



UNIVERSIDADE FEDERAL DE MINAS GERAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA
ELÉTRICA

LUIZ ALBERTO QUEIROZ CORDOVIL JÚNIOR
RODRIGO FARIAS ARAÚJO

SISTEMAS NEBULOSOS: EXERCÍCIO COMPUTACIONAL 3

Belo Horizonte - MG, 2017

LUIZ ALBERTO QUEIROZ CORDOVIL JÚNIOR
RODRIGO FARIAS ARAÚJO

APLICAÇÃO DO ALGORITMO C-MEANS PARA SEGMENTAÇÃO DE IMAGENS

Relatório apresentado como requisito parcial para obtenção de aprovação na disciplina Sistemas Nebulosos do Programa de Pós-Graduação em Engenharia Elétrica, na Universidade Federal de Minas Gerais.

Prof. Dr. André Paim Lemos

Belo Horizonte - MG, 2017

Sumário

1	Introdução	1
2	Descrição do Experimentos	2
3	Resultados	3
3.1	Agrupamento de dados	3
3.2	Segmentação de imagens	5
4	Conclusões	6
	Referências	7

Lista de Figuras

1	Conjunto de dados <code>fcmdata.dat</code>	3
2	Imagens usadas na aplicação do algoritmo c-means para segmentação de imagens.	4
3	Comparação dos centros dos agrupamentos, considerando 4 agrupamentos, obtidos por meio do algoritmo c-means implementado e da função <code>fcm</code> do MATLAB.	4

1 Introdução

Um algoritmo de clusterização organiza itens em grupos, baseado no critério de similaridade. O algoritmo *fuzzy c-Means* realiza agrupamentos de tal forma que determinado item pode pertencer a mais de um grupo afim, em que o grau de pertinência para cada item é dado por uma distribuição de probabilidade sobre os *clusters*.

Fuzzy c-means, é um algoritmo para análise de agrupamentos associado à representação de uma classe ou grupo singular em um dado conjunto de dados. c agrupamentos são representados por um vetor de centros C .

Considerada como uma técnica de aprendizado não-supervisionado e raciocínio aproximado, possui aplicações diversas em tratativas de classificação, dentre as quais aplicação à tarefa de segmentação de imagens e inferência de dados abordadas neste trabalho.

Para que se alcance este objetivo ou *clusterização*, realiza-se um número de iterações no sentido de minimizar $J_c M(\mu_h, C)$ dada por:

$$J = \sum_{i=1}^N \sum_{j=1}^C \mu_{ij} \|x_i - c_j\|^2 \quad (1.1)$$

onde:

- $N \longrightarrow$ número de dados;
- $C \longrightarrow$ número de centros (*clusters*);
- $c_j \longrightarrow$ vetor de centro do j ;
- $\mu_{ij} \longrightarrow$ grau de pertinência do i -ésimo dado x_i no *cluster* j .

A norma $\|x_i - c_j\|$ mede a similaridade (ou proximidade) de um dado x_i para vetor de centro c_j do j . Em cada iteração, o algoritmo mantém o vetor centro para cada *cluster*.

Para cada ponto x_i , o grau de pertinência para cada c_j é calculado da seguinte forma:

$$\mu_{ij} = \frac{1}{\sum_{k=1}^C \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}} \quad (1.2)$$

onde m é coeficiente de fuzzificação.

O vetor centro c_j é calculado como:

$$c_j = \frac{\sum_{i=1}^N \mu_{ij}^m x_i}{\sum_{i=1}^N \mu_{ij}^m} \quad (1.3)$$

No início do algoritmo, o grau de pertinência para os dados é inicializado com um valor randômico $\theta_{ij} \in [0, 1] \mid \sum_j^C \mu_{ij} = 1$.

O coeficiente de *fuzzificação* $m \in [1, \infty]$, mede a tolerância do *cluster*. Este valor determina o quanto um *cluster* pode sobrepor outro. Quanto maior este valor, maior a sobreposição entre os agrupamentos, ao passo que também se usa uma maior quantidade de dados que podem estar inseridos em subconjunto fuzzy em que o grau de pertinência não é nem 0 ou 1, mas sim, algo entre estes valores.

O critério de parada é expresso em termos da acurácia dos graus de pertinência aplicados ao conjunto de dados, no sentido de determinar o número de iterações durante o processo de minimização. Esta é calculada utilizando o grau de pertinência de uma iteração para outra, considerando o maior valor de μ_{ij} para todos os dados x_i em todos os (clusters) c_j . A representação desta medida entre a iteração k e $k + 1$ é dada por:

$$\epsilon = \Delta_i^N \Delta_j^C |\mu_{ij}^{k+1} - \mu_{ij}^k| \quad (1.4)$$

onde μ_{ij}^k e μ_{ij}^{k+1} são os graus de pertinência nas respectivas iterações e o operador Δ , retorna o maior valor do vetor em análise.

Algoritmo 1: Algoritmo Fuzzy c-Means

Agrupamento

Determinar a quantidade de partições c ;

Determinar o erro máximo e ;

Inicializar os centros aleatoriamente;

Inicializar o contador de iterações $t = 0$;

repita

Incrementar t ;

Atualizar μ_h ;

Atualizar C ;

até que $\|C^{(t)} - C^{(t-1)}\| < (e)$

2 Descrição do Experimentos

Com o objetivo de avaliar o algoritmo *c-means* foram elaborados os seguintes experimentos:

1. Aplicação do algoritmo *fuzzy c-means* desenvolvido ao conjunto de dados `fcmdat.dat` para 4 (quatro) grupos de agrupamentos e comparação dos resultados obtidos com a

função f_{cm} do MATLAB.

2. Adaptação do algoritmo anterior para o problema de segmentação de imagens e comparação dos resultados obtidos com a função f_{cm} do MATLAB para o espaço de cores RGB utilizando diferentes quantidades de grupos de agrupamento.

A Figura 1 e 2 ilustram os dados do conjunto `fcmdata.dat` e as imagens utilizadas na aplicação do algoritmo *c-means* para segmentação de imagens, respectivamente.

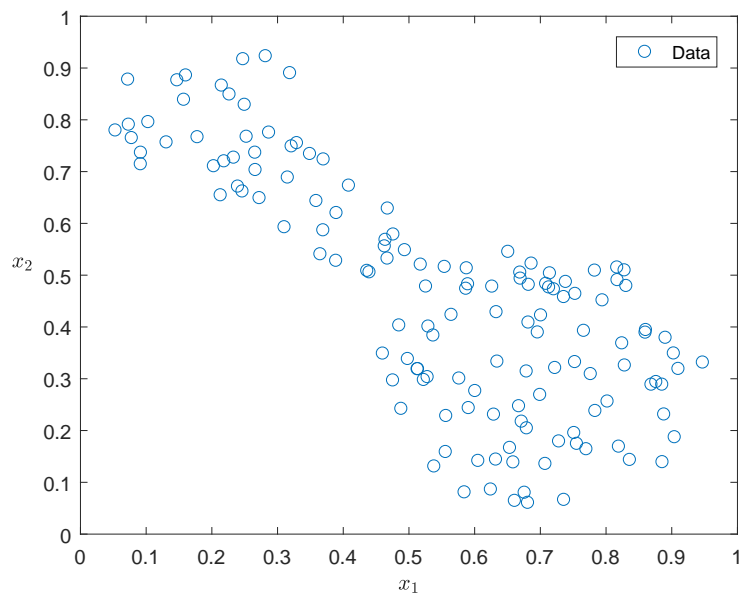


Figura 1: Conjunto de dados `fcmdata.dat`.

3 Resultados

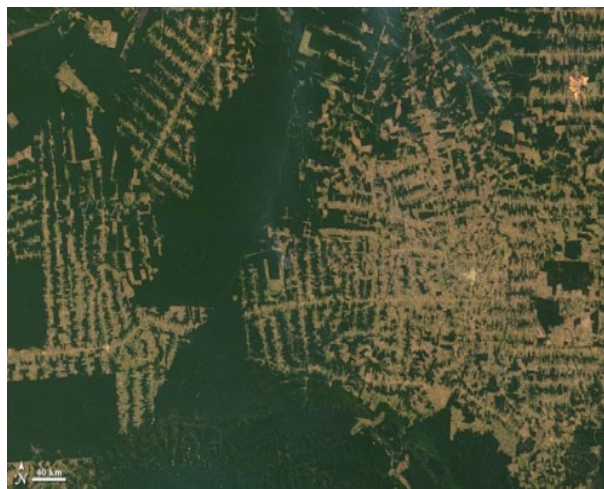
Nesta seção serão apresentados os resultados dos experimentos descritos anteriormente na Seção 2.

3.1 Agrupamento de dados

O algoritmo *c-Means* implementado foi utilizado para agrupamento de dados do conjunto de dados `fcmdata.dat`, considerando 4 agrupamentos. As Figuras 3(a) e 3(b) ilustram os centros dos agrupamentos obtidos quando utilizado o algoritmo implementado e a função f_{cm} do MATLAB, respectivamente.



(a) Imagem 01

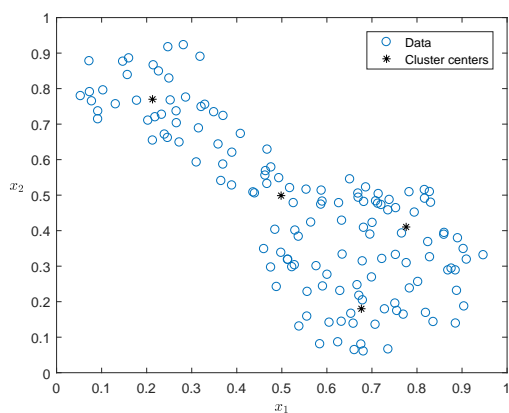


(b) Imagem 02

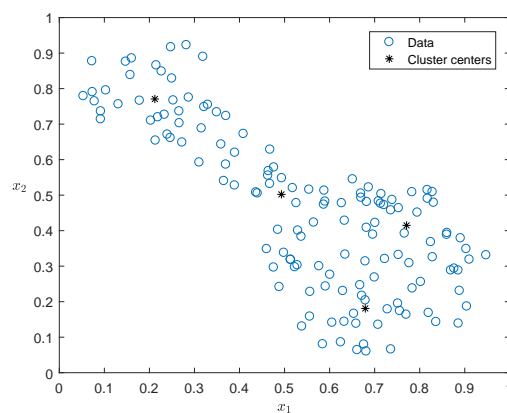


(c) Imagem 03

Figura 2: Imagens usadas na aplicação do algoritmo c-means para segmentação de imagens.



(a) Algoritmo desenvolvido.



(b) Função f_{cm} do MATLAB.

Figura 3: Comparação dos centros dos agrupamentos, considerando 4 agrupamentos, obtidos por meio do algoritmo c-means implementado e da função f_{cm} do MATLAB.

3.2 Segmentação de imagens

Na abordagem computacional, por meio da técnica de validação cruzada, cada conjunto de espécies foi separado de forma aleatória em subconjuntos de treinamento e validação representados, com dimensão de 70% e 30%, respectivamente de cada classe de padrões.

Para todas as classes, no total, quatro dados de entrada aplicados a três espécies, com 50 amostras. Como ferramenta de aprendizagem, os conjuntos de treinamento e validação, foram determinados observando-se os percentuais em 35 e 15 amostras, respectivamente, para cada classe C_j .

Neste sentido, na etapa de treinamento, a partir de aprendizado supervisionado, houve a classificação inicial das amostras, como função da quantidade de regras e dos graus de certeza, no sentido de se estabelecer as condições e parâmetros iniciais do sistema de classificação.

A partir destes, houve a implementação das características de contexto adquiridas na etapa de treinamento, no sentido de aferir-se o desempenho do sistema de classificação.

Os experimentos foram realizados 25 vezes, para cada número de funções de pertinência de 2 até 15, com o propósito de se observar o limiar de convergência e paridade entre o erro do conjunto de treinamento e erro no conjunto de validação e eliminar a natureza estocástica da escolha do dados de treinamento de validação.

Nota-se que, para ambas as *t-normas* o número dito ideal de funções de pertinência é 8. De modo que, para um número menor de funções de pertinência ocorre o chamado *underfitting*, e para um número maior de funções de pertinência ocorre o chamado *overfitting*.

4 Conclusões

A utilização de métodos de classificação baseados em inferência, oferece uma solução baseada em conhecimento adquirido por aprendizagem ativa. Nesta abordagem, a classificação de pontos afins com relação aos seus respectivos centros, é relevante para análise de similaridade entre conjuntos de dados não-rotulados.

Com a aplicação de algoritmo *fuzzy c-Means* observou-se na realização dos experimentos, o efeito da aplicação das relações nebulosas e avaliação baseada em acurácia, em processos de inferência, propiciando distinção entre os dados de acordo com suas características de contexto. Nos experimentos de segmentação de imagem e no agrupamento de um conjunto de dados não-rotulados, a rotina implementada apresentou resultado bastante similar tomando-se como referência a função f_{cm} do MATLAB©, apresentado desempenho satisfatório, conforme indicado na seção 3.

Referências

- [1] CANNON, Robert L.; DAVE, Jitendra V.; BEZDEK, James C. Efficient implementation of the fuzzy c-means clustering algorithms. IEEE transactions on pattern analysis and machine intelligence, n. 2, p. 248-255, 1986.
- [2] COUTINHO, P. H. S. Proposta de Novos Algoritmos Híbridos de Clusterização Fuzzy e suas aplicações. Trabalho de Conclusão de Curso. Departamento de Ciências Exatas e Tecnológicas. Universidade Estadual de Santa Cruz. Ilhéus-Bahia, 2017.