Universidade Federal do Piauí
Centro de Ciências da Natureza
Programa de Pós-Graduação em Ciência da Computação

Rotulação de grupos utilizando conjuntos fuzzy

Vilmar Pereira Ribeiro Filho

Número de Ordem PPGCC: M001

Teresina-PI, 31 de agosto de 2015

Vilmar Pereira Ribeiro Filho

Rotulação de grupos utilizando conjuntos fuzzy

Universidade Federal do Piauí – UFPI

Centro de Ciências da Natureza

Programa de Pós-Graduação em Ciência da Computação

Orientador: Vinícius Ponte Machado

Coorientador: Ricardo de Andrade Lira Rabêlo

Teresina-PI

31 de agosto de 2015

Vilmar Pereira Ribeiro Filho
Rotulação de grupos utilizando conjuntos *fuzzy*/ Vilmar Pereira Ribeiro Filho. – Teresina-PI, 31 de agosto de 2015-
25 p. : il ; 30 cm.

Orientador: Vinícius Ponte Machado

Dissertação (Mestrado) – Universidade Federal do Piauí – UFPI
Centro de Ciências da Natureza
Programa de Pós-Graduação em Ciência da Computação, 31 de agosto de 2015.

1. *Cluster*. 2. *Fuzzy*. I. Vinícius Ponte Machado. II. Universidade Federal do Piauí. III. Rotulação de grupos utilizando conjuntos *fuzzy*

CDU 02:141:005.7

Vilmar Pereira Ribeiro Filho

Rotulação de grupos utilizando conjuntos fuzzy

Trabalho aprovado. Teresina-PI, XX de janeiro de 20XX:

Vinícius Ponte Machado
Orientador

Ricardo de Andrade Lira Rabêlo
Co-Orientador

Rodrigo de Melo Souza Veras

Teresina-PI
31 de agosto de 2015

Aos meus pais Rozângela Maria e Vilmar Ribeiro,
por sempre estarem comigo em todos os momentos.

Agradecimentos

Agradeço em primeiro lugar, a Deus, e a minha família.

Agradeço a meus pais, Vilmar Ribeiro e Rozângela Maria, por todo o carinho, atenção, amor, confiança, ensino e inspiração em toda a minha vida.

Agradeço ao meu orientador, Vinícius Ponte Machado e co-orientador, Ricardo de Andrade Lira Rabêlo, pela paciência e confiança. Foi um grande prazer e uma honra tê-los durante essa jornada. Meu eterno agradecimento.

Agradeço a todo os novos e velhos amigos do PPgCC, especialmente ao Lucas, Jonathas, Thiago e Kalyf.

Agradeço a minha irmã, Sara, à minha namorada, Nathylla, pela compreensão, carinho e paciência.

Agradeço a cada um de meus professores, de colégio e faculdade, pelos conhecimentos adquiridos, em especial aos professores Francisco Araújo, José Ferreira, Ricardo Sekeff, Ricardo Queiroz e Harilton Araújo, Rosianni e Amélia pela confiança e pelos ensinamentos além do curso.

Agradeço a todos que contribuíram direta ou indiretamente com a realização deste trabalho.

"A persistência é o caminho do êxito."
(Charles Chaplin)

Resumo

O agrupamento (*clustering*) de dados tem sido considerado como um dos tópicos mais relevantes dentre aqueles existentes na área de aprendizagem de máquina não-supervisionada. Embora o desenvolvimento e aprimoramento de algoritmos que tratam esse problema tenham sido o principal foco de muitos pesquisadores, a compreensão da definição dos grupos (*clusters*) é tão importante quanto sua formação. Uma boa definição de um grupo pode ajudar na interpretação dos dados. Frente ao problema de compreender a definição dos grupos este trabalho descreve uma solução que utiliza a teoria de conjuntos *fuzzy* para identificar os elementos mais relevantes do agrupamento e modelar faixas de valores que sejam capazes de identificar cada um dos grupos, baseando-se em características únicas. Os experimentos realizados demostram que o modelo proposto é bastante factível e capaz de construir faixas de valores para a identificação dos grupos, assim como classificar novos elementos utilizando as definições fornecidas.

Palavras-chaves: Aprendizagem. *Cluster*. *Fuzzy*. Rotulação.

Abstract

The clustering of data has been regarded as one of the most relevant topics among those existing on unsupervised machine learning area. Although the development and algorithms improvement that address this issue have been the main focus of many researchers, understanding the definition of the clusters is as important as your training. A good cluster definition can help in interpreting the data. Facing the problem of understanding the definition of the clusters this work describes a solution which uses the theory of fuzzy sets to identify the most relevant elements of the group and model ranges of values capable of identifying each group, based on unique characteristics . The experiments performed demonstrate that the model is quite feasible and able to build ranges of values for the identification of clusters, and classifying new elements using the definitions provided.

Keywords: Learning. *Cluster*. *Fuzzy*, Labeling.

Lista de ilustrações

Lista de tabelas

Sumário

Introdução

Com o surgimento crescente de novas tecnologia, o numero de dados produzido é cada vez maior. Uma das maneiras de lidar com esse volume de dados é por meio de agrupamentos. As pessoas sempre buscam agrupar dados com a finalidade de extrair características que sejam capazes de descrevê-los e ainda compará-los (XU; WUNSCH II, 2005). O problema básico do agrupamento pode ser declarado como segue: Dado um conjunto de dados, particionar em grupos os elementos que são tão semelhantes quanto possível. Note que esta é uma definição muito simples, e as variações na definição do problema podem ser significativas, dependendo do modelo utilizado (AGGARWAL; REDDY, 2013).

Motivação

O agrupamento (*clustering*) de dados tem sido considerado como um dos tópicos mais relevantes dentre aqueles existentes na área de aprendizagem de máquina e mineração de dados (AGGARWAL; REDDY, 2013). Assim, embora o desenvolvimento e aprimoramento de algoritmos que enfrentam esse problema tenham sidos foco de muitos pesquisadores, poucos trabalhos lidam explicitamente com a interpretação dos *clusters* formados. Segundo Tzerpos (2001), muitos pesquisadores se preocuparam com os demais problemas e não têm demonstrado atenção necessária ao problema específico de melhor compreender os *clusters*.

Uma boa interpretação de um grupo é capaz de fornecer um bom entendimento sobre o os dados estudados. A interpretação de um rótulo pode implicar em diversas soluções ou otimização de um problema abordado e ainda sobre a forma que os dados estão distribuídos.

Proposta

Este trabalho descreve um modelo capaz de analisar os *clusters* fornecidos e produzir uma definição de fácil interpretação para cada *cluster*. Para isto é utilizando poucos parâmetros de entrada e bases de dados contínuos. Os grupos são formados por meio do algoritmo *Fuzzy C-means*, que atribui um grau de pertinência para cada elemento, em cada um dos grupos formados. A definição dos grupos é composta pela identificação de uma ou mais características distintas capazes de identificar a maioria dos elementos pertencentes ao *cluster*. Assim o conjunto dessas características forma uma definição,

também chamada de rótulo. O modelo se mostra eficaz, uma vez que realiza uma extração de conhecimento onde o resultado pode ser facilmente interpretado.

Objetivos

O objetivo desse trabalho é criar rótulos, ou seja definições, para os *clusters*. Para isto é proposto um modelo, que possa ser capaz de identificar características únicas nos *clusters*, por meio da seleção de atributos relevantes e elaboração de faixas de valores. O rótulo gerado pelo sistema é composto por um conjunto de pares de valores associados a seus respectivos atributos, de forma a definir cada um dos *clusters*.

Estrutura Organizacional

Este trabalho está organizado como segue: o Capítulo 1 apresenta os trabalhos relacionados e as principais teorias utilizadas no modelo, o Capítulo 2 apresenta o modelo proposto. O Capítulo 3 apresenta os detalhes da implementação, os resultados obtidos com o modelo, uma avaliação do desempenho dos rótulos gerados e por fim as conclusões e os trabalhos futuros.

1 Referencial Teórico

1.1 Trabalhos relacionados

Durante o levantamento bibliográfico alguns trabalhos pesquisados mantiveram algumas semelhanças com esta pesquisa e somente um trata o mesmo problema dessa pesquisa.

Os trabalhos de Glover et al. (2002), Chuang e Chien (2004), Popescul e Ungar (2000) abordam o mesmo problema com abordagens diferentes. Estes trabalhos abordam o problema de extrair tópicos de textos e organiza-los de forma hierárquica. A exemplo disso temos estes algoritmos processando textos de biologia e sendo capazes de organizar de forma arvores hierárquicas contendo tópicos como: ciência, biologia e botânica colocando cada uma dessas palavras em um nível hierárquico diferente, juntamente com outros termos encontrado nos textos. Apesar destas pesquisas trabalharem com dados textuais, o que não é o foco desse trabalho, elas mostram a importância de condensar dados em estaturas que possam ser facilmente interpretadas.

Outros trabalhos como mostrados em Eltoft e Figueiredo (1998) e Chen, Chuang e Chen (2008) se referem à rotulação de *clusters* como o problema de atribuir um elemento desconhecido a um *cluster*. Estes trabalhos apresentam bons resultados ao atribuir um novo elemento a um *cluster* porem não demostram as regras utilizadas para esta atribuição.

Algoritmos como C4.5 e ID3 propostos por Quinlan (1986), que constroem árvores de decisão, uma vez aplicados para encontrar regras para os *clusters*, tem uma proposta similar, já que o caminho até as folhas pode der considerado um rótulo, porém esses algoritmos geram extensas árvores de decisão, dificultando a interpretação do *cluster* pelo especialista.

Pode ser visto em Cintra et al. (2011), Liu, Feng e Pedrycz (2013) que alguns trabalhos se propõem a construir árvores de decisão por meio da extração de regras com lógica *fuzzy*. Entre eles os trabalhos de Setnes (1999) constrói regras de lógica *fuzzy* para classificar elementos em um *clusters*, este trabalho se assemelha a proposta dessa pesquisa porém gera árvores de decisão. Como foi dito, grandes árvores de decisão podem dificultar o entendimento do *cluster* por um especialista.

Os trabalhos Vargas, Bedregal e Filho (2009) e Vargas e Bedregal (2009) utilizam de matemática intervalar para estender as funcionalidades do algoritmo *Fuzzy C-Means* porém seu objetivo é agrupar dados que estão no formato de intervalos e não realizam uma descoberta de conhecimento sobre os *cluster* gerados.

O trabalho de Lopes, Machado e Rabelo (2014), aborda o mesmo problema desta pesquisa. O modelo proposto no trabalho utiliza um processo de discretização para a construção das faixas de valores e posteriormente a utilização de redes neutrais artificiais para montar os rótulos adequados para cada *cluster*. Durante o processo de discretização, é necessário estabelecer algumas faixas de valores para cada atributo da base de dados, isto necessita que ao utilizar o modelo deva-se conhecer previamente a disposição dos dados, dificultando a operação do modelo. Outra característica desse modelo e a utilização de redes neurais artificiais para a montagem dos rótulos, o que pode levar a ter um grande custo computacional quando a base de dados possuí um grande numero de atributos. O modelo desta pesquisa se propõem a montar faixas de valores dinamicamente e realizando a montagem dos rótulos baseando-se em características únicas para cada *cluster*, utilizando um custo computacional menor.

Portanto, poucos trabalhos foram analisados envolvendo a rotulação de *clusters* no que se diz respeito a apresentar uma definição de fácil interpretação dos *clusters* e que possa ser útil para classificar de novos elementos.

1.2 Aprendizagem de maquina

Segundo Mitchell (1997), a área da aprendizagem de máquina está preocupada em construir programas de computador que possam melhorar seus desempenho de forma automática por meio de experiências. Para atingir seus resultados a Aprendizagem de Maquina utiliza-se de outras áreas de conhecimento como: Inteligencia Artificial, Probabilidade, Estatística, Psicologia e Biologia .

Os programas de computador que utilizam modelos de Aprendizagem de Maquina são capazes de formular hipóteses por meio de um conjunto de experiências anteriormente adquiridas. Durante o processo, o programa deve passar por uma aprendizagem pelo qual ele irá adquirir tal experiência. Podemos dividir o processo de aquisição da experiencia em duas grandes paradigmas: supervisionada e não-supervisionada. Ambos paradigmas realizam a busca por um modelo capaz de generalizar os dados. A aprendizagem supervisionada se destaca pela busca de um modelo preciso em relação à predição de valores para novos dados enquanto que na não-supervisionada o objetivo é encontrar características que podem resumir os dados. Existe também uma abordagem semi-supervisionada na qual existe uma tentativa de aprimorar um classificador criado a partir de dados rotulados com o uso de amostras não rotuladas (BARBER, 2012). Não abordaremos a aprendizagem supervisionada e semi-supervisionada já que não são utilizadas neste trabalho. A proposta apresentada nesse trabalho utiliza-se de um algoritmo de aprendizagem não-supervisionada, para realizar a tarefa de agrupamento de dados. Em geral estes algoritmos utilizam o conceito de similaridade para a construção de um grupo(*cluster*). Quando dois padrões

são similares de acordo com o critério escolhido eles são agrupados em um mesmo *cluster*, caso contrário, serão agrupados em *clusters* diferentes.

1.3 Agrupamento (*Clustering*)

Segundo Oliveira e Pedrycz (2007), *Clustering* é uma tarefa de aprendizagem não-supervisionada que visa a decomposição de um dado conjunto de objetos em subgrupos ou grupos com base na similaridade. O objetivo é o de dividir os dados de tal maneira que os objetos ou conjunto de dados que pertencem ao mesmo grupo são tão semelhantes quanto possível, enquanto objetos pertencentes a diferentes grupos são tão diferentes quanto possível. A motivação para encontrar ou construir grupos podem ser várias (BOCK, 1974). A análise de agrupamento é principalmente uma ferramenta para descobrir estrutura escondida em um conjunto de objetos. Neste caso, supõe-se que um "verdadeiro"agrupamento existe nos dados. No entanto, a atribuição de objetos para as classes e a descrição dessas classes são desconhecidos. Ao organizar objetos semelhantes em *clusters* se tenta reconstruir a estrutura desconhecida na esperança de que cada aglomerado encontrado representa um tipo real ou uma categoria de objetos. Métodos de *Clustering* também podem ser utilizados para fins de redução de dados. Representando de forma simplificada o conjunto de objetos que permite lidar com um número razoável de grupos homogêneos em vez de um grande número de objetos individuais. Alguns critérios matemáticos podem decidir sobre a composição de *clusters* para classificar conjuntos de dados automaticamente. Portanto, os métodos de *Clustering* são dotados de funções de distância que medem a dissimilaridade de casos de exemplo apresentados, o que é equivalente a medir a sua semelhança. O agrupamento de dados pode ser realizado utilizando-se de várias técnicas como: Cobweb (FISHER, 1987), Self Organizing Maps (SOM) (FIGUEIREDO S. BOTELHO; HAFFELE, 2012), Redes Neurais Artificiais (AZIZ et al., 2012), *K-means* (KANUNGO et al., 2002), *Hard C-means* (PERES et al., 2012), *fuzzy C-means* (RAMATHILAGA; LEU; HUANG, 2011) entre outras.

Algoritmos de particionamento visam encontrar a melhor partição dos grupos de dados com base na medida de dissimilaridade dada. Métodos de particionamento de agrupamento são diferentes das técnicas hierárquicas. Este último organizar os dados numa sequência aninhada dos grupos, que podem ser visualizados na forma de um dendrograma ou árvore. Com base em um dendrograma pode decidir sobre o número de grupos na qual os dados são melhor representados para uma dada finalidade. Normalmente, o número de aglomerados (verdadeiro) nos dados indicados é conhecida com antecedência. No entanto, geralmente ao usar um métodos de particionamento e necessário especificar o número de *clusters* como um parâmetro de entrada. Estimar o número real de *clusters* é, portanto, uma questão importante (OLIVEIRA; PEDRYCZ, 2007).

Um conceito comum em todas as abordagens de agrupamento descritos é que eles são baseados em protótipos, ou seja, os grupos são representados por protótipos de *cluster* $C_i, i = 1, ..., c$. Os protótipos são usados para capturar a estrutura (distribuição) de dados em cada grupo. Com essa representação dos *clusters* denotamos formalmente o conjunto de protótipos $C = \{C_i, ..., C_c\}$. Cada protótipo C_i é um n-tuplo de parâmetros que consiste de um conjunto $\mathbf{c}_i$ (parâmetro de localização) e talvez alguns parâmetros adicionais sobre o tamanho e a forma do agrupamento. O centro do agrupamento c_i é uma instanciação dos atributos utilizados para descrever o domínio. Os parâmetros de tamanho e forma de um protótipo determina a extensão do *cluster* em direções diferentes do domínio subjacente. Os protótipos são construídos pelos algoritmos de agrupamento e servem como representações dos elementos (pontos de dados) em cada *cluster*.

Será apresentados os algoritmos de agrupamento *Fuzzy C-means* e *Hard C-means*. O último serviu como um ponto de partida para as suas extensões *fuzzy*. O protótipo será apresentado na sua forma mais simples. Cada protótipo consiste apenas dos vetores do centro, $C_i = (c_i)$, de tal modo que os pontos de dados atribuído a um *cluster* são representados por um ponto prototípico no espaço de dados. Os algoritmos descritos são baseados em funções objetivo J. Funções objetivo devem ser minimizadas para obter as melhores soluções para cada *cluster*. Tendo definido um critério desse tipo, a tarefa de agrupamento pode ser formulada como um problema de otimização de função.

1.4 *Hard C-means*

No modelo clássico *Hard C-means* cada ponto de dados x_j em um conjunto de dados $X = \{x_1, ..., X_n\}$, onde $X \subseteq \mathbb{R}^p$ é atribuído a um *cluster*. Cada *cluster* Γ_i é assim, um subconjunto do conjunto de dados $\Gamma_i \subset X$. O conjunto de *clusters* $\Gamma = \{\Gamma_1, ..., \Gamma_c\}$ é uma partição dos conjunto de dados X em c, não vazio com pares de subconjuntos disjuntos onde $\Gamma_i, 1 < c < n$. No *Hard C-means* uma partição de dados é dito ser ótima quando a soma dos quadrados das distâncias entre os centros dos grupos e os pontos de dados que lhes são atribuídas é mínima (KRISHNAPURAM; KELLER, 1996). A função objetivo do *Hard C-means* podem ser escritas da seguinte forma:

$$J_n(X, U_h, C) = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij} d_{ij}^2 \tag{1.1}$$

onde $C = \{C_i, ..., C_c\}$ é o conjunto de protótipos, d_{ij} é a distância entre x_j e o centro do agrupamento c_i, U é uma matriz $c \times n$ binária chamada matriz de partição, tal que.

$$U_{ij} \in \{0, 1\} \tag{1.2}$$

A atribuição de um dados para o *clusters* é dada como: $u_{ij} = 1$ se o ponto de dados x_j é atribuída ao protótipo C_{uj}, ou seja,$x_j \in \Gamma_i$; e $u_{ij} = 0$ caso contrario. Para assegurar que cada ponto de dados é atribuída exatamente a um conjunto, é necessário que:

$$\sum_{i=1}^{c} u_{ij} = 1, \forall j \in \{1, ..., n\} \tag{1.3}$$

Esta restrição impõe partições completas e também serve o propósito de evitar a solução trivial minimizando J_h, que é, não atribuir dados a *clusters*: $u_{ij} = 0, \forall i, j$. Juntamente com isso, uma outra restrição também é necessária. A restrição (1.4) evita que seja possível ter *clusters* vazios.

$$\sum_{j=1}^{n} u_{ij} > 0, \forall i \in \{1, ..., c\} \tag{1.4}$$

O problema de encontrar parâmetros que minimizem a função objetivo é *NP-hard* (DRINEAS et al., 2004). Portanto não é garantido que o ótimo global será alcançado. No caso do *Hard C-means* o regime de otimização iterativa funciona da seguinte maneira: primeiramente os centros dos grupos iniciais são escolhidos. Isto pode ser feito de forma aleatória, isto é, escolhendo c vetores aleatórios que se encontram os dados; ou inicializar os centros dos *clusters* com pontos de dados escolhidos aleatoriamente. Outros métodos de inicialização mais sofisticados também podem ser usados (MCKAY; BECKMAN; CONOVER, 1979). Em seguida, os parâmetros C são mantidos fixos e atribuições em U são determinadas para minimizar a quantidade de J_h. Nesta etapa, cada ponto de dados é atribuída ao seu centro mais próximo do*cluster*:

$$u_{ij} = \begin{cases} 1, & \text{se } i = \text{argmin}_{l=1}^{c} d_{ij} \\ 0, & \text{senão} \end{cases} \tag{1.5}$$

Quando a atribuição aos ponto de dados parar de minimizar J_h, com centro dos *clusters* fixos, a partição de dados U é fixada e novos centros dos *clusters* são calculados como a média de todos os vetores de dados atribuídos a eles, uma vez que a média minimiza a soma dos quadrados das distâncias em J_h. O cálculo da média de cada *cluster*(para o qual o algoritmo tem o seu nome) é indicado mais formalmente como:

$$c_i = \frac{\sum_{j=1}^{n} u_{ij} x_j}{\sum_{j=1}^{n} u_{ij}}. \tag{1.6}$$

As duas etapas (1.5) e (1.6) são iteradas até que nenhuma alteração em C ou U possa ser observada. Em seguida, o *Hard C-means* termina, dando origem a centros de *clusters* finais e partição de dados que são possivelmente ótimos locais. Concluindo o *Hard C-means*, expressa a tendencia de ficar preso em mínimos locais, o que torna necessário a

realização de várias execuções do algoritmo com diferentes inicializações (DUDA; HART, 1973). Em seguida, o melhor resultado de agrupamentos podem ser escolhidos com base nos valores de J_h.Passaremos agora para as abordagens *fuzzy*, que relaxar a exigência $u_{ij} \in \{0, 1\}$ que é colocada sobre as atribuições nas abordagens clássicas de *Clustering*. As extensões são baseadas nos conceitos de conjuntos *fuzzy* e grau de pertinência.

1.5 *Fuzzy C-means*

A análise de agrupamento *fuzzy* permite que os dados tenham grau de pertinência em relação a seus *clusters*, variando de $[0, 1]$. Isto dá a flexibilidade para expressar pontos de dados que pode pertencer a mais de um conjunto. Além disso, estes graus de pertinência oferecem um ajuste mais fino e detalhado do modelo de dados. Além de atribuir um ponto de dados ao *clusters*, os graus de pertinência também podem expressar o quão ambígua ou diferente um ponto de dados é em um *cluster*. O conceito destes graus de pertinência é definido pela interpretação dos conjuntos *fuzzy* (ZADEH, 1965). Assim, agrupamentos *fuzzy* permitem espaços de solução de granulação fina em forma de partições do conjunto de dados por exemplo $X = \{x_1, ..., X_n\}$. Considerando o *cluster*Γ_i de partições um subconjuntos clássico, a sua representação pelos conjuntos *fuzzy* μ_{Γ_i} dos dados do conjunto X é definido a seguir. Cumprindo com a teoria dos conjuntos *fuzzy*, o *cluster* é atribuído u_{ij} é agora o grau de pertinência de um x_j a um *cluster*Γ_i, de tal forma que: $u_{ij} = \mu_{\Gamma_i}(x_j) \in [0, 1]$. As associações dos elementos aos *clusters* são difusas, ou seja, não há nada que indique que o ponto de dados pertence a um *cluster*. Em vez disso, os métodos de agrupamento *fuzzy* associar um vetor para cada ponto de dados x_j que indica suas associações para os c *clusters*: $u_j = (u_{1j}, ..., u_{cj})^T$

A matriz $c \times n$ $U = (u_{ij}) = (u_1, ..., u_n)$ é chamada matriz de partição *fuzzy*. Com base na notação de conjuntos *fuzzy* os *clusters* são mais adequado para lidar com a ambiguidade de atribuições.

Sendo $X = \{x_1, ..., x_n\}$ o conjunto de dados e c o número de *clusters* $(1 < c < n)$ representado pelos conjuntos *fuzzy* $\mu_{\Gamma_i}, (i = 1, ..., c)$. Então chamamos $U = (u_i j) = (\mu_{\Gamma_i}(x_j))$ partição de X se:

$$\sum_{j=1}^{n} u_{ij} > 0, \forall i \in \{1, ..., c\} \tag{1.7}$$

e

$$\sum_{i=1}^{c} u_{ij} = 1, \forall j \in \{1, ..., n\} \tag{1.8}$$

Os $u_{ij} \in [0, 1]$ são interpretados como o grau de pertinência de um x_j para o *cluster* Γ_i relativo a todos os outros *clusters*.

A restrição (1.8) garante que não exista um agrupamento vazio. A condição (1,9), garante que a soma dos graus de pertinência para cada ponto de dados de um *cluster* seja igual a 1. Isto significa que cada ponto de referência recebe o mesmo peso em comparação com todos os outros dados e, portanto, de que todos os dados são (igualmente) incluído na partição do *cluster*. Além disso, a condição (1.9) corresponde a uma normalização das associações. Assim, os graus de pertinência para um determinado dado assemelham formalmente as probabilidades de que seja um membro do *cluster* correspondente.

Depois de definir partições podemos voltar para o desenvolvimento de uma função objetivo para a tarefa de agrupamento *fuzzy*. Certamente, quanto mais próximo um ponto de dados encontra-se do centro de um *cluster*, maior o seu grau de pertinência deve ser para este *cluster*. Seguindo esse raciocínio, pode-se dizer que as distâncias entre os centros dos grupos e os pontos de dados que lhe são atribuídas deve ser mínima. O problema de dividir um dado conjunto de dados em c *clusters* pode ser indicado como a tarefa de minimizar as distâncias ao quadrado dos pontos de dados para seus centros de *cluster*, ou seja, nós queremos maximizar os graus de pertinência. A função objetivo J_f baseia-se assim na menor soma das distâncias ao quadrado.

Formalmente, um modelo de grupos *fuzzy* é dado como um conjunto de dados X em c *clusters* é definido como ótimo quando minimiza a função objetivo:

$$J_f(X, U_f, C) = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^m d_{ij}^2 \tag{1.9}$$

As restrições (1.8) e (1.9) têm de ser satisfeitas para graus de pertinência em U. A condição (1.8) evita a solução trivial do problema de minimização, ou seja, $u_{ij} = 0, \forall i, j$. A restrição de normalização (1.9) conduz a uma "distribuição"do peso de cada ponto de dados sobre os diferentes conjuntos. Uma vez que todos os pontos de dados têm a mesma quantia fixa de associação para compartilhar entre os *clusters*. A condição de normalização implementa a propriedade de particionamento conhecido de qualquer algoritmo de agrupamento *fuzzy*. O parâmetro $m, m > 1$, é chamado de *fuzzifier* ou expoente de ponderação. A exponenciação das associações com m em J_f pode ser visto como um função g nos graus de pertinência, $g(u_{ij}) = u_{ij}^m$. Para o caso de $m = 1$ (quando J_h e J_f sao idêntico), que as atribuições do *cluster* permanecem semelhando ao *Hard C-means* quando minimizando a função objetivo, mesmo que eles estejam limitados a $\{0, 1\}$ (DUNN, 1973). Para alcançar o fuzificação desejado da partição de dados, o resultante da função $g(u_{ij}) = u_{ij}^2$ foi proposto pela primeira vez (DUNN, 1973). A generalização para exponentes $m > 1$ tem sido proposto em (BEZDEK, 1973). Com valores mais elevados para o m os limites entre aglomerados se tornam mais suaves, com valores mais baixos que se tornam *crisp*. Normalmente $m = 2$ é escolhido. Além da ponderação padrão das associações u_{ij}^m outras funções g para servir como *fuzzifiers* podem ser exploradas.

A função objetivo J_f é alternadamente otimizado, ou seja, em primeiro lugar os graus de pertinência são otimizados para valores parâmetros fixos de *cluster*, em seguida, os protótipos de fragmentação são otimizados para graus de pertinência fixos:

Os graus de adesão têm de ser escolhidos de acordo com a seguinte fórmula de atualização que é independente da medida de distância escolhida (BEZDEK, 1981).

$$u_{ij} = \frac{1}{\sum_{l=1}^{c} \left(\frac{d_{ij}^2}{d_{lj}^2} \right)^{\frac{1}{m-1}}} = \frac{d_{ij}^{-\frac{2}{m-1}}}{\sum_{l=1}^{c} d_{lj}^{-\frac{2}{m-1}}} \tag{1.10}$$

No caso de existe um *cluster* i com zero de distância de uma x_j, $u_{ij} = 1$ e $u_{lj} = 0$ para todos os outros grupos $l \neq i$. A equação (1.10) mostra claramente o caráter relativo ao grau de pertinência. Ele não depende apenas da distancia do ponto de referencia x_j para o *cluster* i, mas também as distâncias entre este os ponto de dados e os outros *clusters*.

A formula de atualização de j_c para os parâmetros do *cluster* dependem, claramente, dos parâmetros utilizados para descrever um *cluster* (localização, forma, tamanho) e a medida de distância escolhida. Portanto, uma fórmula de atualização geral não pode ser dada. A formula para calcular centros é dada como (BEZDEK, 1981):

$$c_i = \frac{\sum_{j=1}^{n} u_{ij}^m x_j}{\sum_{j=1}^{n} u_{ij}^m}. \tag{1.11}$$

2 Proposta

2.1 Modelo Proposto

Esta seção descreve passo a passo o funcionamento do modelo e a sua aplicação em uma base de dados fictícia. O modelo tem como objetivo encontrar as diferenças existentes em cada um dos *clusters*, por meio da seleção de elementos representativos e em seguida a construção de faixas de valores.

Na Figura 1 é demostrado o modelo proposto utilizando um fluxograma. As etapas e termos demostrados na figura são explicados posteriormente.

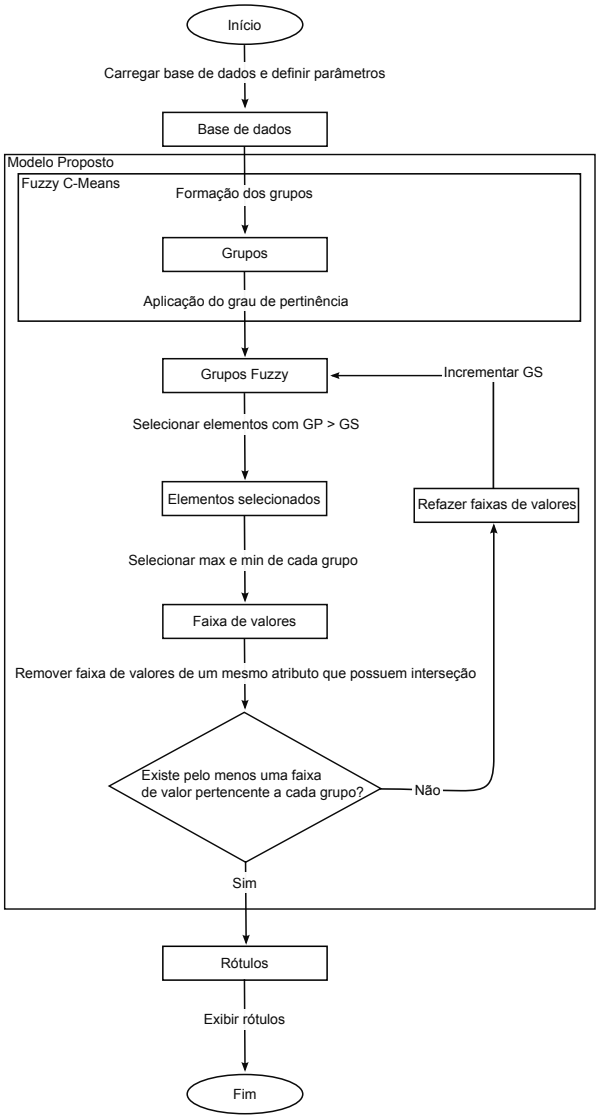

Figura 1 – Fluxograma do Modelo Proposto

O modelo dá início com o carregamento da base de dados, a definição dos parâmetros GS (Grau de Seleção) e definição do IGS (Incremento do Grau de Seleção). O parâmetro GS consiste em um número que serve de base para a seleção dos elementos mais significativos na formulação do rótulo e pode variar entre 0 e 1. O IGS consiste em um valor de incremento do parâmetro GS a cada iteração e também varia entre 0 e 1.

A base de dados fictícia é demostrada na Tabela 1. Os elementos são compostos pelos atributos At 1 e At 2 e possuem valores definidos nos conjuntos dos números reais. Os valores foram escolhidos de forma arbitraria para ilustrar a aplicação do modelo. O parâmetro GS foi definido como 0.5 por representar um grau de pertinência intermediário, no qual elementos que estão abaixo desse número têm grande chances de pertencer a dois grupos. Já o parâmetro IGS foi definido como 0.0001, pois provoca no GS um ajuste de 4 casas decimais, uma vez que a matriz U possui também graus de pertinência de 4 casas decimais.

Id	**At.1**	**At.2**	**Id**	**At.1**	**At.2**
1	4.3	6.0	16	8.9	4.0
2	9.7	6.5	17	7.8	7.7
3	4.7	5.7	18	7.8	4.6
4	7.0	8.0	19	3.9	3.6
5	4.3	4.7	20	8.5	5.1
6	4.9	4.8	21	7.6	6.9
7	3.1	5.7	22	5.9	8.5
8	4.1	5.5	23	9.1	6.4
9	5.8	7.5	24	7.0	7.3
10	9.4	6.0	25	6.3	7.4
11	9.8	3.5	26	5.3	3.8
12	8.0	4.0	27	7.5	8.2
13	9.0	6.7	28	5.9	6.9
14	4.5	3.7	29	3.8	4.6
15	8.8	4.5	30	6.8	7.9

Tabela 1 – Base de Dados Fictícia

Inicialmente é aplicado o algoritmo de agrupamento não-supervisionado *Fuzzy C-Means*. Este algoritmo é inicializado com os parâmetros que correspondem a base de dados, e a quantidade de grupos a serem formados. Como parâmetro do *Fuzzy C-Means* foi utilizada a quantidade de grupos igual a 3.

O algoritmo *Fuzzy C-Means* fornece como saída uma matriz U. Esta matriz atribui, a cada elemento, um grau de pertinência em cada um dos grupos formados. O grau de pertinência é atribuído de tal forma que, quanto mais próximo o elemento estiver do centróide de um grupo, maior é seu grau de pertinência em relação a ele.

A Matriz U da base de dados fictícia pode ser vista na Tabela 2. Nela o elemento de

id igual a 10, por exemplo, possui grau de pertinência maior no grupo 2, isso significa que este elemento está mais próximo do centróide do grupo 2. De maneira análoga podemos perceber que este elemento possui grau de pertinência menor no grupo 3, significando que ele está mais distante do centróide do grupo 3.

	Grau de Pertinência		
Id	**Grup.1**	**Grup.2**	**Grup.3**
1	0.1269	0.0535	0.8196
2	0.2383	0.6920	0.0697
3	0.0973	0.0458	0.8569
4	0.9758	0.0140	0.0102
5	0.0009	0.0007	0.9984
6	0.0286	0.0217	0.9497
7	0.1062	0.0565	0.8373
8	0.0401	0.0213	0.9386
9	0.8411	0.0598	0.0991
10	0.1329	0.8230	0.0441
11	0.0983	0.8230	0.0787
12	0.0890	0.8220	0.0890
13	0.3388	0.5891	0.0721
14	0.0549	0.0565	0.8886
15	0.0142	0.9764	0.0094
16	0.0476	0.9148	0.0376
17	0.8693	0.0920	0.0387
18	0.0954	0.8273	0.0773
19	0.0589	0.0561	0.8850
20	0.0135	0.9797	0.0068
21	0.7885	0.1567	0.0548
22	0.8425	0.0681	0.0894
23	0.2381	0.7011	0.0608
24	0.9791	0.0126	0.0083
25	0.9436	0.0261	0.0303
26	0.0944	0.1165	0.7891
27	0.9095	0.0576	0.0329
28	0.7602	0.0859	0.1539
29	0.0162	0.0117	0.9721
30	0.9871	0.0071	0.0058

Tabela 2 – Matriz U

Com a formação da matriz U, o modelo escolhe elementos que possuem um GP (Grau de pertinência) maior que o parâmetro GS. Com isto, em cada grupo são extraídos o máximo e o mínimo de cada atributo. Estes valores correspondem as faixas de valores de cada grupo.

Os elementos escolhidos durante a primeira iteração do modelo, com o grau de seleção (GS) igual a 0.5, podem ser vistos nas Tabelas 3, 4 e 5. Em negrito pode-se

perceber os valores máximos e mínimos de cada atributo utilizado na formação das faixas de valores. A Tabela 6 representa as faixas de valores geradas, tendo por base os elementos selecionados.

			Grau de Pertinência		
Id	**At.1**	**At.2**	**Grup.1**	**Grup.2**	**Grup.3**
4	7.0	8.0	0.9758	0.0140	0.0102
9	**5.8**	7.5	0.8411	0.0598	0.0991
17	**7.8**	7.7	0.8693	0.0920	0.0387
21	7.6	**6.9**	0.7884	0.1567	0.0548
22	5.9	**8.5**	0.8425	0.0681	0.0894
24	7.0	7.3	0.9792	0.0126	0.0083
25	6.3	7.4	0.9436	0.0261	0.0303
27	7.5	8.2	0.9095	0.0576	0.0329
28	5.9	6.9	0.7602	0.0859	0.1539
30	6.8	7.9	0.9872	0.0071	0.0058

Tabela 3 – Elementos escolhidos no grupo 1

			Grau de Pertinência		
Id	**At.1**	**At.2**	**Grup.1**	**Grup.2**	**Grup.3**
2	9.7	6.5	0.2383	0.6920	0.0697
10	9.4	6.0	0.1329	0.8229	0.0441
11	**9.8**	**3.5**	0.0983	0.8229	0.0787
12	8.0	4.0	0.0890	0.8220	0.0890
13	9.0	**6.7**	0.3388	0.5891	0.0721
15	8.8	4.5	0.0142	0.9764	0.0094
16	8.9	4.0	0.0476	0.9148	0.0376
18	**7.8**	4.6	0.0954	0.8273	0.0773
20	8.5	5.1	0.0135	0.9797	0.0068
23	9.1	6.4	0.2381	0.7011	0.0608

Tabela 4 – Elementos escolhidos no grupo 2

Por fim, é verificado se existem interseções entre as faixas de valores pertencentes a um mesmo atributo. Caso exista interseção entre as faixas de valores, como mostrado na Tabela 6 pelas faixas $7.80 \sim 5.80$ e $9.80 \sim 7.80$, que compartilham o número 7.80 em comum nas duas faixas, estas faixas são descartadas e a análise parte para outro conjunto de faixas de valores. Isto é necessário pois as faixas de valores que compõem interseção são ambíguas e incapazes de representar um único grupo.

Caso nenhum atributo possua pelo menos uma faixa de valor capaz de representar cada um dos grupos, como mostrado na Tabela 6, o parâmetro GS é incrementado pelo parâmetro IGS e o processo de seleção de elementos é refeito utilizando um novo valor para GS. Por fim são geradas novas faixas de valores a serem analisadas. Este processo é necessário para remover a interseção entre as faixas de valores, tornado-as únicas. Os valores

			Grau de Pertinência		
Id	**At1**	**At.2**	**Grup.1**	**Grup.2**	**Grup.3**
1	4.3	**6.0**	0.1269	0.0535	0.8196
3	4.7	5.7	0.0973	0.0458	0.8569
5	4.3	4.7	0.0009	0.0007	0.9984
6	4.9	4.8	0.0286	0.0217	0.9498
7	**3.1**	5.7	0.1062	0.0565	0.8372
8	4.1	5.5	0.0401	0.0213	0.9386
14	4.5	3.7	0.0549	0.0565	0.8886
19	3.9	**3.6**	0.0589	0.0561	0.8850
26	**5.3**	3.8	0.0944	0.1165	0.7891
29	3.8	4.6	0.0162	0.0117	0.9721

Tabela 5 – Elementos escolhidos no grupo 3

	Grup.1	**Grup.2**	**Grup. 3**
At.1	**5.8 ∼ 7.8**	**7.8 ∼ 9.8**	3.1 ∼ 5.3
At.2	6.9 ∼ 8.5	**3.5 ∼ 6.7**	**3.6 ∼ 6.0**

Tabela 6 – Rótulos da Interação #1

únicos apresentam características capazes de distinguir cada um dos grupos, representando assim os seus identificadores.

Caso exista pelo menos uma faixa de valores sem interseção em cada grupo, como mostrado na Tabela 7, o processo é encerrado e estas faixas de valores servem como rótulo para seus grupos.

Depois de 319 iterações, utilizando-se de um grau de seleção de 0.8290. As faixas de valores criadas ainda possuem interseção, porém cada grupo possui pelo menos uma faixa de valor que não tem interseção, satisfazendo a condição de parada do modelo.

	Grup.1	**Grup.2**	**Grup.3**
At.1	5.8 ∼ 7.8	8.5 ∼ 8.9	3.1 ∼ 4.9
At.2	7.3 ∼ 8.5	**4.0 ∼ 5.1**	**3.6 ∼ 5.7**

Tabela 7 – Rótulos da Iteração #319

Por fim, os rótulos são exibidos na Tabela 8 que mostra as faixas de valores correspondentes a seu grupo e atributo. Podem existir várias faixas de valores relacionadas a um grupo, porém cada uma delas em atributos diferentes. Assim também podem existir faixas de valores em um mesmo atributo, porém não deve existir interseção de valores entre elas.

Os rótulos exibidos na Tabela 8 mostram somente faixas de valores únicas em relação a um atributo. O conjunto das faixas de valores compõem uma identificação para o grupo e representa a maioria dos elementos contidos nele.

	Grup.1	**Grup.2**	**Grup.3**
At.1	$5.8 \sim 7.8$	$8.5 \sim 8.9$	$3.1 \sim 4.9$
At.2	$7.3 \sim 8.5$	-	-

Tabela 8 – Rótulos Finais

.

3 Implementação e Resultados

3.1 Detalhes da Implementação

O modelo proposto foi implementado na ferramenta MATLAB. A escolha dessa ferramenta se dá por ter algoritmos já validados pela comunidade científica. Os algoritmos fornecidos apresentam uma boa performance e um ambiente de desenvolvimento fácil para a prototipação de novos modelos.

Para o procedimento de criação dos *clusters* foi utilizado o algoritmo *Fuzzy C-means* dado pelo comando FCM. Este algoritmo tem como saída a matriz U que representa os elementos relacionados com seus respectivos grau de penitencia em cada grupo formado, formando assim grupos que são representados por conjuntos *fuzzy*, por fim realizando as etapas de agrupamento e formação de grupos *fuzzy*. Foram utilizados os parâmetros *default* do comando FCM. Isto corresponde aos valores 2.0 para o expoente para a matriz de partição U, 100 para o numero máximo de interações e quantidade minima de melhoria 1e-5. Para os parâmetros do modelo GS e IGS, foram utilizados 0.5 e 0.0001 receptivamente.

Para saber o nível de precisão dos rótulos foi realizado um teste. Com isso verificou se os valores dos elementos estavam presente nos intervalos das faixas de valores construídas pelo modelo. Uma vez feito isto, o elemento era atribuído ao *cluster* correspondente às faixas obedecidas. Esta quantidade de elementos que se encaixam nos rótulos podem ser vistas nas tabelas 9, 11 e 12 com os resultados presente na coluna Elementos.

3.2 Base de Dados Iris

O modelo foi aplicado na base de dados "Iris", disponível no repositório UCI *Machine Learning* [1]. A base de dados refere-se a uma amostra de plantas e contém 150 elementos, cada um deles possui 4 atributos definidos por valores reais. Os atributos são: o comprimento da sépala (CS), a largura da sépala (LS), o comprimento da pétala (CP), a largura da pétala (LP). Esta base de dados possui dados coletados de 3 classes de plantas, *Setosa*, *Versicolor* e *Virginica*.

Segundo o trabalho de Runkler, as classes podem ser divididas em (RUNKLER, 2012):

1) 50 elementos da classe Iris Setosa;

2) 50 elementos da classe Iris Virginica;

[1] https://archive.ics.uci.edu/ml/

3) 50 elementos da classe Iris Versicolor;

Como sugerido no trabalho de Runkler, a base de dados foi dividida em 3 grupos (RUNKLER, 2012). A Tabela 9 mostra o resultado da execução do modelo na base de dados Iris. Nesta tabela e mostrado os *clusters* associados aos seus respectivos rótulos, a coluna Elementos mostra a quantidade de elementos que obedecem ao rotulo.

Cluster	**Rótulos**		**Elementos**
	Atributos	**Intervalos (cm)**	
1	CP	$5.1 \sim 6.9$	42
2	CP LP	$1.0 \sim 1.9$ $0.1 \sim 0.6$	50
3	CP	$3.5 \sim 5.0$	55

Tabela 9 – Resultados da base de dados Iris

Pode-se ver na Tabela 9 os rótulos gerados pelo modelo, sendo compostos de faixas de valores associadas a alguns atributos de cada *cluster*. Percebe-se também que o atributo comprimento da pétala (CP) está presente em todos os *clusters*, isso mostra que os três *clusters* se diferem principalmente pela faixa de valor desse atributo. Um especialista que eventualmente quiser atribuir um novo elemento a um grupo qualquer teria no comprimento da pétala a principal característica para a identificação do novo elemento. Os demais atributos e suas faixas de valores representam características secundarias, mas também caracterizam os grupos de forma única.

Ao comparar o total de elementos e a soma dos elementos, percebe-se que 3 elementos não foram rotulados em nenhum dos *clusters*. Os elementos que não foram rotulados pelos rótulos gerados pelo modelo, podem ser vistos na Tabela 10.

CS	**LS**	**CP**	**LP**
5.0	2.3	3.3	1.0
5.1	2.5	3.0	1.1
4.9	2.4	3.3	1.0

Tabela 10 – Elementos não rotulados da base de dados Iris

Os 3 elementos da Tabela 10 possuem valores no atributo CP que não existe em nenhuma das faixas de valores. Com isto o modelo conseguiu rotular 98% dos elementos baseando-se somente em poucas características de cada grupo.

3.3 Base de Dados *Seed*

Esta base de dados se refere a identificação de sementes de trigo, também pode ser encontrada no repositório UCI *Machine Learning* como *Seed Data Set*[1]. Esta base de

dados contém 210 elementos, cada elemento possui 7 atributos como: área (A), perímetro (P), densidade (C), comprimento da semente (LK), largura da semente (WK), coeficiente de assimetria (AC), comprimento do sulco da semente (LKG). Cada uma das amostras de sementes são classificadas em 3 diferentes tipos, que são:

1) 70 elementos do tipo Kama;

2) 70 elementos do tipo Rosa;

3) 70 elementos do tipo Canadian.

	Rótulos		
Cluster	**Atributos**	**Intervalos**	**Elementos**
1	A	10.5 ∼13.3	84
2	A	13.5 ∼16.1	56
3	A P	17.2 ∼21.1 15.6 ∼17.2	54

Tabela 11 – Resultados da base de dados Seed

Os resultados mostrados na tabela 11 mostra que os tipos de sementes são fortemente diferenciadas pela área, uma vez que o atributo área esta presente nos rótulos de todos os grupos. O grupo 3 apresenta não só uma faixa de valor para a área mais também uma faixa de valor no atributo perímetro, isto mostra que o os elementos do grupo 3 apresentam elementos bem definidos na faixa de valor mostrada.

Para um especialista ciente dos resultados do modelo poderia classificar uma nova amostra focando-se principalmente nas faixas de valores presente no resultado. Durante os testes 194 elementos obedeceram os rótulos e 16 elementos não obedeceram nenhum dos rótulos. A classificação realizada pelos rótulos e a classificação utilizando o algoritmo *Fuzzy C-Means* obteve 87.14 % de semelhanças, ou seja, na maioria dos casos os rótulos atribuíram aos elementos os mesmos grupos que o algoritmo *Fuzzy C-Means*.

3.4 Base de Dados *Glass*

O modelo foi aplicado a base de dados *Glass Identification* que pode ser encontrada no repositório de dados UCI *Machine Learning*[1]. O estudo desta base de dados tem aplicação na área forense onde a identificação do tipo de vidro pode ajudar a solucionar crimes.

A base de dados contém 214 amostras de vidro. Cada amostra e formada pelos atributos: índice de refração (IR), e a porcentagens do óxido referente aos elementos Na, Mg, Al, Si, K, Ca, Ba, e Fe. Os valores dos atributos são contínuos e formam dois grandes grupos que são: 163 elementos de amostra de vidros de janela e 51 de outros tipos de objetos de vidro. No primeiro grupo temos amostras de janelas de veículos, float e no-float,

e janelas de construção, do tipo float e no-float. No segundo grupo temos amostras de utensílios de cozinha, recipientes e faróis.

	Rótulos		
Cluster	**Atributos**	**Intervalos**	**Elementos**
1	Mg	$0 \sim 1.8$	53
2	Mg	$2.1 \sim 4.5$	160

Tabela 12 – Resultados da base de dados Glass

A tabela 12 mostra que que a principal característica que diferencia os grupos é a presença de Magnésio (Mg) nas faixas de valores mostradas no rotulo. Pelos rótulos mostrados percebe-se que o grupo 1 consiste das amostras de utensílios de cozinha, recipientes e faróis e o grupo 2 consistem nas amostra de vidros de janela. Segundo (NAVARRO, 2003) a quantidade de Mg na fabricação do vidro é relacionada a resistência mecânica que o vidro possuí. Com isto a quantidade de Mg nas amostras de Janela são maiores devido a necessidade dos vidros de janelas serem mais resistentes a impactos, já que sua geometria plana e fina facilita sua quebra. Durante os testes 213 elementos obedeceram os rótulos e 1 elemento não obedeceu o rotulo. A comparação da classificação utilizando os rótulos e a classificação utilizando o algoritmo *Fuzzy C-Means* obteve 98% de semelhança.

Conclusões

Conclusão e Trabalhos Futuros

Neste trabalho foi proposto um modelo capaz de estimar uma definição para os *clusters*, de forma a facilitar sua interpretação. A definição se dá pela elaboração de faixas de valores de cada atributo para cada *cluster*. As faixas de valores são associadas a atributos capazes de distinguir cada *cluster*. As faixas de valores geradas contribuem para o entendimento dos *clusters*, que geralmente são formados por algoritmos de aprendizagem não-supervisionada.

O modelo apresentado mostrou-se promissor conseguindo rotular 98% dos elementos da base de dados, como mostrado nos testes. Esta abordagem se diferencia de outras por fornecer dados sobre os *clusters* de forma simples, que podem ser facilmente interpretados. Durante os testes foi percebido que a posição dos *clusters* influencia diretamente a formação das faixas de valores.

No contexto de inteligência empresarial, um especialista de uma empresa, ao observar os rótulos fornecidos pode, por exemplo, identificar quais são as principais características que definem os diferentes grupos de clientes. Por exemplo, uma vez conhecida as características, a empresa pode elaborar modelos de negócio ou estratégias específicas para um determinado grupo.

Como trabalhos futuros espera-se fazer testes para saber a aplicabilidade do modelo em grandes volumes de dados, a adaptação do modelo para outros algoritmos de agrupamento e o desenvolvimento de novas abordagens sobre o problema.

Referências

AGGARWAL, C. C.; REDDY, C. K. *Data Clustering: Algorithms and Applications*. 1st. ed. [S.l.]: Chapman & Hall/CRC, 2013. ISBN 1466558210, 9781466558212. Citado na página 1.

AZIZ, D. et al. Initialization of adaptive neuro-fuzzy inference system using fuzzy clustering in predicting primary triage category. In: IEEE. *Intelligent and Advanced Systems (ICIAS), 2012 4th International Conference on*. [S.l.], 2012. v. 1, p. 170–174. Citado na página 5.

BARBER, D. *Bayesian Reasoning and Machine Learning*. New York, NY, USA: Cambridge University Press, 2012. ISBN 0521518148, 9780521518147. Citado na página 4.

BEZDEK, J. *Fuzzy mathematics in pattern classification*. [S.l.]: Cornell University, August, 1973. Citado na página 9.

BEZDEK, J. C. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Norwell, MA, USA: Kluwer Academic Publishers, 1981. ISBN 0306406713. Citado na página 10.

BOCK, H. H. Automatische klassifikation. vandenhoeck & ruprecht. Göttingen, Zürich, 1974. Citado na página 5.

CHEN, H.-L.; CHUANG, K.-T.; CHEN, M.-S. On data labeling for clustering categorical data. *IEEE Transactions on Knowledge and Data Engineering*, IEEE Computer Society, Los Alamitos, CA, USA, v. 20, n. 11, p. 1458–1472, 2008. ISSN 1041-4347. Citado na página 3.

CHUANG, S.-L.; CHIEN, L.-F. A practical web-based approach to generating topic hierarchy for text segments. In: *Proceedings of the Thirteenth ACM International Conference on Information and Knowledge Management*. New York, NY, USA: ACM, 2004. (CIKM '04), p. 127–136. ISBN 1-58113-874-1. Citado na página 3.

CINTRA, M. E. et al. An approach for the extraction of classification rules from fuzzy formal contexts. *Computer Science and Mathematics Institute Technical Reports*, p. 1–28, 2011. Citado na página 3.

DRINEAS, P. et al. Clustering large graphs via the singular value decomposition. *Machine Learning*, v. 56, p. 9–33, 2004. Citado na página 7.

DUDA, R. O.; HART, P. E. *Pattern Classification and Scene Analysis*. New Yotk, USA: John Willey & Sons, 1973. Citado na página 8.

DUNN, J. C. A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*, v. 3, n. 3, p. 32–57, 1973. Citado na página 9.

ELTOFT, T.; FIGUEIREDO, R. de. A self-organizing neural network for cluster detection and labeling. In: IEEE. *Neural Networks Proceedings, 1998. IEEE World Congress on*

Computational Intelligence. The 1998 IEEE International Joint Conference on. [S.l.], 1998. v. 1, p. 408–412. Citado na página 3.

FIGUEIREDO S. BOTELHO, P. D. M.; HAFFELE, C. Self-organizing mapping of robotic environments based on neural networks. *Brazilian Symposium on Artificial Neural Networks*, p. 136–141, 2012. Citado na página 5.

FISHER, D. Improving inference through conceptual clustering. In: *Proceedings of the Sixth National Conference on Artificial Intelligence - Volume 2*. [S.l.]: AAAI Press, 1987. (AAAI'87), p. 461–465. ISBN 0-934613-42-7. Citado na página 5.

GLOVER, E. et al. Inferring hierarchical descriptions. In: *Proceedings of the Eleventh International Conference on Information and Knowledge Management*. New York, NY, USA: ACM, 2002. (CIKM '02), p. 507–514. ISBN 1-58113-492-4. Citado na página 3.

KANUNGO, T. et al. An efficient k-means clustering algorithm: Analysis and implementation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 24, n. 7, p. 881–892, 2002. Citado na página 5.

KRISHNAPURAM, R.; KELLER, J. The possibilistic c-means algorithm: insights and recommendations. *Fuzzy Systems, IEEE Transactions on*, v. 4, n. 3, p. 385–393, Aug 1996. ISSN 1063-6706. Citado na página 6.

LIU, X.; FENG, X.; PEDRYCZ, W. Extraction of fuzzy rules from fuzzy decision trees: An axiomatic fuzzy sets (afs) approach. *Data Knowl. Eng.*, Elsevier Science Publishers B. V., Amsterdam, The Netherlands, The Netherlands, v. 84, p. 1–25, mar. 2013. ISSN 0169-023X. Disponível em: <http://dx.doi.org/10.1016/j.datak.2012.12.001>. Citado na página 3.

LOPES, L.; MACHADO, V.; RABELO, R. Automatic cluster labeling through artificial neural networks. In: *Neural Networks (IJCNN), 2014 International Joint Conference on*. [S.l.: s.n.], 2014. p. 762–769. Citado na página 4.

MCKAY, M. D.; BECKMAN, R. J.; CONOVER, W. J. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, Taylor & Francis, Ltd. on behalf of American Statistical Association and American Society for Quality, v. 21, n. 2, p. pp. 239–245, 1979. ISSN 00401706. Citado na página 7.

MITCHELL, T. M. *Machine Learning*. 1. ed. New York, NY, USA: McGraw-Hill, Inc., 1997. ISBN 0070428077, 9780070428072. Citado na página 4.

NAVARRO, J. *El vidrio*. Consejo Superior de Investigaciones Científicas, 2003. ISBN 9788400081584. Disponível em: <https://books.google.com.br/books?id=4GsNCPQRaTwC>. Citado na página 20.

OLIVEIRA, J. V. d.; PEDRYCZ, W. *Advances in Fuzzy Clustering and Its Applications*. New York, NY, USA: John Wiley & Sons, Inc., 2007. ISBN 0470027606. Citado na página 5.

PERES, S. M. et al. Tutorial sobre fuzzy-c-means e fuzzy learning vector quantization: Abordagens híbridas para tarefas de agrupamento e classificação. *Revista de Informática Teórica e Aplicada*, v. 19, n. 1, p. 120–163, 2012. Citado na página 5.

POPESCUL, A.; UNGAR, L. H. Automatic labeling of document clusters. Unpublished MS, U. Pennsylvania. 2000. Disponível em: <http://www.cis.upenn.edu/~popescul/Publications/popescul00labeling.pdf>. Citado na página 3.

QUINLAN, J. Induction of decision trees. *Machine Learning*, Kluwer Academic Publishers, v. 1, n. 1, p. 81–106, 1986. ISSN 0885-6125. Citado na página 3.

RAMATHILAGA, S.; LEU, J.-Y.; HUANG, Y.-M. Adapted mean variable distance to fuzzy-cmeans for effective image clustering. In: IEEE. *Robot, Vision and Signal Processing (RVSP), 2011 First International Conference on*. [S.l.], 2011. p. 48–51. Citado na página 5.

RUNKLER, T. A. *Data Analytics: Models and Algorithms for Intelligent Data Analysis*. [S.l.]: Vieweg, 2012. ISBN 3834825883, 9783834825889. Citado 2 vezes nas páginas 17 e 18.

SETNES, M. Supervised fuzzy clustering for rule extraction. In: *Fuzzy Systems Conference Proceedings, 1999. FUZZ-IEEE '99. 1999 IEEE International*. [S.l.: s.n.], 1999. v. 3, p. 1270–1274 vol.3. ISSN 1098-7584. Citado na página 3.

TZERPOS, V. *Comprehension-Drive Software Clustering*. Tese (Doutorado) — Universisity of Toronto, 2001. Citado na página 1.

VARGAS, R.; BEDREGAL, B. Uma extensão intervalar do algoritmo fuzzy c-means. In: SBMAC (BRAZILIAN SOCIETY OF APPLIED AND COMPUTATIONAL MATH), CURITIBA, BRAZIL. *Proceedings of CNMAC 2009 (32th Brazilian Conference on Applied and Computational Math)*. [S.l.], 2009. Citado na página 3.

VARGAS, R.; BEDREGAL, B.; FILHO, I. O. Algoritmo fuzzy c-means adaptado para aplicações com dados intervalares simbólicos. In: *Proceedings of ERMAC 2009 (IX Encontro Regional de Matemática Aplicada e Computacional)*. [S.l.: s.n.], 2009. Citado na página 3.

XU, R.; WUNSCH II, D. Survey of clustering algorithms. *Transactions on Neural Networks*, IEEE Press, Piscataway, NJ, USA, v. 16, n. 3, p. 645–678, maio 2005. ISSN 1045-9227. Citado na página 1.

ZADEH, L. A. Fuzzy sets. *Information and control*, Elsevier, v. 8, n. 3, p. 338–353, 1965. Citado na página 8.