
Agrupamento de Dados Semissupervisionado
na Geração de Regras Fuzzy

Priscilla de Abreu Lopes

Agrupamento de Dados Semissupervisionado na Geração de Regras Fuzzy

Priscilla de Abreu Lopes

Orientadora: *Profa. Dra. Heloisa de Arruda Camargo*

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de São Carlos, para obtenção do título de Mestre em Ciência da Computação.

UFSCar – São Carlos
Agosto/2010

**Ficha catalográfica elaborada pelo DePT da
Biblioteca Comunitária da UFSCar**

L864ad Lopes, Priscilla de Abreu.
Agrupamento de dados semissupervisionado na geração
de regras fuzzy / Priscilla de Abreu Lopes. -- São Carlos :
UFSCar, 2015.
74 f.

Dissertação (Mestrado) -- Universidade Federal de São
Carlos, 2010.

1. Inteligência artificial. 2. Aprendizado de máquina. 3.
Aprendizado semi-supervisionado. 4. Agrupamento fuzzy de
dados. 5. Sistemas baseados em regras fuzzy. I. Título.

CDD: 006.3 (20ª)

Universidade Federal de São Carlos

Centro de Ciências Exatas e de Tecnologia

Programa de Pós-Graduação em Ciência da Computação

“Agrupamento de Dados semi-supervisionado na Geração de Regras Fuzzy”

PRISCILLA DE ABREU LOPES

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de São Carlos, como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação

Membros da Banca:



Profa. Dra. Heloisa de Arruda Camargo
(Orientadora - DC/UFSCar)



Prof. Dr. Estevam Rafael Hruschka Júnior
(DC/UFSCar)



Profa. Dra. Maria do Carmo Nicoletti
(PPG-CC/UFSCar)



Prof. Dr. Alexandre Evsukoff
(UERJ)

São Carlos
Agosto/2010

Resumo

O aprendizado indutivo é, tradicionalmente, dividido em supervisionado e não supervisionado. No aprendizado supervisionado é fornecido ao método de aprendizado um conjunto de dados rotulados (dados que tem a classe conhecida). Estes dados são adequados para problemas de classificação e regressão. No aprendizado não supervisionado são analisados dados não rotulados, com o objetivo de identificar estruturas embutidas no conjunto.

Tipicamente, métodos de agrupamento não se utilizam de conhecimento prévio, como rótulos de classes, para desempenhar sua tarefa. A característica de conjuntos de dados atuais, grande volume e estruturas de atributos mistas, contribui para a busca de melhores soluções para tarefas de aprendizado de máquina.

É neste contexto em que se encaixa esta proposta de pesquisa. Trata-se da aplicação de métodos de agrupamento fuzzy semi-supervisionados na geração de bases de regras fuzzy. Os métodos de agrupamento semi-supervisionados realizam sua tarefa incorporando algum conhecimento prévio a respeito do conjunto de dados. O resultado do agrupamento é, então, utilizado para rotulação do restante do conjunto. Em seguida, entram em ação algoritmos de aprendizado supervisionado que tem como objetivo gerar regras fuzzy.

Este documento contém conceitos teóricos para compreensão da proposta de trabalho e uma discussão a respeito do contexto onde se encaixa a proposta. Alguns experimentos foram realizados a fim de mostrar que esta pode ser uma solução interessante para tarefas de aprendizado de máquina que encontram dificuldades devido à falta de informação disponível sobre dados.

Palavras-chave: Aprendizado Semi-Supervisionado, Agrupamento Fuzzy de Dados, Geração de Regras Fuzzy

Abstract

Inductive learning is, traditionally, categorized as supervised and unsupervised. In supervised learning, the learning method is given a labeled data set (classes of data are known). Those data sets are adequate for problems of classification and regression. In unsupervised learning, unlabeled data are analyzed in order to identify structures embedded in data sets.

Typically, clustering methods do not make use of previous knowledge, such as classes labels, to execute their job. The characteristics of recently acquired data sets, great volume and mixed attribute structures, contribute to research on better solutions for machine learning jobs.

The proposed research fits into this context. It is about semi-supervised fuzzy clustering applied to the generation of sets of fuzzy rules. Semi-supervised clustering does its job by embodying some previous knowledge about the data set. The clustering results are, then, useful for labeling the remaining unlabeled data in the set. Following that, come to action the supervised learning algorithms aimed at generating fuzzy rules.

This document contains theoretic concepts, that will help in understanding the research proposal, and a discussion about the context wherein is the proposal. Some experiments were set up to show that this may be an interesting solution for machine learning jobs that have encountered difficulties due to lack of available information about data.

Keywords: Semi-Supervised Learning, Fuzzy Data Clustering, Fuzzy Rules Generation

Lista de Figuras

2.1	Representação gráfica das funções de pertinência: (a) Triangular, (b) Trapezoidal e (c) Gaussiana	7
2.2	Variável Linguística: Temperatura	10
3.1	Exemplo de estrutura resultante de um agrupamento hierárquico, com possíveis divisões de níveis (linhas tracejadas). Cada retângulo representa um exemplo do conjunto de dados.	28
3.2	Exemplo de resultado de um agrupamento particional, que dividiu o conjunto de dados em quatro grupos. Cada ponto representa um exemplo do conjunto de dados.	29
3.3	Ilustração do processo de agrupamento. (a) conjunto de dados e agrupamento desejado. (b) Grupos formados em uma iteração t . (c) Grupo menos bem-definido (destacado) é o que tem menor densidade. (d) Exemplos à fronteira do grupo menos bem-definido (círculos cheios) são aqueles com menor grau de pertinência entre os dados do grupo. (e) Para cada exemplo de fronteira, é selecionado (fronteira tracejada) o grupo mais próximo (segundo maior grau de pertinência para o exemplo). São verificadas as restrições para o exemplo de fronteira. (f) adaptação dos grupos para não violação da restrição <i>must-link</i> entre os exemplos. (Girra et al., 2005)	40

Lista de Algoritmos

3.1 K-Means	30
3.2 Fuzzy C-Means	32
3.3 semi-supervised FCM	37
3.4 partially supervised Gustafson and Kessel	39
3.5 Active Fuzzy Constrained Clustering	41
3.6 SEEDED K-Means	42
3.7 CONSTRAINED K-Means	42
3.8 COP K-Means	43
3.9 PCK-Means	44
3.10 Semi-Supervised Point Prototype Clustering	45

Lista de Tabelas

5.1	Conjuntos de dados utilizados nos experimentos para validação	55
5.2	Índice de heterogeneidade geral (R) para os métodos FCM	57
5.3	Índice de heterogeneidade geral (R) para os métodos ssFCM	58
5.4	TCC para conjuntos de dados O	59
5.5	TCC para conjuntos de dados agrupado P	59
5.6	TCC para conjuntos de dados agrupado A	60
5.7	Teste de Friedman com post-hoc de Dunn para os conjuntos O	60
5.8	Teste de Friedman com post-hoc de Dunn para os conjuntos P	61
5.9	Teste de Friedman com post-hoc de Dunn para os conjuntos A	61
5.10	Somas de Rank por algoritmo para cada tipo de conjunto (O, P e A) . .	61
5.11	Teste de Friedman com post-hoc de Dunn para algoritmo Ishib-Hybrid	62
5.12	Teste de Friedman com post-hoc de Dunn para o algoritmo SGERD . .	62
5.13	Teste de Freidman com post-hoc de Dunn para o algoritmo SLAVE . .	62

Lista de Abreviaturas

- AFCC:** *Active Fuzzy Constrained Clustering*
- AG:** *Algoritmo Genético*
- AM:** *Aprendizado de Máquina*
- BC:** *Base de Conhecimento*
- BD:** *Base de Dados*
- BR:** *Base de Regras*
- FCM:** *Fuzzy C-means*
- GK:** *Gustafson and Kessel*
- psGK:** *Partially Supervised Gustafson and Kessel*
- RF:** *Regras Fuzzy*
- SCF:** *Sistemas de Classificação Fuzzy*
- SF:** *Sistema Fuzzy*
- SBRF:** *Sistemas Baseados em Regras Fuzzy*
- SFGBR:** *Sistemas Fuzzy Genéticos Baseados em Regras*
- ssFCM:** *Semi-Supervised Fuzzy C-Means*
- ssPPC:** *Semi-Supervised Point Prototype Clustering*
- SSQ:** *Sum of Squares*

Sumário

1	Introdução	1
1.1	Motivações e Objetivos	2
1.2	Organização do Trabalho	4
2	Sistemas Baseados em Regras Fuzzy	5
2.1	Conjuntos Fuzzy	5
2.1.1	Funções de Pertinência	6
2.1.2	Operações Sobre Conjuntos Fuzzy	7
2.1.3	Relações Fuzzy	9
2.1.4	Varáveis Linguísticas	10
2.2	Computação com Regras Fuzzy	10
2.2.1	Regras Fuzzy	11
2.2.2	Semântica das Regras e Inferência Fuzzy	12
2.3	Sistemas Fuzzy	13
2.3.1	Sistemas de Classificação Fuzzy	15
2.4	Considerações Finais	17
3	Agrupamento de Dados Semissupervisionado	18
3.1	Aprendizado de Máquina	18
3.1.1	Aprendizado Supervisionado	19
3.1.2	Aprendizado Não Supervisionado	20
3.1.3	Aprendizado Semissupervisionado	20
3.2	Processo de Agrupamento	22
3.2.1	Preparação de Exemplos	23
3.2.2	Medidas de Similaridade	24
3.2.3	Técnicas de Agrupamento Não Supervisionado	28

3.2.4	Validação	31
3.2.5	Interpretação dos Resultados	35
3.3	Agrupamento de Dados Semissupervisionado	35
3.3.1	Semi-Supervised Fuzzy C-Means	36
3.3.2	Partially Supervised Gustafson and Kessel	37
3.3.3	Active Fuzzy Constrained Clustering	38
3.3.4	Outros Algoritmos	39
3.4	Considerações Finais	44
4	Agrupamento de Dados Semissupervisionado na Geração de Regras Fuzzy	46
4.1	Agrupamento de Dados	46
4.2	Sistemas Baseados em Regras Fuzzy	47
4.2.1	Sistemas Fuzzy Genéticos	49
4.3	Agrupamento de Dados na Geração de Regras Fuzzy	50
4.4	Trabalho Proposto	51
4.5	Considerações Finais	53
5	Experimentos e Resultados	54
5.1	Considerações Preliminares	54
5.1.1	Validação de Agrupamento	55
5.1.2	Validação de Classificação	56
5.2	Resultados de Experimentos e Análise	57
5.3	Considerações Finais	63

Introdução

Nos últimos anos, questões relacionadas ao aprendizado de máquina e aquisição do conhecimento têm sido focalizadas com frequência nas pesquisas relacionadas a Sistemas Fuzzy. Os Sistemas Fuzzy são sistemas fundamentados na teoria de conjuntos fuzzy e caracterizam-se principalmente por utilizar variáveis linguísticas para representação de pelo menos uma parte do conhecimento. A utilidade desses sistemas é reconhecida nos trabalhos reportados na literatura e, ainda nos dias de hoje, motiva investigações sobre metodologias para construção e otimização automática dos mesmos, a partir de conjuntos de dados.

A área de aprendizado de máquina, que tem como objetivo o estudo de métodos para extrair conhecimento de dados, distingue entre aprendizado supervisionado, que utiliza dados com classes conhecidas, ditos rotulados e não supervisionado, que utiliza dados sem classe conhecida, ou não rotulados. Repetidamente nos deparamos com a argumentação, nos trabalhos científicos, sobre a necessidade de investimentos de pesquisa para o tratamento do grande volume de dados disponíveis nas mais diversas áreas do conhecimento, devido à grande facilidade de aquisição, acompanhada pelo alto custo de interpretação e rotulação manual desses dados. Essa situação nos remete ao estudo de mecanismos de aprendizado que considerem tanto dados rotulados quanto dados não rotulados, que tem sido chamado de aprendizado semissupervisionado.

Especificamente na construção de sistemas baseados em regras fuzzy, o tipo particular de sistemas fuzzy de interesse para este trabalho, têm sido aplicadas técnicas tanto de aprendizado supervisionado quando de aprendizado não-supervisionado.

Entre as técnicas de aprendizado supervisionado, destacam-se as abordagens que envolvem métodos híbridos que combinam a teoria de conjuntos fuzzy com outras metodologias, especialmente algoritmos evolutivos e redes neurais. Na combinação com redes neurais, surgem os sistemas neuro-fuzzy, que codificam um sistema baseado em regras fuzzy em uma rede neural e aplicam mecanismos de aprendizado para treinar a rede e assim definir parâmetros do sistema fuzzy. Na combinação com algoritmos evolutivos, uma grande variedade de abordagens pode ser encontrada, desde a otimização de parâmetros dos conjuntos fuzzy, passando pela construção ou seleção das regras fuzzy, até a geração do sistema fuzzy completo.

Ainda nesta categoria de aprendizado supervisionado encontram-se também as variações de árvore de decisão, que são formas estendidas da árvore de decisão tradicional para geração de regras fuzzy. Entre as técnicas de aprendizado não supervisionado, encontramos a análise de grupos como um mecanismo explorado para gerar regras e conjuntos fuzzy, a partir de dados não rotulados. A abordagem mais frequente é aplicar um algoritmo de agrupamento e transformar o resultado em uma base de regras fuzzy, em que cada grupo é uma regra. Os conjuntos fuzzy em cada dimensão são obtidos por meio de projeções dos grupos. Muitas vezes essas são projeções são aproximadas para conjuntos da forma triangular ou trapezoidal.

O estágio atual das pesquisas evidencia que a combinação de metodologias traz benefícios para o processo de aprendizado como um todo, por permitir o balanceamento de vantagens entre métodos complementares. Este trabalho investiga a combinação de algoritmos de agrupamento semissupervisionado na construção de regras fuzzy.

1.1 Motivações e Objetivos

O principal objetivo deste trabalho é investigar a combinação de agrupamento semissupervisionado com sistemas baseados em regras fuzzy.

A organização de dados em grupos é uma das formas mais fundamentais para analisar características intrínsecas dos dados. O objetivo do processo de agrupamento é encontrar uma estrutura dentro de conjuntos de dados. A análise de grupos é o estudo de algoritmos e métodos de agrupamento de objetos de acordo com a similaridade entre eles.

Em variados domínios podemos identificar a importância da divisão de dados em classes que tenham algum significado. Tarefas de agrupamento de dados ocorrem em diversificados tipos de aplicação, como processamento de imagens, classi-

ificação e organização de textos, gerenciamento de recursos, estudo de genoma. A aplicação de agrupamento fuzzy visa encontrar estruturas dentro de conjuntos de dados que sejam mais maleáveis quanto ao particionamento dos dados, resultado que pode ser aproveitado para o desenvolvimento de outros sistemas. É frequente a associação de algoritmos de agrupamento a outros processos, como a geração de modelos de inferência ou outras categorias de classificadores.

Tipicamente, métodos de agrupamento não se utilizam de conhecimento prévio, como rótulos de classes, para desempenhar sua tarefa. A característica de conjuntos de dados atuais, grande volume e estruturas mistas com atributos de tipos variados (como texto, *streamings* de vídeo e áudio), contribui para a busca de melhores soluções para tarefas de aprendizado de máquina.

Pesquisas que envolvem agrupamento semissupervisionado tem o objetivo de gerar estruturas mais fiéis à realidade dos conjuntos de dados. Estruturas sobre as quais, ou conjuntamente, possam ser aplicados outros métodos de aprendizado.

Enquanto a geração de regras fuzzy é, na maioria das vezes, feita por mecanismos que requerem dados rotulados, a crescente disponibilidade de dados não rotulados evidencia a necessidade de pesquisar métodos adequados às situações como essas, nas quais não se dispõe de informação prévia relativa às classes. Os métodos de agrupamento, por outro lado, adequados à análise de dados não rotulados, apresentam dificuldades relativas à validação e interpretação dos grupos resultantes do processo, o que conseqüentemente dificulta sua utilização. O método de geração de regras proposto visa estudar alternativas de superação de tais dificuldades.

O objetivo deste trabalho é estudar e analisar o uso de agrupamento semissupervisionado na geração de regras fuzzy. A abordagem utilizada diferencia da mais usual, e consiste de duas etapas: a primeira, é uma etapa de agrupamento semissupervisionado dos dados e rotulação dos mesmos e a segunda etapa é a geração das regras propriamente dita, usando os dados rotulados na etapa anterior, aplicando métodos próprios para classificação, que requerem que todos os dados estejam rotulados.

A motivação para o estudo de métodos de agrupamento na geração de regras fuzzy decorre também da tentativa de ampliar o foco de trabalho de pesquisa do grupo, incluindo outra metodologia para o estudo de sistemas híbridos na modelagem fuzzy, além daquelas que têm sido investigadas. O trabalho do grupo de pesquisa focaliza principalmente a geração automática de bases de conhecimento por meio de metodologias como algoritmos genéticos e agrupamento. Os resultados obtidos até o momento incluem algumas propostas de métodos de geração e otimização de bases de conhecimento fuzzy por meio de algoritmos genéticos,

que tratam questões como a geração de regras fuzzy pela abordagem de Pittsburgh (Castro; Camargo, 2005), otimização da base de dados (Pires; Camargo, 2004), geração genética com pré-seleção de regras candidatas (Cintra; Camargo, 2007; Cintra et al., 2008), geração de classificadores intervalares tipo-2 (Pimenta; Camargo, 2010b,a), entre outras. Mais recentemente foi também explorado o uso de agrupamento fuzzy condicional com a finalidade de gerar regras que contém tanto variáveis como grupos no antecedente (Nogueira; Camargo, 2010) além das árvores de decisão fuzzy (Cintra; Camargo, 2010; Cintra et al., 2010). O trabalho descrito aqui visa ampliar os resultados obtidos com a exploração dos métodos de agrupamento semissupervisionados na geração de regras.

1.2 Organização do Trabalho

Esta pesquisa foi organizada com base nos principais assuntos focalizados: Sistemas Baseados em Regras Fuzzy, Agrupamento de Dados Semissupervisionado e sua aplicação no contexto de Geração de Regras. O embasamento teórico dos assuntos mencionados possibilitou o desenvolvimento da proposta e sua validação por meio de experimentos.

Neste documento são descritos os resultados deste trabalho de pesquisa e está organizado como segue:

Capítulo 2 - Sistemas Baseados em Regras Fuzzy. Neste capítulo são descritos conceitos básicos da teoria de conjuntos e lógica fuzzy, destacando sua aplicação a sistemas de classificação fuzzy.

Capítulo 3 - Agrupamento de Dados Semissupervisionado. Neste capítulo são abordados conceitos básicos de agrupamento de dados não supervisionado e semissupervisionado dentro do contexto de aprendizado de máquina e, posteriormente descritos algoritmos relevantes para o trabalho.

Capítulo 4 - Agrupamento de Dados Semissupervisionado na Geração de Regras Fuzzy. Este é o capítulo onde é apresentada a contextualização e descrição do projeto de pesquisa, citando uma série de abordagens que podem ser empregadas a tarefas de agrupamento de dados, geração de regras fuzzy e, ainda, extensões desenvolvidas para o tratamento de aprendizado semissupervisionado.

Capítulo 5 - Experimentos e Resultados. Este capítulo apresenta detalhes dos experimentos realizados para validação da proposta de trabalho, seus resultados e análises feitas a respeito deles.

Capítulo 6 - Conclusão. Este capítulo é a finalização do trabalho e inclui sugestões para melhorar e refinar a pesquisa iniciada no tema deste trabalho.

Sistemas Baseados em Regras Fuzzy

O trabalho de Zadeh (1965), apresentou pela primeira vez a teoria de conjuntos fuzzy e a lógica fuzzy. Os sistemas baseados nestas teorias são chamados Sistemas Fuzzy. Estes sistemas utilizam um mecanismo de raciocínio baseado no raciocínio aproximado, permitindo a expressão da ambiguidade e subjetividade presentes no raciocínio humano (Pedrycz; Gomide, 1998).

Neste capítulo são descritos alguns conceitos sobre conjuntos fuzzy, funções de pertinência, operações sobre conjuntos fuzzy e relações fuzzy, além de uma introdução à inferência utilizando regras fuzzy.

As definições da teoria de conjuntos fuzzy apresentadas neste capítulo foram extraídas de Klir; Yuan (1995) e Nicoletti; Camargo (2004).

2.1 Conjuntos Fuzzy

Na teoria clássica dos conjuntos, a pertinência de um elemento a um conjunto é representada por uma função característica que associa o valor 1 (pertence) e 0 (não pertence) aos elementos de um conjunto universo. Para muitas das situações do mundo real, a representação por conjuntos clássicos é insuficiente para a descrição de elementos, pois essa teoria força a pertinência absoluta ou não absoluta de um elemento a um conjunto.

A imprecisão associada às situações reais pode ser representada de maneira mais flexível por meio dos conjuntos fuzzy, que possibilitam a transição gradual de significado. A teoria de conjuntos fuzzy generaliza a teoria clássica de conjuntos

(conjuntos *crisp*), e permite que um elemento pertença, com diferentes graus, a mais de um conjunto.

A generalização da função característica dos conjuntos clássicos associa, a cada elemento do conjunto universo, um valor, dentro de um intervalo, que reflete o quanto o elemento pertence ao conjunto sendo definido. Conjuntos fuzzy são definidos por esse tipo de função, chamada função de pertinência. Seja X um conjunto (*crisp*) universo e A um conjunto fuzzy, cuja função de pertinência mapeia X no intervalo de números reais $[0, 1]$, essa função pode ser notada por:

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

onde $A(x)$ representa o grau com que o elemento $x \in X$ pertence ao conjunto fuzzy A .

2.1.1 Funções de Pertinência

As funções de pertinência mais utilizadas nas aplicações práticas geralmente são apresentadas nos formatos triangular, trapezoidal ou gaussiano devido à sua simplicidade de representação e utilização. No entanto, o conjunto fuzzy está bastante ligado ao contexto do domínio, o que pode fazer com que tenha formas específicas, diferentes das mencionadas. A seguir são definidos os formatos mais comuns e sua representação gráfica é apresentada na Figura 2.1.

- *Triangular*: parâmetros (a, m, b) , tal que $a \leq m \leq b$

$$A(x) = \begin{cases} 0 & \text{se } x \leq a \\ \frac{x-a}{m-a} & \text{se } x \in (a, m) \\ 1 & \text{se } x = m \\ \frac{b-x}{b-m} & \text{se } x \in (m, b) \\ 0 & \text{se } x \geq b \end{cases}$$

- *Trapezoidal*: parâmetros (a, m, n, b) , tal que $a \leq m < n \leq b$

$$A(x) = \begin{cases} 0 & \text{se } x \leq a \\ \frac{x-a}{m-a} & \text{se } x \in (a, m) \\ 1 & \text{se } x \in [m, n] \\ \frac{b-x}{b-n} & \text{se } x \in (n, b) \\ 0 & \text{se } x \geq b \end{cases}$$

- *Gaussiana*: parâmetros (m, k) , tal que $k > 0$

$$A(x) = e^{-k(x-m)^2}$$

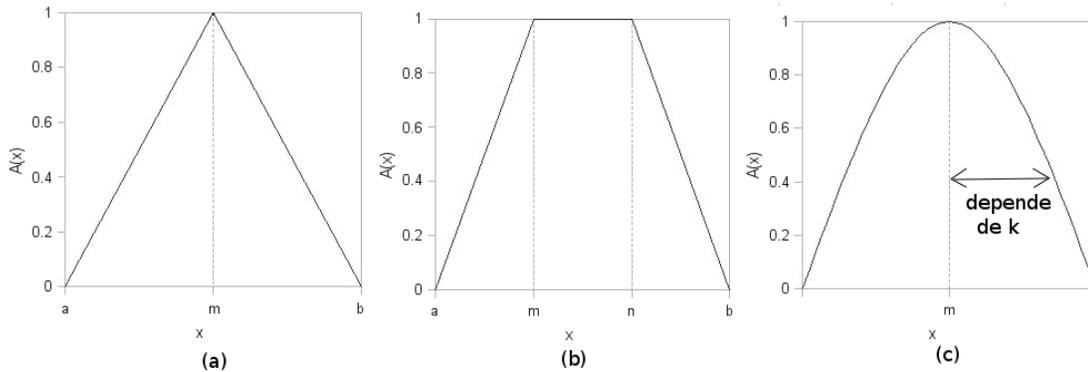


Figura 2.1: Representação gráfica das funções de pertinência: (a) Triangular, (b) Trapezoidal e (c) Gaussiana

2.1.2 Operações Sobre Conjuntos Fuzzy

Na teoria clássica de conjuntos são definidas as operações de complemento, união e intersecção de conjuntos *crisp*. Essas operações deixam de ser únicas quando estendidas a conjuntos fuzzy, podendo ser implementadas por uma classe de operações. Funções que se qualificam como complemento fuzzy estão dentro da classe chamada complementos fuzzy, as que se qualificam como intersecção fuzzy pertencem à classe chamada t-normas, e aquelas que se qualificam como união são encontradas na classe chamada s-normas.

A generalização das operações e alguns membros de suas classes são apresentadas a seguir:

Complemento Fuzzy

Denota-se cA o complemento generalizado fuzzy de A do tipo c , onde c é definida pela função $c : [0, 1] \rightarrow [0, 1]$. Logo, $cA(x)$ pode ser interpretado como o grau com que x pertence a cA e, também, como o grau com que x não pertence a A . A função c deve satisfazer os seguintes axiomas:

1. **Condições Limites:** $c(0) = 1$ e $c(1) = 0$
2. **Mono-tonicidade:** para todo $a, b \in [0, 1]$, se $a \leq b$, então $c(a) \geq c(b)$

3. **Continuidade:** c é uma função contínua

4. **Involução:** c é involutiva, ou seja $c(c(a)) = a$, para cada $a \in [0, 1]$

Seja X o conjunto universo e x um elemento de X . O complemento padrão do conjuntos fuzzy A sobre X é definido para todo $x \in X$ por:

$$\bar{A}(x) = 1 - A(x) \quad (2.1)$$

Intersecção Fuzzy

Sejam A e B conjuntos fuzzy definidos em um universo X . A intersecção fuzzy, notada por $A \cap B$, é definida para todo $x \in X$ pela equação

$$(A \cap B)(x) = i_{x \in X}\{A(x), B(x)\} \quad (2.2)$$

onde i é uma t-norma, ou seja, i é uma operação binária da forma $i : [0, 1]^2 \rightarrow [0, 1]$ que deve satisfazer os axiomas abaixo, para todo $a, b, c \in [0, 1]$:

1. **Condições Limites:** $i(a, 1) = a$
2. **Mono-tonicidade:** $b \leq c$ implica $i(a, b) \leq i(a, c)$
3. **Comutatividade:** $i(a, b) = i(b, a)$
4. **Associatividade:** $i(a, i(b, c)) = i(i(a, b), c)$

Operações t-norma frequentemente utilizadas:

- *Intersecção padrão:* $i(a, b) = \min(a, b)$
- *Produto algébrico:* $i(a, b) = ab$
- *Diferença limitada:* $i(a, b) = \max(0, a + b - 1)$

União Fuzzy

Sejam A e B conjuntos fuzzy definidos em um universo X . A união fuzzy, notada por $A \cup B$, é definida para todo $x \in X$ pela equação

$$(A \cup B)(x) = u_{x \in X}\{A(x), B(x)\} \quad (2.3)$$

onde u é uma s-norma (ou t-co-norma), ou seja, u é uma operação binária da forma $u : [0, 1]^2 \rightarrow [0, 1]$ que deve satisfazer os axiomas abaixo, para todo $a, b, c \in [0, 1]$:

1. **Condições Limites:** $u(a, 0) = a$
2. **Mono-tonicidade:** $b \leq c$ implica $u(a, b) \leq u(a, c)$
3. **Comutatividade:** $u(a, b) = u(b, a)$
4. **Associatividade:** $u(a, u(b, c)) = u(u(a, b), c)$

Operações s-norma frequentemente utilizadas:

- *União padrão:* $u(a, b) = \max(a, b)$
- *Soma algébrica:* $a + b - ab$
- *Soma limitada:* $u(a, b) = \min(1, a + b)$

2.1.3 Relações Fuzzy

Assim como operações fuzzy generalizam o conceito de operações sobre conjuntos convencionais, relações fuzzy generalizam o conceito de relações da teoria clássica e representam o grau da associação entre elementos de dois ou mais conjuntos. Estas relações, definidas em universos que são produtos cartesianos, são conjuntos fuzzy.

Sejam X e Y conjuntos *crisp*, a relação fuzzy binária $R(X, Y)$, é um conjunto fuzzy caracterizado pela função de pertinência $R(x, y)$, com $x \in X$ e $y \in Y$, que assume valores no intervalo $[0, 1]$. Estes valores definem o grau com que x e y estão relacionados, pela relação R , ou seja:

$$R : X \times Y \rightarrow [0, 1]$$

Quando o conjunto universo é discreto, uma relação binária fuzzy pode ser representada por lista de pares ordenados precedidos dos graus de pertinência relativos a cada par.

A representação da relação binária fuzzy entre os conjuntos $A = \{a_1, a_2, a_3\}$ e $B = \{b_1, b_2\}$ é descrita abaixo usando a notação de lista, na qual os números antes da barra representam o grau de pertinência de cada par:

$$R(A, B) = 0,5/(a_1, b_1) + 1/(a_1, b_2) + 0,9/(a_2, b_1) + 0/(a_2, b_2) + 0,2/(a_3, b_1) + 0,7/(a_3, b_2)$$

2.1.4 Varáveis Linguísticas

A representação de conhecimento e de experiências de seres humanos é feita na forma linguística, o que dificulta sua utilização por computadores, incapazes de lidar com a imprecisão desta representação. Com o conceito de variáveis linguísticas, informações complexas e imprecisas passam a ser representadas de forma matemática, tornando o conhecimento manipulável por computadores.

Variáveis linguísticas são aquelas cujos valores são palavras ou sentenças em linguagem natural em vez de números (Zimmerman, 1991). Estas variáveis são definidas sobre um domínio, o qual é granularizado em termos linguísticos definidos por conjuntos fuzzy. O processo de granularização de um domínio de uma variável em conjuntos fuzzy define a chamada partição fuzzy.

Na Figura 2.2, pode-se observar uma partição fuzzy, cujo domínio da variável temperatura é granularizado nos termos linguísticos *alta*, *média-alta*, *média*, *média-baixa* e *baixa*. Assim, a variável temperatura também poderia assumir valores como *baixa*, *muito alta*, *mais ou menos média-baixa*, etc.

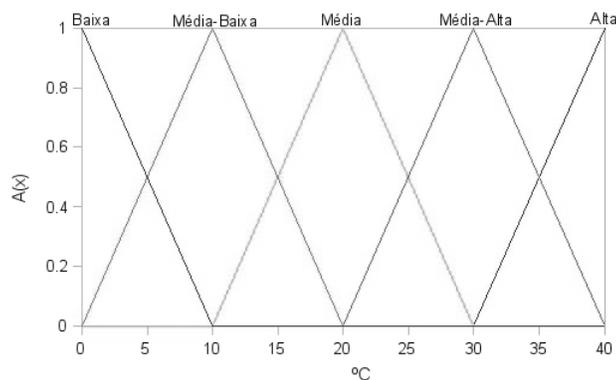


Figura 2.2: Variável Linguística: Temperatura

2.2 Computação com Regras Fuzzy

Regras podem ser utilizadas para representar conhecimento de um determinado domínio. Regras fuzzy permitem a representação de conhecimento impreciso, o que, em muitos casos, torna a representação mais fiel e adequada à realidade do domínio. Processos computacionais envolvendo regras, fuzzy ou não, consistem em aplicar mecanismos de raciocínio que farão inferências, baseados em um conjunto de regras e um conjunto de fatos sobre um domínio.

2.2.1 Regras Fuzzy

Uma técnica das mais antigas para representação de conhecimento são as bases de regras, que continuam a ser amplamente utilizadas por sua capacidade de expressar diretivas e estratégias, de captar o conhecimento de especialistas humanos e de possuir um formato linguístico de fácil compreensão.

A expressão de conhecimento impreciso por estas regras dá ao sistema fuzzy maior transparência e compreensibilidade.

Em geral, as regras fuzzy possuem o seguinte formato:

Se antecedente, então consequente

e tem a finalidade de estabelecer relações entre as variáveis que aparecem no antecedente e as que aparecem no consequente.

O antecedente e o consequente das regras fuzzy são sempre proposições, que representam informação do tipo “o (atributo) do (objeto) é (valor)”. Essas proposições podem ser atômicas ou compostas. As proposições atômicas especificam um valor linguístico A (valor) para uma variável linguística V (atributo) e tem a forma: V é A .

Uma regra fuzzy assume o formato de uma proposição condicional, que no caso mais simples fica:

Se X é A então Y é B

onde X e Y são variáveis linguísticas sobre os domínios \mathbb{X} e \mathbb{Y} , respectivamente, e A e B são termos linguísticos, que rotulam conjuntos fuzzy definidos sobre os domínios \mathbb{X} e \mathbb{Y} , respectivamente.

As proposições compostas podem ser formadas por conjunções ou disjunções de proposições atômicas. Por exemplos, as proposições:

$$\begin{array}{c} X_1 \text{ é } A \text{ e } X_2 \text{ é } B \\ \text{e} \\ X_1 \text{ é } C \text{ ou } X_2 \text{ é } D \end{array}$$

podem representar conhecimento do tipo “*A temperatura está alta e a umidade está baixa*” e “*A água está boa para consumo humano ou o trabalho é difícil*”.

Em uma regra fuzzy, tanto o antecedente quanto o consequente podem ser formados por proposições compostas, o que dá origem, por exemplo, a regras da forma:

Se X_1 é A_1 e X_2 é A_2 então Y_1 é B_1 ou Y_2 é B_2

onde A_1, A_2, B_1 e B_2 são conjuntos fuzzy nos universos $\mathbb{X}_1, \mathbb{X}_2, \mathbb{Y}_1, \mathbb{Y}_2$, respectivamente, e X_1, X_2, Y_1 e Y_2 são variáveis sobre os domínios $\mathbb{X}_1, \mathbb{X}_2, \mathbb{Y}_1, \mathbb{Y}_2$, respectivamente.

A parte consequente de uma regra fuzzy pode ser uma proposição fuzzy ou uma função, dependendo do modelo de regra utilizado. Na maioria das aplicações as regras fuzzy utilizadas tem um formato restrito, com conjunções no antecedente e uma única proposição no consequente, como:

$$\mathbf{Se } X_1 \text{ é } A_1 \mathbf{ e } \dots X_n \text{ é } A_n \mathbf{ então } Y \text{ é } B$$

Os conceitos apresentados na sequência referem-se a esse formato específico de regra, que será o formato tratado pela proposta de trabalho apresentada.

2.2.2 Semântica das Regras e Inferência Fuzzy

Uma proposição condicional de forma

$$\mathbf{Se } X \text{ é } A \mathbf{ então } Y \text{ é } B$$

descreve uma relação entre as variáveis X e Y , o que significa que uma regra desse tipo pode ser definida por uma relação fuzzy R . O grau de pertinência de um par (x, y) em R , representa o quanto esse par é compatível com a relação entre as variáveis X e Y da regra. A definição da relação R consiste na definição da semântica da regra.

Assim, podemos definir:

Seja A um conjunto fuzzy sobre \mathbb{X} , e B um conjunto fuzzy sobre \mathbb{Y} . A regra **Se** X é A **então** Y é B induz uma relação fuzzy $R(x, y)$ sobre \mathbb{X} e \mathbb{Y} definida por

$$R(x, y) = f((A(x), B(y)))$$

em que f é uma função da forma $f : [0, 1]^2 \rightarrow [0, 1]$. Em geral, f pode ser de uma entre três classes de funções: conjunção fuzzy, disjunção fuzzy e implicação fuzzy. Definições detalhadas dessas classes de funções fogem do escopo deste trabalho e não serão apresentadas aqui. As definições e discussões relativas a esses conceitos podem ser encontradas em Pedrycz; Gomide (1998).

A computação com regras fuzzy envolve a aplicação de um método de raciocínio, o qual está baseado em regras de inferência, que buscam tirar conclusões a partir de um conjunto de regras fuzzy. A regra de inferência composicional é a regra de inferência básica para os processos de raciocínio nos sistemas fuzzy.

A regra de inferência composicional, também chamada de regra da composição, será apresentada para o caso simples, isto é, quando a regra tem apenas uma variável no antecedente e uma no conseqüente.

O processo de inferência é aplicado a partir de uma regra, a qual estabelece um conhecimento sobre algum domínio, e de um fato, o qual estipula uma informação específica sobre a situação da qual deseja-se obter alguma conclusão.

Considere a regra **Se** X **é** A **então** Y **é** B definida anteriormente e a relação R induzida. Para usar essa regra em uma inferência, é necessário conhecer algum fato que estabeleça um valor específico para a variável do antecedente X .

Dado o fato X **é** A' , pode-se inferir a conclusão Y **é** B' , sendo que B' é definido pela equação 2.4 (Pedrycz; Gomide, 1998), para todo $y \in \mathbb{Y}$, onde sup denota o operador de supremo e t uma t-norma.

$$B'(y) = sup_{x \in X} [A'(x)tR(x, y)] \quad (2.4)$$

A aplicação da regra de inferência composicional pode ser facilmente estendida para regras com múltiplas variáveis, e também para situações em que o conhecimento do problema é descrito por um conjunto de regras. A regra da inferência composicional pode se tornar excessivamente custosa para regras com mais de uma condição no antecedente.

Como nas aplicações práticas o número de variáveis e o número de regras podem ser elevados, métodos simplificados foram definidos, baseados na regra de inferência composicional. Entre os mais conhecidos pode-se citar o método de inferência de Mamdani e o método de inferência escalonada. Outras modificações podem ser adicionadas a esses métodos simplificados dependendo do tipo de conseqüente da regra e do tipo de aplicação focalizado.

Aspectos relativos aos métodos de inferência são discutidos na próxima seção.

2.3 Sistemas Fuzzy

Sistemas fuzzy são aqueles que utilizam a lógica fuzzy em seu processo de raciocínio e as variáveis do problema são representadas por valores linguísticos definidos por conjuntos fuzzy.

Neste trabalho, o tipo específico de SF de interesse é o Sistema Baseado em Regras Fuzzy (SBRF), o qual é composto por dois componentes principais: a Base de Conhecimento (BC) e o Mecanismo de Inferência (MI).

A Base de Conhecimento é composta pela Base de Dados (BD), que contém as definições dos conjuntos fuzzy relacionados aos termos linguísticos usados nas

regras fuzzy e pela Base de Regras (BR), que armazena o conjunto de regras que modelam um determinado problema.

O Mecanismo de Inferência é responsável pelo processamento das regras, realizado por algum método de raciocínio. Este consiste da aplicação de um procedimento de inferência para derivar conclusões a partir das regras e de fatos conhecidos.

No trabalho proposto aqui, os métodos de inferência de interesse baseiam-se no processo de inferência escalonada, cujas descrições são definidas, genericamente, a seguir (Pedrycz; Gomide, 1998):

1. *Compatibilidade de antecedentes*: para cada regra, calcular a relação entre cada proposição atômica do antecedente da regra e a proposição atômica correspondente à proposição de entrada.
2. *Agregação de antecedentes*: para cada regra, calcular o nível de ativação da regra de forma conjuntiva ou disjuntiva sobre o grau correspondente à combinação, dependendo da operação entre proposições atômicas do antecedente da regra.
3. *Derivação do resultado da regra*: para cada regra, calcular o valor de inferência correspondente, baseado na agregação de antecedentes e na semântica da regra escolhida.
4. *Agregação de regras*: calcular o valor inferido por todo o conjunto de regras, agregando o resultado dos valores inferidos de cada regra.

O tipo de conseqüente usado nas regras distingue dois tipos de sistemas fuzzy: Mamdani (1977) e Takagi; Sugeno (1985). O primeiro, possui proposições fuzzy no antecedente e no conseqüente da regra onde X_1 a X_n representam as variáveis linguísticas de entrada, Y é uma variável linguística de saída e A_{i1_1} a A_{in_n} e B representam os termos linguísticos.

Se X_1 é A_{i1_1} **e** ... **e** X_n é A_{in_n} **então** Y é B .

No sistema fuzzy do tipo Mamdani o processo de inferência é executado com o operador de mínimo para o passo de conjunção de antecedentes e com a função de conjunção para a definição da semântica das regras. A agregação das regras é calculada usando o operador de máximo.

Os sistemas do tipo Mamdani obtêm como resultado do seu processamento de inferência um conjunto fuzzy que deve ser defuzificado para a obtenção de uma saída numérica precisa.

Já os sistemas do tipo Takagi-Sugeno possuem proposições fuzzy no antecedente da regra e o consequente da regra é uma função f aplicada sobre os valores de entrada x_1, \dots, x_n onde X_1 a X_n representam as variáveis linguísticas e A_{i1} a A_{in} representam os termos linguísticos.

$$\text{Se } X_1 \text{ é } A_{i1} \text{ e } \dots \text{ e } X_n \text{ é } A_{in} \text{ então } f(x_1, \dots, x_n).$$

No sistema fuzzy do tipo Takagi-Sugeno a etapa de derivação de resultado de cada regra é o cálculo do valor definido pela função do consequente. A agregação dos resultados individuais é feita por um operados do tipo s-norma.

Os sistemas do tipo Takagi-Sugeno já possuem uma saída numérica precisa como resultado da aplicação de uma função sobre os valores de entrada, não sendo necessária, portanto, a defuzificação.

A seguir é destacado um tipo particular de SBRF de interesse para este trabalho: os Sistemas de Classificação Fuzzy.

2.3.1 Sistemas de Classificação Fuzzy

Classificação é uma importante tarefa encontrada nas áreas de reconhecimento de padrões, tomada de decisão, mineração de dados e modelagem. Dado um objeto, o objetivo da classificação é atribuir uma classe de um conjunto conhecido de classes a esse objeto, o qual é descrito por n atributos.

Muitos métodos têm sido utilizados para a tarefa de classificação de padrões. Entre eles: árvores de decisão (Quinlan, 1986), métodos estatísticos (Duda; Hart, 1973), redes neurais (Bishop, 1995) e sistemas fuzzy (Kecman, 2001).

Sistemas de Classificação Fuzzy (SCF) são SBRF projetados com a finalidade específica de realizar a tarefa de classificação. Assim, os SCF utilizam métodos de raciocínio próprios para essa tarefa.

O formato da regra fuzzy de classificação segue o formato geral das regras fuzzy descrito na seção anterior e pode ser expressa por:

$$\text{Se } X_1 \text{ é } A_{1l_1} \text{ e } \dots \text{ e } X_n \text{ é } A_{nl_n} \text{ então } Classe = C_j.$$

no qual X_1, \dots, X_n são os atributos do dado considerado no problema (representados aqui por variáveis linguísticas), $A_{1l_1}, \dots, A_{nl_n}$ são os valores linguísticos usados para representar os valores de tais atributos e C_j é a classe a qual o dado pertence.

Métodos de raciocínio fuzzy para classificação são procedimentos de inferência que tiram conclusões de um conjunto de regras **se...então** e um dado de entrada, com o objetivo de classificar esse dado. Os métodos mais utilizados são: Método de Raciocínio Fuzzy Clássico e Método de Raciocínio Fuzzy Geral.

O método de Raciocínio Fuzzy Clássico classifica um dado de entrada de acordo com a regra que possui maior grau de compatibilidade com o mesmo.

Seja $e_p = (a_{p1}, a_{p2}, \dots, a_{pn})$ um dado a ser classificado e R_1, R_2, \dots, R_S um conjunto de S regras do sistema de classificação, cada uma com n antecedentes. Seja $A_{il_i}(a_{pi}), i = 1, \dots, n$, o grau de pertinência do atributo a_{pi} no i -ésimo conjunto fuzzy da regra R_k .

O Método de Raciocínio Fuzzy Clássico segue as seguintes etapas para a classificação do dado e_p :

1. Calcular o grau de compatibilidade entre o dado e_p e cada regra R_k , para $k = 1, \dots, S$.

$$Compat(R_k, e_p) = t(A_{1l_1}(a_{p1}), A_{2l_2}(a_{p2}), \dots, A_{nl_n}(a_{pn}))$$
 no qual t denota uma t-norma.
2. Encontrar a regra que possui o maior grau de compatibilidade com o dado.

$$Compat(R_{max}, e_p) = max(Compat(R_k, e_p), k = 1, 2, \dots, S)$$
3. Atribuir a classe C_j ao dado e_p , onde C_j é a classe do consequente da regra R_k que possui o maior grau de compatibilidade com o dado.

O método de Raciocínio Fuzzy Geral, o qual combina as informações providas por todas as regras de uma mesma classe para classificar o dado, é definido a seguir.

O Método de Raciocínio Fuzzy Geral segue as seguintes etapas para a classificação do dado e_p :

1. Calcular o grau de compatibilidade entre o dado e_p e cada regra R_k , para $k = 1, \dots, S$.

$$Compat(R_k, e_p) = t(A_{1l_1}(a_{p1}), A_{2l_2}(a_{p2}), \dots, A_{nl_n}(a_{pn}))$$
 no qual t denota uma t-norma.
2. Calcular para cada classe C o valor de $Classe_C$, ou seja, o grau de classificação do dado na classe, agregando os graus de compatibilidade do passo anterior de todas as regras cuja classe predita é C :

$$Classe_C = f\{Compat(R_k, e_p) | C \text{ é a classe de } R_k\}$$
 sendo f um operador de agregação tal que $min \leq f \leq max$.
3. A classe C será atribuída ao dado e_p , sendo C a classe correspondente ao valor máximo de $Classe_C$ encontrado no passo anterior.

Um método alternativo de raciocínio fuzzy é descrito em Cordón et al. (1999), cujas duas primeiras etapas seguem o padrão dos métodos anteriores, sendo diferenciado pela maneira como são definidos os graus de associação entre as regras e o dado de entrada.

2.4 Considerações Finais

Este capítulo abordou alguns aspectos fundamentais a respeito da teoria de conjuntos fuzzy e sistemas fuzzy, com enfoque em sistemas de classificação baseados em regras fuzzy.

No próximo capítulo será abordado o tema de agrupamento de dados, especificando o processo de agrupamento e descrevendo algoritmos de agrupamento particionais que podem ser adotados dentro deste processo.

Agrupamento de Dados Semissupervisionado

O agrupamento de dados tem como objetivo agrupar exemplos semelhantes para refletir a forma como um conjunto de dados está estruturado. Estes exemplos são agrupados de acordo com suas similaridades ou dissimilaridades. O grupo que representa um exemplo de um conjunto de dados é um conjunto de exemplos que são mais similares entre si, de acordo com uma determinada métrica, do que os exemplos contidos em outros grupos diferentes.

O processo de agrupamento envolve distintas etapas que vão da preparação dos exemplos até a interpretação dos grupos obtidos. A etapa de interpretação dos grupos pode ser omitida, de acordo com o objetivo que se deseja atingir com o agrupamento.

Neste capítulo são colocadas ideias gerais a respeito de Aprendizado de Máquina, a descrição do processo geral de agrupamento e uma seleção de algoritmos de agrupamento particionais que podem ser adotados durante este processo.

3.1 Aprendizado de Máquina

Aprendizado de Máquina (AM), de acordo com Mitchell (1997), refere-se à investigação de métodos computacionais capazes de adquirir conhecimento de forma automática.

A inferência indutiva é um dos principais mecanismos utilizados para derivar

conhecimento novo e prever escritos futuros no contexto de AM. No aprendizado indutivo o conhecimento é aprendido por meio de inferência indutiva sobre um conjunto de dados ou exemplos.

O aprendizado indutivo é, tradicionalmente, dividido em supervisionado e não supervisionado. No aprendizado supervisionado é fornecido ao método de aprendizado um conjunto de dados rotulados, que são dados que tem a classe conhecida. Estes dados são adequados para problemas de classificação e regressão. No aprendizado não supervisionado são analisados dados não rotulados, com o objetivo de identificar estruturas embutidas no conjunto.

Mais recentemente surgiu o conceito de aprendizado semissupervisionado, que visa considerar tanto os dados rotulados como não rotulados (Witten; Frank, 2005), que pode ser aplicado tanto à classificação quanto à regressão e ao agrupamento.

3.1.1 Aprendizado Supervisionado

O aprendizado supervisionado é, de longe, o paradigma mais comum de aprendizagem. Este tipo de aprendizado se baseia em um conjunto de dados cujos exemplos são rotulados, ou seja, possuem um atributo que identifica a classe à qual pertencem. O aprendizado supervisionado vem resolver o problema de generalização das regularidades encontradas em um conjunto de dados. Métodos de aprendizado supervisionado podem ser divididos em métodos de classificação e regressão.

Grande parte dos métodos de classificação utilizam-se de um conjunto de dados de treinamento para a construção de um classificador (etapa 1), que usualmente consiste de conjunto de regras ou de uma estrutura da qual possam ser extraídas regras de classificação. Um conjunto de testes com dados independentes do conjunto de treinamento é aplicado ao classificador com a intenção de verificar a qualidade do resultado obtido na etapa de construção (etapa 2). Caso o resultado seja satisfatório, o classificador pode ser aplicado a conjunto de novos dados (passo 3). Alguns métodos requerem um ajuste do classificador após determinado período de tempo ou aumento do volume de dados. Esse ajuste é, muitas vezes, realizado pela reaplicação dos passos 1 e 2 até que seja obtido um classificador melhor do que o anterior.

Tentativas de resolver o problema de classificação incluem a aplicação de árvores de decisão, redes neurais e métodos estatísticos, como redes Bayesianas (Mitchell, 1997). Métodos como o *K-Nearest Neighbors* (k vizinhos mais próximos) (Cover; Hart, 1967) não geram classificadores, mas utilizam as informações de rótulos no conjunto de dados atual para classificar novos exemplos, atribuindo rótulos a estes exemplos por meio de medidas similaridade.

Uma variação dos métodos de classificação realiza a indução de regras fuzzy a partir de um conjunto de dados. Uma maneira popular é a utilização de sistemas neuro-fuzzy, que se utilizam de algoritmos de aprendizado derivados da teoria de redes neurais para gerar regras fuzzy (Klose et al., 2001). Outro grupo de abordagens é baseado em árvores de decisão, que podem ser eficientemente induzidas a partir de heurísticas *greedy divide-and-conquer*, e regras pode ser extraídas da árvore posteriormente (Quinlan, 1993). Algumas propostas podem ser encontradas na literatura, com extensões chamadas árvores de decisão fuzzy.

Algoritmos genéticos são utilizados na otimização e criação de sistemas fuzzy (Seção 4.2.1). A vantagem de estratégias evolutivas é a habilidade de modificar e otimizar a estrutura e parâmetros de modelos, enquanto a maior parte das estratégias de otimização é capaz de apenas adaptar os parâmetros de um modelo (Klose; Kruse, 2005).

3.1.2 Aprendizado Não Supervisionado

O aprendizado não supervisionado ocorre a partir de um conjunto de dados com exemplos não rotulados. O processo de aprendizado a partir de conjunto de dados não rotulados é, na maioria das vezes, feita por meio de análise de grupos ou agrupamento. A análise de grupos tenta encontrar subconjuntos dentro do conjunto de dados, tais que exemplos pertencentes a um mesmo grupo sejam similares.

A noção de similaridade pode ser definida de diferentes maneiras, de acordo com a finalidade do processo e as características dos dados. Quando os atributos são todos numéricos, a medida de similaridade mais comumente utilizada baseia-se em medidas de distância.

O maior problema do aprendizado não supervisionado está na forte dependência de medidas de distância e normalização dos dados. Além disso, um problema crucial é encontrar o número correto de grupos para a divisão, pois, de uma forma geral, não podemos garantir que os grupos correspondem a subconjuntos de exemplos significativos dentro do conjunto de dados.

O processo de agrupamento é melhor descrito na Seção 3.2.

3.1.3 Aprendizado Semissupervisionado

Os métodos de aprendizado mencionados anteriormente, apesar de úteis para determinadas aplicações, possuem algumas desvantagens. No aprendizado supervisionado há a necessidade de que a base de dados esteja já rotulada para possibilitar a extração de regras. Já o aprendizado não-supervisionado depende da escolha apropriada de funções de medidas e distribuição, número de grupos a

serem criados e também há o problema de representação apropriada dos grupos significativos do conjunto de dados. Os resultados do aprendizado não supervisionado são, muitas vezes, insatisfatórios, e o aprendizado supervisionado acaba sendo mais comum na prática (Klose; Kruse, 2005).

Com o crescimento dos conjuntos de dados em muitos domínios do conhecimento, a rotulação de dados total e manual torna-se cada vez mais difícil e onerosa. A exploração de ambos os tipos de dados, rotulados e não rotulados, torna-se mais interessante. Com o passar dos anos, um número crescente de publicações e conferências sobre aprendizado semissupervisionado podem ser observadas, sendo aplicadas com sucesso, especialmente, em processamento de imagens (Bensaid et al., 1996; Grira et al., 2006; Pedrycz et al., 2008) e classificação de textos (Liu; Huang, 2003; Geng et al., 2009)

As publicações sugerem e analisam modificações de métodos já conhecidos a fim de considerar sua aplicação a um conjunto com maioria de dados não rotulados e uma pequena parte de dados rotulados. A obra de Zhu (2005) apresenta de forma resumida algumas tendências e características para classificação semi-supervisionada, como *self-training*, *co-training* e *generative models* (Chapelle et al., 2006), e apontamentos a respeito de outras formas de aprendizado semissupervisionado, como por agrupamento e regressão.

A utilização de métodos de agrupamento em aprendizado semissupervisionado pode ocorrer de duas formas: colaboração na rotulação do conjunto de dados ou agrupamento considerando informação prévia. No primeiro caso, algoritmos de agrupamento são aplicados ao conjunto de dados não rotulado para gerar grupos que, posteriormente, serão rotulados por algum outro método, com base na porção rotulada do conjunto. No segundo caso, métodos consagrados de agrupamento são modificados a fim de implementar a semi-supervisão já no processo de geração de grupos e, em alguns casos, poder definir rótulos para estes grupos.

Chama-se de agrupamento semissupervisionado aquele realizado por métodos que incluem mecanismos para a consideração da informação dos rótulos pré-existentes no processo de geração de grupos. Os mecanismos utilizados incluem: modificação da função objetivo, para que inclua satisfação de restrições (Pedrycz; Waletzky, 1997); Reforço de restrições durante o processo de agrupamento (Wagstaff et al., 2001; Grira et al., 2005); Inicialização e restrição do agrupamento com base nos exemplos rotulados (Bensaid et al., 1996; Basu et al., 2002).

Os métodos desta classe podem ser divididos em duas abordagens para incorporação de semi-supervisão, dependendo do conhecimento disponível: abordagem por sementes e abordagem por restrições entre pares. As sementes compõem uma parte do conjunto de dados que pode ser utilizada para estabelecer restrições ao

algoritmo, restrições entre pares de dados e para definição de rótulos de grupos. As restrições entre pares podem ser da forma *must-link*, indicando que um par de dados deve pertencer ao mesmo grupo, ou *cannot-link*, indicando que os exemplos do par devem pertencer a grupos distintos.

O agrupamento fuzzy semissupervisionado ocorre quando são incluídos mecanismos de semi-supervisão em métodos de agrupamento fuzzy. A ideia de exploração de informações rotuladas e não rotuladas para agrupamento fuzzy data da década de 80. A maior parte das publicações coloca a abordagem de Pedrycz (1985) como o primeiro trabalho na área de agrupamento semissupervisionado. Anos após a primeira publicação, Pedrycz; Waletzky (1997) comenta sobre a falta de atenção dada aos mecanismos de supervisão parcial e discute mais amplamente suas ideias de 1985.

Exemplos de algoritmo de agrupamento fuzzy semissupervisionado são descritos na Seção 3.3.

3.2 Processo de Agrupamento

Uma atividade de agrupamento exemplo, tipicamente, segue os passos abaixo (Jain et al., 1999):

1. **Preparação de exemplos** (opcionalmente inclui extração e/ou seleção de características): determina como os exemplos serão representados. Há casos em que se faz necessária a aplicação de algum tipo de transformação nos dados, como normalizações e seleção e/ou extração de características. A preparação de exemplos é detalhada na Seção 3.2.1.
2. **Medida de similaridade**: consiste em definir uma medida de similaridade apropriada ao domínio da aplicação. Geralmente é fornecida por uma função de distância definida entre pares de exemplos. É possível incluir na medida de distância aspectos conceituais (qualitativos) ou numéricos (quantitativos). Algumas medidas de similaridade principais são abordadas na Seção 3.2.2.
3. **Agrupamento**: consiste em aplicar um algoritmo com o intuito de agrupar dados de acordo com um objetivo específico. Existem inúmeros algoritmos que podem ser aplicados nesta etapa que podem ter como resposta um exemplo pertencente ou não a um dado grupo ou atribuição de um grau de pertinência a cada exemplo para cada um dos grupos (fuzzy). Alguns algoritmos de agrupamento são detalhados na Seção 3.2.3.

4. **Validação:** nesta fase ocorre a validação dos resultados obtidos, após aplicado o algoritmo de agrupamento. Tem como objetivo determinar se o resultado é significativo. As principais formas de validação são apresentadas na Seção 3.2.4.
5. **Interpretação dos resultados:** consiste em examinar resultados com relação a seus exemplos, com o objetivo de descrever a natureza dos grupos.

3.2.1 Preparação de Exemplos

A preparação de dados envolve diversificados aspectos relacionados ao pré-processamento e à forma de representação apropriada para a sua utilização por um algoritmo de agrupamento.

O pré-processamento pode envolver, por exemplo, normalizações, conversão de tipos e redução do número de atributos por meio de seleção ou extração de características. Este processo é importante, pois define os exemplos disponíveis e o número, tipo e escala das características disponíveis para o algoritmo de agrupamento.

Um exemplo pode medir um objeto físico (por exemplo, uma cadeira) ou uma noção abstrata (por exemplo, um estilo de escrita). Geralmente os exemplos são representados como vetores multidimensionais, onde cada dimensão é uma única característica.

No que se refere às características, um problema é encontrar um conjunto delas que melhor representa a similaridade que está sendo trabalhada. Para resolver esta questão podem ser utilizadas técnicas de seleção e/ou extração de características.

A seleção de característica é o processo de identificar o subconjunto mais efetivo das características disponíveis para ser utilizado no agrupamento. A extração de características é o uso de uma ou mais transformações das características de entrada disponíveis para salientar características presentes nos dados.

O tipo e a escala das características são informações importantes na escolha da medida de similaridade e do algoritmo de agrupamento a ser empregado. Para cada tipo/escala de atributo existem medidas de similaridade adequadas. O tipo de um atributo diz respeito ao grau de quantização nos dados. A escala indica a significância relativa dos números. Os possíveis tipos de atributos:

1. **Binários:** são atributos que representam apenas dois valores;
2. **Discretos:** representam um número finito de valores;
3. **Contínuos:** podem assumir um número infinito de valores.

Em relação à escala, as características podem ser quantitativas e qualitativas:

1. Qualitativa

- (a) *Nominal*: os valores são apenas nomes distintos. Exemplos: CEP, cores, sexo.
- (b) *Ordinal*: os valores apenas refletem uma ordenação. Exemplos: Ruim, Regular, Bom ou cores ordenadas pelo espectro.

2. Quantitativa

- (a) *Intervalo*: a diferença entre os valores tem significado, isto é, existe uma unidade de medida. Exemplos: Em uma escala de 1 a 10 para dar notas as provas de alunos, duração de um evento.
- (b) *Razão*: uma escala que possui zero absoluto. Exemplos: altura, comprimento, largura de um objeto.

Algumas vezes os exemplos apresentam atributos de escalas diferentes ou a representação dos dados não é adequada para aplicação do algoritmo de agrupamento. Assim, é necessária a aplicação de algumas transformações antes de iniciar a utilização dos dados. Quando os limites dos intervalos de valores de atributos são muito diferentes, um atributo pode dominar o resultado do agrupamento. Para solucionar este problema, é comum a padronização dos dados de forma que os atributos estejam na mesma escala.

Os dados são representados, na maioria dos casos, por uma matriz de exemplos $X_{n \times d}$, em que n é o número de exemplos e d é o número de atributos que representam os exemplos. Cada elemento desta matriz, X_{ij} , contém o valor da j -ésima característica para o i -ésimo exemplo. Cada exemplo pode ser visto como um ponto neste espaço e um grupo como um conjunto de exemplos próximos ou que satisfazem uma relação espacial.

Padrões também podem ser representados por uma matriz e um grafo de similaridade ou proximidade. Uma matriz de similaridade $S_{n \times n}$, contém os valores da similaridade/dissimilaridade entre dois exemplos i e j , representados respectivamente na linha i e coluna j da matriz. Este valor é geralmente calculado por uma medida de similaridade.

3.2.2 Medidas de Similaridade

Medidas de similaridade indicam o grau de semelhança entre dois exemplos. A medida de similaridade tem um papel fundamental para a maioria dos algoritmos

de agrupamento, por isso deve ser cuidadosamente escolhida, considerando os diferentes tipos e as diferentes escalas mensuradas dos dados.

Uma vez que agrupamento consiste em agrupar exemplos de tal forma que os exemplos pertencentes ao mesmo grupo sejam mais semelhantes entre si, do que exemplos pertencentes a grupos diferentes, de acordo com alguma medida de similaridade, é importante definir antecipadamente qual medida será utilizada.

De uma forma geral, as medidas de similaridades devem satisfazer algumas propriedades:

1. Para dissimilaridade: $S_{ii} = 0$, para todo i
Para similaridade: $S_{ii} \geq \max S_{ij}, i \neq j$
2. $S_{ij} = S_{ji}$ (Simetria)
3. $S_{ij} \geq 0$ para todo i e j (Positividade)
4. $S_{ij} = 0$ somente se $i = j$
5. $S_{ik} \leq S_{ij} + S_{jk}$ para todo i, j e k (Desigualdade triangular)

As medidas que satisfazem todas as propriedades acima são chamadas de métricas. Se a medida de similaridade não satisfizer as propriedades 4 e 5, elas não são consideradas métricas.

De acordo com o tipo e escala das características, um conjunto de medidas de similaridade pode ser empregado. Para conjuntos de dados em que todas as características são contínuas e a escala é do tipo relacional, as medidas mais comumente utilizadas são as distâncias baseadas na métrica *Minkowski*, como a distância *Euclidiana*, de *Manhattan* e *supremum*. Para todas as características binárias é comum a utilização da distância de *Manhattan* (distância de *Manhattan* entre dois vetores é chamada de distância de *Hamming*). Para características binárias e nominais, existem coeficientes de casamento (*matching*), como coeficiente de casamento simples e coeficiente de *Jaccard*. A seguir serão apresentadas algumas das medidas citadas, maiores informações quanto a métricas e suas características podem ser encontradas na obra de Jain et al. (1999).

Métricas de *Minkowski*

Métricas de *Minkowski* são aquelas derivadas da equação geral 3.1, de acordo com um valor definido para p , sendo $1 \leq p < \infty$. Chamadas de L_p , medem a dissimilaridade entre exemplos, e, para os menores valores de p , correspondem a estimativas mais robustas. Uma desvantagem das métricas de *Minkowski* é que

estas são sensíveis às variações de escala dos atributos, ou seja, atributos representados em uma escala maior tendem a dominar os outros. Uma solução para este problema é realizar a normalização dos atributos para um intervalo comum, ou outros esquemas de ponderação.

$$S_{ij} = \left(\sum_{k=1}^d |x_{ik} - x_{jk}|^p \right)^{1/p} \quad (3.1)$$

Alguns valores de p definem métricas bem conhecidas:

- $p = 1$: Distância de *Manhattan*, dada pela equação 3.2:

$$S_{ij} = \sum_{k=1}^d |x_{ik} - x_{jk}| \quad (3.2)$$

- $p = 2$: Distância Euclidiana, muito utilizada, é apropriada para conjuntos de dados que possuem grupos compactos e isolados. É dada pela equação 3.3:

$$S_{ij} = \left(\sum_{k=1}^d |x_{ik} - x_{jk}|^2 \right)^{1/2} \quad (3.3)$$

- $p = \infty$: Distância *supremum*, calcula o máximo da diferença absoluta em coordenadas, ou seja, é a diferença máxima entre quaisquer componentes de vetores. É dada pela equação 3.4:

$$S_{ij} = \max_{1 \leq k \leq d} |x_{ik} - x_{jk}| \quad (3.4)$$

Distância de Mahalanobis

Dada pela equação 3.5, em que C_{kl} é o elemento da k -ésima linha e l -ésima coluna da inversa da matriz de covariância.

$$S_{ij} = \left(\sum_{k=1}^d \sum_{l=1}^d (x_{ik} - x_{jk}) C_{kl} \right)^{1/2} \quad (3.5)$$

Esta distância incorpora a correlação entre os atributos e padroniza cada característica para média zero e variância um. A ideia é associar diferentes pesos à diferentes atributos com base em suas variâncias e a correlação linear entre pares de padrões. A aplicação dessa medida de distância pode minimizar as distorções causadas pelas medidas de correlações lineares entre atributos.

Medidas para atributos nominais e ordinais

Estas medidas tem foco na determinação da contribuição de cada atributo. As medidas de similaridade entre pares de padrões são obtidas pela soma das contribuições individuais de todas as variáveis.

Similaridade nominal/ordinal geral

A similaridade é dada pela equação 3.6, em que S_{ijk} é a contribuição de cada exemplo baseada em índices de discordância entre pares de estados dos atributos.

$$S_{ij} = \sum_{k=1}^d S_{ijk} \quad (3.6)$$

Medidas para atributos mistos

A medida para atributos mistos é adequada para obter a similaridade entre dados que contenham atributos de tipos diferentes, pois tem a capacidade de se adequar a qualquer tipo de atributo individualmente.

Coefficiente geral de similaridade

Obtido por meio da equação 3.7, onde s_{ijk} é a contribuição do do k -ésimo atributo para a similaridade e w_{ijk} é 0 ou 1, dependendo se a comparação para a variável k é válida ou não. O valor de s_{ijk} pode ser definido para atributos de tipos diferentes, dependendo da complexidade apresentada quanto aos tipos do conjunto de dados.

$$S_{ij} = \frac{\sum_{k=1}^d w_{ijk} S_{ijk}}{\sum_{k=1}^d w_{ijk}} \quad (3.7)$$

3.2.3 Técnicas de Agrupamento Não Supervisionado

Neste passo ocorre, efetivamente, a aplicação do algoritmo de agrupamento. Os métodos de agrupamento são divididos em dois conjuntos: agrupamento hierárquico e agrupamento particional.

Métodos de agrupamento hierárquico constroem uma sequência de partições aninhadas. Isso significa que, ao final do agrupamento, obtemos uma estrutura que pode ser dividida em vários níveis e, a cada divisão, há uma possível partição do conjunto de exemplos de entrada (Figura 3.1).

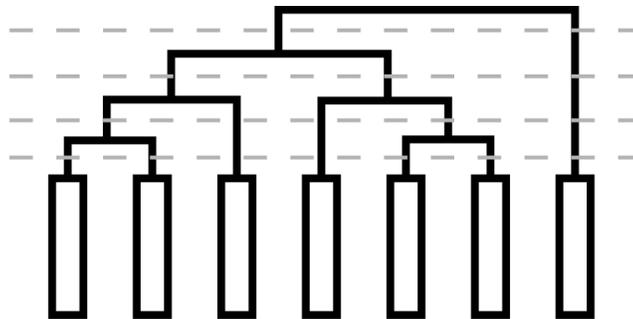


Figura 3.1: Exemplo de estrutura resultante de um agrupamento hierárquico, com possíveis divisões de níveis (linhas tracejadas). Cada retângulo representa um exemplo do conjunto de dados.

Os diferentes métodos de agrupamento hierárquico podem ser divididos em duas abordagens: aglomerativa (*bottom-up*) e divisiva (*top-down*). Na primeira, inicia-se com agrupamento disjunto, onde cada um dos n exemplos do conjunto de dados é um grupo. Por uma matriz de similaridade os grupos individuais são aninhados em novos grupos, até que haja apenas um grupo contendo todos os exemplos. Métodos da segunda abordagem são iniciados como um único grupo que contém todos os n exemplos do conjunto de dados, por uma matriz de similaridade os grupos que contém mais de um exemplo são divididos em sub-grupos, até que haja um grupo para cada um dos n exemplos.

Devido ao resultado final ser uma taxonomia, os métodos de agrupamento hierárquico são bastante populares nas ciências biológica, social e comportamental. No entanto, esse tipo de agrupamento passa a ter desvantagens, dependendo do

domínio em que será aplicado.

A maior dificuldade de visualização do conjunto de partições aninhadas é proporcional ao aumento do número de exemplos dentro de um conjunto de dados. A execução desses algoritmos gera grupos que contem exemplos não agrupáveis (não similares), por exemplo, em ambas as abordagens temos um grupo maior que contém todos os exemplos. Além disso, grande parte dos algoritmos de agrupamento hierárquico não melhoram os grupos.

Agrupamentos particionais definem um procedimento para encontrar todos os grupos de um conjunto de dados simultaneamente, como uma única partição dos dados (Figura 3.2). Além disso, é comum que este tipo de método tenha como parte do resultado elementos típicos para cada grupo, calculados a partir dos membros do grupo. Esta técnica pode gerar grupos melhor definidos, mostrando uma situação mais próxima do real, para conjuntos de dados de certos domínios.

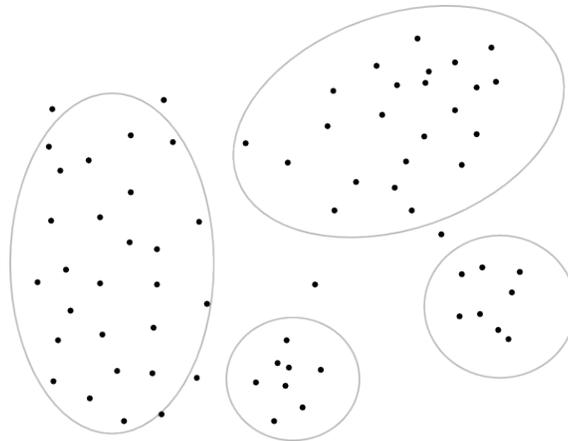


Figura 3.2: Exemplo de resultado de um agrupamento particional, que dividiu o conjunto de dados em quatro grupos. Cada ponto representa um exemplo do conjunto de dados.

A seguir, são apresentados dois exemplos de algoritmos simples de agrupamento particional.

K-Means

O algoritmo K-Means (MacQueen, 1967), descrito em pseudo-código 3.1, é um dos mais populares e simples algoritmos de agrupamento. Apesar do grande número de publicações com diferentes algoritmos de agrupamento, o K-Means ainda é amplamente utilizado, muitas vezes servindo de base ao desenvolvimento de novos algoritmos.

O objetivo do K-Means é agrupar os dados em k grupos disjuntos, de maneira que a soma das distâncias entre os exemplos pertencentes a um grupo e seu respectivo centro seja mínima. O centro de grupo ou protótipo representa o ponto médio dos pontos pertencentes a um determinado agrupamento.

Algoritmo 3.1: K-Means

Entrada: X : conjunto de dados, k : número de grupos.

Saída: P : partição dos dados em k grupos, C : conjunto de exemplos protótipos (centros de grupos).

início

$t := 0$;

$C :=$ conjunto de k exemplos $x \in X$;

$P^{t+1} :=$ partição de X com relação a C ;

repita

$P^t := P^{t+1}$;

 Atualização de C ;

 Atualização de P^{t+1} ;

até $P^t == P^{t+1}$;

Fuzzy C-Means

O algoritmo Fuzzy C-Means (FCM) (Bezdek, 1981) (pseudo-código 3.2) é uma extensão fuzzy do algoritmo K-Means. Extensões fuzzy para agrupamento representam os grupos por conjuntos fuzzy multidimensionais (Höppner et al., 1999), ou seja, cada exemplo do conjunto de dados pertence aos grupos com um determinado grau. O resultado da aplicação do FCM é uma pseudo-partição fuzzy do conjunto de dados.

Seja $X = \{x_1, x_2, \dots, x_n\}$ um conjunto de dados. Uma pseudo-partição fuzzy de X é uma família de subconjuntos de X , denotados por $\wp = \{A_1, A_2, \dots, A_c\}$, onde A_i é o conjunto dos graus de pertinência de x_k para cada partição e

$$\sum_{i=1}^c A_i(x_k) = 1$$

para todo $k \in \mathbb{N}_n$ e

$$0 < \sum_{k=1}^n A_i(x_k) < n$$

para todo $i \in \mathbb{N}_c$, onde c é um inteiro positivo e n é o número de elementos em X , ou seja, uma partição não deve ter n elementos com grau de pertinência máximo (equivalente ao conjunto X) ou mínimo (partição vazia). Por exemplo, seja $X =$

$\{x_1, x_2, x_3\}$ e

$$A_1 = 0,6/x_1 + 1/x_2 + 0,1/x_3$$

$$A_2 = 0,4/x_1 + 0/x_2 + 0,9/x_3$$

então $\{A_1, A_2\}$ é uma pseudo-partição fuzzy.

Dado um conjunto de dados $X = \{x_1, x_2, \dots, x_n\}$, onde cada elemento, em geral, é um vetor

$$x_k = [x_{k1}, x_{k2}, \dots, x_{kp}] \in \mathbb{R}^p$$

para todo $k \in \mathbb{N}_n$, o objetivo do agrupamento FCM é encontrar uma pseudo-partição fuzzy e os protótipos de grupos associados a ela de forma a representar a estrutura dos dados da melhor forma possível. Para isso é utilizado um critério de associação, que neste caso é formulado em termos de um índice de desempenho, normalmente baseado nos protótipos. Dada uma pseudo-partição $\varphi = \{A_1, A_2, \dots, A_n\}$, os c centros, $U = v_1, v_2, \dots, v_c$ associados à partição são calculados pela fórmula:

$$v_i = \frac{\sum_{k=1}^n [A_i(x_k)]^m x_k}{\sum_{k=1}^n [A_i(x_k)]^m} \quad (3.8)$$

para todo $i \in \mathbb{N}^c$, onde $m > 1$ é um número real que expressa a influência dos graus de pertinência.

O índice de performance de uma pseudo-partição fuzzy φ , $J_m(\varphi)$, é definido pela fórmula

$$J_m(\varphi) = \sum_{k=1}^n \sum_{i=1}^c [A_i(x_k)]^m \|x_k - v_i\|^2 \quad (3.9)$$

onde $\|x_k - v_i\|^2$ representa a distância euclidiana entre x_k e v_i . Este índice de desempenho mede a soma ponderada das distâncias entre os protótipos e os elementos dentro do grupo correspondente. Quanto menor o valor de $J_m(\varphi)$, melhor a partição fuzzy φ . Pode-se dizer que o objetivo do agrupamento FCM é encontrar a pseudo-partição φ que minimize o índice de desempenho.

3.2.4 Validação

As seções anteriores apresentaram alguns algoritmos de agrupamento que dividiram o conjunto de dados baseado em critérios diferentes. No entanto, diferentes algoritmos de agrupamento, ou até um único algoritmo de agrupamento utilizando diferentes parâmetros, geralmente resultam em diferentes conjuntos de grupos.

Assim, é importante comparar os vários resultados gerados e selecionar um que seja adequado a melhor distribuição verdadeira dos dados. Ao processo de avaliar a qualidade e a confiabilidade dos grupos gerados dos vários processos de

Algoritmo 3.2: Fuzzy C-Means

Entrada: X : conjunto de dados, c : número de grupos, d : medida de distância, m : $(1, \infty)$, é o nível de fuzzificação, ϵ : erro.

Saída: φ : pseudo-partição fuzzy dos dados em c grupos, U : conjunto de protótipos para os grupos.

início

$t := 0$ Inicialização randômica de φ^{t+1} ;

repita

$\varphi^t := \varphi^{t+1}$;

Atualização de U a partir da equação 3.8, utilizando a medida de distância d ;

Atualização de φ^{t+1} ;

até $|\varphi^{t+1} - \varphi^t| \leq \epsilon$;

agrupamento, dá-se o nome de validação dos grupos (Jiang et al., 2004).

A avaliação do resultado de um agrupamento deve ser objetiva, tendo como propósito determinar se a solução encontrada é representativa para o conjunto de dados analisado. A validação do resultado de um agrupamento é realizada com base em dados estatísticos, que julgam, de maneira qualitativa, o mérito dos grupos encontrados.

Um índice quantifica alguma informação a respeito da qualidade do agrupamento. A maneira pela qual um índice é aplicado para validar um agrupamento é dada pelo critério de validação. Existem três tipos de critérios para analisar a qualidade de um agrupamento: internos, externos e relativos.

Critérios internos avaliam o agrupamento com base nos dados originais (matriz de exemplos ou matriz de similaridade) sem nenhum conhecimento externo ao agrupamento. Critérios externos avaliam um agrupamento de acordo com uma estrutura pré-estabelecida que reflita a intuição do pesquisador sobre a estrutura presente nos dados. Esta estrutura pré-estabelecida pode ser uma partição construída por um especialista da área com base em conhecimento prévio. Critérios relativos comparam diversos agrupamentos para decidir qual deles é melhor em um determinado aspecto (agrupamento mais estável ou mais adequado, por exemplo). Podem ser usados para determinar o valor mais apropriado para um parâmetro, como número de grupos, ou para comparar diferentes algoritmos de agrupamento.

Carvalho et al. (2006) sugere índices de interpretação baseados em somas quadráticas (*sum of squares* - SSQ) para avaliação do resultado de agrupamentos fuzzy aplicados a conjuntos de dados quantitativos. Os índices descritos neste trabalho utilizam-se da pseudo-partição fuzzy (φ) e do conjunto de protótipos finais ($U = (v_1, \dots, v_c)$), resultantes do algoritmo de agrupamento, além da medida

de distância utilizada pela algoritmo e um vetor $z = (z^1, \dots, z^p)$, que representa um protótipo geral para os n exemplos de um conjunto de dados, e é calculado como segue:

$$\mathbf{z} = (z^1, \dots, z^p) = \frac{\sum_{i=1}^c \sum_{k=1}^n (u_{ik})^m \mathbf{x}_k}{\sum_{i=1}^c \sum_{k=1}^n (u_{ik})^m} = \frac{\sum_{i=1}^c \mu_i \mathbf{v}_i}{\sum_{i=1}^c \mu_i}, \quad \mu_i = \sum_{k=1}^n (u_{ik})^m \quad (i = 1, \dots, c)$$

com

$$z^j = \frac{\sum_{i=1}^c \sum_{k=1}^n (u_{ik})^m x_k^j}{\sum_{i=1}^c \sum_{k=1}^n (u_{ik})^m} = \frac{\sum_{i=1}^c \mu_i g_i^j}{\sum_{i=1}^c \mu_i} \quad (j = 1, \dots, p)$$

As métricas e índices adaptados por Carvalho et al. (2006) são caracterizadas a seguir.

Métricas Baseadas em Somas Quadráticas

São definidas três métricas, *overall SSQ*, *within-cluster SSQ* e *between-cluster SSQ*, que são as bases para a definição dos índices de interpretação apresentados.

A métrica *overall SSQ*, representada por T , avalia a heterogeneidade geral para todos os n exemplos do conjunto de dados, de acordo com a função de distância usada pelo algoritmo de agrupamento. Esta métrica pode ser decomposta para indicar a heterogeneidade específica de um grupo (T_i) e, ainda, para indicar a heterogeneidade específica com relação a um determinado atributo do conjunto (T_j). Esta métrica relaciona cada exemplo do conjunto de dados com o protótipo geral z .

A heterogeneidade dentro de cada grupo é calculada pela métrica *within-cluster SSQ*, representada por J_i . Esta métrica relaciona os exemplos do conjunto de dados com cada um dos protótipos de grupos v_i obtidos pelo algoritmo de agrupamento. Partindo desta métrica, também é possível obter uma avaliação heterogeneidade geral por grupos (J) e a heterogeneidade relacionada a atributos do conjunto (J_j).

Between-cluster SSQ, representada por B , é a métrica que avalia a dispersão dos protótipos de grupos e, portanto, a diferença entre todos os grupos obtidos com a aplicação do algoritmo de agrupamento. A métrica relaciona os protótipos de grupos v_i com o protótipo geral z e utiliza-se dos graus de pertinência a grupos obtidos por meio do algoritmo de agrupamento. A métrica pode ser decomposta para encontrar a dispersão para cada protótipo (B_i) ou a dispersão relacionada a atributos do conjunto (B_j).

Índices de Interpretação

Os índices de interpretação para agrupamento fuzzy são definidos de forma a avaliar a qualidade da partição, a homogeneidade e excentricidade de cada grupo,

além do papel dos diferentes atributos do conjunto de dados no processo de formação de grupos. O valor dos índices está dentro do intervalo de 0 a 1.

1. Índices de interpretação de partição fuzzy

Índice de heterogeneidade geral: obtido por

$$R = \frac{B}{T}$$

onde valores altos para R implicam que os grupos são mais homogêneos e melhor representados pelos protótipos de grupos.

Índice de heterogeneidade geral relacionado a atributos: obtido por

$$COR(j) = \frac{B_j}{T_j}$$

quando realizada a comparação entre $COR(j)$ e R pode-se avaliar se o poder discriminativo de um atributo está acima ou abaixo da média. A razão

$$CTR(j) = \frac{B_j}{B}$$

indica a contribuição do atributo para a dispersão. Altos valores para esta razão indicam que o atributo tem contribuição importante para a separação dos protótipos de grupos.

2. Índices de interpretação de grupo

Índice de heterogeneidade de grupo: são definidas as proporções de cada métrica SSQ para cada grupo i :

$$T(i) = \frac{T_i}{T}, B(i) = \frac{B_i}{B} \text{ e } J(i) = \frac{J_i}{J}$$

onde valores altos obtidos de $B(i)$ indicam que o grupo específico está distante do centro global do conjunto, comparativamente com os outros grupos gerados. Valores altos para $J(i)$ indicam que um grupo específico é relativamente heterogêneo em comparação com os demais grupos.

Índice de heterogeneidade de grupo relacionado a atributos: pode ser avaliado com ajuda do índice de heterogeneidade geral relacionado a atributos. A proporção de um atributo j com relação a um grupo i é dada por:

$$COR(j, i) = \frac{B_{ij}}{T_j}$$

onde altos valores de $COR(j, i)$ mostram que o atributo j tem comportamento relativamente homogêneo dentro do grupo i . A contribuição relativa do atributo j para a heterogeneidade do grupo i é dada por:

$$CTR(j, i) = \frac{B_{ij}}{B_i}.$$

E a contribuição relativa do atributo j e grupo i para a métrica *between-cluster* SSQ é dada por:

$$CE(j, i) = \frac{B_{ij}}{B}$$

e quando $CE(j, i)$ é próximo de 1 o atributo j tem grande contribuição para a excentricidade do grupo i .

Há outros índices de validação definidos de forma diferente, com relação aos resumidamente descritos nesta seção. Por exemplo, Campello; Hruschka (2006) define a métrica *Fuzzy Silhouette*, que é uma generalização fuzzy da métrica *Average Silhouette Width Criterion*, originalmente criada para aplicação a partições de dados *crisp*.

3.2.5 Interpretação dos Resultados

Nesta etapa os resultados do agrupamento são examinados com base nos seus exemplos, com o objetivo de descrever a natureza do grupo gerado. A interpretação dos grupos pode permitir avaliações subjetivas que tenham significados práticos, ou seja, especialistas podem ter interesse em encontrar características semânticas de acordo com os exemplos e valores de seus atributos em cada grupo.

3.3 Agrupamento de Dados Semissupervisionado

Durante o processo de agrupamento ocorre o passo de aplicação de uma técnica de agrupamento. O algoritmo não supervisionado aplicado neste passo pode ser substituído por uma técnica que se utiliza de semi-supervisão.

As subseções que seguem trazem a descrição de três algoritmos de agrupamento fuzzy semissupervisionado, relevantes para este trabalho, e também são colocados, de forma mais simplificada, outros algoritmos de agrupamento semissupervisionado que poderiam ser utilizados no passo de aplicação de técnica de agrupamento.

3.3.1 Semi-Supervised Fuzzy C-Means

Bensaid et al. (1996) tomam como base o algoritmo FCM, descrito em 3.2, e sugerem o algoritmo *semi-supervised Fuzzy C-Means* (ssFCM) (pseudo-código 3.3). As adaptações propostas incluem exemplos rotulados para melhorar o desempenho do FCM.

As modificações sugeridas foram realizadas no intuito de resolver três problemas da aplicação do FCM original:

1. a escolha do número de grupos c : a falta de conhecimento sobre o conjunto de dados que será agrupado pode fazer com que o número c escolhido gere um agrupamento que não reflete a realidade do conjunto;
2. a associação de rótulos aos grupos definidos pelo algoritmo: após o agrupamento, não sabemos qual o rótulo apropriado para cada grupo;
3. funções de desempenho que tendem a igualar o número de membros de cada grupo: nem todos os conjuntos de dados estão divididos em grupos com o mesmo número de elementos, podendo existir grupos com mais (ou menos) exemplos do que outros grupos do mesmo conjunto. A figura ilustra este problema.

São sugeridas três propostas principais para a minimização dos problemas citados. As duas primeiras solucionariam os problemas 1 e 2, já o problema 3 seria resolvido pela última adaptação sugerida:

1. Introdução de exemplos rotulados: o conjunto de dados inicial do FCM seria substituído por uma união dos conjuntos de dados rotulados, X^l , e não rotulados, X^u . Seja n o número de classes representadas em X^l , então ao número de grupos c é atribuído o valor n .
2. Definição das pseudo-partições fuzzy (φ): a pertinência dos exemplos rotulados será igual a 1 e não será alterada durante a atualização da pseudo-partição. A pertinência para exemplos não rotulados é calculada como no FCM.
3. Cálculo dos protótipos de grupos: é incorporado um peso atribuído a cada exemplo rotulado (w_k) para realizar o cálculo dos protótipos. Este peso é definido segundo o grau de influência de cada dado rotulado. Caso nada se conheça a respeito da influência destes exemplos, $w_k = w$ para todo $x_k \in X^l$, onde w pode ser um valor definido de forma aleatória. O cálculo do protótipo

$v_i^{(t)}$ do grupo i na iteração t do algoritmo ssFCM é

$$v_i^{(t)} = \left(\frac{\sum_{k=1}^{n_l} w_k [A_i^l(x_k)]^m x_k^l + \sum_{k=1}^{n_u} [A_i^u(x_k)]^m x_k^u}{\sum_{k=1}^{n_l} w_k [A_i^l(x_k)]^m + \sum_{k=1}^{n_u} [A_i^u(x_k)]^m} \right) \quad (3.10)$$

Algoritmo 3.3: semi-supervised FCM

Entrada: X^l : é um conjunto de sementes (exemplos rotulados), X^u : é o restante do conjunto de dados (exemplos não rotulados), d : medida de distância, m : $(1, \infty)$, é o nível de fuzzificação, ϵ : erro, $W = w_1, w_2, \dots, w_{n_l}$ ou w : peso para exemplos rotulados.

Saída: \wp : pseudo-partição fuzzy dos dados em grupos, U : conjunto de protótipos para os grupos e seus possíveis rótulos.

início

$t := 0$ $c :=$ número de rótulos distintos em X^l Inicialização de $\wp^t + 1$;

repita

$\wp^t := \wp^t + 1$;

Atualização de U a partir da equação 3.10, utilizando a medida de distância d ;

Atualização de \wp^{t+1} ;

até $|\wp^{t+1} - \wp^t| \leq \epsilon$;

3.3.2 Partially Supervised Gustafson and Kessel

O método *partially supervised Gustafson and Kessel* (Gustafson e Kessel parcialmente supervisionado) (pseudo-código 3.4) é a extensão do método descrito na publicação de Gustafson; Kessel (1978) (GK), que utiliza uma distância quadrática adaptativa no lugar da simples distância euclidiana do FCM. Esta extensão foi desenvolvida por Pedrycz (1985); Pedrycz; Waletzky (1997).

Como a abordagem tratada na Seção 3.3.1, esta também é baseada no algoritmo FCM. Neste método também trata-se o conjunto de dados de entrada X como a união de dados rotulados, X^l , e não rotulados, X^u . Adicionalmente, existe um vetor de valores binários (b) que indica quais são os dados rotulados e não-rotulados, respectivamente representados pelos valores 1 e 0.

A função de desempenho para o método, sugerida na extensão, contém um termo de penalidade que tem o objetivo de balancear a influência dos dados rotulados. Isto significa que dados rotulados que não têm um alto grau de pertinência a uma classe devem ser “desconsiderados”, pois sua influência ocasionaria a distorção do agrupamento resultante da aplicação do método. A função de desempenho

seria modificada da seguinte forma

$$J_m(\varphi) = J_m^{GK}(\varphi) + \alpha \sum_{i=1}^k \sum_{j=1}^n ([A_i(x_j)] - f_{ij} b_j)^m \|x_j - v_i\|_{A_i}^2 \quad (3.11)$$

onde $J_m^{GK}(\varphi)$ é semelhante à função 3.9 de desempenho do FCM original.

A função de atualização da partição φ , representada por $A_i^{(t+1)}(x_k)$, é estendida para

$$A_i^{(t+1)}(x_k) = \frac{1}{1 + \alpha} \cdot \frac{1 + \alpha \left(1 - b_j \sum_{i'=1}^k f_{i'j}\right)}{\sum_{i'=1}^k \left(\frac{d_{M_i}^2(x_j, v_i)}{d_{M_{i'}}^2(x_j, v_{i'})}\right)^m} + \frac{\alpha}{1 + \alpha} \cdot f_{ij} b_j \quad (3.12)$$

onde $f_{ij} \in F$, F é uma matriz que contém o grau de pertinência de cada exemplo rotulado para cada grupo, b é o vetor indicador de dados rotulados e não-rotulados e a distância entre x_k e o protótipo $v_i^{(t)}$ é calculado como segue

$$d_{M_i}^2(x_k, v_i) = (x_k - v_i)^T M_i (x_k - v_i) \quad (3.13)$$

onde M_i é a matriz de covariância para a partição A_i , atualizada a cada execução do algoritmo, entre o cálculo dos protótipos de grupos e a atualização da pseudo-partição por meio do procedimento

$$M_i = [\det(C_i)]^{-1/p} C_i^{-1} \text{ com } C_i = \sum_{k=1}^n A_i(x_k)^m (x_k - v_i)(x_k - v_i)^T$$

Quando a penalidade $\alpha = 0$, os rótulos de X^l são totalmente ignorados e o algoritmo tem desempenho similar ao GK original. Para o caso de $\alpha \rightarrow \infty$, o psGK tem um desenvolvimento parecido com o do ssFCM.

3.3.3 Active Fuzzy Constrained Clustering

O *Active Fuzzy Constrained Clustering* (AFCC) (pseudo-código 3.5), apresentado por Grira et al. (2008), é um método baseado no FCM que implementa semi-supervisão pela utilização de restrições do tipo must-link e cannot-link colocadas para uma parte dos exemplos do conjunto de dados. O algoritmo proposto é baseado em uma realocação iterativa que particiona o conjunto de dados em um número ótimo de grupos na tentativa de minimizar a soma de distâncias intra-grupo sem deixar de considerar as restrições disponibilizadas pelo usuário.

De forma geral, pode-se dizer que o AFCC alterna entre os dois passos: atualização de grau de pertinência e estimativa de protótipo. O algoritmo depende de

Algoritmo 3.4: partially supervised Gustafson and Kessel

Entrada: c : número de grupos, X : conjunto de dados, b : indicador de exemplos rotulados, F : matriz de graus de pertinência dos exemplos rotulados, ϵ : erro.

Saída: φ : pseudo-partição fuzzy dos dados em grupos, U : conjunto de protótipos para os grupos e seus possíveis rótulos.

início

Inicialização de φ' ;

repita

$\varphi := \varphi'$;

 Cálculo de U ;

 Cálculo das matrizes P_i de covariância para cada grupo i ;

$\varphi' :=$ atualização de φ ;

até $\|\varphi' - \varphi\| \leq \epsilon$;

dois fatores importantes, que são calculados a cada iteração. O fator α garante o equilíbrio da influência entre restrições e exemplos não rotulados e o fator β que determina qual o termo da atualização de graus de pertinência será dominante.

A aplicação do AFCC considera que as similaridades entre exemplos de dados é informação confiável com relação ao objetivo do agrupamento, tendo o conjunto de restrições como um auxílio na busca de grupos mais relevantes dentro do conjunto de dados. Assim, há pouca incerteza com de que o algoritmo não identificará grupos compactos e bem separados.

Uma diferença notável neste método é a adaptação do número de protótipos e grupos durante a execução do algoritmo. Os métodos apresentados anteriormente definiam um valor para o número de grupos e trabalhavam com este valor até o fim da aplicação do método. O AFCC inicia um valor máximo para o número de grupos e, ao longo das iterações, vai aperfeiçoando as partições e pode eliminar grupos que sejam considerados não significativos após avaliação de densidade (Figura 3.3). A re-distribuição dos exemplos do conjunto de entrada e o cálculo para atualização dos protótipos de grupo são realizados a cada iteração do algoritmo, até que os protótipos sejam considerados estáveis.

3.3.4 Outros Algoritmos

Nesta seção são descritos quatro algoritmos de agrupamento semissupervisionado não fuzzy. Estes algoritmos são importantes por terem sido os primeiros que surgiram no contexto de AM explorando a ideia de semi-supervisão para agrupamento. Ainda nesta seção é apresentado um algoritmo de agrupamento semis-

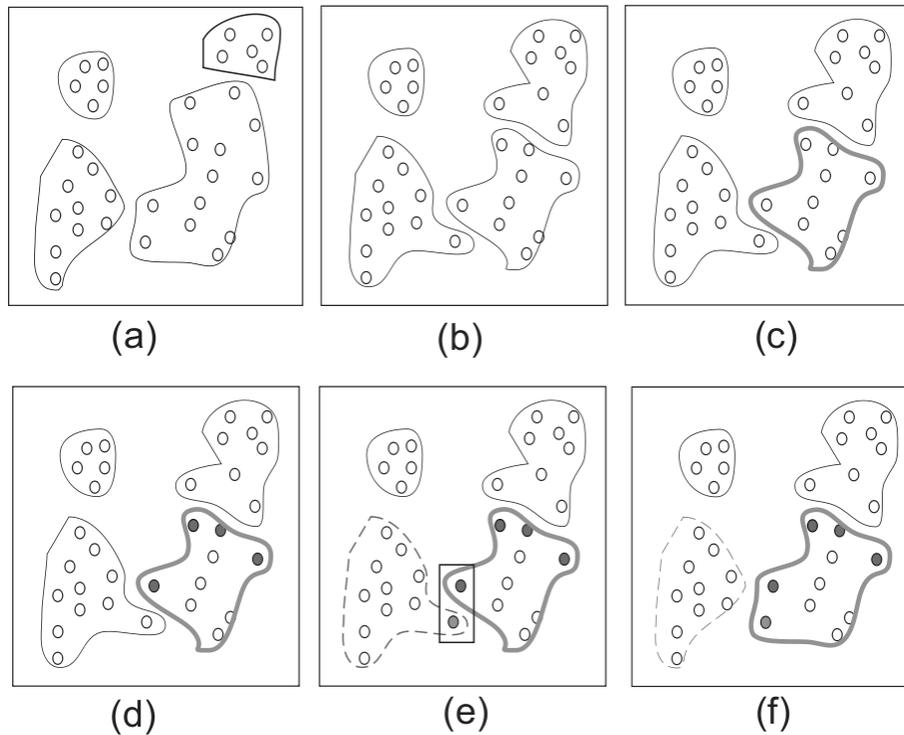


Figura 3.3: Ilustração do processo de agrupamento. (a) conjunto de dados e agrupamento desejado. (b) Grupos formados em uma iteração t . (c) Grupo menos bem-definido (destacado) é o que tem menor densidade. (d) Exemplos à fronteira do grupo menos bem-definido (círculos cheios) são aqueles com menor grau de pertinência entre os dados do grupo. (e) Para cada exemplo de fronteira, é selecionado (fronteira tracejada) o grupo mais próximo (segundo maior grau de pertinência para o exemplo). São verificadas as restrições para o exemplo de fronteira. (f) adaptação dos grupos para não violação da restrição *must-link* entre os exemplos. (Gira et al., 2005)

Algoritmo 3.5: Active Fuzzy Constrained Clustering

Entrada: C : número máximo de clusters, X : conjunto de dados, $Const$: conjunto restrições para exemplos de X

Saída: P : partição do dados, U : conjunto de protótipos para grupos.

início

Inicialização de C protótipos;

Inicialização de P mesma probabilidade de pertencer a cada grupo;

Computar cardinalidade inicial para os C grupo;

repita

 Computar β ;

 Computar α ;

 Computar grau de pertinência;

 Computar cardinalidades para os C grupos;

para cada $k \in C$ **fazer**

se cardinalidade de $k < limite$ **então**

 descartar o grupo k ;

 Atualizar C ;

 Atualizar P ;

até obter protótipos estáveis;

supervisionado fuzzy com uma proposta diferenciada com relação aos algoritmos colocados anteriormente.

SEDED K-Means

O algoritmo SEDED-K-Means (pseudo-código 3.6), proposto por Basu et al. (2002), é uma variante do K-Means, ou seja, também particiona o conjunto de dados em k grupos. Este algoritmo utiliza exemplos rotulados (sementes - *seeds*) para calcular os protótipos iniciais dos grupos ao invés de escolhê-los aleatoriamente.

Considerando um conjunto de exemplos X , toma-se um subconjunto $S \subset X$ como sendo o conjunto de sementes. Na inicialização do algoritmo, o usuário é responsável por atribuir cada $x_i \in S$ a um dos k grupos a serem encontrados, dividindo o conjunto S em k subconjuntos S_l , de tal forma que $S = \bigcup_k^{l=1} S_l$. O algoritmo exige que para cada grupo seja atribuído, no mínimo, uma semente. Desta maneira, cada grupo terá seu protótipo inicializado com a média das sementes atribuídas pelo usuário. A partição definida pelas sementes é usada apenas para inicialização e as sementes não são usadas nos passos seguintes do algoritmo.

Algoritmo 3.6: SEEDED K-Means

Entrada: X : conjunto de dados, k : número de grupos, S : conjunto de sementes iniciais.

Saída: P : partição dos dados em k grupos, C : conjunto de exemplos protótipos (centros de grupos).

início

$C :=$ protótipos obtidos por meio do conjunto S ;

repita

 Atualizar P ;

 Atualizar C ;

até nenhum $x \in X$ mudar de grupo;

CONSTRAINED K-Means

Também proposto por Basu et al. (2002), o algoritmo CONSTRAINED K-Means (pseudo-código 3.7) é uma melhoria do algoritmo SEEDED-K-Means. A diferença está nos passos seguintes a inicialização dos protótipos, nos quais os exemplos que fazem parte do conjunto das sementes, e que foram inicialmente associados a um dado grupo pelo usuário, não poderão ser associados a um outro grupo. Assim, apenas os exemplos não selecionados como sementes serão reagrupados, diferentemente do SEEDED-K-Means em que as sementes podem vir a pertencer a grupos diferentes daqueles inicialmente associados. Desta maneira, o CONSTRAINED K-Means é mais adequado quando as sementes, relacionados aos exemplos rotulados, estão livres de ruídos.

Algoritmo 3.7: CONSTRAINED K-Means

Entrada: X : conjunto de dados, k : número de grupos, S : conjunto de sementes iniciais.

Saída: P : partição dos dados em k grupos, C : conjunto de exemplos protótipos (centros de grupos).

início

$C :=$ protótipos obtidos por meio do conjunto S ; **repita**

 Atualizar P , mantendo $x_i \in S$ em seus grupos iniciais;

 Atualizar C ;

até nenhum $x \in X$ mudar de grupo;

COP K-Means

Proposto por Wagstaff et al. (2001), o COP K-Means (pseudo-código 3.8) é uma variante do algoritmo não supervisionado K-Means. Utiliza-se de conhecimento

prévio, descrito na forma de relações entre os exemplos, que é utilizado no processo de formação dos grupos. As restrições são dos tipos *must-link* e *cannot-link*.

Os grupos encontrados pelo COP K-Means devem respeitar todas as relações *must-link* e *cannot-link* impostas pelo usuário nos exemplos rotulados. Durante a construção dos k grupos, cada exemplo do conjunto de exemplos não rotulados é associado ao grupo mais próximo.

Algoritmo 3.8: COP K-Means

Entrada: X : conjunto de dados, M : conjunto de restrições *must-link*, CN : conjunto de restrições *cannot-link*.

Saída: P : partição dos dados em k grupos, C : conjunto de exemplos protótipos (centros de grupos).

início

$C :=$ conjunto de k exemplos $x \in X$;

repita

 Atualizar P , sem violar restrições M e CN ;

 Atualizar C ;

até nenhum $x \in X$ mudar de grupo;

PCK-Means

Este algoritmo também é uma variante do algoritmo não supervisionado K-Means e foi proposto por Basu et al. (2004). Este algoritmo utiliza um cenário de restrições entre pares dos tipos *must-link* e *cannot-link* no conjunto de dados, assim como o COP K-Means. À medida de similaridade utilizada nesse algoritmo é adicionada de dois fatores que avaliam o custo de violação das restrições conhecidas. O custo de violação de uma restrição do tipo *must-link* é dado se os exemplos ligados por *must-link* forem associados a dois diferentes grupos. Similarmente, o custo de violação de uma restrição do tipo *cannot-link* é dado se os exemplos ligados por *cannot-link* forem associados ao mesmo grupo.

O PCK-Means (pseudo-código 3.9) adota o conceito de vizinhança para a inicialização dos protótipos de grupos. A aplicação do fecho transitivo no conjunto de restrições *must-link* aumenta este conjunto pela adição de novas restrições. Uma vizinhança contém os exemplos que estão conectados entre si no conjunto de restrições *must-link*. Seja x_1 um elemento de uma vizinhança λ_i . Seja y tal que x_1 e y estão relacionados dentro do conjunto de restrições *cannot-link*. Para cada elemento x_j de λ_i será adicionada uma restrição entre x_j e y no conjunto *cannot-link*.

Se o número n de vizinhanças for maior que o número k de grupos em que deve ser dividido o conjunto de dados, são selecionadas as k maiores vizinhanças e

seus centróides inicializarão os protótipos de grupos. Se $n < k$, n protótipos serão inicializados com os centróides das vizinhanças e os $k - n$ protótipos restantes serão inicializados com base em exemplos mantêm restrições do tipo *cannot-link* com todas as n vizinhanças. Se ainda houver protótipos não inicializados, estes serão iniciados aleatoriamente.

Algoritmo 3.9: PCK-Means

Entrada: X : conjunto de dados, M : conjunto de restrições *must-link*, CN : conjunto de restrições *cannot-link*, w : peso das restrições, k : número de grupos.

Saída: P : partição dos dados em k grupos, C : conjunto de exemplos protótipos (centros de grupos).

início

$M :=$ aplicação do fecho transitivo no conjunto M ;

 Complemento de CN a partir de M ;

 Inicialização e adaptação de C para k grupos;

repita

 | Atualizar P , considerando custo de violação de restrições;

 | Atualizar C ;

até nenhum $x \in X$ mudar de grupo;

Semi-Supervised Point Prototype Clustering

O algoritmo *semi-supervised Point Prototype Clustering* (ssPPC) (pseudo-código 3.10) foi apresentado por Bensaid; Bezdek (1998); Labzour et al. (1998) e utiliza exemplos rotulados para definir variáveis de entrada de agrupamento não supervisionado e para realizar o ajuste do grau de pertinência dos exemplos não rotulados aos grupos.

Este método, como outros apresentados neste texto, separa o conjunto de dados X em X^l (dados rotulados) e X^u (dados não rotulados). A principal característica do ssPPC é a possibilidade de aplicação de qualquer algoritmo de agrupamento não supervisionado do tipo *point-prototype*. Os algoritmos K-Means e FCM são exemplos deste tipo de algoritmo, pois realizam a divisão de grupos a partir de exemplos (pontos) protótipos.

3.4 Considerações Finais

Este capítulo descreveu conceitos gerais de agrupamento não supervisionado e semissupervisionado contextualizados na área de aprendizado de máquina.

Algoritmo 3.10: Semi-Supervised Point Prototype Clustering

Entrada: X^l : conjunto de dados rotulado, X^u : conjunto de dados não rotulados ppc : agrupamento do tipo point-prototype a ser utilizado.

Saída: P : partição dos dados X^u em n grupos (n é o número de rótulos que existem em X^l), C : conjunto de exemplos protótipos (centros de grupos).

início

$k := |X^l|;$

Aplicação do agrupamento ppc ao conjunto X^u para divisão em k grupos;

$C :=$ protótipos finais obtidos após ppc ;

$P_{aux} :=$ partição final obtida após ppc ;

Rotulação dos k grupos baseando-se na distância entre $c_i \in C$ e $x_j \in X^l$;

Cálculo de P com base em P_{aux} ;

No próximo capítulo são especificados os objetivos e o método de geração de regras fuzzy combinado com agrupamento semissupervisionado, proposto neste projeto de dissertação.

Agrupamento de Dados Semissupervisionado na Geração de Regras Fuzzy

O objetivo deste trabalho é explorar o uso de algoritmos de agrupamento na geração de regras fuzzy. Este capítulo se inicia com uma breve discussão que resgata as principais motivações para se utilizar métodos de agrupamento em sistemas de aprendizado em geral e na geração de regras em particular. Em seguida, são descritos, resumidamente, os mecanismos mais utilizados na geração automática de regras fuzzy e o papel que os métodos de agrupamento têm ou tiveram nesse processo, principalmente por meio da versão não supervisionada. Na sequência, é descrita a proposta de abordagem de geração de regras fuzzy desenvolvida aqui, utilizando os algoritmos de agrupamento fuzzy semissupervisionados para a prévia rotulação de dados não rotulados.

4.1 Agrupamento de Dados

A geração de regras fuzzy, assim como acontece com a geração de regras simbólicas, é feita, na maioria das vezes, por meio de métodos de classificação ou regressão, ou seja, aqueles que utilizam dados rotulados. O grande volume de dados não-rotulados disponíveis, porém, exige investimentos em pesquisas que busquem utilizá-los nos processos de aprendizado de máquina, ao invés de descartá-los.

Este fato evidencia a necessidade e a importância dos métodos de agrupamento. A organização de dados em grupos é uma das formas mais fundamentais para entender e aprender Jain (2010). Em variados domínios podemos identificar a importância da divisão de dados em classes que tenham algum significado. Tarefas de agrupamento de dados ocorrem em diversificados tipos de aplicação, como processamento de imagens, classificação e organização de textos, gerenciamento de recursos, estudo de genoma.

A dificuldade em realizar este tipo de tarefa pode servir de evidência a favor da popularidade de algoritmos antigos e simples, como o K-Means, e do esforço empregado na evolução de algoritmos de agrupamento que possam lidar com distintos tipos de dados.

Tipicamente, métodos de agrupamento não se utilizam de conhecimento prévio, como rótulos de classes, para desempenhar sua tarefa. A característica de conjuntos de dados atuais, grande volume e estruturas mistas com atributos de tipos variados (como texto, *streamings* de vídeo e áudio), contribui para a busca de melhores soluções para tarefas de aprendizado de máquina.

Em decorrência disso, o agrupamento semissupervisionado passou a ser considerado, nos últimos anos, um importante processo a ser explorado com a finalidade de tratar o problema de dispormos de muitos dados, sem dispormos de recursos para rotulá-los manualmente. Pesquisas que envolvem agrupamento semissupervisionado tem o objetivo de gerar estruturas mais fiéis à realidade dos conjuntos de dados. Estruturas sobre as quais, ou conjuntamente, possam ser aplicados outros métodos de aprendizado. O trabalho desenvolvido aqui visa investigar os benefícios que podem ser obtidos no processo de geração automática de regras, com a prévia identificação de tais estruturas contidas nos conjuntos de dados.

4.2 Sistemas Baseados em Regras Fuzzy

Os Sistemas Baseados em Regras Fuzzy (SBRF) são compostos por dois componentes principais: a Base de Conhecimento (BC) e o Mecanismo de Inferência (MI). A geração automática desses sistemas é um tópico frequente nas pesquisas atuais e caracteriza a aproximação entre as áreas de modelagem fuzzy e aprendizado de máquina (Hullermeier, 2005). Como a BC de um SBRF inclui a Base de Dados (BD), que contém os conjuntos fuzzy, e a Base de Regras (BR), que contém as regras, os métodos para geração desses sistemas variam desde o ajuste de conjuntos fuzzy, passando pela geração de um dos componentes da BC, até a geração simultânea de todos os componentes, incluindo parâmetros do MI.

Várias abordagens têm sido usadas para a geração automática da BC a partir

de dados numéricos que representam amostras ou exemplos do problema. Dentre as técnicas mais bem sucedidas estão os algoritmos de agrupamento (*clustering*) (Liao et al., 2003), os métodos baseados em gradiente (Nomura et al., 1992), as árvores de decisão fuzzy (Janikow, 1998), as redes neurais (Jang et al., 1997) e os algoritmos genéticos (Cordón et al., 2004a).

O uso de agrupamento com a finalidade de construir SBRF consiste em identificar regiões no espaço de entrada que podem formar o antecedente de uma regra. Os resultados de agrupamentos fuzzy podem ser transformados em uma BR fuzzy e podem, assim, ser utilizados para extração não supervisionada de regras fuzzy de um conjunto de dados. Cada grupo ou grupo representa uma regra fuzzy **se...então** (Seção 2.2.1). Os conjuntos fuzzy são derivados da projeção dos grupos em determinada dimensão. A base de regras fuzzy pode ser obtida por meio da projeção de todos os grupos.

Devido à aproximação da projeção dos grupos, usualmente para conjuntos fuzzy triangulares ou trapezoidais, as regras fuzzy geradas podem não representar a realidade dos grupos originais. A projeção dos grupos pode trazer, ainda, um problema de interpretabilidade linguística, pois não há restrições quanto a semântica dos conjuntos fuzzy gerados (Klose; Kruse, 2005).

Alternativamente pode-se definir uma partição fuzzy fixa para cada atributo, o que determina regiões fuzzy regulares do espaço de busca, e considerar cada uma dessas regiões como o antecedente de uma regra em potencial (Wang; Mendel, 1992). Essa abordagem é vantajosa do ponto de vista de interpretabilidade dos conjuntos fuzzy mas pode definir modelos imprecisos quando a partição inicial define regiões que não correspondem à estrutura dos dados.

A indução de árvores de decisão, uma das formas mais utilizadas de aprendizado de máquina supervisionado, também tem sido explorada para a geração de regras fuzzy, por meio de variações fuzzy do método original (Janikow, 1998; Olaru; Wehenkel, 2003). Depois que a árvore de decisão foi construída, cada caminho pode ser considerado uma regra. A maneira mais simples de estender uma árvore de decisão é utilizar predicados fuzzy nos nós internos das árvores para flexibilizar os limites de decisão. Essas abordagens são, de acordo com Hullermeier (2005), um exemplo típico de como pode ocorrer a fuzzificação de métodos tradicionais de aprendizado de máquina indutivo.

As abordagens híbridas, que combinam outras metodologias, especialmente redes neurais e computação evolutiva, com a teoria de conjuntos fuzzy, são particularmente importantes para o campo de aprendizado de regras fuzzy. Os métodos neuro-fuzzy incluem entre as abordagens mais utilizadas, a de codificar um sistema fuzzy como uma rede neural e aplicar métodos conhecidos de aprendi-

zado, como *backpropagation*, para treinar essa rede. Assim, os sistemas neuro-fuzzy combinam o poder de representação dos sistemas fuzzy com a capacidade de aprendizado das redes neurais.

Os algoritmos evolutivos em geral e os algoritmos genéticos em particular, têm sido extensivamente utilizados para o aprendizado de SFBR. Nessa abordagem encontra-se uma grande variedade de mecanismos para otimizar ou construir um SFBR ou partes dele, sequencialmente, simultaneamente ou ainda de forma combinada com outros métodos. Por serem de maior interesse para este trabalho, os sistemas fuzzy genéticos são descritos a seguir.

4.2.1 Sistemas Fuzzy Genéticos

Um Sistema Fuzzy Genético (SFG) é o nome utilizado para designar Sistemas Fuzzy combinados com um processo de aprendizado baseado em Algoritmos Genéticos. Os Sistemas Fuzzy buscam a representação do conhecimento impreciso, por meio do uso da lógica fuzzy, cujos conceitos foram descritos no Capítulo 2. Algoritmos Genéticos são aqueles de busca e otimização que se baseiam nos princípios da genética e da seleção natural para encontrar a solução de um determinado problema Goldberg (1989).

A pesquisa sobre os SFGBR já deu origem a um número expressivo de trabalhos e continua em atividade nos dias de hoje. Embora seja difícil classificar as abordagens mais recentes que surgem a cada ano, é possível identificar algumas como clássicas e, também, novas tendências. Importantes contribuições que caracterizam o estado da arte e as principais tendências podem ser encontradas nas diversas edições especiais sobre o assunto, como Cordón et al. (2004b); Casillas et al. (2007); Carse; Pipe (2007); Herrera (2008); Casillas; Carse (2009).

Os dois grupos principais de métodos que combinam as abordagens fuzzy e genética para geração de BC são: Adaptação Genética (Bonissone et al., 1996; Cordón; Herrera, 1997; Casillas et al., 2001; Castro; Camargo, 2004), que contém os métodos que fazem o ajuste ou sintonia de componentes da BC; e Construção Genética (Ishibuchi et al., 1994; Hoffmann; Pfister, 1997; Ishibuchi et al., 1999), que contém os métodos que constroem componentes da BC.

Os SFG são particularmente relevantes para este trabalho, devido a esse ser um tema que tem sido investigado há alguns anos por outras pesquisas do grupo. Assim, alguns dos métodos supervisionados para geração de regras utilizados nesta proposta foram selecionados entre os que se enquadram nessa proposta, visando possíveis trabalhos futuros de comparação dos métodos.

4.3 Agrupamento de Dados na Geração de Regras Fuzzy

A análise de agrupamento passou a ser explorada como um importante mecanismo para geração de SBRF em função do crescimento do interesse por técnicas que permitam a construção automática de tais sistemas a partir de conjuntos de dados.

A abordagem mais comum para gerar regras por meio de métodos de agrupamento fuzzy é considerar cada grupo como uma possível regra. Assim, após encontrar grupos por meio de algum algoritmo, o antecedente da regra é determinado projetando-se o grupo em cada uma das dimensões do espaço de entrada para obter proposições fuzzy da forma “V é A”. A conjunção dessas proposições forma, portanto, o antecedente da regra e cada grupo é associado com uma classe. A desvantagem dessa abordagem é que cada regra faz uso de seus próprios conjuntos fuzzy, gerados pela projeção, o que pode prejudicar a interpretabilidade. Exemplos de trabalhos pioneiros seguindo essa abordagem são Sugeno; Yasukawa (1993); Yager (1993); Babuska et al. (1994); Skarmeta et al. (1999); Abonyi et al. (2002). Em (Delgado et al., 1997) pode ser encontrada uma proposta baseada em formas diferentes de induzir conjuntos fuzzy a partir dos grupos.

Em trabalhos mais recentes encontra-se, frequentemente, abordagens que combinam os métodos de agrupamento com outros mecanismos de aprendizado para a geração de regras fuzzy. Em (Juang, 2005), por exemplo, é proposto um método que combina agrupamento *on-line* e algoritmo genético para a geração de SFBR, com reforço. Outras abordagens, combinado ag com agrupamento pode ser encontradas em (Saez et al., 2008). Em (Lee et al., 2008) é apresentado um novo algoritmo de agrupamento fuzzy iterativo que incorpora um esquema de supervisão em um processo de agrupamento fuzzy não supervisionado. O processo de aprendizado começa de uma maneira não supervisionada e constrói gradualmente partições significativas no espaço de entrada. As regras fuzzy são obtidas por um processo iterativo de selecionar grupos com supervisão baseado nas noções de pureza e separabilidade dos grupos.

Finalmente, vale destacar que os métodos de agrupamento são muito utilizados para geração de conjuntos fuzzy, de forma combinada ou complementar ao processo de geração das regras fuzzy, feito por outro método (Liao et al., 2003).

Embora não focalizem a geração explícita de regras fuzzy, alguns trabalhos adicionais relacionam-se com a proposta descrita aqui, por tratarem a questão de classificação por meio de agrupamento fuzzy. Stutz; Runkler (2002) apresentam extensões ao método de agrupamento fuzzy que permitem a incorporação de conhecimento adicional ao processo de agrupamento, que resultam em sistemas de

classificação e predição específicos para o problema de tráfego. Geng et al. (2009) apresentaram uma proposta para melhorar a classificação de páginas da Web usando um algoritmo de agrupamento fuzzy estendido. O mecanismo de agrupamento fuzzy não supervisionado e semissupervisionado para análise e classificação de imagens é estudado em (Pedrycz et al., 2008).

A proposta de trabalho desenvolvida aqui, aborda a combinação de métodos de agrupamento com sistemas fuzzy de uma maneira diferente das usualmente encontradas na literatura. Essa proposta é apresentada na próxima seção.

4.4 Trabalho Proposto

O objetivo do trabalho proposto aqui é explorar e avaliar o uso de métodos de agrupamento semi supervisionados na geração de regras fuzzy para classificação. A ideia principal é rotular previamente os dados não rotulados usando métodos de agrupamento fuzzy e, na sequência, aplicar métodos de geração de regras fuzzy conhecidos, que utilizam dados rotulados. A principal motivação para esse trabalho foi investigar mecanismos híbridos para utilização de dados não rotulados, pela combinação de duas metodologias que podem se complementar. Enquanto a geração de regras fuzzy é, na maioria das vezes, feita por mecanismos que requerem dados rotulados, a crescente disponibilidade de dados não rotulados evidencia a necessidade de pesquisar métodos adequados à situações como essas, nas quais não se dispõe de informação prévia relativa às classes. Os métodos de agrupamento, por outro lado, adequados à análise de dados não rotulados, apresentam dificuldades relativas à validação e interpretação dos grupos resultantes do processo, o que conseqüentemente dificulta sua utilização. O método de geração de regras proposto visa estudar alternativas de superação de tais dificuldades.

O método desenvolvido neste trabalho pode ser dividido em duas etapas: (a) agrupamento e rotulação; (b) geração de regras . Estas etapas são descritas nas próximas seções.

Etapa de Agrupamento e Rotulação

Para esta etapa foram selecionados quatro algoritmos de agrupamento fuzzy de dados. O algoritmo não supervisionado que serviu de base para as comparações é o popular FCM, detalhado na Seção 3.2.3.

Os algoritmos de agrupamento semissupervisionado ssFCM (Seção 3.3.1), psGK (Seção 3.3.2) e AFCC (Seção 3.3.3) foram selecionados de acordo com suas características e mecanismos de semi-supervisão empregados, de maneira a compreender

alguma diversidade. Os três métodos são adaptações do FCM, sendo que os dois primeiros encaixam-se na abordagem por sementes e o terceiro na abordagem por restrições.

Ainda durante esta etapa, os exemplos não rotulados do conjunto de dados de entrada para os algoritmos de agrupamento semissupervisionado recebem um rótulo.

Esta etapa visa avaliar o desempenho de algoritmos de agrupamento fuzzy que incorporam algum tipo de supervisão parcial com relação a um algoritmo que realiza o agrupamento fuzzy de forma não supervisionada. Além disso, preparar, pelo processo de rotulação, o conjunto de dados parcialmente rotulado, que serviu de entrada, para a aplicação de métodos de geração de regras fuzzy.

Etapa de Geração de Regras

Nesta etapa ocorre o aprendizado supervisionado de regras fuzzy, a partir do conjunto de dados rotulados resultante da etapa anterior. Os algoritmos aplicados são de classificação e tem como resultado uma BR fuzzy para o conjunto de dados de entrada. O objetivo é comparar os resultados obtidos após aplicação da geração de BR ao conjunto parcialmente rotulado que foi particionado por um algoritmo de agrupamento semi-supervisionando e passou por um processo de rotulação de seus exemplos não classificados, com aqueles obtidos pela geração de regras a partir do conjunto rotulado original

Para a execução dessa etapa foram utilizados um método de geração de regras fuzzy não-evolutivo, o Fuzzy Rule Learning Model (Chi et al., 1996; Cordón et al., 1999; Ishibuchi; Yamamoto, 2005). O Fuzzy Rule Learning Model, que será referido neste trabalho por Chi-RW, realiza a construção de um sistema fuzzy baseado em regras, que são obtidas por meio de uma técnica adaptada a partir da técnica de Wang; Mendel (1992), e contam, ainda, com pesos que ajudam na classificação. Os outros algoritmos selecionados executam a tarefa de geração de regras fuzzy por métodos evolutivos. O Hybrid Fuzzy GBML (Ishibuchi et al., 2005) (referido por Ishib-Hybrid neste trabalho) é um método que implementa um hibridismo entre as abordagens de algoritmos genéticos consagradas: Michigan (Ishibuchi et al., 1999) e Pittsburgh (Hoffmann; Pfister, 1997). Já o SGERD (Mansoori et al., 2008) é um algoritmo genético com número de gerações finitas e ligadas à dimensão do problema avaliado. O objetivo do SGERD é extrair uma BR mais compacta e mais legível. O método SLAVE (González; Perez, 2001) é um algoritmo genético que se encaixa na abordagem iterativa. Os métodos selecionados foram implementados com a ferramenta Keel – *Knowledge Extraction based on Evolutionary Learning* (Alcala-Fdez

et al., 2009, 2011) (<http://www.keel.es>). A avaliação do resultado foi feita com base na porcentagem de acertos da classificação.

4.5 Considerações Finais

Neste capítulo foi retomada a motivação para desenvolvimento do projeto de dissertação e foram descritas as diferentes etapas do projeto. O capítulo seguinte detalha a validação do projeto, os experimentos realizados e apresenta avaliações dos resultados obtidos.

Experimentos e Resultados

O objetivo deste capítulo é apresentar e analisar os resultados dos experimentos realizados com os algoritmos de agrupamento e de geração de regras fuzzy. A colaboração entre agrupamento semissupervisionado e a geração de regras fuzzy, foco do trabalho, também será analisada.

A Seção 5.1 descreve conjuntos de dados, características dos experimentos realizados, o processo de validação, entre outras considerações iniciais. A Seção 5.2 coloca os resultados obtidos e suas análises.

5.1 Considerações Preliminares

Para validação da proposta, foram utilizados 13 conjuntos de dados bastante populares para avaliar os resultados obtidos pelos métodos de agrupamento e classificação. Os conjunto de atributos numéricos, descritos na Tabela 5.1, estão disponíveis no repositório UCI para aprendizado de máquina (Frank; Asuncion, 2010).

O método *cross-validation* para *5-fold* (5-CV) foi adotado para comparar a performance de algoritmos de agrupamento e classificação. As performances são baseadas em múltiplos conjuntos de treinamento e teste. Para 5-CV, cada conjunto de dados é randomicamente dividido em cinco conjuntos disjuntos de igual cardinalidade. Cada conjunto é utilizado como conjunto de teste, tendo como conjunto de treinamento os quatro outros conjuntos restantes. Os métodos de agrupamento e classificação são treinados 5 vezes, cada vez com um conjunto diferente de teste.

Tabela 5.1: Conjuntos de dados utilizados nos experimentos para validação

Conjunto de Dados	Número de Exemplos	Dimensão	Número de Classes
Balance	625	4	3
Bupa	345	6	2
Haberman	306	3	2
Ionosphere	351	33	2
Iris	150	4	3
Monk 2	432	6	2
New Thyroid	215	5	3
Pima	768	8	2
Sonar	208	60	2
SPECTF Heart	267	44	2
Vehicle	846	18	4
WDBC	569	30	2
Wine	178	13	3

A performance estimada está relacionada a média dos 5 resultados.

Os objetivos da proposta permitem que a validação seja dividida em duas partes: a validação dos resultados obtidos pela aplicação dos métodos de agrupamento e a validação das regras fuzzy geradas por métodos de classificação. As especificações de cada uma das partes da validação são descritas de forma mais detalhada nas seções 5.1.1 e 5.1.2.

5.1.1 Validação de Agrupamento

Para validação dos resultados dos métodos de agrupamento, foram realizados experimentos considerando a aplicação de quatro algoritmos de agrupamento fuzzy distintos, sendo um deles não supervisionado, dois semissupervisionados baseados em sementes e um semissupervisionado baseado em restrições. Especificações dos algoritmos são descritas a seguir:

1 - FCM - Método não supervisionado, descrito pelo algoritmo 3.2. Para os experimentos, foram retirados os rótulos das bases de dados e a métrica utilizada para comparação entre padrões de uma mesma base foi a distância euclidiana (Equação 3.3).

2 - ssFCM - Método semissupervisionado baseado em sementes, descrito pelo algoritmo 3.3. Para os experimentos, foram sorteados conjuntos de exemplos para servirem de sementes.

3 - psGK - Método semissupervisionado baseado em sementes, descrito pelo algoritmo 3.4. Para os experimentos, também foram sorteados conjuntos de exem-

plos que são as sementes para o algoritmo.

4 - AFCC - Método semissupervisionado baseado em restrições, descrito pelo algoritmo 3.5. Para os experimentos, foram sorteados conjuntos de exemplos (conjunto de sementes), a partir dos quais foram definidas as restrições *must-link* e *cannot-link*. O conjunto de sementes é uma porcentagem dos padrões do conjunto de treinamento.

O sorteio de exemplos para os conjuntos de sementes foram realizados de duas maneiras diferentes. Foram utilizados conjuntos de sementes de 10%, 20% e 30% do total de exemplos no conjunto de treinamento. Para cada porcentagem, foram gerados dois tipos de conjuntos. Um deles foi definido por meio de sorteio entre todos os exemplo do conjunto de dados, sem qualquer restrição. O outro conjunto, também obtido por meio de sorteio, respeitava a restrição de que, do total de sementes, todas as classes do conjunto de dados deveriam estar representadas no conjunto de sementes de forma igualitária. Assim, supondo que a cardinalidade do conjunto de sementes assumi-se o valor s , cada classe existente no conjunto de treinamento teria s/c , onde c é o número de classes do conjunto.

Para as implementações dos algoritmos de agrupamento foram utilizadas algumas estruturas do pacote de software para mineração de dados WEKA (Hall et al., 2009).

A validação dos métodos de agrupamento é dada com relação a uma parte dos índices descritos na Seção 3.2.4.

5.1.2 Validação de Classificação

A validação desta etapa dá-se pela comparação dos resultados obtidos com a aplicação dos métodos de classificação sobre os conjuntos de dados originais selecionados do repositório UCI e sobre os mesmo conjuntos de dados, após passar pelo processo de agrupamento semissupervisionado e rotulação.

Para a comparação dos resultados, o teste de Friedman (Demšar, 2006; Lei; Ren-hou, 2008) foi utilizado. O teste de Friedman faz a atribuição de rankings (um por conjunto de dados) para o resultados da aplicação de algoritmos de classificação. Este teste é baseado em uma hipótese nula que assume que todos os algoritmos tem desempenho semelhante e, então, teriam rankings semelhantes. Se possível provar que há diferença entre o desempenho de um ou mais algoritmos, é possível aplicar testes chamados post-hoc para saber se a performance dos algoritmos é significativamente diferente.

De uma forma geral, o teste de Friedman é aplicado para avaliação entre algoritmos. Neste trabalho, a ideia é que se possa mostrar se a aplicação de algoritmos

Tabela 5.2: Índice de heterogeneidade geral (R) para os métodos FCM

Conjunto	FCM(2)	FCM(3)	FCM(4)
Balance	0,0042	0,0366	0,0818
Bupa	0,0001	0,0187	0,0717
Haberman	0,8935	0,9432	0,9607
Ionosphere	0	0	0
Iris	0,5224	0,6476	0,7188
Monk2	0,0174	0,0436	0,0697
New Thyroid	0,6579	0,7499	0,8184
Pima	0	0,0006	0,0072
Sonar	0	0	0
SPECTF Heart	0	0	0
Vehicle	0,0026	0,0041	0,0058
WDBC	0,0025	0,0056	0,0161
Wine	0,0038	0,0116	0,1196

de classificação a conjuntos que passaram por processos de agrupamento semissupervisionado tem alguma influência no desempenho do algoritmo de classificação.

Os cálculos para o teste de Friedman, e eventual teste post-hoc, foram realizados pelo software de análise de dados GraphPad (<http://www.graphpad.com/>).

A próxima seção apresenta os resultados dos experimentos realizados.

5.2 Resultados de Experimentos e Análise

O objetivo de um algoritmo de agrupamento é encontrar a melhor partição possível para um determinado conjunto de dados. Isso implica em particionar o conjunto de forma a obter grupos o mais homogêneos e compactos possível.

Um dos inimigos do agrupamento pode ser, justamente, o conjunto de dados. Conjuntos com alta dispersão dos dados podem influenciar negativamente o resultado do agrupamento. As tabelas 5.2 e 5.3 indicam o índice R , apresentado na Seção 3.2.4, para o resultado do agrupamento de 13 conjuntos de dados, com a aplicação dos algoritmos FCM¹ e ssFCM².

Os baixos valores para R indicam que o conjunto é consideravelmente disperso. É interessante verificar que ao aumentar o número de grupos para o particionamento com o algoritmo FCM, o valor de R também aumentou. A partição em muitos grupos pode produzir melhores índices para R mas não significa, necessaria-

¹em FCM(i), i é o número de grupos em que o algoritmo particionou o conjunto

²em ssFCM(i), i é a porcentagem relativa ao total do conjunto que foi utilizada como conjunto de sementes

Tabela 5.3: Índice de heterogeneidade geral (R) para os métodos ssFCM

Conjunto	ssFCM(10%)	ssFCM(20%)	ssFCM(30%)
Balance	0,0192	0,0168	0,0121
Bupa	0	0	0
Haberman	0,9049	0,8778	0,8628
Ionosphere	0	0	0
Iris	0,6800	0,6815	0,6825
Monk2	0,0221	0,0114	0,007
New Thyroid	0,6985	0,6842	0,6931
Pima	0	0	0
Sonar	0	0	0
SPECTF Heart	0	0	0
Vehicle	0,0019	0,0012	0,0005
WDBC	0,0020	0,0016	0,0012
Wine	0,0003	0,0003	0,0003

mente, que trata-se de uma boa partição.

Nota-se que o valor de R para este conjunto, com a aplicação do agrupamento ssFCM, diminui com o aumento do número de sementes iniciais. A natureza do conjunto, quanto à dispersão de dados, faz com que um maior número de sementes “confunda” o algoritmo.

As tabelas 5.4, 5.5 e 5.6 apresentam a média taxa de classificação correta (TCC) para os experimentos realizados com os algoritmos Chi-RW, Ishib-Hybrid, SGERD e SLAVE e os diferentes tipos de conjuntos de dados. A tabela 5.4 contém os resultados para as bases de regras geradas a partir dos conjuntos de dados rotulados originais (O). As tabelas 5.5 e 5.6 apresentam os resultados para conjuntos de sementes planejadas (P) e sementes aleatórias (A), que foram submetidos a um agrupamento semissupervisionado (ssFCM) e, então, rotulados.

De uma forma geral, os resultados para geração de regras não parecem muito diferentes quanto à aplicação a conjuntos rotulados e conjuntos agrupados. Para tentar avaliar esta observação, é aplicado o teste de Friedman.

O teste está dividido em dois casos: no primeiro, verificamos o teste para um único tipo de dados (por exemplo, o conjunto original rotulado) e varia-se o método de geração de regras. Esse procedimento é realizado para os três tipos de conjuntos.

Ao realizar o teste de Friedman para os tipos de conjuntos de dados individualmente, tem-se o objetivo de verificar se os algoritmos de classificação apresentam comportamento semelhante, independente do tipo de conjunto de dados. Ou seja,

Tabela 5.4: TCC para conjuntos de dados O

Conjunto	Chi-RW	Ishib-Hybrid	SGERD	SLAVE
Balance	90,70	84,20	75,00	81,00
Bupa	59,13	70,43	60,11	63,80
Haberman	73,77	75,20	74,14	73,94
Ionosphere	81,39	65,39	76,71	94,59
Iris	92,83	95,17	93,92	96,83
Monk 2	69,43	95,91	80,55	97,22
New Thyroid	85,52	94,53	89,77	91,80
Pima	73,99	76,50	71,76	75,22
Sonar	79,59	48,95	72,86	78,73
SPECTF Heart	81,47	73,03	78,89	79,87
Vehicle	63,39	52,50	51,89	71,47
WDBC	94,40	91,78	91,81	92,05
Wine	95,70	91,13	90,81	95,62

Tabela 5.5: TCC para conjuntos de dados agrupado P

Conjunto	Chi-RW	Ishib-Hybrid	SGERD	SLAVE
Balance	85,56	84,52	85,00	84,82
Bupa	80,87	88,99	85,11	90,94
Haberman	67,79	90,12	64,05	83,10
Ionosphere	74,29	61,50	71,66	83,76
Iris	91,50	93,83	92,58	93,50
Monk 2	53,07	88,52	83,58	91,06
New Thyroid	58,84	58,84	58,84	58,84
Pima	77,38	85,94	78,74	82,03
Sonar	74,28	47,24	71,41	77,29
SPECTF Heart	81,47	73,03	78,89	79,87
Vehicle	81,09	77,69	84,18	84,62
WDBC	92,42	91,24	95,45	96,38
Wine	95,70	91,13	90,81	95,62

Tabela 5.6: TCC para conjuntos de dados agrupado A

Conjunto	Chi-RW	Ishib-Hybrid	SGERD	SLAVE
Balance	79,32	80,32	77,86	82,96
Bupa	80,18	86,26	82,57	89,34
Haberman	70,26	88,85	29,68	87,75
Ionosphere	72,4	53,91	71,66	84,67
Iris	91,33	93,25	91,75	92,91
Monk 2	54,14	91,49	89,8	92,42
New Thyroid	77,15	88,37	78,43	81,33
Pima	78,22	87,5	79,07	84,98
Sonar	74,27	47,24	71,4	77,29
SPECTF Heart	83,8	68,93	71,67	82,25
Vehicle	79,03	76,48	79,09	83,89
WDBC	92,42	93,31	95,71	96,85
Wine	76,64	80,78	83,34	85,82

Tabela 5.7: Teste de Friedman com post-hoc de Dunn para os conjuntos O

Comparação	Diferença entre Soma de Rank	Valor de P
Chi-RW vs. Ishib-Hybrid	0,000	ns P>0,05
Chi-RW vs. SGERD	10,000	ns P>0,05
Chi-RW vs. SLAVE	-8,000	ns P>0,05
Ishib-Hybrid vs. SGERD	10,000	ns P>0,05
Ishib-Hybrid vs. SLAVE	-8,000	ns P>0,05
SGERD vs. SLAVE	-18,000	* P<0,05

um algoritmo a_1 , que apresenta melhor desempenho, com alguma significância³, que um outro algoritmo a_2 , quando aplicado a um conjunto do tipo O, continuará apresentando desempenho semelhante, quanto aos demais algoritmos, quando aplicado a conjuntos P ou A.

Avaliando os resultados⁴ expressos nas tabelas 5.7, 5.8 e 5.9, verifica-se que a distribuição de desempenho entre os algoritmos é próxima para qualquer tipo de conjunto de dados. Percebe-se que o algoritmo SLAVE apresenta alguma vantagem sobre os demais. Entretanto, com exceção do caso da tabela 5.9, não é significativo.

Para o caso da tabela 5.9, é interessante verificar uma saída complementar⁵ para o teste de Friedman (tabela 5.10).

³A significância, neste contexto, esta relacionada ao valor de P que aparece nas tabelas para o teste de Friedman e determina o grau de relevância da diferença de desempenho entre dois algoritmos. Nas tabelas, as letras "ns" indicam que a diferença é "não significativa".

⁴Os resultados são mostrados no formato de saída oferecido pelo software GraphPad.

⁵Trazida no relatório da execução do teste no software GraphPad

Tabela 5.8: Teste de Friedman com post-hoc de Dunn para os conjuntos P

Comparação	Diferença entre Soma de Rank	Valor de P
Chi-RW vs. Ishib-Hybrid	2,000	ns P>0,05
Chi-RW vs. SGERD	3,000	ns P>0,05
Chi-RW vs. SLAVE	-13,000	ns P>0,05
Ishib-Hybrid vs. SGERD	1,000	ns P>0,05
Ishib-Hybrid vs. SLAVE	-15,000	ns P>0,05
SGERD vs. SLAVE	-16,000	ns P>0,05

Tabela 5.9: Teste de Friedman com post-hoc de Dunn para os conjuntos A

Comparação	Diferença entre Soma de Rank	Valor de P
Chi-RW vs. Ishib-Hybrid	-9,000	ns P>0,05
Chi-RW vs. SGERD	-5,000	ns P>0,05
Chi-RW vs. SLAVE	-24,000	** P<0,01
Ishib-Hybrid vs. SGERD	4,000	ns P>0,05
Ishib-Hybrid vs. SLAVE	-15,000	ns P>0,05
SGERD vs. SLAVE	-19,000	* P<0,05

Tabela 5.10: Somas de Rank por algoritmo para cada tipo de conjunto (O, P e A)

Comparação	Soma de Ranks (O)	Soma de Ranks (P)	Soma de Ranks (A)
Chi-RW	33,00	30,50	23,00
Ishib-Hybrid	33,00	28,50	32,00
SGERD	23,00	27,50	28,00
SLAVE	41,00	43,50	47,00

Tabela 5.11: Teste de Friedman com post-hoc de Dunn para algoritmo Ishib-Hybrid

Comparação	Diferença entre Soma de Rank	Valor de P
Ishib-Hybrid (O) vs. Ishib-Hybrid (P)	0,5000	ns P>0,05
Ishib-Hybrid (O) vs. Ishib-Hybrid (A)	5,500	ns P>0,05
Ishib-Hybrid (P) vs. Ishib-Hybrid (A)	5,000	ns P>0,05

Tabela 5.12: Teste de Friedman com post-hoc de Dunn para o algoritmo SGERD

Comparação	Diferença entre Soma de Rank	Valor de P
SGERD (O) vs. SGERD (P)	-2,500	ns P>0,05
SGERD (O) vs. SGERD (A)	2,500	ns P>0,05
SGERD (P) vs. SGERD (A)	5,000	ns P>0,05

Nota-se que o desempenho do algoritmo Chi-RW caiu consideravelmente quando aplicado aos conjuntos do tipo A. A investigação do motivo, talvez seja interessante, porém envolveria conhecimento mais aprofundado sobre os mecanismos específicos adotados pelo algoritmo de geração de BR.

Pode-se dizer, devido a semelhança de desempenho ilustrada pelas tabelas 5.7, 5.8 e 5.9, que, de forma geral, o resultado da rotulação pós agrupamento semissupervisionado deve estar próxima aos rótulos originais dos conjuntos.

No segundo caso para análise do experimento, o método é fixo e são avaliadas as combinações com os diferentes tipos de dados. Aqui o objetivo é avaliar o impacto dos tipos de conjuntos com relação a um único algoritmo. Considera-se que o foco do agrupamento semissupervisionado com binado com a rotulação é chegar o mais próximo possível dos rótulos originais. Assim, se o objetivo foi atingido, não se deve esperar diferenças significativas no desempenho do algoritmo.

As tabelas 5.11, 5.12 e 5.13 reforçam, mais uma vez, a observação que foi colocado anteriormente. As relação de desempenho entre algoritmos de geração de BR com aplicação a tipos de dados diferentes: o resultado do agrupamento e rotulação deve estar bastante próximo ao original.

Tabela 5.13: Teste de Friedman com post-hoc de Dunn para o algoritmo SLAVE

Comparação	Diferença entre Soma de Rank	Valor de P
SLAVE (O) vs. SLAVE (P)	-0,5000	ns P>0,05
SLAVE (O) vs. SLAVE (A)	-2,500	ns P>0,05
SLAVE (P) vs. SLAVE (A)	-2,000	ns P>0,05

5.3 Considerações Finais

Neste capítulo foram apresentados e discutidos os experimentos realizados com os métodos de agrupamento de dados e geração de base de regras fuzzy sobre diferentes domínios. O capítulo seguinte traz conclusões sobre os experimentos e sobre o projeto como um todo, além da indicação de trabalhos futuros.

Este trabalho apresentou uma proposta de aprendizado semissupervisionado que se utiliza de algoritmos de agrupamento fuzzy semissupervisionado, para a rotulação de dados e posterior aplicação de métodos de classificação, cujo objetivo é gerar uma base de regras fuzzy.

A proposta está direcionada à tentativa de solução do problema de aprendizado de máquina em domínios que se caracterizam pelo grande volume e a mistura de tipos de dados, porém contam com um conjunto relativamente pequeno (ou nenhum conjunto) de exemplos já classificados.

Primeiramente a ideia foi apresentada para, então, descrever a fundamentação teórica. Esta descrição teve início no Capítulo 2, que abordou conceitos teóricos básicos a respeito de teoria de conjuntos fuzzy e sistemas fuzzy.

No Capítulo 3 foi apresentada uma visão do problema de aprendizado de máquina, abordando conceitos gerais de agrupamento não supervisionado e semissupervisionado de dados, incluindo a descrição do processo geral de agrupamento e variados algoritmos de agrupamento semissupervisionado.

No Capítulo 4 foi reforçada a motivação para a exploração de algoritmos de agrupamento na geração de regras fuzzy. O detalhamento da proposta de trabalho foi descrito neste capítulo.

Subsequente, no Capítulo 5, foram apresentados os modelos seguidos para a realização de experimentos, detalhes sobre a validação da proposta, resultados obtidos e sua análise. Os experimentos realizados sugerem que o algoritmo de agrupamento semissupervisionado, de uma forma geral, alcançou o objetivo de gerar partições dos conjuntos avaliados que fossem semelhantes ao conjunto rotulado original. Esta observação foi feita em vista dos resultados apresentados pelo teste de Friedman. A discussão do capítulo anterior confirmou a relevância e importância da proposta de trabalho.

O desenvolvimento deste trabalho possibilitou a consideração de questões relevantes que podem ser exploradas em trabalhos futuros:

1. Confirmação do observado sobre a aproximação (quanto aos rótulos) dos conjuntos agrupados e conjuntos rotulados originais pela aplicação medidas de avaliação multi-objetivo;
2. Adaptação e proposta de métodos de agrupamento semissupervisionado, visando a questão da rotulação dos dados parcialmente classificados;
3. Investigação de outras soluções para rotulação de dados agrupados de forma semi-supervisionada;

4. Investigação de métodos semissupervisionados para geração de regras fuzzy, que não ocorram por agrupamento;
5. Comparação de resultados com distintas abordagens de aprendizado semisupervisionado;
6. Análise de resultados da aplicação do método proposto para a solução de problemas em domínios específicos.

Referências Bibliográficas

- Abonyi, J.; Babuska, R.; Szeifert, F. Modified gath-geva fuzzy clustering for identification of takagi-sugeno fuzzy models. **IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics**. v. 32, n. 5, p. 612–621, 2002.
- Alcala-Fdez, J.; Fernández, A.; Luengo, J.; Derrac, J.; García, S.; Sánchez, L.; Herrera, F. Keel data-mining software tool: Data set repository and integration of algorithms and experimental analysis framework. **Journal of Multiple-Valued Logic and Soft Computing**. v. 17, n. 2–3, p. 255–287, 2011.
- Alcala-Fdez, J.; Sánchez, L.; García, S.; del Jesus, M.; Ventura, S.; Garrell, J.; Otero, J.; Romero, C.; Bacardit, J.; Rivas, V.; Fernández, J.; Herrera, F. Keel: A software tool to assess evolutionary algorithms for data mining problems. **Soft Computing**. v. 13, n. 3, p. 307–318, 2009.
- Babuska, R.; Jager, R.; Verbruggen, H., 1994. Interpolation issues in sugeno-takagi reasoning. In: **Proceedings of the Third IEEE Conference on Fuzzy Systems** p. 859–863.
- Basu, S.; Banerjee, A.; Mooney, R. J., 2002. Semi-supervised clustering by seeding. In: **ICML '02: Proceedings of the Nineteenth International Conference on Machine Learning** p. 27–34.
- Basu, S.; Banerjee, A.; Mooney, R. J., 2004. Active semi-supervision for pairwise constrained clustering. In: **Proceedings of the 2004 SIAM International Conference on Data Mining** p. 333–344.
- Bensaid, A.; Bezdek, J. C. Semi-supervised point prototype clustering. **International Journal of Pattern Recognition and Artificial Intelligence**. v. 12, n. 4, p. 625–643, 1998.

Bensaid, A. M.; Hall, L. O.; Bezdek, J. C.; Clarke, L. P. Partially supervised clustering for image segmentation. **Pattern Recognition**. v. 29, n. 5, p. 859–871, 1996.

Bezdek, J. C. **Pattern Recognition with Fuzzy Objective Function Algorithms**. Norwell, MA, USA: Kluwer Academic Publishers, 1981.

Bishop, C. M. **Neural Networks for Pattern Recognition**. New York, NY, USA: Oxford University Press, 1995.

Bonissone, P.; Khedkar, P. S.; Chen, Y., 1996. Genetic algorithms for automated tuning of fuzzy controllers: A transportation application. In: **Proceedings of the 5th IEEE International Conference on Fuzzy Systems** p. 674–680.

Campello, R. J. G. B.; Hruschka, E. R. A fuzzy extension of the silhouette width criterion for cluster analysis. **Fuzzy Sets and Systems**. v. 157, n. 21, p. 2858–2875, 2006.

Carse, B.; Pipe, A. G. Introduction: Genetic fuzzy systems: Introduction. **International Journal on Intelligent Systems**. v. 22, n. 9, p. 905–907, 2007.

Carvalho, F. A. T.; Tenorio, C. P.; Junior, N. L. C. Partitional fuzzy clustering methods based on adaptive quadratic distances. **Fuzzy Sets and Systems**. v. 157, n. 21, p. 2833–2857, 2006.

Casillas, J.; Carse, B. Special issue on genetic fuzzy systems: Recent developments and future directions. **Soft Computing**. v. 13, n. 5, p. 417–418, 2009.

Casillas, J.; Cordón, O.; Jesus, M. J. D.; Herrera, F. Genetic feature selection in a fuzzy rule-based classification system learning process for high-dimensional problems. **Information Sciences**. v. 136, n. 1–4, p. 135–157, 2001.

Casillas, J.; Herrera, F.; Pérez, R.; del Jesús, M. J.; Villar, P. Special issue on genetic fuzzy systems and the interpretability-accuracy trade-off. **International Journal of Approximate Reasoning**. v. 44, n. 1, p. 1–3, 2007.

Castro, P. A. D.; Camargo, H. A., 2004. A study of the reasoning methods impact on genetic learning and optimization of fuzzy rules. In: **Advances in Artificial Intelligence – SBIA 2004**, v. 3171 of **Lecture Notes in Computer Science** p. 414–423.

Castro, P. A. D.; Camargo, H. A., 2005. Focusing on interpretability and accuracy of a genetic fuzzy system. In: **The 14th IEEE International Conference on Fuzzy Systems, 2005** p. 696–701.

- Chapelle, O.; Schlökopf, B.; Zien, A., Eds. **Semi-Supervised Learning**. Cambridge, MA, USA: MIT Press, 2006.
- Chi, Z.; Yan, H.; Pham, T. **Fuzzy Algorithms: With Applications To Image Processing and Pattern Recognition**. World Scientific, 1996.
- Cintra, M. E.; Camargo, H. A., 2007. Fuzzy rules generation using genetic algorithms with self-adaptive selection. In: **IEEE International Conference on Information Reuse and Integration**, v. 1 p. 261–266.
- Cintra, M. E.; Camargo, H. A., 2010. Feature subset selection for fuzzy classification methods. In: **International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems. Proceedings of IPMU2010**, v. 7 p. 33–43.
- Cintra, M. E.; Camargo, H. A.; Hruschka Júnior, E. R.; Nicoletti, M. C. Automatic construction of fuzzy rule bases: a further investigation into two alternative inductive approaches. **Journal of Universal Computer Science**. v. 14, p. 2456–2470, 2008.
- Cintra, M. E.; Monard, M. C.; Camargo, H. A., 2010. An evaluation of rule-based classification models induced by a fuzzy method and two classic learning algorithms. In: **Brazilian Symposium on Neural Networks - SBRN** p. 188–193.
- Cordón, O.; del Jesús, M. J.; Herrera, F. A proposal on reasoning methods in fuzzy rule-based classification systems. **International Journal of Approximate Reasoning**. v. 20, n. 1, p. 21–45, 1999.
- Cordón, O.; Gomide, F.; Herrera, F.; Hoffmann, F.; Magdalena, L. Special issue on genetic fuzzy systems. **Fuzzy Sets and Systems**. v. 141, 2004a.
- Cordón, O.; Gomide, F.; Herrera, F.; Hoffmann, F.; Magdalena, L. Ten years of genetic fuzzy systems: current framework and new trends. **Fuzzy Sets and Systems**. v. 141, n. 1, p. 5–31, 2004b.
- Cordón, O.; Herrera, F. A three-stage evolutionary process for learning descriptive and approximate fuzzy-logic-controller knowledge bases from examples. **International Journal of Approximate Reasoning**. v. 17, n. 4, p. 369–407, 1997.
- Cover, T. M.; Hart, P. E. Nearest neighbor pattern classification. **IEEE Transactions in Information Theory**. v. IT-13, n. 1, p. 21–27, 1967.

- Delgado, M.; Gomez-Skarmeta, A.; Martin, F. A fuzzy clustering-based rapid prototyping for fuzzy rule-based modeling. **IEEE Transactions on Fuzzy Systems**. v. 5, n. 2, p. 223–233, 1997.
- Demšar, J. Statistical comparisons of classifiers over multiple data sets. **Journal of Machine Learning Research**. v. 7, p. 1–30, 2006.
- Duda, R. O.; Hart, P. E. **Pattern classification and scene analysis**. John Wiley and Sons, 1973.
- Frank, A.; Asuncion, A., 2010. UCI machine learning repository. University of California, Irvine, School of Information and Computer Sciences. <http://archive.ics.uci.edu/ml>.
- Geng, C.; Yuquan, Z.; Jianing, T.; Tianhan, H., 2009. An algorithm of semi-supervised web-page classification based on fuzzy clustering. In: **International Forum on Information Technology and Applications**, v. 1 p. 3–7.
- Goldberg, D. E. **Genetic Algorithms in Search, Optimization, and Machine Learning**. Addison-Wesley Professional, 1989.
- González, A.; Perez, R. Selection of relevant features in a fuzzy genetic learning algorithm. **IEEE Transactions on Systems, Man and Cybernetics, Part B: Cybernetics**. v. 31, n. 3, p. 417–425, 2001.
- Gira, N.; Crucianu, M.; Boujemaa, N., 2005. Active semi-supervised fuzzy clustering for image database categorization. In: **MIR '05: Proceedings of the 7th ACM SIGMM international workshop on Multimedia information retrieval** p. 9–16: ACM.
- Gira, N.; Crucianu, M.; Boujemaa, N. Fuzzy clustering with pairwise constraints for knowledge-driven image categorisation. **IEE Proceedings - Vision, Image and Signal Processing**. v. 153, n. 3, p. 299–304, 2006.
- Gira, N.; Crucianu, M.; Boujemaa, N. Active semi-supervised fuzzy clustering. **Pattern Recognition**. v. 41, n. 5, p. 1851–1861, 2008.
- Gustafson, D. E.; Kessel, W. C., 1978. Fuzzy clustering with a fuzzy covariance matrix. In: **Proceedings of the IEEE Conference on Decision and Control including the 17th Symposium on Adaptive Processes**, v. 17 p. 761–766.
- Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; Witten, I. H. The weka data mining software: an update. **SIGKDD Explorations Newsletter**. v. 11, n. 1, p. 10–18, 2009.

- Herrera, F. Genetic fuzzy systems: Taxonomy and current research trends and prospects. **Evolutionary Intelligence**. v. 1, p. 27–46, 2008.
- Hoffmann, F.; Pfister, G. Evolutionary design of a fuzzy knowledge base for a mobile robot. **International Journal of Approximate Reasoning**. v. 17, n. 4, p. 447–469, 1997.
- Höppner, F.; Klawonn, F.; Kruse, R.; Runkler, T. **Fuzzy Cluster Analysis**. John Wiley and Sons, Inc., 1999.
- Hullermeier, E. Experience-based decision making: A satisficing decision tree approach. **IEEE Transactions on Systems, Man, and Cybernetics, Part A**. v. 35, n. 5, p. 641–653, 2005.
- Ishibuchi, H.; Nakashima, T.; Murata, T. Performance evaluation of fuzzy classifier systems for multidimensional pattern classification problems. **IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics**. v. 29, n. 5, p. 601–618, 1999.
- Ishibuchi, H.; Nozaki, K.; Tanaka, H.; Hosaka, Y.; Matsuda, M. Empirical study on learning in fuzzy systems by rice taste analysis. **Fuzzy Sets and Systems**. v. 64, n. 2, p. 129–144, 1994.
- Ishibuchi, H.; Yamamoto, T. Rule weight specification in fuzzy rule-based classification systems. **IEEE Transactions on Fuzzy Systems**. v. 13, n. 4, p. 428–435, 2005.
- Ishibuchi, H.; Yamamoto, T.; Nakashima, T. Hybridization of fuzzy gbml approaches for pattern classification problems. **IEEE Transactions on Systems, Man and Cybernetics, Part B: Cybernetics**. v. 35, n. 2, p. 359–365, 2005.
- Jain, A. K. Data clustering: 50 years beyond k-means. **Pattern Recognition Letters**. v. 31, n. 8, p. 651–666, 2010.
- Jain, A. K.; Murty, M. N.; Flynn, P. J. Data clustering: a review. **ACM Computing Survey**. v. 31, n. 3, p. 264–323, 1999.
- Jang, J. S. R.; Sun, C. T.; Mizutani, E. **Neuro-Fuzzy and Soft Computing**. Matlab Curriculum Series. Prentice Hall, 1997.
- Janikow, C. Z. Fuzzy decision trees: issues and methods. **IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics**. v. 28, n. 1, p. 1–14, 1998.

- Jiang, D.; Tang, C.; Zhang, A. Cluster analysis for gene expression data: A survey. **IEEE Transactions on Knowledge and Data Engineering**. v. 16, n. 11, p. 1370–1386, 2004.
- Juang, C. F. Combination of online clustering and q-value based ga for reinforcement fuzzy system design. **IEEE Transactions on Fuzzy Systems**. v. 13, n. 3, p. 289–302, 2005.
- Kecman, V. **Learning and Soft Computing**. Cambridge, MA, USA: MIT Press, 2001.
- Klir, G. J.; Yuan, B. **Fuzzy Sets and Fuzzy Logic: Theory and Applications**. Upper Saddle River, NJ, USA: Prentice Hall, 1995.
- Klose, A.; Kruse, R. Semi-supervised learning in knowledge discovery. **Fuzzy Sets and Systems**. v. 149, p. 209–233, 2005.
- Klose, A.; Nürnberger, A.; Nauck, D.; Kruse, R. Data mining with neuro-fuzzy models. **Data mining and computational intelligence**, p. 1–35, 2001.
- Labzour, T.; Bensaid, A.; Bezdek, J., 1998. Improved semi-supervised point-prototype clustering algorithms. In: **IEEE International Conference on Fuzzy Systems** p. 1383–1387.
- Lee, H. E.; Park, K. H.; Bien, Z. Z. Iterative fuzzy clustering algorithm with supervision to construct probabilistic fuzzy rule base from numerical data. **IEEE Transactions on Fuzzy Systems**. v. 16, n. 1, p. 263–277, 2008.
- Lei, Z.; Ren-hou, L. Designing of classifiers based on immune principles and fuzzy rules. **Information Sciences**. v. 178, n. 7, p. 1836–1847, 2008.
- Liao, T. W.; Celmins, A. K.; Hammell II, R. J. A fuzzy c-means variant for the generation of fuzzy term sets. **Fuzzy Sets and Systems**. v. 135, n. 2, p. 241–257, 2003.
- Liu, H.; Huang, S. T., 2003. A genetic semi-supervised fuzzy clustering approach to text classification. In: **Proceedings of the 4th International Conference on Advances in Web-Age Information Management (WAIM'03)** p. 173–180.
- MacQueen, J. B., 1967. Some methods for classification and analysis of multivariate observations. In: **Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability**, v. 1 p. 281–297.

- Mamdani, E. H. Application of fuzzy logic to approximate reasoning using linguistic synthesis. **IEEE Transactions on Computers**. v. 26, n. 12, p. 1182–1191, 1977.
- Mansoori, E.; Zolghadri, M.; Katebi, S. Sgerd: A steady-state genetic algorithm for extracting fuzzy classification rules from data. **IEEE Transactions on Fuzzy Systems**. v. 16, n. 4, p. 1061–1071, 2008.
- Mitchell, T. **Machine Learning**. McGraw-Hill Education (ISE Editions), 1997.
- Nicoletti, M. C.; Camargo, H. A. **Fundamentos da Teoria de Conjuntos Fuzzy**. Apontamentos. São Carlos, SP, BRA: EdUFSCar, 2004.
- Nogueira, T.; Camargo, H. A. Fuzzy-ccm: A context-sensitive approach to fuzzy modeling. **International Journal of Hybrid Intelligent Systems**. v. 7, p. 33–43, 2010.
- Nomura, H.; Hayashi, L.; Wakami, N., 1992. A learning method of fuzzy inference rules by descent method. In: **IEEE International Conference on Fuzzy Systems** p. 203–210.
- Olaru, C.; Wehenkel, L. A complete fuzzy decision tree technique. **Fuzzy Sets and Systems**. v. 138, n. 2, p. 221–254, 2003.
- Pedrycz, W. Algorithms of fuzzy clustering with partial supervision. **Pattern Recognition Letters**. v. 3, n. 1, p. 13–20, 1985.
- Pedrycz, W.; Amato, A.; Di Lecce, V.; Piuri, V. Fuzzy clustering with partial supervision in organization and classification of digital images. **IEEE Transactions on Fuzzy Systems**. v. 16, n. 4, p. 1008–1026, 2008.
- Pedrycz, W.; Gomide, F. **An Introduction to Fuzzy Sets: Analysis and Design**. Cambridge, MA, USA: MIT Press, 1998.
- Pedrycz, W.; Waletzky, J. Fuzzy clustering with partial supervision. **IEEE Transactions on Systems, Man, and Cybernetics, Part B**. v. 27, n. 5, p. 787–795, 1997.
- Pimenta, A. H. M.; Camargo, H. A., 2010a. Genetic interval type-2 fuzzy classifier generation: A comparative approach. In: **Brazilian Symposium on Neural Network - SBRN** p. 194–199.
- Pimenta, A. H. M.; Camargo, H. A., 2010b. Interval type-2 fuzzy classifier design using genetic algorithms. In: **IEEE International Conference on Fuzzy Systems**.

- Pires, M. G.; Camargo, H. C., 2004. Genetic learning and optimization of fuzzy sets in fuzzy rule-based systems. In: **IEEE International Conference On Information Reuse and Integration** p. 623–628.
- Quinlan, J. R. Induction of decision trees. **Machine Learning**. v. 1, n. 1, p. 81–106, 1986.
- Quinlan, R. J. **C4.5: Programs for Machine Learning**. Machine Learning. Morgan Kaufmann, 1993.
- Saez, Y.; Zazo, F.; Isasi, P., 2008. A study of the effects of clustering and local search on radio network design: Evolutionary computation approaches. In: **Eighth International Conference on Hybrid Intelligent Systems** p. 951–954.
- Skarmeta, A.; Barbera, H.; Alonso, M., 1999. A fuzzy logic based language to model autonomous mobile robots. In: **IEEE International Fuzzy Systems Conference Proceedings**, v. 1 p. 550–555.
- Stutz, C.; Runkler, T. A. Classification and prediction of road traffic using application-specific fuzzy clustering. **IEEE Transactions on Fuzzy Systems**. v. 10, n. 3, p. 297–308, 2002.
- Sugeno, M.; Yasukawa, T. A fuzzy-logic-based approach to qualitative modeling. **IEEE Transactions on Fuzzy Systems**. v. 1, n. 1, p. 7–31, 1993.
- Takagi, T.; Sugeno, M. Fuzzy identification of systems and its applications to modeling and control. **IEEE Transactions on Systems, Man, and Cybernetics**. v. 15, n. 1, p. 116–132, 1985.
- Wagstaff, K.; Cardie, C.; Rogers, S.; Schrödl, S., 2001. Constrained k-means clustering with background knowledge. In: **ICML '01: Proceedings of the Eighteenth International Conference on Machine Learning** p. 577–584.
- Wang, L.; Mendel, J. Generating fuzzy rules by learning from examples. **IEEE Transactions on Systems, Man and Cybernetics**. v. 22, p. 414–427, 1992.
- Witten, I. H.; Frank, E. **Data Mining: Practical Machine Learning Tools and Techniques**. Morgan Kaufmann, 2005.
- Yager, R. On a hierarchical structure for fuzzy modeling and control. **IEEE Transactions on Systems, Man and Cybernetics**. v. 23, n. 4, p. 1189–1197, 1993.
- Zadeh, L. A. Fuzzy sets. **Information and Control**. v. 8, n. 3, p. 338–353, 1965.

Zhu, X., 2005. **Semi-Supervised Learning Literature Survey**. Relatório Técnico 1530, Computer Sciences, University of Wisconsin-Madison.

Zimmerman, H. J. **Fuzzy Set Theory and Its Applications**. Kluwer, 1991.