Gráfico de EC

A figura 5.11 ilustra o gráfico da função EA dada pela equação de EA , onde deixamos q_{i1} e w_{i1} variando.

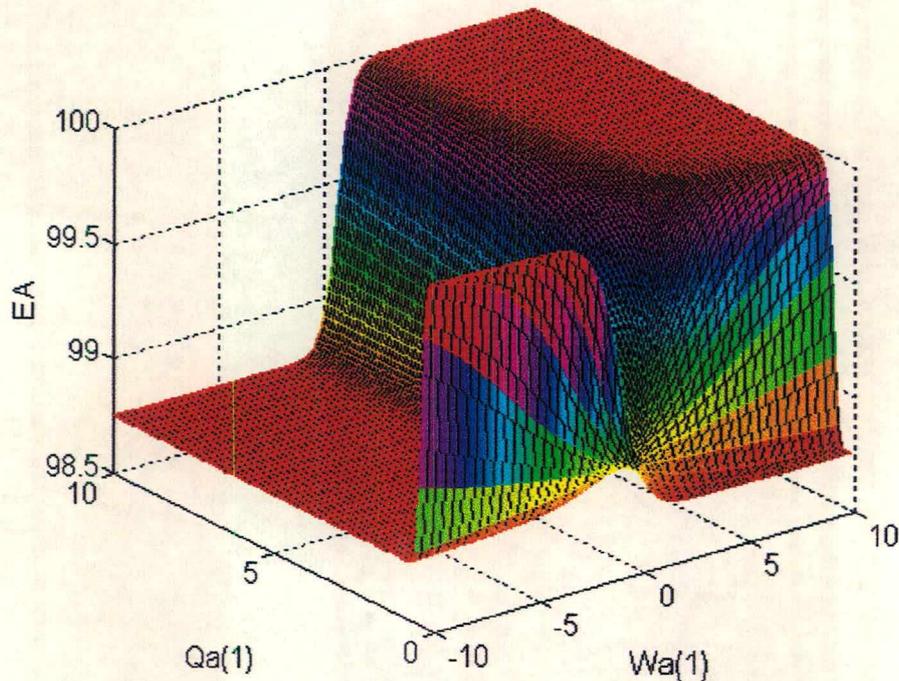


Figura 5.11: Gráfico de EA com dois parâmetros variando

5.6.7 Conclusões das subseções acima

No pior caso, o resultado que obtivemos foi de 93,33% e no melhor 100%. Com a relação de resultados de classificadores publicados por Woods [34], para o conjunto de dados íris e outros, ficamos acima de dois deles. Como os outros três que deram resultados melhores do que o nosso, também tiveram resultados bem inferiores nos outros conjuntos de dados, acreditamos que devemos aplicar nosso modelo nos conjuntos de dados do artigo citado, para então tirarmos uma conclusão final quanto à comparação com outros classificadores.

5.7 Aproximação de funções

Finalmente, uma propriedade que consideramos extremamente relevante para a nossa escolha das funções de pertinência e operadores está contida na conjectura que enunciaremos a seguir. Em termos simples quer dizer que: Dada $\varphi: (0, +\infty) \rightarrow [-1, 1]$, $\varphi \in C^0$ tal que

$$\lim_{x \rightarrow +\infty} \varphi(x) = \pm 1 \quad (5.35)$$

e

$$\lim_{x \rightarrow 0^+} \varphi(x) = \pm 1 \quad (5.36)$$

φ pode ser aproximada por uma função obtida a partir das funções da família $\mu_{p,k}$ com um número finito de operações de conjunção e disjunção. Este resultado mostra que podemos *gerar* qualquer função de pertinência satisfazendo (5.35) e (5.36) e contínua, a partir dos elementos de nossa família mais a ação dos operadores conjuntivo e disjuntivo. Para enunciá-lo, precisamos fixar uma notação. Sejam

$$\mathcal{P} = \left\{ \varphi: (0, +\infty) \rightarrow [-1, 1], C^0 \mid \lim_{x \rightarrow +\infty} \varphi(x) = \pm 1 \text{ e } \lim_{x \rightarrow 0^+} \varphi(x) = \pm 1 \right\}$$

e

$$\mathcal{P}_o = \{ \mu_{p,k}; p > 0, k \neq 0 \}.$$

Definimos indutivamente a família de subconjuntos de \mathcal{P} :

$$\mathcal{P}_1^d = \{ \mathcal{D}(\varphi_1, \dots, \varphi_r); \varphi_1, \dots, \varphi_r \in \mathcal{P}_o, r > 0 \}$$

$$\mathcal{P}_1^c = \{ \mathcal{C}(\varphi_1, \dots, \varphi_r); \varphi_1, \dots, \varphi_r \in \mathcal{P}_o, r > 0 \}$$

Suponhamos definidos $\mathcal{P}_k^d, \mathcal{P}_k^c$. Então definimos

$$\mathcal{P}_{k+1}^d = \{ \mathcal{D}(\varphi_1, \dots, \varphi_r); \varphi_1, \dots, \varphi_r \in \mathcal{P}_k^c \}$$

$$\mathcal{P}_{k+1}^c = \{ \mathcal{C}(\varphi_1, \dots, \varphi_r); \varphi_1, \dots, \varphi_r \in \mathcal{P}_k^d \}$$

Conjetura 5.1 *Seja*

$$\Pi = \bigcup_{k=1}^{\infty} (\mathcal{P}_k^d \cup \mathcal{P}_k^c) \cup \mathcal{P}_o \quad (5.37)$$

então

$$\overline{\Pi} = \mathcal{P} \quad (5.38)$$

em que o fecho é tomado relativamente à topologia do supremo e \mathcal{P} foi definido no capítulo 4.

Capítulo 6

APLICAÇÃO: Rede Neural

6.1 Introdução

Redes neurais classificadoras, que utilizam neurônios AND e OR, têm sido desenvolvidas para efetuar operações de conjunção e disjunção sobre os graus de pertinência a conceitos difusos que entram na rede. Em particular, os trabalhos que têm por objetivo a classificação e geração de regras com base em Sistemas Especialistas Conexionistas Difusos, como os desenvolvidos em Brasil [29], Machado [5], Mitra [104], Fu [77] e Gupta [42], utilizam as t-normas e t-conormas do MIN e do MAX, respectivamente, propostas em Zadeh [65], ou os operadores produto algébrico e soma probabilística.

Neste capítulo apresentamos a base para uma versão difusa do Perceptron de Múltiplas Camadas (PMC), composto por neurônios lógicos que efetuam operações de conjunção (AND) e disjunção (OR) dadas pelos operadores descritos nos capítulos anteriores. Estes operadores vão processar a imprecisão e/ou incerteza na entrada da rede. São apresentadas as derivadas da função de erro, usadas para a atualização dos parâmetros livres da rede com base na implementação do gradiente descendente sobre a função de erro. Para o treinamento da rede é possível implementar o algoritmo de retropropagação (*back-propagation*) de erros (Zanusso [72]).

6.2 A Versão Difusa do PMC

Considere a rede de três camadas da figura 6.1. No lugar de neurônios simples, que implementam a soma ponderada e de funções sigmóides, como desenvolvido em um dos artigos de Rumelhart [33] sobre o aprendizado de representações internas à rede, por retropropagação do erro, utilizando a regra delta generalizada, o modelo aqui proposto utiliza neurônios AND(\wedge) e OR(\vee), que empregam pares conjugados de t-norma e t-conorma.

As entradas da rede são graus de pertinência a conceitos difusos correspondentes, a cada atributo observado no objeto a ser classificado. Elas são denotadas por O_{pl} , em que p e l se referem ao p -ésimo padrão de treino e à l -ésima unidade de entrada, respectivamente. Consideramos $p = 1, \dots, N$, em que N é o número de padrões de treino e $l = 1, \dots, n_1$, com n_1 , o número de unidades na camada de entrada que é igual ao número de atributos observados. A saída das unidades de

entrada são os próprios O_{pl} , observando-se que o peso da conexão de uma unidade l desta camada com uma unidade i da camada intermediária é denotada por w_{il} .

A camada intermediária é formada por n_2 neurônios AND. A saída das unidades desta camada é denotada por O_{pi} , com $i = 1, \dots, n_2$. Para facilitar a comparação deste trabalho com a dedução da regra delta generalizada dada em Rumelhart [33], pode-se considerar a função de transferência como sendo igual à identidade I . Assim

$$O_{pi} = I(\text{net}_{pi}) \quad (6.1)$$

$$\text{net}_{pi} = \bigwedge_{l=1}^{n_1} (O_{pl} \vee w_{il}) \quad (6.2)$$

A camada de saída é formada por n_3 neurônios OR, sendo este número igual ao número de classes em que os exemplos de treino são classificados. A saída destas unidades é denotada por O_{pj} , com $j = 1, \dots, n_3$ e representa valores de pertinência dos padrões às classes. Para facilitar a comparação com Rumelhart [33] escrevemos:

$$O_{pj} = I(\text{net}_{pj}) \quad (6.3)$$

$$\text{net}_{pj} = \bigvee_{i=1}^{n_2} (O_{pi} \wedge w_{ji}) \quad (6.4)$$

6.3 T-Norma e T-Conorma

As t-normas e t-conormas que foram propostas no capítulo 4 são utilizadas aqui. Empregam-se nas equações (6.2) e (6.4), $\wedge = U_c$ e $\vee = U_d$, donde se obtém

$$O_{pi} = U_{c_{l=1}}^{n_1} U_d(O_{pl}, w_{il}) \quad (6.5)$$

$$O_{pi} = U_c(U_d(O_{p1}, w_{i1}), \dots, U_d(O_{pl}, w_{il}), \dots, U_d(O_{pn_1}, w_{in_1})), \quad (6.6)$$

com $i = 1, \dots, n_2$. e

$$O_{pj} = U_{c_{i=1}}^{n_2} U_d(O_{pi}, w_{ji}) \quad (6.7)$$

$$O_{pj} = U_d(U_c(O_{p1}, w_{j1}), \dots, U_c(O_{pi}, w_{ji}), \dots, U_c(O_{pn_2}, w_{jn_2})) \quad (6.8)$$

com $j = 1, \dots, n_3$.

6.4 A medida de erro

O objetivo do algoritmo de treinamento é minimizar a distância entre o valor de pertinência de um padrão à classe j , calculado pela rede, denotado por O_{pj} , e o desejado T_{pj} , que para cada exemplo de treino é dado pela equação

$$E_p = \frac{1}{2} \sum_{j=1}^{n_3} (T_{pj} - O_{pj})^2 \quad (6.9)$$

6.4.1 Derivada das t-normas

Para encontrar as derivadas do erro são necessárias as derivadas das t-norma e t-conorma em um caso particular. Para a uma constante e $\varphi(x)$ uma função real qualquer

$$\frac{dU_d(\varphi(x), a)}{dx} = \varphi'(x) \frac{(4a^2 - 8a + 4)}{[(3 - a) - \varphi(x)(1 + a)]^2} \quad (6.10)$$

$$\frac{dU_c(\varphi(x), a)}{dx} = \varphi'(x) \frac{(4a^2 + 8a + 4)}{[(3 + a) + \varphi(x)(1 - a)]^2} \quad (6.11)$$

6.4.2 Derivadas do erro

A função de erro médio da rede é

$$E = \sum E_p \quad (6.12)$$

que depende de todos os w_{ji} 's e w_{il} 's.

Para minimizá-la usa-se a regra delta generalizada [33], que implementa a heurística da descida do gradiente sobre E . Portanto, é preciso encontrar

$$\frac{\partial E_p}{\partial w_{ji}} \quad (6.13)$$

$$\frac{\partial E_p}{\partial w_{il}} \quad (6.14)$$

Através de derivações sucessivas, chega-se aos seguintes resultados:

$$\frac{\partial E_p}{\partial w_{ji}} = (T_{pj} - O_{pj}) \frac{[4a_1^2 - 8a_1 + 4]}{[(3 - a_1) - U_c(O_{pi}, w_{ji})(1 + a_1)]^2} \frac{(4O_{pi}^2 + 8O_{pi} + 4)}{[(3 + O_{pi}) + w_{ji}(1 - O_{pi})]^2} \quad (6.15)$$

$$\frac{\partial E_p}{\partial w_{il}} = \frac{[4O_{pl}^2 - 8O_{pl} + 4]^2}{[(3 - O_{pl}) - w_{il}(1 + O_{pl})]^4} \frac{(4a_2^2 + 8a_2 + 4)}{[(3 + a_2) + U_d(O_{pl}, w_{il})(1 - a_2)]^2} \sum_{j=1}^{n_3} (T_{pj} - O_{pj}) \frac{[4a_1^2 - 8a_1 + 4]}{[(3 - a_1) - U_c(O_{pi}, w_{ji})(1 + a_1)]^2} \frac{[4w_{ji}^2 - 8w_{ji} + 4]}{[(3 - w_{ji}) - O_{pi}(1 + w_{ji})]^2}$$

sendo

$$a_1 = U_{d_{i' \neq i}^{n_2}} U_c(O_{pi'}, w_{ji'}) \quad (6.16)$$

$$a_2 = U_{c_{l' \neq l}^{n_1}} U_d(O_{pl'}, w_{il'}) \quad (6.17)$$

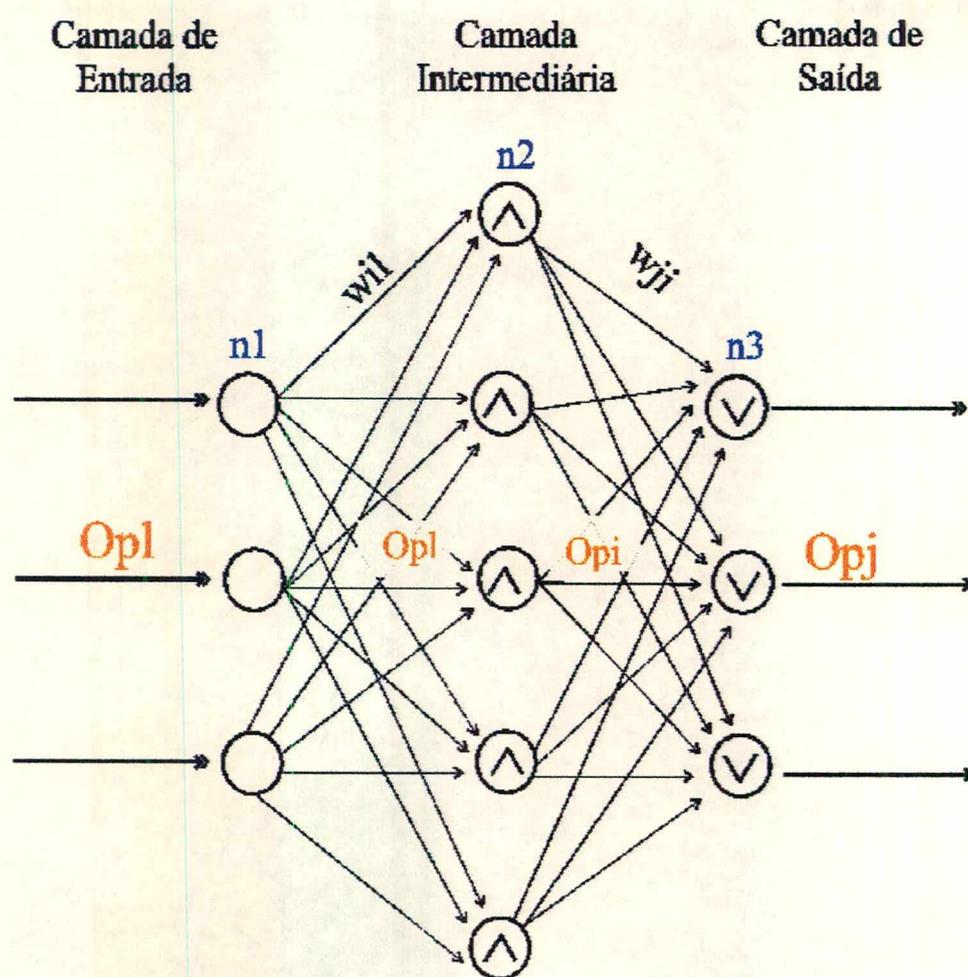


Figura 6.1: Rede neural AND e OR para a classificação e geração de regras

Capítulo 7

Conclusões

Temos em mãos uma família de t-normas e t-conormas diferenciáveis e estritamente relacionadas com uma família particular de funções de pertinência. A qualidade de diferenciabilidade insere grande vantagem em relação a estas t-normas, visto que as principais t-normas que se usam no dia-a-dia são dependentes do máximo e do mínimo, o que dificulta a dedução da regra delta para treinamento de redes neurais, e também o estudo da convergência da rede e de outros algoritmos.

Dentre todas as funções de pertinência possíveis, consideramos que a família proposta na tese tem uma base experimental pelo fato de ser uma curva em S e ter muita semelhança com as propostas pelos pesquisadores que realizaram experimentos empíricos. A função é fácil de calcular e parece ajustar-se bem a diversos contextos. Os parâmetros que a definem têm influência direta sobre a transição entre pertencer e não pertencer ao conceito considerado e sobre a própria região da transição. Além disso tudo, ela tem uma estreita relação com as t-normas e t-conormas, o que parece garantir que a informação contida nas entradas (*inputs*) de um sistema que usa estas t-normas seja mantida nas saídas (*outputs*).

O Modelo de Classificador que desenvolvemos surge de uma maneira natural, podemos dizer, pois parece ser a sistematização de um processo de abstração para a formação de conceitos, que ocorrem no dia-a-dia de todos. Ou seja, observamos todas as flores à nossa volta, de todas as formas, cores e jardins. Abstraímos de suas diferenças e semelhanças as classes de equivalência, sem *buracos* em seu interior, que podem ser representadas pela conjunção convexa de conceitos associados aos atributos observados. Por exemplo, as classes de equivalência das margaridas, das dalias e dos girassóis. Unidas, formam a família das *asteraceae*, que pode ser representada pela disjunção, que é não convexa, significando que há *buracos* entre margaridas, dalias e girassóis. A extensão destes *buracos* poderia ser definida pelo estudo da separação de conjuntos difusos convexos cujos primeiros conceitos foram introduzidos por Zadeh.

Na tese foi considerada a representação canônica, em que os valores dos atributos entram na forma de um vetor de características; este, por sua vez, passa através de um conjunto de módulos que calculam os valores de funções discriminantes, chamadas caracterizadoras, os quais, depois, entram num módulo seletor do valor máximo que produz a decisão. Observamos que este módulo seletor não foi representado nas figuras do capítulo 5, exceto no modelo básico.

Treinar o classificador para produzir funções caracterizadoras envolve a minimização de uma função de erro, que, no caso, consistia de uma função não

linear a ser minimizada numa determinada região. Existem inúmeros programas que podem dar a solução quase exata para estes problemas. Mas sua precisão é desnecessária para aplicações difusas, que necessitam apenas de soluções difusas. Em nosso trabalho faltou estudar a minimização na região especificada pelos *conjuntos referenciais* de cada atributo.

Além do classificador descrito, desenvolvemos as derivadas de uma rede neural, com neurônios *and* e *or*, que podem fazer classificação e geração de regras para um sistema especialista híbrido, com a diferença de que as t-normas e t-conormas utilizadas são as que propomos na tese. Na verdade, na tese, sugerimos dois modelos de classificador.

No contexto dos classificadores que propomos, o que de fato acontece em seu interior é a aproximação de uma função $f: A \subset R^n \rightarrow R^m$, através de um treinamento que utiliza exemplos $(x_1, y_1), (x_2, y_2), \dots, (x_k, y_k), \dots$ onde $y_k = f(x_k)$. É suposto que tais exemplos são gerados selecionando-se os vetores x_k aleatoriamente de A , de acordo com alguma distribuição de probabilidade. Neste sentido, qual seria a função associada aos nossos classificadores? Quanto ao do capítulo 5, na tese lançamos uma conjectura que pode chegar a ser demonstrada.

A reflexão sobre um sistema lógico difuso para Taxionomia ficou em aberto, requerendo mais estudos sobre técnicas de agrupamento e hierarquização que são fundamentais para a formação de taxionomias.

Capítulo 8

Sugestões

Quanto às t-normas:

1. Para reconhecer experimental e teoricamente as vantagens de nossas t-normas seria preciso aplicá-las a vários contextos, como finanças, transportes, diagnósticos, taxionomias, etc.
2. Fazer o gráfico das t-normas e t-conormas para vários valores de parâmetros e dar sua interpretação.
3. Desenvolver estudos no sentido da criação de uma regra *modus ponens* para desenvolver uma rede neural que faça inferência.
4. Estudar a representatividade das famílias de t-normas e funções de pertinência como modelos de conceitos.
5. Transformar a conjectura em teorema demonstrado, ou encontrar um contra-exemplo.
6. Fazer estudo sobre a escalas e informações que são mantidas ao aplicar os operadores sobre as medidas dos atributos.

Quanto ao classificador:

Comparar os resultados deste sistema classificador com outros da literatura, em vários conjuntos de dados, como, por exemplo, o método do vizinho mais próximo, redes neurais, árvores de decisão, bayes quadrático e bayes linear.

Quanto à Rede Neural:

1. Seria necessário treinar a rede em vários conjuntos de dados e verificar sua capacidade de classificação.
2. Também pode ser feito um estudo da convergência da rede com a função de erro médio dada em termos de nosso *and* e *or*.
3. *A implementação em Hardware:* Espera-se que sejam desenvolvidos circuitos analógicos capazes de realizar as funções conceituais. A disponibilidade desses circuitos analógicos, aliada a técnicas de matrizes lógicas programáveis, pode levar à construção de sistemas eficazes que aproveitem efetivamente o paralelismo verdadeiro inerente aos dispositivos analógicos.

Referências Bibliográficas

- [1] HERSH, H.M. , CARAMAZZA, A. A fuzzy set approach to modifiers and vagueness in natural language. *J. Exper. Psychology: General*, v. 105, n. 1, p. 254–276, 1976.
- [2] SCHWEIZER, B. , SKLAR, A. Associative functions and statistical inequalities. *Publ. Math. Debrecen*, v. 8, n. 1, p. 169–186, 1961.
- [3] JENSEN, A.A. *Sistemas indígenas de classificação de aves: aspectos comparativos, ecológicos e evolutivos*. Museu Paraense Emílio Goeldi, Belém- Pará, 1988.
- [4] GATH, I. , GEVA, A.B. Unsupervised optimal fuzzy clustering. *IEEE Trans. Pattern Anal. Machine Intelligence*, v. PAMI-11, n. 7, p. 773–781, 1989.
- [5] MACHADO, R.J., ROCHA, A.F. The combinatorial network: a connectionist model for knowledge based systems. In: *INTERNATIONAL CONFERENCE ON INFORMATION PROCESSING AND MANAGEMENT OF UNCERTAINTY IN KNOWLEDGE-BASED SYSTEMS, 3.*, p. 578–587, Paris, France, 1990.
- [6] PEDRYCZ, W. , ROCHA, A.F. Fuzzy-set based models of neurons and knowledge-based networks. *IEEE Trans. On Fuzzy Systems*, v. 1, n. 4, p. 254–266, 1993.
- [7] KANDEL, A., et al. Foundations of rule-based computations in fuzzy models. In: *Fuzzy Modelling: Paradigms and Practice*. Kluwer Academic, Netherlands, 1996.
- [8] RODRIGUES, A.M. *Artificial intelligence techniques for modeling financial analysis*. Tese de Doutorado, UFSC-EPS, Florianópolis, 1996.
- [9] KAMP, H., PARTEE, B. Prototype theory and compositionality. *Cognition*, v. 57, n. 1, p. 129–191, 1995.
- [10] SIY, P. , CHEN, C.S. Fuzzy logic for handwritten numerical character recognition. *IEEE, Trans. Syst. Man*, v. SMC-4, n. 1, p. 570–575, 1974.
- [11] TEODOROVIĆ, D. Fuzzy sets theory applications in traffic and transportation. *European J. Operational Research*, v. 74, n. 3, p. 379–390, 1994.
- [12] PAL, S.K. , MAJUMDER, D.D. Fuzzy sets and decisionmaking approaches in vowel and speaker recognition. *IEEE Transactions on Systems, Man, and Cybernetics*, v. 7, n. 1, p. 625–629, August 1977.
- [13] BALL, G.H. , HALL, D.J. A clustering technique for summarizing multivariate data. *Behavioral Science*, v. 12, n. 1, p. 153–155, 1967.

- [14] KELLER, J.M., HUNT, D.J. Incorporating fuzzy membership functions into the perceptron algorithm. *IEEE Trans. Pattern Anal. Machine Intelligence*, v. PAMI-7, n. 6, p. 693–699, november 1985.
- [15] MEDIN, D.L. and SMITH, E.E. Concepts and concept formation. *Annual Review of Psychology*, v. 35, n. 1, p. 113–138, 1984.
- [16] ANDERSON, E. The irises of the gaspe peninsula. *Bull. Amer. Iris Soc.*, v. 59, n., p. 2–5, 1935.
- [17] ANDERSON, E. The species problem in iris. *Ann. of the Missouri Botanical Garden*, v. 23, n. 1, p. 457–509, 1936.
- [18] HISDAL, E. Infinite-valued logic based on two-valued logic and probability, part 1.2. different sources of fuzzines. *Internat. J. Man-Machine Stud.*, v. 25, n. 1, p. 113–138, 1986.
- [19] HISDAL, E. A flexible classification structure. In: *Fuzzy logic in knowledge-based systems, decision and control*, p. 11–67. North-Holand, New York, 1988.
- [20] LEACH, E. Anthropological aspects of language: animal categories and verbal abuse. In: *New directions in the study of language*, p. 23–63. Mit-Press, Cambridge, Massachussets, 1964.
- [21] MERVIS, C.B. , ROSCH, E. Categorization of natural objects. *Annual Review of Psychology*, v. 32, n., p. 89–115, 1981.
- [22] ROSCH, E. On the internal structure of perceptual and semantic categories. In: *Cognitive development and the acquisition of language*. Academic Press, New York, 1973.
- [23] OSHERSON, D.N. , SMITH, E.E. On the adequacy of prototype theory as a theory of concepts. *Cognition*, v. 9, n. 1, p. 35–58, 1981.
- [24] ASSILIAN, S. MAMDANI, E.H. *Artificial intelligence in the control of real dynamic systems*. Dep. Elec. Eng. Queen Mary College, Londres-Aug., 1974.
- [25] KING, P.J. , MAMDANI, E.H. The application of fuzzy control systems to industrial processes. In: *TRIENNIAL IFAC WORLD CONGRES*, 6,, v. 6, p. 24–30, 1975.
- [26] LAPOLLI, E.M. *Processamento de imagens digitais utilizando conjuntos difusos*. Tese de Doutorado, UFSC-EPS, Florianópolis, Brasil, 1994.
- [27] AZEVEDO, F.M. , et al. Teaching medical diagnosis with associative memories, (abstract). In: *SYSTEMS SCIENCE INTERNATIONAL CONFERENCE*, 11, p. 28–29, Wroclaw, Poland, 1992.
- [28] BARRETO, J.M. , et al. A neural network with fuzzy elements for a medical diagnosis. In: *LATICON'92, IEEE LATINOAMERICAN CONFERENCE*, p. 335–339, Santiago, Chile, 1992.
- [29] BRASIL, L.M. , et al. A methodology for implementing hybrid expert systems. In: *MEDITERRANEAN ELECTROTECHNICAL CONFERENCE*, v. 2, p. 661–664, Bary, Italy, May 1996.

- [30] CASTILLO, P., et al. *The plant parasitic nematode-rotylechus*. C.S.I.C., Córdoba-Spain, 1993.
- [31] DREYFUSS-RAIMI, G., et al. On the psycho-linguistic reality of fuzzy sets: effect of context and set. In: *CHICAGO LINGUISTIC CIRCLE*, Eds. Chicago: Univ. Chicago, 1975.
- [32] MICHALSKI, R.S., et al. A recent advance in data analysis: clustering objects into classes characterized by conjunctive concepts. In: *Progress in pattern recognition*, p. 33–56. North-Holland, Amsterdam, 1981.
- [33] RUMELHART, D.E., HINTON, G.E., et al. Learning internal representation by error propagation. In: *Parallel distributed processing*, v. 1, p. 319–362. MIT Press, Cambridge, Massachusetts, 1987.
- [34] WOODS, K., et al. Combination of multiple classifiers using local accuracy estimates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 19, n. 4, p. 405–410, 1997.
- [35] SOKAL, R.R., ROHLF, F.J. The comparison of dendograms by objective methods. *Taxon*, v. 11, n. 1, p. 33–40, 1962.
- [36] KLIR, G. *Fuzzy sets and fuzzy logics*. Prentice Hall, New Jersey, 1995.
- [37] LAKOFF, G. Hedges: a study in meaning criteria and the logic of fuzzy concepts. *Journal of Philosophical Logic*, v. 2, n. 1, p. 458–508, 1973.
- [38] SIMPSON, G.G. *Principles of animal taxonomy*. Columbia University Press, New York, 1961.
- [39] COHEN B., MURPHY, G.L. Models of concepts. *Cognitive Science*, v. 8, n., p. 27–58, 1984.
- [40] DUBOIS, D., PRADE, H. *Fuzzy sets and systems-theory and applications*. Academic Press, New York-NY, 1980.
- [41] DUBOIS, D., PRADE, H. A review of fuzzy set aggregation connectives. *Information Science*, v. 36, n., p. 85–121, 1985.
- [42] GUPTA, M.M., RAO, D. H. On the principles of fuzzy neural networks. *Fuzzy Sets and Systems*, v. 61, n. 1, p. 1–18, 1994.
- [43] KELLER, J.M., TAHANI, H. Implementation of conjunctive and disjunctive fuzzy logic rules with neural networks. *Int. Journal of Approximate Reasoning*, v. 6, n., p. 221–240, 1992.
- [44] KICKERT, W.J.M., KOPPELAAR, H. Application of fuzzy set theory to syntactic pattern recognition of handwritten capitals. *IEEE Trans. Syst. Man. Cybern.*, v. 6, n., p. 148–151, 1976.
- [45] LING, C. H. Representation of associative functions. *Publ. Math. Debrecen*, v. 12, n., p. 189–212, 1965.
- [46] ZIMMERMANN, H.J. Quantifying vagueness in decision models. *European Journal of Operational Research*, v. 22, n., p. 148–158, 1985.

- [47] ZIMMERMANN, H.J. *Fuzzy sets, decision making, and expert systems*. Kluwer Academic Publishers, Norwell, MA, 1987.
- [48] JOSEPH, H.W.B. *An introduction to logic*. Oxford, 1916.
- [49] NORWICH, A.M. , TURKSEN, I.B. A model for the measurement of membership and the consequences of its empirical implementation. *Fuzzy Sets and Systems*, v. 12, n., p. 1–25, 1984.
- [50] TURKSEN, I.B. Measurement of membership functions and their acquisition. *Fuzzy Sets and Systems*, v. 40, n. 1, p. 5–38, 1991.
- [51] DOMBI, J. A general class of fuzzy operators, the morgan class of fuzzy operators and fuzziness measures induced by fuzzy operators. *Fuzzy Sets and Systems*, v. 8, n., p. 149–163, 1982.
- [52] DOMBI, J. Membership functions as an evaluation. *Fuzzy Sets and Systems*, v. 35, n., p. 1–21, 1990.
- [53] FRIEDMAN, H.P., RUBIN, J. On some invariant criteria for grouping data. *JASA*, v. 62, n., p. 1159–1178, 1967.
- [54] PIAGET, J. *A Epistemologia Genética*. Victor Civita, São Paulo, 1983.
- [55] WOLFE, J. Pattern clustering by multivariate mixture analysis. *Mult. Behav. Research*, v. 5, n., p. 329–350, 1970.
- [56] GOGUEN, J.A. The logic of inexact concepts. *Syntese*, v. 19, n., p. 325–373, 1968.
- [57] BEZDEK, J.C. Numerical taxonomy with fuzzy sets. *Mathematical Biology*, v. 1, n. 1, p. 57–71, 1974.
- [58] BEZDEK, J.C. Mathematical models for systematics and taxonomy. In: *INTERNATIONAL CONFERENCE ON NUMERICAL TAXONOMY*, 8, p. 143–166, West Yorkshire, UK, 1975.
- [59] BEZDEK, J.C. *Pattern recognition with fuzzy objective function algorithms*. Plenum Press, New York, 1987.
- [60] BALDWIN, J.F. Fuzzy sets and expert systems. *Information Sciences*, v. 36, n. 1 e 2, p. 123–156, 1985.
- [61] HIROTA, K. *Industrial applications of fuzzy technology*. Springer-Verlag, Tokyo, 1993.
- [62] MENGER, K. Statistical metrics. *Proc. Nat. Acad. Sci.*, v. 28, n., p. 535–537, 1942.
- [63] KÖRNER. Experiment and theory. In: *Philosophy of Logic*. Black Well, 1966.
- [64] UHR, L. *Pattern recognition learning and thought, computers-programed: Models of Higher Mental Processes*. Prentice-Hall, INC, Englewood Cliffs, New Jersey, 1973.
- [65] ZADEH, L.A. Fuzzy sets. *Information and Control*, v. 8, n. 1, p. 338–353, 1965.

- [66] ZADEH, L.A. Fuzzy algorithms. *Information and Control*, v. 12, n., p. 94, 1968.
- [67] ZADEH, L.A. Similarity relations and fuzzy orderings. *Information Science*, v. 3, n., p. 177, 1971.
- [68] ZADEH, L.A. A note on prototype theory and fuzzy sets. *Cognition*, v. 12, n., p. 291–297, 1982.
- [69] ZADEH, L.A. Making computers think like people. *IEEE SPECTRUM*, p. 26–32, 1984.
- [70] ARAÚJO, A. L. A. , ZANUSSO, M.B., MAIA, L.F.J. Novos operadores conjuntivos e disjuntivos para conjuntos difusos. In: *XVII-Congresso da Sociedade Brasileira de Computação - I ENIA*, Brasília - DF, 1997.
- [71] MAIA, L.F.J. *Caracterização e reconhecimento de conceitos*. Tese de Doutorado, UNICAMP-FEE, Campinas, Brasil, 1991.
- [72] ZANUSSO, M.B. , ARAÚJO, A. L. A. , MAIA, L.F.J. Um pmcd baseado em novos operadores lógicos difusos. In: *SIMPÓSIO DE REDES NEURAI*S, Goiania - GO, 1997.
- [73] ZANUSSO, M.B., MAIA, L.F.J. Caracterização e reconhecimento de conceitos. In: *CONGRESSO BRASILEIRO DE REDES NEURAI*S,3, p. 231–218, Florianópolis - Brasil, 1997.
- [74] ZANUSSO, M.B., ARAÚJO, A. L. A. , MAIA, L.F.J. T-normas e t-conormas-projeto de tese. In: *XVII-Congresso da Sociedade Brasileira de Computação - I ENIA*, Brasília - DF, 1997.
- [75] LINNAEUS, C. *Systema naturae per regna tria naturae*, v. 1. Holmiae-Stockholm, 10 edição, 1758.
- [76] FU, L.M. Connectionism for fuzzy learning in rule-based expert systems. In: *INTERNATIONAL CONFERENCE ON INDUSTRIAL AND ENGINEERING APPLICATIONS OF ARTIFICIAL INTELLIGENCE AND EXPERT SYSTEMS, IEA/AIE'92*, 5, p. 337–340, Paderborn, Germany, 1992.
- [77] FU, L.M. Rule generation from neural networks. *IEEE Transactions on Systems, Man and Cybernetics*, v. 28, n. 8, p. 1114–1124, 1994.
- [78] BELLMAN, R. , GIERTZ, M. On the analytic formalism of the theory of fuzzy sets. *Information Sciences*, v. 5, n., p. 149–156, 1973.
- [79] SCOTT, A. , SYMONS, M. Clustering methods based on likelihood ratio criteria. *Biometrics*, v. 27, n., p. 387–397, 1971.
- [80] TERANO, T., ASAI, K. SUGENO, M. *Applied fuzzy systems*. Academic Press, Boston, 1994.
- [81] YAGER, R.R., KACPRZYK, J. , FEDRIZZI, M. *Advances in the dempster-shafer theory of evidence*. J. Wiley, New York, 1994.
- [82] PAPPIS, C. , MANDANI, E. A fuzzy controller for a traffic junction. *IEEE Transactions on Systems, Man, and Cybernetics*, v. 7, n., p. 707–717, 1977.

- [83] ROSCH, E. , MERVIS, C.B. Family resemblances: studies in the internal structure of categories. *Cognitive Psychology*, v. 7, n., p. 573–605, 1975.
- [84] KENDALL, M.G. Discrimination and classification. In: *Multivariate analysis*, p. 165–185. Academic Press, New York, 1966.
- [85] SUGENO, M. , MURAKAMI, K. An experimental study on fuzzy parking control using a model car. In: *Industrial applications of fuzzy control*. Elsevier Science Publishers, Amsterdam, 1985.
- [86] MARINOS, N.P. Fuzzy logic and its applications to switching systems. *IEEE Trans. on Computers*, v. C-18, n. 4, p. 345, 1969.
- [87] HUNTSBERGER, T. , AJJIMARANGSEE, P. Parallel self-organizing feature maps for unsupervised pattern recognition. In: *Fuzzy models for pattern recognition*, p. 483–495. IEEE Press, New York, 1990.
- [88] DUDA, R.O., HART, P.E. *Pattern classification and scene analysis*. John Wiley, New York, NY, 1973.
- [89] SNEATH, P.H.A. Thirty years of numerical taxonomy. *Systematic Biological*, v. 44, n. 3, p. 281–298, 1995.
- [90] MACVICAR-WHELAN, P.J. Fuzzy sets, the concepts of height and height very. *IEEE Transactions on Systems, Man, and Cybernetics*, v. 8, n., p. 507–511, 1978.
- [91] JARDINE, N., SIBSON, R. A model for taxonomy. *Math. Biosci.*, v. 2, n., p. 465–482, 1968a.
- [92] FISHER, R.A. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, v. 7, n., p. 179–188, 1936.
- [93] PACHECO, R.C.S. *A hybrid intelligent system for diagnosing and solving financial problems*. Tese de Doutorado, UFSC-EPS, Florianópolis-SC, 1996.
- [94] MICHALSKI, R.S. , STEPP, R.E. Automated construction of classifications: conceptual clustering versus numerical taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. PAMI-5, n. 4, p. 396–410, July 1983.
- [95] MICHALSKI, R.S. , STEPP, R.E. A theory and methodology of inductive learning. In: *Machine learning: an artificial intelligence approach*, p. 83–134. Tioga, Palo Alto, CA, 1983.
- [96] SNEATH, P.H.A. , SOKAL, R.R. *Numerical taxonomy*. W. H. Freeman And Company, San Francisco, 1973.
- [97] SOKAL, R.R. Classification: purposes, principles, progress, prospects. *Science*, v. 185, n., p. 1115–23, 1974.
- [98] YAGER, R.R. Building fuzzy systems models. In: *Applied general systems research*, p. 313–317. Plenum Press, New York, 1978.
- [99] YAGER, R.R. A measurement-informational discussion of fuzzy union and intersection. *Int. J. Man-Machine Studies*, v. 11, n., p. 189–200, 1979.

- [100] DIETTERICH, T.G. , MICHALSKI, R.S. A comparative review of selected methods for learning from examples. In: *Machine learning: an artificial intelligence approach*, p. 40–81. Tioga, Palo Alto, CA, 1983.
- [101] MICHALSKI, R.S. Knowledge acquisition through conceptual clustering: a theoretical framework and an algorithm for partitioning data into conjunctive concepts(a special issue on knowledge acquisition and induction). *Policy Analysis and Information Systems*, v. 4, n. 3, p. 219–244, 1980.
- [102] TAMURA, S., HIGUCHI, S. and TANAKA, K. Pattern classification based on fuzzy relations. In: *Fuzzy models for pattern recognition*, p. 169–174. IEEE Press, New York, 1990.
- [103] JEPSON, S.B. *Identification of root-knot nematodes-meloidogyne species*. C.A.B. INTERNATIONAL, Wallingford, Oxon, U.K., 1987.
- [104] MITRA, S. , PAL, S.K. Logical operation based fuzzy mlp for classification and rule generation. *Neural Networks*, v. 7, n. 2, p. 353–373, 1994.
- [105] WINOGRAD, T. *Understanding natural language*. Academic, New York, 1972.
- [106] MAMDANI, E.H. , PROCYK, T.J. *Application of fuzzy logic to controller: design based on linguistic protocol*. Dep. Elec. and electron. Eng. Quenn Mary College, London, 1976.
- [107] LABOV, W. The boundaries of words and their meanings. In: *New ways of analyzing variation in English*, p. 340–73. Georgetown Univ. Press, Washington, 1973.
- [108] WEE W.G. A formulation of fuzzy automata and its applications as a model of learning systems. *IEEE Trans.*, v. SSC-5, n., p. 215, 1969.
- [109] FORTUNER, R. , WONG, Y. Review of the genus *helicotylenchus steiner*, 1945. 1. a computer program for identification of the species. *Revue Nématologie*, p. 385–392, 1984.
- [110] ZYSNO,P. Modelling membership functions. In: *Empirical semantics*, p. 350–375. Brockmeyer, Bochum, 1981.

Apêndice 1
Medidas de 4 Atributos de 150 Flores Íris

Iris Setosa				Iris Versicolor				Iris Virginica			
SL	SW	PL	PW	SL	SW	PL	PW	SL	SW	PL	PW
5,1	3,5	1,4	0,2	7	3,2	4,7	1,4	6,3	3,3	6	2,5
4,9	3	1,4	0,2	6,4	3,2	4,5	1,5	5,8	2,7	5,1	1,9
4,7	3,2	1,3	0,2	6,9	3,1	4,9	1,5	7,1	3	5,9	2,1
4,6	3,1	1,5	0,2	5,5	2,3	4	1,3	6,3	2,9	5,6	1,8
5	3,6	1,4	0,2	6,5	2,8	4,6	1,5	6,5	3	5,8	2,2
5,4	3,9	1,7	0,4	5,7	2,8	4,5	1,3	7,6	3	6,6	2,1
4,6	3,4	1,4	0,3	6,3	3,3	4,7	1,6	4,9	2,5	4,5	1,7
5	3,4	1,5	0,2	4,9	2,4	3,3	1	7,3	2,9	6,3	1,8
4,4	2,9	1,4	0,2	6,6	2,9	4,6	1,3	6,7	2,5	5,8	1,8
4,9	3,1	1,5	0,1	5,2	2,7	3,9	1,4	7,2	3,6	6,1	2,5
5,4	3,7	1,5	0,2	5	2	3,5	1	6,5	3,2	5,1	2
4,8	3,4	1,6	0,2	5,9	3	4,2	1,5	6,4	2,7	5,3	1,9
4,8	3	1,4	0,1	6	2,2	4	1	6,8	3	5,5	2,1
4,3	3	1,1	0,1	6,1	2,9	4,7	1,4	5,7	2,5	5	2
5,8	4	1,2	0,2	5,6	2,9	3,6	1,3	5,8	2,8	5,1	2,4
5,7	4,4	1,5	0,4	6,7	3,1	4,4	1,4	6,4	3,2	5,3	2,3
5,4	3,9	1,3	0,4	5,6	3	4,5	1,5	6,5	3	5,5	1,8
5,1	3,5	1,4	0,3	5,8	2,7	4,1	1	7,7	3,8	6,7	2,2
5,7	3,8	1,7	0,3	6,2	2,2	4,5	1,5	7,7	2,6	6,9	2,3
5,1	3,8	1,5	0,3	5,6	2,5	3,9	1,1	6	2,2	5	1,5
5,4	3,4	1,7	0,2	5,9	3,2	4,8	1,8	6,9	3,2	5,7	2,3
5,1	3,7	1,5	0,4	6,1	2,8	4	1,3	5,6	2,8	4,9	2
4,6	3,6	1	0,2	6,3	2,5	4,9	1,5	7,7	2,8	6,7	2
5,1	3,3	1,7	0,5	6,1	2,8	4,7	1,2	6,3	2,7	4,9	1,8
4,8	3,4	1,9	0,2	6,4	2,9	4,3	1,3	6,7	3,3	5,7	2,1
5	3	1,6	0,2	6,6	3	4,4	1,4	7,2	3,2	6	1,8
5	3,4	1,6	0,4	6,8	2,8	4,8	1,4	6,2	2,8	4,8	1,8
5,2	3,5	1,5	0,2	6,7	3	5	1,7	6,1	3	4,9	1,8
5,2	3,4	1,4	0,2	6	2,9	4,5	1,5	6,4	2,8	5,6	2,1
4,7	3,2	1,6	0,2	5,7	2,6	3,5	1	7,2	3	5,8	1,6
4,8	3,1	1,6	0,2	5,5	2,4	3,8	1,1	7,4	2,8	6,1	1,9
5,4	3,4	1,5	0,4	5,5	2,4	3,7	1	7,9	3,8	6,4	2
5,2	4,1	1,5	0,1	5,8	2,7	3,9	1,2	6,4	2,8	5,6	2,2
5,5	4,2	1,4	0,2	6	2,7	5,1	1,6	6,3	2,8	5,1	1,5
4,9	3,1	1,5	0,2	5,4	3	4,5	1,5	6,1	2,6	5,6	1,4
5	3,2	1,2	0,2	6	3,4	4,5	1,6	7,7	3	6,1	2,3
5,5	3,5	1,3	0,2	6,7	3,1	4,7	1,5	6,3	3,4	5,6	2,4
4,9	3,6	1,4	0,1	6,3	2,3	4,4	1,3	6,4	3,1	5,5	1,8
4,4	3	1,3	0,2	5,6	3	4,1	1,3	6	3	4,8	1,8
5,1	3,4	1,5	0,2	5,5	2,5	4	1,3	6,9	3,1	5,4	2,1
5	3,5	1,3	0,3	5,5	2,6	4,4	1,2	6,7	3,1	5,6	2,4
4,5	2,3	1,3	0,3	6,1	3	4,6	1,4	6,9	3,1	5,1	2,3
4,4	3,2	1,3	0,2	5,8	2,6	4	1,2	5,8	2,7	5,1	1,9
5	3,5	1,6	0,6	5	2,3	3,3	1	6,8	3,2	5,9	2,3
5,1	3,8	1,9	0,4	5,6	2,7	4,2	1,3	6,7	3,3	5,7	2,5
4,8	3	1,4	0,3	5,7	3	4,2	1,2	6,7	3	5,2	2,3
5,1	3,8	1,6	0,2	5,7	2,9	4,2	1,3	6,3	2,5	5	1,9
4,6	3,2	1,4	0,2	6,2	2,9	4,3	1,3	6,5	3	5,2	2
5,3	3,7	1,5	0,2	5,1	2,5	3	1,1	6,2	3,4	5,4	2,3
5	3,3	1,4	0,2	5,7	2,8	4,1	1,3	5,9	3	5,1	1,8

Apêndice 2

Código do MATLAB para Construir Matriz de Treino e Matriz de Teste

```
PROGRAMA: callmt.m
%Este programa gera os dados para o programa calibra.m
%Dá entrada na matriz de dados iris = [SL SW PL PW] que tem 150 linhas
%e 4 colunas, constrói a matriz de treino, Da, e de teste Tes.

% Entrada da matriz de dados:
load iris.dat
iris;

%Gerando números aleatórios entre 1 e 150:

for i = 1:150
M(i) = fix(rand*150) + 1
end

%Matriz com os 105 primeiros números diferentes

M =[ 105; 18; 115; 79; 84; 89; 50; 106; 22; 25;
122; 111; 48; 21; 80; 47; 78; 65; 39; 56;
59; 41; 72; 69; 42; 12; 56; 88; 101; 102;
109; 142; 49; 70; 100; 61; 91; 148; 23; 66;
52; 82; 125; 3; 97; 138; 146; 127; 123; 4;
38; 149; 114; 26; 1; 75; 136; 118; 2; 16;
11; 33; 36; 95; 99; 81; 43; 103; 117; 34;
45; 8; 67; 129; 5; 57; 74; 87; 53; 93;
81; 6; 19; 147; 29; 83; 144; 54; 126; 64;
37; 73; 130; 110; 113; 107; 9; 126; 92; 135;
120; 132; 140; 14; 31];

%Construção da matriz de treino.
%São 35 flores de cada espécie classificadas anteriormente %por Anderson

for i = 1:105
Obs(i,:) = iris(M(i),:);
end
Da = Obs;

%T é a matriz dos índices que vão entrar para teste.
%São 15 flores de cada espécie classificadas anteriormente %por Anderson.

T =[ 7; 10; 13; 15; 17; 20; 24; 27; 28; 30; 32;
35; 40; 44; 46; 51; 55; 58; 60; 62; 68; 71;
76; 77; 85; 86; 90; 94; 96; 98; 108; 112; 116;
119; 121; 124; 128; 131; 133; 134; 137; 139; 143; 145;
150];

% Construção da matriz de teste.

for i = 1:50
Tes(i,:) = iris(T(i),:);
end
```

```
%Classificações de Anderson para as 150 flores.  
%Definição de mTA, mTB e mTC.
```

```
for Obs = 1:150  
    if Obs < 51,  
        mTA(Obs) = 1;  
        mTB(Obs) = -1;  
        mTC(Obs) = -1;  
    elseif 51 <= Obs & Obs < 101,  
        mTA(Obs) = -1;  
        mTB(Obs) = 1;  
        mTC(Obs) = -1;  
    else  
        mTA(Obs) = -1;  
        mTB(Obs) = -1;  
        mTC(Obs) = 1;  
    end  
end
```

```
%Classificações de Anderson para cada linha da matriz de  
%treino Da.
```

```
%Definição da matriz TA, TB e TC.
```

```
for i = 1:105  
    TA(i) = mTA(M(i));  
    TB(i) = mTB(M(i));  
    TC(i) = mTC(M(i));  
end
```

Apêndice 3:

Código do MATLAB para Algoritmo de Descida do Gradiente

```
%PROGRAMA:calibra.m
%Implementa o algoritmo para ajustar os parâmetros três funções
%de caracterização correspondente a três espécies de flores iris:
%RA = Íris setosa,
%RB = Íris versicolor,
%RC = Íris virgínica

%DADOS%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
callmt.m %gera os dados
%IRIS=[SL SW PL PW] matriz das 150 flores
%Matriz de treino Da:105 linhas extraídas aleatoriamente de
%IRIS
Da;
%Vetores de atributos observados nas 105 flores
x1=Da(:,1); %SL = comprimento da sépala
x2=Da(:,2); %SW = largura da sépala
x3=Da(:,3); %PL = comprimento da pétala
x4=Da(:,4); %PW = largura da pétala
%Matriz de teste Tes:45 linhas o restante da matriz IRIS
Tes;

%DEFINIÇÕES DAS FUNÇÕES%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Definição da função RA(Wa,Qa;Obs)=Uc(z1:qa1:wa1,z1:qa5:wa5,
%
%           z2:qa2:wa2,z2:qa6:wa6,
%           z3:qa3:wa3,z3:qa7:wa7,
%           z4:qa4:wa4,z4:qa8:wa8).
%Definição da função RB(Wb,Qb;Obs)=Ud(RB1:1:1,RB2:1:1)
%Definição da função RB1
RB1(Wb1,Qb1;Obs)=Uc(z1:qb11:wb11,z1:qb15:wb15,
%           z2:qb12:wb12,z2:qb16:wb16,
%           z3:qb13:wb13,z3:qb17:wb17,
%           z4:qb14:wb14,z4:qb18:wb18).
%Definição da função RB2(Wb2,Qb2;Obs) é igual a RB1.
%Definição da função RC(Wc,Qc;Obs)=Ud(RC1:1:1,RC2:1:1)
%Definições das funções RC1(Wc1,Qc1;Obs) e RC2(Wc2,Qc2;Obs)
%são iguais a RB1

%INICIALIZAÇÃO%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Parâmetros de treino
eta=0.5; %Taxa de aprendizado
errobjetivo=0.001; %Erro objetivo
numepocas=2000; %Número de vezes que passa pelo
%conjunto de treino
%Parâmetros dos conceitos primários (são os mesmos para A,
%B e C)
k1=4.05;p1=5.78;
k2=3.96;p2=3.09;
k3=1.97;p3=3.78;
k4=1.96;p4=1.197;
%Vetores dos parâmetros
P=[p1 p2 p3 p4];
K=[k1 k2 k3 k4];
%Parâmetros modificadores iniciais
Qa=[1; 1; 1; 1; 1; 1; 1; 1];
Wa=[1; 1; 1; 1; -1; -1; -1; -1];
Qb1=[1; 1; 1; 1; 1; 1; 1; 1];
Wb1=[1; 1; 1; 1; -1; -1; -1; -1];
Qb2=[1; 1; 1; 1; 1; 1; 1; 1];
Wb2=[1; 1; 1; 1; -1; -1; -1; -1];
Qc1=[1; 1; 1; 1; 1; 1; 1; 1];
Wc1=[1; 1; 1; 1; -1; -1; -1; -1];
Qc2=[1; 1; 1; 1; 1; 1; 1; 1];
Wc2=[1; 1; 1; 1; -1; -1; -1; -1];
```

```

%%FASE DE TREINO%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
for n=1:numepocas %para cada n passa todo conjunto de dados
%%Cálculo de EA(n):
    for Obs=1:105
Sa(Obs)=
(Qa(1)^Wa(1))* (p1^(k1*Wa(1))) * ((1/Da(Obs,1))^(k1*Wa(1))) +
(Qa(5)^Wa(5))* (p1^(k1*Wa(5))) * ((1/Da(Obs,1))^(k1*Wa(5))) +
(Qa(2)^Wa(2))* (p2^(k2*Wa(2))) * ((1/Da(Obs,2))^(k2*Wa(2))) +
(Qa(6)^Wa(6))* (p2^(k2*Wa(6))) * ((1/Da(Obs,2))^(k2*Wa(6))) +
(Qa(3)^Wa(3))* (p3^(k3*Wa(3))) * ((1/Da(Obs,3))^(k3*Wa(3))) +
(Qa(7)^Wa(7))* (p3^(k3*Wa(7))) * ((1/Da(Obs,3))^(k3*Wa(7))) +
(Qa(4)^Wa(4))* (p4^(k4*Wa(4))) * ((1/Da(Obs,4))^(k4*Wa(4))) +
(Qa(8)^Wa(8))* (p4^(k4*Wa(8))) * ((1/Da(Obs,4))^(k4*Wa(8)));
RA(Obs)=(1-Sa(Obs))/(1+Sa(Obs));
EAo(Obs)=(1/2)*((RA(Obs)-TA(Obs))^2);
    end
EA(n)=(1/105)*sum(EAo);

%%Cálculo de EB(n):
    for Obs=1:105
Sb1(Obs)=
(Qb1(1)^Wb1(1))* (p1^(k1*Wb1(1))) * (1/Da(Obs,1))^(k1*Wb1(1)) +
(Qb1(5)^Wb1(5))* (p1^(k1*Wb1(5))) * (1/Da(Obs,1))^(k1*Wb1(5)) +
(Qb1(2)^Wb1(2))* (p2^(k2*Wb1(2))) * (1/Da(Obs,2))^(k2*Wb1(2)) +
(Qb1(6)^Wb1(6))* (p2^(k2*Wb1(6))) * (1/Da(Obs,2))^(k2*Wb1(6)) +
(Qb1(3)^Wb1(3))* (p3^(k3*Wb1(3))) * (1/Da(Obs,3))^(k3*Wb1(3)) +
(Qb1(7)^Wb1(7))* (p3^(k3*Wb1(7))) * (1/Da(Obs,3))^(k3*Wb1(7)) +
(Qb1(4)^Wb1(4))* (p4^(k4*Wb1(4))) * (1/Da(Obs,4))^(k4*Wb1(4)) +
(Qb1(8)^Wb1(8))* (p4^(k4*Wb1(8))) * (1/Da(Obs,4))^(k4*Wb1(8));
RB1(Obs)=(1-Sb1(Obs))/(1+Sb1(Obs));
Yb1(Obs)=(1+RB1(Obs))/(1-RB1(Obs));
Sb2(Obs)=
(Qb2(1)^Wb2(1))* (p1^(k1*Wb2(1))) * (1/Da(Obs,1))^(k1*Wb2(1)) +
(Qb2(5)^Wb2(5))* (p1^(k1*Wb2(5))) * (1/Da(Obs,1))^(k1*Wb2(5)) +
(Qb2(2)^Wb2(2))* (p2^(k2*Wb2(2))) * (1/Da(Obs,2))^(k2*Wb2(2)) +
(Qb2(6)^Wb2(6))* (p2^(k2*Wb2(6))) * (1/Da(Obs,2))^(k2*Wb2(6)) +
(Qb2(3)^Wb2(3))* (p3^(k3*Wb2(3))) * (1/Da(Obs,3))^(k3*Wb2(3)) +
(Qb2(7)^Wb2(7))* (p3^(k3*Wb2(7))) * (1/Da(Obs,3))^(k3*Wb2(7)) +
(Qb2(4)^Wb2(4))* (p4^(k4*Wb2(4))) * (1/Da(Obs,4))^(k4*Wb2(4)) +
(Qb2(8)^Wb2(8))* (p4^(k4*Wb2(8))) * (1/Da(Obs,4))^(k4*Wb2(8));
RB2(Obs)=(1-Sb2(Obs))/(1+Sb2(Obs));
Yb2(Obs)=(1+RB2(Obs))/(1-RB2(Obs));
RB(Obs)=(Yb1(Obs)+Yb2(Obs)-1)/(Yb1(Obs)+Yb2(Obs)+1);
EBo(Obs)=(1/2)*((RB(Obs)-TB(Obs))^2);
    end
EB(n)=(1/105)*sum(EBo);

%%Cálculo de EC(n):
    for Obs=1:105
Sc1=
(Qc1(1)^Wc1(1))* (p1^(k1*Wc1(1))) * (1/Da(Obs,1))^(k1*Wc1(1)) +
(Qc1(5)^Wc1(5))* (p1^(k1*Wc1(5))) * (1/Da(Obs,1))^(k1*Wc1(5)) +
(Qc1(2)^Wc1(2))* (p2^(k2*Wc1(2))) * (1/Da(Obs,2))^(k2*Wc1(2)) +
(Qc1(6)^Wc1(6))* (p2^(k2*Wc1(6))) * (1/Da(Obs,2))^(k2*Wc1(6)) +
(Qc1(3)^Wc1(3))* (p3^(k3*Wc1(3))) * (1/Da(Obs,3))^(k3*Wc1(3)) +
(Qc1(7)^Wc1(7))* (p3^(k3*Wc1(7))) * (1/Da(Obs,3))^(k3*Wc1(7)) +
(Qc1(4)^Wc1(4))* (p4^(k4*Wc1(4))) * (1/Da(Obs,4))^(k4*Wc1(4)) +
(Qc1(8)^Wc1(8))* (p4^(k4*Wc1(8))) * (1/Da(Obs,4))^(k4*Wc1(8));
RC1(Obs)=(1-Sc1(Obs))/(1+Sc1(Obs));
Yc1(Obs)=(1+RC1(Obs))/(1-RC1(Obs));
Sc2 =
(Qc2(1)^Wc2(1))* (p1^(k1*Wc2(1))) * (1/Da(Obs,1))^(k1*Wc2(1)) +
(Qc2(5)^Wc2(5))* (p1^(k1*Wc2(5))) * (1/Da(Obs,1))^(k1*Wc2(5)) +
(Qc2(2)^Wc2(2))* (p2^(k2*Wc2(2))) * (1/Da(Obs,2))^(k2*Wc2(2)) +
(Qc2(6)^Wc2(6))* (p2^(k2*Wc2(6))) * (1/Da(Obs,2))^(k2*Wc2(6)) +
(Qc2(3)^Wc2(3))* (p3^(k3*Wc2(3))) * (1/Da(Obs,3))^(k3*Wc2(3)) +
(Qc2(7)^Wc2(7))* (p3^(k3*Wc2(7))) * (1/Da(Obs,3))^(k3*Wc2(7)) +
(Qc2(4)^Wc2(4))* (p4^(k4*Wc2(4))) * (1/Da(Obs,4))^(k4*Wc2(4)) +
(Qc2(8)^Wc2(8))* (p4^(k4*Wc2(8))) * (1/Da(Obs,4))^(k4*Wc2(8));
RC2(Obs)=(1-Sc2(Obs))/(1+Sc2(Obs));

```

```

Yc2(Obs)=(1+RC2(Obs))/(1-RC2(Obs));
RC(Obs)=(Yc1(Obs)+Yc2(Obs)-1)/(Yc1(Obs)+Yc2(Obs)+1);
ECo(Obs)=(1/2)*((RC(Obs)-TC(Obs))^2);
    end
EC(n)=(1/105)*sum(ECo);

%ERRO MÉDIO AO FINAL DE CADA CICLO COM OS PARÂMETROS
%ATUAIS
    Emedio(n)=EA(n) + EB(n) + EC(n);
%    if Emedio(n)<errobjetivo ,break, end
%Não usei esta condição, preferi deixar até um certo num. de ciclos

%CÁLCULO DO GRADIENTE DE Emedio NO
%PONTO%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%(Qa,Qb1,Qb2,Qc1,Qc2,Wa,Wb1,Wb2,Wc1,Wc2) atual%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%Definição dos fatores F's comuns%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
for Obs=1:105,
    Fa(Obs)=(RA(Obs)-TA(Obs)) * (-2/((1+Sa(Obs))^2));
    Fb1(Obs)=(RB(Obs)-TB(Obs)) * (-2/((1+Sb1(Obs))^2)) *
        (2/((Yb1(Obs)+Yb2(Obs)+1)^2)) * (2/((1-RB1(Obs))^2));
    Fb2(Obs)=(RB(Obs)-TB(Obs)) * (-2/((1+Sb2(Obs))^2)) *
        (2/((Yb1(Obs)+Yb2(Obs)+1)^2)) * (2/((1-RB2(Obs))^2));
    Fc1(Obs)=(RC(Obs)-TC(Obs)) * (-2/((1+Sc1(Obs))^2)) *
        (2/((Yc1(Obs)+Yc2(Obs)+1)^2)) * (2/((1-RC1(Obs))^2));
    Fc2(Obs)=(RC(Obs)-TC(Obs)) * (-2/((1+Sc2(Obs))^2)) *
        (2/((Yc1(Obs)+Yc2(Obs)+1)^2)) * (2/((1-RC2(Obs))^2));
end %são vetores 105 x 1.

%%% 1-DERIVADA do Emedio(n) em relação a Qa=dvEAQa(n) e a
%%% Wa=dvEAWa(n)%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%%% Em relação a Qa:%%%
for i=1:4
    if Qa(i)>0
        for Obs=1:105
            dvEAQao(Obs,i)=Fa(Obs)*(Qa(i)^Wa(i))*(Wa(i)/Qa(i))*
                ((P(i)*(1/Da(Obs,i)))^(K(i)*Wa(i)));
        end
        dvEAQa(i)=(1/105)*sum(dvEAQao(:,i));
    else
        dvEAQa(i)=0;           %só para dvEAQa permanecer 8x1
        Qa(i)=0.001;         %para não dar log0 e continuar
                            %ajustando em Wa(i)
    end
end

for j=5:8
    if Qa(j)>0
        for Obs=1:105
            dvEAQao(Obs,j)=Fa(Obs)*(Qa(j)^Wa(j))*(Wa(j)/Qa(j))*
                ((P(j-4)*(1/Da(Obs,j-4)))^(K(j-4)*Wa(j)));
        end
        dvEAQa(j)=(1/105)*sum(dvEAQao(:,j));
    else
        dvEAQa(j)=0;
        Qa(j)=0.001;
    end
end

%%% Em relação a Wa:%%%
for i=1:4
    if Qa(i)>0,
        for Obs=1:105,
            dvEAWao(Obs,i)=Fa(Obs)*(Qa(i)^Wa(i))*
                ((P(i)*(1/Da(Obs,i)))^(K(i)*Wa(i)))*
                log(Qa(i))*((P(i)/Da(Obs,i))^K(i));
        end
        dvEAWa(i)=(1/105)*sum(dvEAWao(:,i));
    end
end

```

```

else
    Qa(i)=0.001;
    for Obs=1:105,
dvEAWao(Obs,i)=Fa(Obs)*(Qa(i)^Wa(i))*
        ((P(i)*(1/Da(Obs,i)))^(K(i)*Wa(i)))*
        log(Qa(i)*((P(i)/Da(Obs,i))^K(i)));
    end
    dvEAWa(i)=(1/105)*sum(dvEAWao(:,i));
end
end

for j=5:8
    if Qa(j)>0,
        for Obs=1:105,
dvEAWao(Obs,j)=Fa(Obs)*(Qa(j)^Wa(j))*
            ((P(j-4)*(1/Da(Obs,j-4)))^(K(j-4)*Wa(j)))*
            log(Qa(j)*((P(j-4)/Da(Obs,j-4))^K(j-4)));
        end
        dvEAWa(j)=(1/105)*sum(dvEAWao(:,j));
    else
        Qa(j)=0.001;
        for Obs=1:105,
dvEAWao(Obs,j)=Fa(Obs)*(Qa(j)^Wa(j))*
            ((P(j-4)*(1/Da(Obs,j-4)))^(K(j-4)*Wa(j)))*
            log(Qa(j)*((P(j-4)/Da(Obs,j-4))^K(j-4)));
        end
        dvEAWa(j)=(1/105)*sum(dvEAWao(:,j));
    end
end

%%% 2-DERIVADA do Emédio(n) em relação a Qb1=dvEBQb1(n) e %%%
%%%a Wb1=dvEBWb1(n) %%%
%%% Em relação a Qb1%%%
for i=1:4
    if Qb1(i)>0
        for Obs=1:105
dvEBQb1o(Obs,i)=Fb1(Obs)*(Qb1(i)^Wb1(i))*(Wb1(i)/Qb1(i))*
            ((P(i)*(1/Da(Obs,i)))^(K(i)*Wb1(i)));
        end
        dvEBQb1(i)=(1/105)*sum(dvEBQb1o(:,i));
    else
        dvEBQb1(i)=0;
        Qb1(i)=0.001;
    end
end

for j=5:8
    if Qb1(j)>0
        for Obs=1:105
dvEBQb1o(Obs,j)=Fb1(Obs)*(Qb1(j)^Wb1(j))*(Wb1(j)/Qb1(j))*
            ((P(j-4)*(1/Da(Obs,j-4)))^(K(j-4)*Wb1(j)));
        end
        dvEBQb1(j)=(1/105)*sum(dvEBQb1o(:,j));
    else
        dvEBQb1(j)=0;
        Qb1(j)=0.001;
    end
end

%%% Em relação a Wb1%%%
for i=1:4
    if Qb1(i)>0,
        for Obs=1:105,
dvEBWb1o(Obs,i)=Fb1(Obs)*(Qb1(i)^Wb1(i))*
            ((P(i)*(1/Da(Obs,i)))^(K(i)*Wb1(i)))*
            log(Qb1(i)*((P(i)/Da(Obs,i))^K(i)));
        end
        dvEBWb1(i)=(1/105)*sum(dvEBWb1o(:,i));
    else
        Qb1(i)=0.001;
    end
end

```

```

        for Obs=1:105,
dvEBWb1o(Obs,i)=Fb1(Obs)*(Qb1(i)^Wb1(i))*
                ((P(i)*(1/Da(Obs,i)))^(K(i)*Wb1(i)))*
                log(Qb1(i)*((P(i)/Da(Obs,i))^K(i)));
        end
        dvEBWb1(i)=(1/105)*sum(dvEBWb1o(:,i));
    end
end

for j=5:8
    if Qb1(j)>0,
        for Obs=1:105,
dvEBWb1o(Obs,j)=Fb1(Obs)*(Qb1(j)^Wb1(j))*
                ((P(j-4)*(1/Da(Obs,j-4)))^(K(j-4)*Wb1(j)))*
                log(Qb1(j)*((P(j-4)/Da(Obs,j-4))^K(j-4)));
        end
        dvEBWb1(j)=(1/105)*sum(dvEBWb1o(:,j));
    else
        Qb1(j)=0.001;
        for Obs=1:105,
                %para não dar log0 e continuar
                %ajustando em Wb1(j)
dvEBWb1o(Obs,j)=Fb1(Obs)*(Qb1(j)^Wb1(j))*
                ((P(j-4)*(1/Da(Obs,j-4)))^(K(j-4)*Wb1(j)))*
                log(Qb1(j)*((P(j-4)/Da(Obs,j-4))^K(j-4)));
        end
        dvEBWb1(j)=(1/105)*sum(dvEBWb1o(:,j));
    end
end

%%% 3- DERIVADA DE Emédio em relação a Qb2=dvEBQb2 e a%%%%%%%%
%%% Wb2=dvEBWb2%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%%% Em relação a Qb2%%%
for i=1:4
    if Qb2(i)>0
        for Obs=1:105
dvEBQb2o(Obs,i)=Fb2(Obs)*(Qb2(i)^Wb2(i))*(Wb2(i)/Qb2(i))*
                ((P(i)*(1/Da(Obs,i)))^(K(i)*Wb2(i)));
        end
        dvEBQb2(i)=(1/105)*sum(dvEBQb2o(:,i));
    else
        dvEBQb2(i)=0;
        Qb2(i)=0.001;
    end
end

for j=5:8
    if Qb2(j)>0
        for Obs=1:105
dvEBQb2o(Obs,j)=Fb2(Obs)*(Qb2(j)^Wb2(j))*(Wb2(j)/Qb2(j))*
                ((P(j-4)*(1/Da(Obs,j-4)))^(K(j-4)*Wb2(j)));
        end
        dvEBQb2(j)=(1/105)*sum(dvEBQb2o(:,j));
    else
        dvEBQb2(j)=0;
        Qb2(j)=0.001;
    end
end

%%% Em relação a Wb2%%%
for i=1:4
    if Qb2(i)>0,
        for Obs=1:105,
dvEBWb2o(Obs,i)=Fb2(Obs)*(Qb2(i)^Wb2(i))*
                ((P(i)*(1/Da(Obs,i)))^(K(i)*Wb2(i)))*
                log(Qb2(i)*((P(i)/Da(Obs,i))^K(i)));
        end

```

```

dvEBWb2(i)=(1/105)*sum(dvEBWb2o(:,i));
else
    Qb2(i)=0.001;
    for Obs=1:105,
dvEBWb2o(Obs,i)=Fb2(Obs)*(Qb2(i)^Wb2(i))*
                ((P(i)*(1/Da(Obs,i)))^(K(i)*Wb2(i)))*
                log(Qb2(i)*((P(i)/Da(Obs,i))^K(i)));
    end
    vEBWb2(i)=(1/105)*sum(dvEBWb2o(:,i));
end
end

for j=5:8
    if Qb2(j)>0,
        for Obs=1:105,
dvEBWb2o(Obs,j)=Fb2(Obs)*(Qb2(j)^Wb2(j))*
                ((P(j-4)*(1/Da(Obs,j-4)))^(K(j-4)*Wb2(j)))*
                log(Qb2(j)*((P(j-4)/Da(Obs,j-4))^K(j-4)));
        end
        dvEBWb2(j)=(1/105)*sum(dvEBWb2o(:,j));
    else
        Qb2(j)=0.001;
        for Obs=1:105,
dvEBWb2o(Obs,j)=Fb2(Obs)*(Qb2(j)^Wb2(j))*
                ((P(j-4)*(1/Da(Obs,j-4)))^(K(j-4)*Wb2(j)))*
                log(Qb2(j)*((P(j-4)/Da(Obs,j-4))^K(j-4)));
        end
        dvEBWb2(j)=(1/105)*sum(dvEBWb2o(:,j));
    end
end

%%% 4- DERIVADA DE Emédio em relação a Qc1=dvECQc1 e%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%% a Wc1=dvECWc1%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%%% Em relação a Qc1%%%
for i=1:4
    if Qc1(i)>0
        for Obs=1:105
dvECQc1o(Obs,i)=Fc1(Obs)*(Qc1(i)^Wc1(i))*(Wc1(i)/Qc1(i))*
                ((P(i)*(1/Da(Obs,i)))^(K(i)*Wc1(i)));
        end
        dvECQc1(i)=(1/105)*sum(dvECQc1o(:,i));
    else
        dvECQc1(i)=0;
        Qc1(i)=0.001;
    end
end

for j=5:8
    if Qc1(j)>0
        for Obs=1:105
dvECQc1o(Obs,j)=Fc1(Obs)*(Qc1(j)^Wc1(j))*(Wc1(j)/Qc1(j))*
                ((P(j-4)*(1/Da(Obs,j-4)))^(K(j-4)*Wc1(j)));
        end
        dvECQc1(j)=(1/105)*sum(dvECQc1o(:,j));
    else
        dvECQc1(j)=0;
        Qc1(j)=0.001;
    end
end

%%% Em relação a Wc1%%%
for i=1:4
    if Qc1(i)>0,
        for Obs=1:105,
dvECWc1o(Obs,i)=Fc1(Obs)*(Qc1(i)^Wc1(i))*
                ((P(i)*(1/Da(Obs,i)))^(K(i)*Wc1(i)))*
                log(Qc1(i)*((P(i)/Da(Obs,i))^K(i)));
        end
        dvECWc1(i)=(1/105)*sum(dvECWc1o(:,i));
    else

```

```

        Qc1(i)=0.001;
        for Obs=1:105,
dvECWc1o(Obs,i)=Fc1(Obs)*(Qc1(i)^Wc1(i))*
                ((P(i)*(1/Da(Obs,i)))^(K(i)*Wc1(i)))*
                log(Qc1(i)*((P(i)/Da(Obs,i))^K(i)));
        end
        dvECWc1(i)=(1/105)*sum(dvECWc1o(:,i));
    end
end

for j=5:8
    if Qc1(j)>0,
        for Obs=1:105,
dvECWc1o(Obs,j)=Fc1(Obs)*(Qc1(j)^Wc1(j))*
                ((P(j-4)*(1/Da(Obs,j-4)))^(K(j-4)*Wc1(j)))*
                log(Qc1(j)*((P(j-4)/Da(Obs,j-4))^K(j-4)));
        end
        dvECWc1(j)=(1/105)*sum(dvECWc1o(:,j));
    else
        Qc1(j)=0.001;
        for Obs=1:105,
dvECWc1o(Obs,j)=Fc1(Obs)*(Qc1(j)^Wc1(j))*
                ((P(j-4)*(1/Da(Obs,j-4)))^(K(j-4)*Wc1(j)))*
                log(Qc1(j)*((P(j-4)/Da(Obs,j-4))^K(j-4)));
        end
        dvECWc1(j)=(1/105)*sum(dvECWc1o(:,j));
    end
end

%%% 5- DERIVADA DE Emédio em relação a Qc2=dvECQc2 e %%%%%%%%%%%
%%% a Wc2=dvECWc2 %%%%%%%%%%%

%%% Em relação a Qc2%%%
for i=1:4
    if Qc2(i)>0
        for Obs=1:105
dvECQc2o(Obs,i)=Fc2(Obs)*(Qc2(i)^Wc2(i))*(Wc2(i)/Qc2(i))*
                ((P(i)*(1/Da(Obs,i)))^(K(i)*Wc2(i)));
        end
        dvECQc2(i)=(1/105)*sum(dvECQc2o(:,i));
    else
        dvECQc2(i)=0;
        Qc2(i)=0.001;
    end
end

for j=5:8
    if Qc2(j)>0
        for Obs=1:105
dvECQc2o(Obs,j)=Fc2(Obs)*(Qc2(j)^Wc2(j))*(Wc2(j)/Qc2(j))*
                ((P(j-4)*(1/Da(Obs,j-4)))^(K(j-4)*Wc2(j)));
        end
        dvECQc2(j)=(1/105)*sum(dvECQc2o(:,j));
    else
        dvECQc2(j)=0;
        Qc2(j)=0.001;
    end
end

%%% Em relação a Wc2%%%
for i=1:4
    if Qc2(i)>0,
        for Obs=1:105,
dvECWc2o(Obs,i)=Fc2(Obs)*(Qc2(i)^Wc2(i))*
                ((P(i)*(1/Da(Obs,i)))^(K(i)*Wc2(i)))*
                log(Qc2(i)*((P(i)/Da(Obs,i))^K(i)));
        end
        dvECWc2(i)=(1/105)*sum(dvECWc2o(:,i));
    else
        Q2(i)=0.001;
        for Obs=1:105,

```

```

ECWc2o(Obs,i)=Fc2(Obs)*(Qc2(i)^Wc2(i))*
                P(i)*(1/Da(Obs,i))^(K(i)*Wc2(i))*
                log(Qc2(i)*((P(i)/Da(Obs,i))^K(i)));
    end
    dvECWc2(i)=(1/105)*sum(dvECWc2o(:,i));
end
end

for j=5:8
    if Qc2(j)>0,
        for Obs=1:105,
dvECWc2o(Obs,j)=Fc2(Obs)*(Qc2(j)^Wc2(j))*
                ((P(j-4)*(1/Da(Obs,j-4)))^(K(j-4)*Wc2(j)))*
                log(Qc2(j)*((P(j-4)/Da(Obs,j-4))^K(j-4)));
            end
            dvECWc2(j)=(1/105)*sum(dvECWc2o(:,j));
        else
            Qc2(j)=0.001;
            for Obs=1:105,
dvECWc2o(Obs,j)=Fc2(Obs)*(Qc2(j)^Wc2(j))*
                ((P(j-4)*(1/Da(Obs,j-4)))^(K(j-4)*Wc2(j)))*
                log(Qc2(j)*((P(j-4)/Da(Obs,j-4))^K(j-4)));
            end
            dvECWc2(j)=(1/105)*sum(dvECWc2o(:,j));
        end
    end
end

%FINAL das derivadas %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%AJUSTE DOS PARÂMETROS NA DIREÇÃO DO
%GRADIENTE%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%Guardando os parâmetros antes de modificá-los
VQa=Qa;VQb1=Qb1;VQb2=Qb2;VQc1=Qc1;VQc2=Qc2;
VWa=Wa;VWb1=Wb1;VWb2=Wb2;VWc1=Wc1;VWc2=Wc2;

%ATUALIZAÇÃO DOS PARÂMETROS%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
for i=1:8
    Wa(i)=Wa(i) + eta*(-dvEAWa(i));
    Wb1(i)=Wb1(i) + eta*(-dvEBWb1(i));
    Wb2(i)=Wb2(i) + eta*(-dvEBWb2(i));
    Wc1(i)=Wc1(i) + eta*(-dvECWc1(i));
    Wc2(i)=Wc2(i) + eta*(-dvECWc2(i));

    if Qa(i)>0
        Qa(i)=Qa(i) + eta*(-dvEAQa(i));
    end
    if Qa(i)<=0
        Qa(i)=0.001;
    end

    if Qb1(i)>0
        Qb1(i)=Qb1(i) + eta*(-dvEBQb1(i));
    end
    if Qb1(i)<=0
        Qb1(i)=0.001;
    end

    if Qb2(i)>0
        Qb2(i)=Qb2(i) + eta*(-dvEBQb2(i));
    end
    if Qb2(i)<=0
        Qb2(i)=0.001;
    end

    if Qc1(i)>0
        Qc1(i)=Qc1(i) + eta*(-dvECQc1(i));
    end
    if Qc1(i)<=0
        Qc1(i)=0.001;
    end
end

```

```

end

if Qc2(i)>0
    Qc2(i)=Qc2(i) + eta*(-dvECQc2(i));
end
if Qc2(i)<=0
    Qc2(i)=0.001;
end
end %final da atualização

end % final do num. de ciclos
%FINAL DA FASE DE
%TREINO%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%% RESULTADOS FINAIS %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Parâmetros depois de completado o num. de épocas
Qa;Qb1;Qb2;Qc1;Qc2;
Wa;Wb1;Wb2;Wc1;Wc2;
%Erros ao final do num. de ciclos
EA(n);
EB(n);
EC(n);
Emedio(n)
%Vetores de erros
EA;
EB;
EC;

%FASE DE TESTE%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%Calcular RA,RB e RC para cada flor do conjunto de teste,%%%%%%%%
%usando os %parâmetros finais%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

for Obs=1:45

%1-RA grau de pertinência de cada flor para a classe A:
Sac(Obs)=
(Qa(1)^Wa(1))* (p1^(k1*Wa(1))) * ((1/Tes(Obs,1))^(k1*Wa(1)))+
(Qa(5)^Wa(5))* (p1^(k1*Wa(5))) * ((1/Tes(Obs,1))^(k1*Wa(5)))+
(Qa(2)^Wa(2))* (p2^(k2*Wa(2))) * ((1/Tes(Obs,2))^(k2*Wa(2)))+
(Qa(6)^Wa(6))* (p2^(k2*Wa(6))) * ((1/Tes(Obs,2))^(k2*Wa(6)))+
(Qa(3)^Wa(3))* (p3^(k3*Wa(3))) * ((1/Tes(Obs,3))^(k3*Wa(3)))+
(Qa(7)^Wa(7))* (p3^(k3*Wa(7))) * ((1/Tes(Obs,3))^(k3*Wa(7)))+
(Qa(4)^Wa(4))* (p4^(k4*Wa(4))) * ((1/Tes(Obs,4))^(k4*Wa(4)))+
(Qa(8)^Wa(8))* (p4^(k4*Wa(8))) * ((1/Tes(Obs,4))^(k4*Wa(8)));
RAC(Obs)=(1-Sac(Obs))/(1+Sac(Obs));

%2-RB grau de pertinência de cada flor para a classe B:
Sb1c(Obs)=
Qb1(1)^Wb1(1)* (p1^(k1*Wb1(1))) * ((1/Tes(Obs,1))^(k1*Wb1(1)))+
Qb1(5)^Wb1(5)* (p1^(k1*Wb1(5))) * ((1/Tes(Obs,1))^(k1*Wb1(5)))+
Qb1(2)^Wb1(2)* (p2^(k2*Wb1(2))) * ((1/Tes(Obs,2))^(k2*Wb1(2)))+
Qb1(6)^Wb1(6)* (p2^(k2*Wb1(6))) * ((1/Tes(Obs,2))^(k2*Wb1(6)))+
Qb1(3)^Wb1(3)* (p3^(k3*Wb1(3))) * ((1/Tes(Obs,3))^(k3*Wb1(3)))+
Qb1(7)^Wb1(7)* (p3^(k3*Wb1(7))) * ((1/Tes(Obs,3))^(k3*Wb1(7)))+
Qb1(4)^Wb1(4)* (p4^(k4*Wb1(4))) * ((1/Tes(Obs,4))^(k4*Wb1(4)))+
Qb1(8)^Wb1(8)* (p4^(k4*Wb1(8))) * ((1/Tes(Obs,4))^(k4*Wb1(8)));
RB1c(Obs)=(1-Sb1c(Obs))/(1+Sb1c(Obs));
Yb1c(Obs)=(1+RB1c(Obs))/(1-RB1c(Obs));
Sb2c(Obs)=
Qb2(1)^Wb2(1)* (p1^(k1*Wb2(1))) * ((1/Tes(Obs,1))^(k1*Wb2(1)))+
Qb2(5)^Wb2(5)* (p1^(k1*Wb2(5))) * ((1/Tes(Obs,1))^(k1*Wb2(5)))+
Qb2(2)^Wb2(2)* (p2^(k2*Wb2(2))) * ((1/Tes(Obs,2))^(k2*Wb2(2)))+
Qb2(6)^Wb2(6)* (p2^(k2*Wb2(6))) * ((1/Tes(Obs,2))^(k2*Wb2(6)))+
Qb2(3)^Wb2(3)* (p3^(k3*Wb2(3))) * ((1/Tes(Obs,3))^(k3*Wb2(3)))+
Qb2(7)^Wb2(7)* (p3^(k3*Wb2(7))) * ((1/Tes(Obs,3))^(k3*Wb2(7)))+
Qb2(4)^Wb2(4)* (p4^(k4*Wb2(4))) * ((1/Tes(Obs,4))^(k4*Wb2(4)))+

```

```

Qb2(8)^Wb2(8)*(p4^(k4*Wb2(8)))*((1/Tes(Obs,4))^(k4*Wb2(8)));
RB2c(Obs)=(1-Sb2c(Obs))/(1+Sb2c(Obs));
Yb2c(Obs)=(1+RB2c(Obs))/(1-RB2c(Obs));
RBc(Obs)=(Yb1c(Obs)+Yb2c(Obs)-1)/(Yb1c(Obs)+Yb2c(Obs)+1);
%3-RC grau de pertinência de cada flor para a classe C:
Sc1c(Obs)=
Qc1(1)^Wc1(1)*(p1^(k1*Wc1(1)))*((1/Tes(Obs,1))^(k1*Wc1(1))+
Qc1(5)^Wc1(5)*(p1^(k1*Wc1(5)))*((1/Tes(Obs,1))^(k1*Wc1(5))+
Qc1(2)^Wc1(2)*(p2^(k2*Wc1(2)))*((1/Tes(Obs,2))^(k2*Wc1(2))+
Qc1(6)^Wc1(6)*(p2^(k2*Wc1(6)))*((1/Tes(Obs,2))^(k2*Wc1(6))+
Qc1(3)^Wc1(3)*(p3^(k3*Wc1(3)))*((1/Tes(Obs,3))^(k3*Wc1(3))+
Qc1(7)^Wc1(7)*(p3^(k3*Wc1(7)))*((1/Tes(Obs,3))^(k3*Wc1(7))+
Qc1(4)^Wc1(4)*(p4^(k4*Wc1(4)))*((1/Tes(Obs,4))^(k4*Wc1(4))+
Qc1(8)^Wc1(8)*(p4^(k4*Wc1(8)))*((1/Tes(Obs,4))^(k4*Wc1(8)));
RC1c(Obs)=(1-Sc1c(Obs))/(1+Sc1c(Obs));
Yc1c(Obs)=(1+RC1c(Obs))/(1-RC1c(Obs));
Sc2c(Obs)=
Qc2(1)^Wc2(1)*(p1^(k1*Wc2(1)))*((1/Tes(Obs,1))^(k1*Wc2(1))+
Qc2(5)^Wc2(5)*(p1^(k1*Wc2(5)))*((1/Tes(Obs,1))^(k1*Wc2(5))+
Qc2(2)^Wc2(2)*(p2^(k2*Wc2(2)))*((1/Tes(Obs,2))^(k2*Wc2(2))+
Qc2(6)^Wc2(6)*(p2^(k2*Wc2(6)))*((1/Tes(Obs,2))^(k2*Wc2(6))+
Qc2(3)^Wc2(3)*(p3^(k3*Wc2(3)))*((1/Tes(Obs,3))^(k3*Wc2(3))+
Qc2(7)^Wc2(7)*(p3^(k3*Wc2(7)))*((1/Tes(Obs,3))^(k3*Wc2(7))+
Qc2(4)^Wc2(4)*(p4^(k4*Wc2(4)))*((1/Tes(Obs,4))^(k4*Wc2(4))+
Qc2(8)^Wc2(8)*(p4^(k4*Wc2(8)))*((1/Tes(Obs,4))^(k4*Wc2(8)));
RC2c(Obs)=(1-Sc2c(Obs))/(1+Sc2c(Obs));
Yc2c(Obs)=(1+RC2c(Obs))/(1-RC2c(Obs));
RCc(Obs)=(Yc1c(Obs)+Yc2c(Obs)-1)/(Yc1c(Obs)+Yc2c(Obs)+1);
end %do for de Obs.
%No final tenho os vetores 45 x 1: RAc, RBc e RCc

```

```

%%%CONTAGEM DE CORRETOS%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Utilizando um critério de classificação%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
for Obs=1:45

```

```

    if RAc(Obs)>=RBc(Obs) & RAc(Obs)>=RCc(Obs)
        classA(Obs)=1;
    else
        classA(Obs)=0;
    end
end

```

```

for Obs=1:45
    if RBc(Obs)>RAc(Obs) & RBc(Obs)>=RCc(Obs)
        classB(Obs)=1;
    else
        classB(Obs)=0;
    end
end

```

```

for Obs=1:45
    if RCc(Obs)>RAc(Obs) & RCc(Obs)>RBc(Obs)
        classC(Obs)=1;
    else
        classC(Obs)=0;
    end
end

```

```

%No final tenho os vetores 45 x 1: classA, classB, classC de 0's e 1's.
%cada coordenada é relativa à observação T(Obs),do conjunto de teste dada
%por Tes(Obs,:).%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

```

```

%Abaixo vamos verificar se há concordância com classificação de Anderson
%dadas por mTA, mTB e mTC nas flores do conjunto de teste.%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%RESULTADOS do teste
corretA=0;
corretB=0;
corretC=0;
for Obs=1:45
    if classA(Obs)== mTA(T(Obs))
        corretA = corretA + 1;
    end
end

```

```

    end
end

for Obs=1:45
    if classB(Obs) == mTB(T(Obs))
        corretB = corretB + 1;
    end
end

for Obs=1:45
    if classC(Obs) == mTC(T(Obs))
        corretC = corretC + 1;
    end
end

%%RESULTADO DA CLASSIFICAÇÃO

    corretA=sprintf('%d',corretA)
    incorretA=sprintf('%d',15-corretA)

    corretB=sprintf('%d',corretB)
    incorretB=sprintf('%d',15-corretB)

    corretC=sprintf('%d',corretC)
    incorretC=sprintf('%d',15-corretC)

%GRÁFICOS DOS ERROS

for i=1:numciclos %para formar N
    N(i)=i;
end

figure
plot(N,EA);
ylabel('EA');
xlabel('Número de ciclos');

figure
plot(N,EB);
ylabel('EB');
xlabel('Número de ciclos');

figure
plot(N,EC);
ylabel('EC');
xlabel('Número de ciclos');

```

Apêndice 4

n = número de épocas								
	n	EA	EB	EC				
	1000	0,0029	0,0778	0,1166				
	2000	0,0012	0,0647	0,0712				
	3000	7,16E-04	0,0587	0,0657				
	4000	5,08E-04	0,0525	0,0636				
	5000	3,91E-04	0,0473	0,0626				
	6000	3,16E-04	0,0434	0,0619				
	7000	2,65E-04	0,0406	0,0615				
	8000	2,15E-04	0,0378	0,0610				
Parâmetros obtidos pelo algoritmo :								
Wa	Wa1	Wa2	Wa3	Wa4	Wa5	Wa6	Wa7	Wa8
3000	1,2503	1,3624	-2,6115	-2,1922	-3,5014	-4,5361	-2,6742	-3,9412
4000	1,2682	1,3638	-2,6608	-2,2422	-3,5786	-4,6924	-2,7213	-4,1061
5000	1,2821	1,3648	-2,6982	-2,2800	-3,6369	-4,8092	-2,7571	-4,2325
6000	1,2935	1,3657	-2,7283	-2,3104	-3,6836	-4,9021	-2,7860	-4,3349
7000	1,3031	1,3665	-2,7535	-2,3357	-3,7225	-4,9790	-2,8102	-4,4211
8000	1,3145	1,3675	-2,7830	-2,3653	-3,7680	-5,0681	-2,8386	-4,5224
Wb1	Wb1(1)	Wb1(2)	Wb1(3)	Wb1(4)	Wb1(5)	Wb1(6)	Wb1(7)	Wb1(8)
3000	1,2505	1,1349	1,3374	3,6526	-3,5473	-3,2900	-3,7726	-7,6630
4000	1,2598	1,1629	1,3400	3,8318	-3,7430	-3,4439	-4,5665	-8,4085
5000	1,2684	1,1871	1,3426	3,9705	-3,9052	-3,5631	-5,4672	-9,0199
6000	1,2757	1,2060	1,3450	4,0785	-4,0348	-3,6570	-6,2765	-9,5249
7000	1,2819	1,2208	1,3473	4,1656	-4,1401	-3,7332	-6,9749	-9,9531
8000	1,2891	1,2364	1,3499	4,2642	-4,2586	-3,8197	-7,7927	-10,4612
Wb2	Wb2(1)	Wb2(2)	Wb2(3)	Wb2(4)	Wb2(5)	Wb2(6)	Wb2(7)	Wb2(8)
3000	1,2505	1,1349	1,3374	3,6526	-3,5473	-3,2900	-3,7726	-7,6630
4000	1,2598	1,1629	1,3400	3,8318	-3,7430	-3,4439	-4,5665	-8,4085
5000	1,2684	1,1871	1,3426	3,9705	-3,9052	-3,5631	-5,4672	-9,0199
6000	1,2757	1,2060	1,3450	4,0785	-4,0348	-3,6570	-6,2765	-9,5249
7000	1,2819	1,2208	1,3473	4,1656	-4,1401	-3,7332	-6,9749	-9,9531
8000	1,2891	1,2364	1,3499	4,2642	-4,2586	-3,8197	-7,7927	-10,4612
Wc1	Wc1(1)	Wc1(2)	Wc1(3)	Wc1(4)	Wc1(5)	Wc1(6)	Wc1(7)	Wc1(8)
3000	1,4405	1,2242	1,7804	7,2927	-3,4693	-3,3182	1,6827	2,3148
4000	1,4412	1,2396	1,7804	7,8665	-3,7403	-3,4588	1,6828	2,3154
5000	1,4418	1,2521	1,7804	8,2966	-3,9233	-3,5605	1,6828	2,3156
6000	1,4423	1,2628	1,7804	8,6385	-4,0600	-3,6400	1,6828	2,3157
7000	1,4428	1,2722	1,7804	8,9203	-4,1682	-3,7050	1,6828	2,3158
8000	1,4436	1,2837	1,7804	9,2439	-4,2893	-3,7797	1,6829	2,3158
Wc2	Wc2(1)	Wc2(2)	Wc2(3)	Wc2(4)	Wc2(5)	Wc2(6)	Wc2(7)	Wc2(8)
3000	1,4405	1,2242	1,7804	7,2927	-3,4693	-3,3182	1,6827	2,3148
4000	1,4412	1,2396	1,7804	7,8665	-3,7403	-3,4588	1,6828	2,3154
5000	1,4418	1,2521	1,7804	8,2966	-3,9233	-3,5605	1,6828	2,3156
6000	1,4423	1,2628	1,7804	8,6385	-4,0600	-3,6400	1,6828	2,3157
7000	1,4428	1,2722	1,7804	8,9203	-4,1682	-3,7050	1,6828	2,3158
8000	1,4436	1,2837	1,7804	9,2439	-4,2893	-3,7797	1,6829	2,3158

Apêndice 4

Qa	Qa(1)	Qa(2)	Qa(3)	Qa(4)	Qa(5)	Qa(6)	Qa(7)	Qa(8)
3000	0,0010	0,0010	1,7973	1,3276	3,2169	6,6018	1,7223	0,3049
4000	0,0010	0,0006	1,8288	1,3593	3,2781	6,7791	1,7550	0,3044
5000	0,0010	0,0006	1,8527	1,3832	3,3240	6,9096	1,7798	0,3040
6000	0,0010	0,0007	1,8719	1,4024	3,3607	7,0121	1,7998	0,3037
7000	0,0010	0,0004	1,8879	1,4184	3,3911	7,0961	1,8164	0,3035
8000	0,0010	0,0006	1,9067	1,4370	3,4266	7,1925	1,8359	0,3032
Qb1								
Qb1	Qb1(1)	Qb1(2)	Qb1(3)	Qb1(4)	Qb1(5)	Qb1(6)	Qb1(7)	Qb1(8)
3000	0,0010	0,0010	0,0001	0,3130	3,8942	3,2136	2,4477	1,7494
4000	0,0010	0,0010	0,0001	0,3088	4,0636	3,3441	2,0338	1,7989
5000	0,0010	0,0010	0,0001	0,3097	4,2112	3,4435	1,9207	1,8254
6000	0,0010	0,0010	0,0002	0,3113	4,3319	3,5212	1,8737	1,8418
7000	0,0010	0,0010	0,0003	0,3128	4,4308	3,5842	1,8510	1,8522
8000	0,0010	0,0010	0,0002	0,3144	4,5423	3,6554	1,8353	1,8615
Qb2								
Qb2	Qb2(1)	Qb2(2)	Qb2(3)	Qb2(4)	Qb2(5)	Qb2(6)	Qb2(7)	Qb2(8)
3000	0,0010	0,0010	0,0001	0,3130	3,8942	3,2136	2,4477	1,7494
4000	0,0010	0,0010	0,0001	0,3088	4,0636	3,3441	2,0338	1,7989
5000	0,0010	0,0010	0,0001	0,3097	4,2112	3,4435	1,9207	1,8254
6000	0,0010	0,0010	0,0002	0,3113	4,3319	3,5212	1,8737	1,8418
7000	0,0010	0,0010	0,0003	0,3128	4,4308	3,5842	1,8510	1,8522
8000	0,0010	0,0010	0,0002	0,3144	4,5423	3,6554	1,8353	1,8615
Qc1								
Qc1	Qc1(1)	Qc1(2)	Qc1(3)	Qc1(4)	Qc1(5)	Qc1(6)	Qc1(7)	Qc1(8)
3000	0,0040	0,0010	0,0000	2,0181	5,4966	4,0884	0,0003	0,0309
4000	0,0007	0,0010	0,0002	2,0083	5,7604	4,2341	0,0001	0,0178
5000	0,0070	0,0010	0,0001	2,0019	5,9354	4,3386	0,0002	0,0116
6000	0,0000	0,0010	0,0007	1,9972	6,0645	4,4194	0,0002	0,0082
7000	0,0003	0,0010	0,0000	1,9937	6,1658	4,4850	0,0000	0,0061
8000	0,0004	0,0010	0,0000	1,9899	6,2780	4,5598	0,0001	0,0043
Qc2								
Qc2	Qc2(1)	Qc2(2)	Qc2(3)	Qc2(4)	Qc2(5)	Qc2(6)	Qc2(7)	Qc2(8)
3000	0,0004	0,0010	0,0000	2,0181	5,4966	4,0884	0,0003	0,0309
4000	0,0007	0,0010	0,0002	2,0083	5,7604	4,2341	0,0001	0,0178
5000	0,0070	0,0010	0,0001	2,0019	5,9354	4,3386	0,0002	0,0116
6000	0,0000	0,0010	0,0007	1,9972	6,0645	4,4194	0,0002	0,0082
7000	0,0003	0,0010	0,0000	1,9937	6,1658	4,4850	0,0000	0,0061
8000	0,0004	0,0010	0,0000	1,9899	6,2780	4,5598	0,0001	0,0043

SOME CLASSES OF FUZZY INTERSECTIONS (t -NORMS)

Apêndice 5

	Formula $i(a, b)$	Decreasing generator $f(a)$	Parameter range	As parameter converges to 0	As parameter converges to 1 or -1	As converg
	$\left\{ 1 + \left[\left(\frac{1}{a} - 1 \right)^\lambda + \left(\frac{1}{b} - 1 \right)^\lambda \right]^{\frac{1}{\lambda}} \right\}^{-1}$	$\left(\frac{1}{a} - 1 \right)^\lambda$	$\lambda > 0$	$i_{\min}(a, b)$	$\frac{ab}{a+b-ab}$ when $\lambda = 1$	
	$\log_s \left[1 + \frac{(s^a - 1)(s^b - 1)}{s - 1} \right]$	$-\ln \left(\frac{s^a - 1}{s - 1} \right)$	$s > 0, s \neq 1$	$\min(a, b)$	ab as $s \rightarrow 1$	$\max(a, b)$
	$\frac{ab}{r + (1-r)(a+b-ab)}$	$-\ln \left(\frac{a}{r + (1-r)a} \right)$	$r > 0$	$\frac{ab}{a+b-ab}$	ab when $r = 1$	
	$(\max(0, a^p + b^p - 1))^{\frac{1}{p}}$	$1 - a^p$	$p \neq 0$	ab	$\max(0, a + b - 1)$, when $p = 1$; $\frac{ab}{a+b-ab}$, when $p = -1$.	$i_{\min}(a, b)$ $\min(a, b)$
ar 2	$1 - \frac{[(1-a)^p + (1-b)^p - (1-a)^p(1-b)^p]^{\frac{1}{p}}}{1}$	$\ln[1 - (1-a)^p]^{\frac{1}{p}}$	$p > 0$	$i_{\min}(a, b)$	ab when $p = 1$	
ar 3	$\exp(-(\ln a ^p + \ln b ^p)^{\frac{1}{p}})$	$ \ln a ^p$	$p > 0$	$i_{\min}(a, b)$	ab when $p = 1$	
ar 4	$\frac{ab}{[a^p + b^p - a^p b^p]^{\frac{1}{p}}}$	$a^{-p} - 1$	$p > 0$	ab	$\frac{ab}{a+b-ab}$, when $p = 1$	
	$1 - \min \left\{ 1, \left[(1-a)^w + (1-b)^w \right]^{\frac{1}{w}} \right\}$	$(1-a)^w$	$w > 0$	$i_{\min}(a, b)$	$\max(0, a + b - 1)$ when $w = 1$	
(1980)	$\frac{ab}{\max(a, b, \alpha)}$		$\alpha \in [0, 1]$	$\min(a, b)$	ab when $\alpha = 1$	
	$\max \left(0, \frac{a + b + \lambda ab - 1}{1 + \lambda} \right)$	$\frac{1}{\lambda} \ln[1 + \lambda(1-a)]$	$\lambda > -1$	$\max(0, a + b - 1)$	$i_{\min}(a, b)$ as $\lambda \rightarrow -1$; $\max\{0, (a + b + ab - 1)/2\}$ when $\lambda = 1$.	
	$\max\{0, (1 + \lambda)(a + b - 1) - \lambda ab\}$	$\frac{1}{\lambda} \ln \frac{1 + \lambda}{1 + \lambda a}$	$\lambda > -1$	$\max(0, a + b - 1)$	ab as $\lambda \rightarrow -1$; $\max\{0, 2(a + b - ab/2 - 1)\}$ when $\lambda = 1$.	

Apêndice 6

3 SOME CLASSES OF FUZZY UNIONS (t-CONORMS)

	Formula $u(a, b)$	Increasing generator $g(a)$	Parameter range	As parameter converges to 0	As parameter converges to 1 or -1	As para converges to
[2]	$\left\{ 1 + \left[\left(\frac{1}{a} - 1 \right)^\lambda + \left(\frac{1}{b} - 1 \right)^\lambda \right]^{-\frac{1}{\lambda}} \right\}^{-1}$	$\left(\frac{1}{a} - 1 \right)^{-\lambda}$	$\lambda > 0$	$u_{\max}(a, b)$	$\frac{a + b - 2ab}{1 - ab}$ when $\lambda = 1$	$\max(a, b)$
[9]	$1 - \log_s \left[1 + \frac{(s^{1-a} - 1)(s^{1-b} - 1)}{s - 1} \right]$	$-\ln \left(\frac{s^{1-a} - 1}{s - 1} \right)$	$s > 0, s \neq 1$	$\max(a, b)$	$a + b - ab$ as $s \rightarrow 1$	$\min(1, a + b)$
[1978]	$\frac{a + b + (r - 2)ab}{r + (r - 1)ab}$	$-\ln \left(\frac{1 - a}{r + (1 - r)(1 - a)} \right)$	$r > 0$	$\frac{a + b - 2ab}{1 - ab}$	$a + b - ab$ when $r = 1$	$u_{\max}(a, b)$
[1963]	$1 - \{\max(0, (1 - a)^p + (1 - b)^p - 1)\}^{\frac{1}{p}}$	$1 - (1 - a)^p$	$p \neq 0$	$a + b - ab$	$\min(1, a + b)$, when $p = 1$; $\frac{a + b - 2ab}{1 - ab}$, when $p = -1$.	$u_{\max}(a, b)$ as $p \rightarrow \infty$; $\min(a, b)$ as $p \rightarrow -\infty$.
[1978] & Sklar 2	$[a^p + b^p - a^p b^p]^{\frac{1}{p}}$	$\ln[1 - a^p]^{\frac{1}{p}}$	$p > 0$	$u_{\max}(a, b)$	$a + b - ab$ when $p = 1$	$\max(a, b)$
[1978] & Sklar 3	$1 - \exp(-(\ln(1 - a) ^p + \ln(1 - b) ^p)^{\frac{1}{p}})$	$ \ln(1 - a) ^p$	$p > 0$	$u_{\max}(a, b)$	$a + b - ab$ when $p = 1$	$\max(a, b)$
[1978] & Sklar 4	$1 - \frac{(1 - a)(1 - b)}{[(1 - a)^p + (1 - b)^p - (1 - a)^p(1 - b)^p]^{\frac{1}{p}}}$	$(1 - a)^{-p} - 1$	$p > 0$	$a + b - ab$	$\min \left(1, \frac{a + b}{1 - ab} \right)$ when $p = 1$	$\max(a, b)$
[1978] of]	$\min \left[1, (a^w + b^w)^{\frac{1}{w}} \right]$	a^w	$w > 0$	$u_{\max}(a, b)$	$\min(1, a + b)$ when $w = 1$	$\max(a, b)$
[1978] rade [1980]	$1 - \frac{(1 - a)(1 - b)}{\max((1 - a), (1 - b), \alpha)}$		$\alpha \in [0, 1]$	$\max(a, b)$	$a + b - ab$ when $\alpha = 1$	
[1978] 3]	$\min \left(1, a + b - \frac{\lambda}{1 - \lambda} ab \right)$	$\frac{1}{\lambda} \ln \frac{1 + \lambda}{1 + \lambda(1 - a)}$	$\lambda > -1$	$\min(1, a + b)$	$u_{\max}(a, b)$ as $\lambda \rightarrow -1$; $\min(1, (a + b - ab)/2)$ when $\lambda = 1$.	$a + b - ab$
[1978]	$\min(1, a + b + \lambda ab)$	$\frac{1}{\lambda} \ln(1 + \lambda a)$	$\lambda > -1$	$\min(1, a + b)$	$a + b - ab$ as $\lambda \rightarrow -1$; $\min(1, a + b + ab)$ when $\lambda = 1$.	$u_{\max}(a, b)$