



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	4
3.1	Agrupamento de dados	4
3.2	Segmentação de imagens	4
4	Conclusões	6
	Referências	7

Lista de Figuras

1	Conjunto de dados <code>fcm_dataset.mat</code>	3
2	Imagens usadas na aplicação do algoritmo <i>c-Means</i> para segmentação de imagens.	3
3	Comparação dos centros dos agrupamentos, considerando 4 agrupamentos, obtidos por meio do algoritmo <i>c-Means</i> implementado e da função <code>fcm</code> do MATLAB.	4
4	Comparação das imagens, considerando 4 agrupamentos, obtidas por meio do algoritmo <i>c-Means</i> implementado e da função <code>fcm</code> do MATLAB.	5
5	Comparação das imagens, considerando 3 agrupamentos, obtidas por meio do algoritmo <i>c-Means</i> implementado e da função <code>fcm</code> do MATLAB.	5
6	Comparação das imagens, considerando 6 agrupamentos, obtidas por meio do algoritmo <i>c-Means</i> implementado e da função <code>fcm</code> do MATLAB.	5

1 Introdução

Um algoritmo de agrupamento organiza itens em grupos, baseado no critério de similaridade. O algoritmo *fuzzy c-Means* realiza agrupamentos de tal forma que determinado dado possuiu um grau de pertinência em relação a cada grupo, de modo que um maior grau de pertinência representa um maior compatibilidade do dado com o respectivo grupo, *cluster*.

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

Para que se alcance o objetivo da *clusterização*, realiza-se determinado número de iterações no sentido de minimizar a função de custo $J_c(\mu_h, C)$ definida na Eq. (1.1), em termos da distância euclidiana entre um ponto de um conjunto de dados e o centro que de cada grupo.

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

onde:

- N : número de dados;
- C : número de grupos (*clusters*);
- c_j : centro do j -ésimo grupo;
- μ_{ij} : grau de pertinência do i -ésimo dado x_i em relação ao j -ésimo grupo.

A norma, $\|x_i - c_j\|$, mede a similaridade (ou proximidade) de um dado elemento x_i para cada centro c_j do j -ésimo grupo. Para cada ponto x_i , o grau de pertinência para cada μ_{ij} é calculado conforme a Eq. (1.2):

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

onde m é coeficiente de fuzzificação e cada centro c_j é calculado de acordo com a Eq. (1.3).

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

Inicialmente, o grau de pertinência para os dados é inicializado com um valor randômico no intervalo $[0, 1]$ tal que $\sum_j^C \mu_{ij} = 1$. O coeficiente de fuzzificação $m \in [1, \infty)$, mede a tolerância do grupo. Este valor determina o quanto um grupo pode sobrepor outro. Quanto maior este valor, maior a sobreposição entre os agrupamentos.

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 pode ser calculada utilizando a norma entre o vetor de centros, C^k da k -ésima iteração, e o vetor C^{k-1} . Assim, o algoritmo *fuzzy c-Means* é proposto conforme o Algoritmo 1 a seguir.

Algoritmo 1: Algoritmo Fuzzy *c-Means*

Determinar a quantidade de grupos c ;

Determinar o erro máximo ϵ ;

Inicializar os centros aleatoriamente;

Inicializar o contador de iterações $k = 0$;

repeat

 Incrementar k ;

 Atualizar μ_{ij} , dado pela Eq. (1.2);

 Atualizar cada centro c_j , dado pela Eq. (1.3);

until $\|C^{(t)} - C^{(t-1)}\| < \epsilon$;

2 Descrição do Experimentos

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

1. Aplicação do algoritmo *c-means* desenvolvido ao conjunto de dados `fcm_dataset.mat` para 4 (quatro) grupos de agrupamentos e comparação dos resultados obtidos com a função `fcm` do MATLAB.
2. Adaptação do algoritmo anterior para o problema de segmentação de imagens e comparação dos resultados obtidos com a função `fcm` 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 `fcm_dataset.mat` e as imagens utilizadas na aplicação do algoritmo *c-means* para segmentação de imagens, respectivamente.

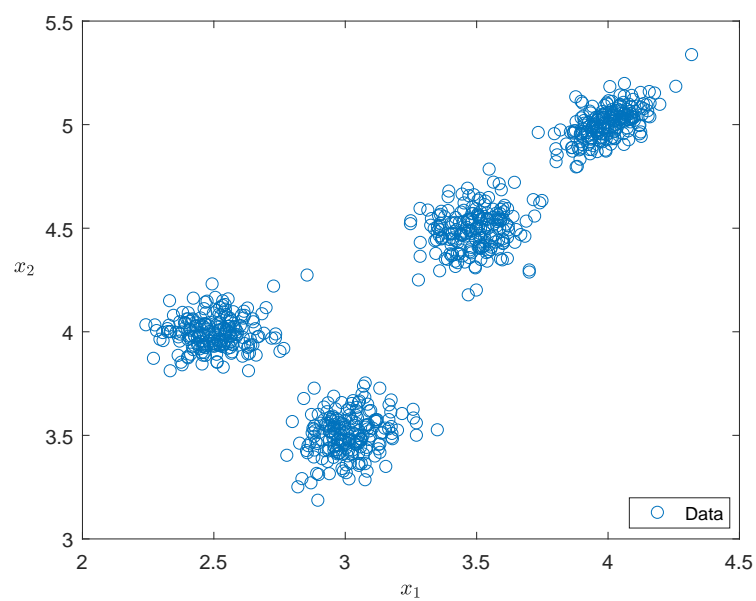


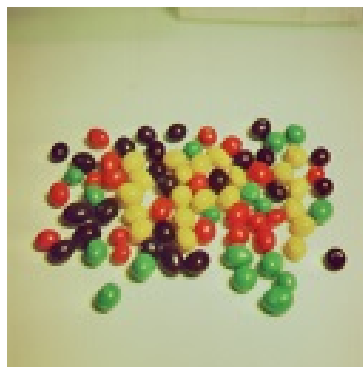
Figura 1: Conjunto de dados `fcm_dataset.mat`.



(a) Imagem 01



(b) Imagem 02



(c) Imagem 03

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

3 Resultados

Nesta seção serão apresentados os resultados dos experimentos descritos anteriormente na Seção 2. Todos os experimentos foram executados utilizando o algoritmo implementado para o erro, $\epsilon = 10^{-3}$, e a função f_{cm} do MATLAB, e para ambos os algoritmos $m = 2$.

3.1 Agrupamento de dados

O algoritmo *c-Means* implementado foi utilizado para agrupamento de elementos do conjunto de dados `fccm_dataset.mat`, considerando 4 grupos. Nas Figuras 3(a) e 3(b) são ilustrados os centros dos agrupamentos obtidos quando utilizado o algoritmo implementado e a função f_{cm} do MATLAB, respectivamente.

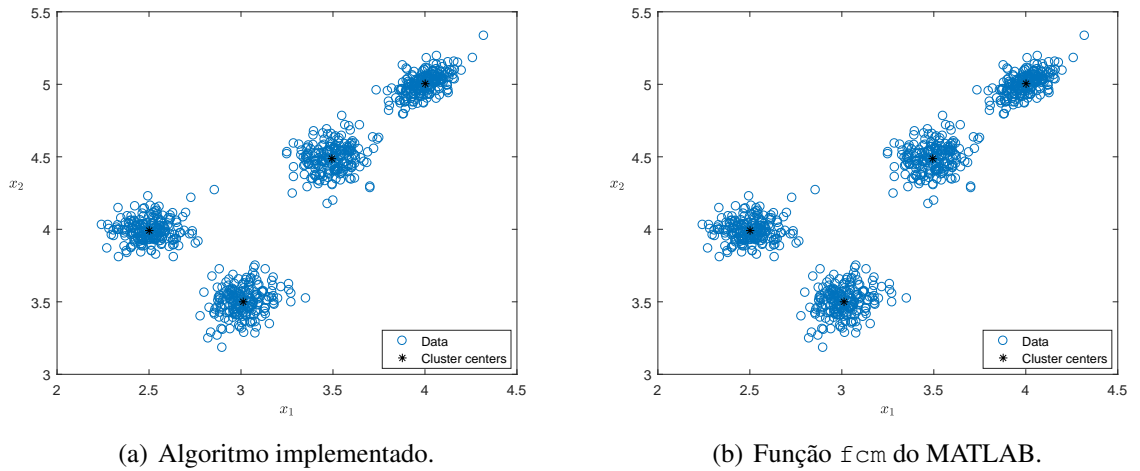


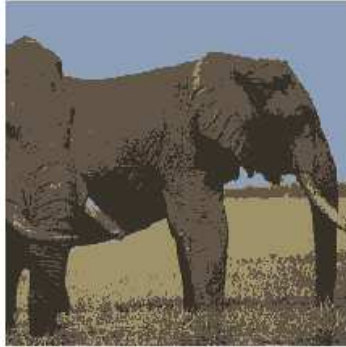
Figura 3: Comparação dos centros dos agrupamentos, considerando 4 agrupamentos, obtidos por meio do algoritmo *c-Means* implementado e da função f_{cm} do MATLAB.

3.2 Segmentação de imagens

O algoritmo utilizado no problema anterior foi adaptado para aplicação ao problema de segmentação de imagens para o espaço de cores RGB utilizando diferentes quantidades de grupos de agrupamento, nesse experimento nós escolhemos o número de grupos de acordo com avaliação visual da imagem.

Dessa maneira, para a Imagem 01, Figura 2(a), foi escolhido 4 grupos, isto é $c = 4$, já para a Imagem 02, Figura 2(b), atribuiu-se $c = 3$, apenas 3 grupos uma vez que a imagem possui poucas cores, e finalmente para a Imagem 03, Figura 2(c), escolheu-se 6 grupos, $c = 6$.

As Figuras 4, 5 e 6 mostram os resultados obtidos pela aplicação dos algoritmos para segmentação das imagens da Figura 2, repectivamente.

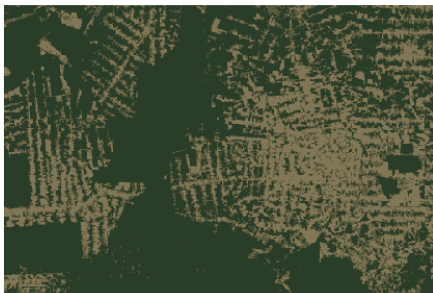


(a) Algoritmo implementado.



(b) Função f_{cm} do MATLAB.

Figura 4: Comparação das imagens, considerando 4 agrupamentos, obtidas por meio do algoritmo *c-Means* implementado e da função f_{cm} do MATLAB.

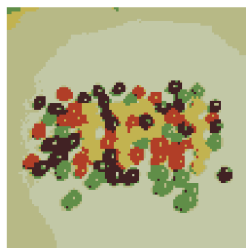


(a) Algoritmo implementado.

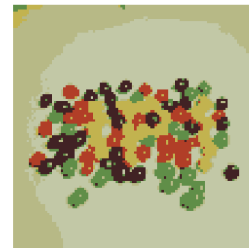


(b) Função f_{cm} do MATLAB.

Figura 5: Comparação das imagens, considerando 3 agrupamentos, obtidas por meio do algoritmo *c-Means* implementado e da função f_{cm} do MATLAB.



(a) Algoritmo implementado.



(b) Função f_{cm} do MATLAB.

Figura 6: Comparação das imagens, considerando 6 agrupamentos, obtidas por meio do algoritmo *c-Means* implementado e da função f_{cm} do MATLAB.

4 Conclusões

A utilização de métodos de classificação 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 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, o algoritmo implementado apresentou resultado bastante similar tomando-se como referência a função f_{cm} do MATLAB©, apresentado desempenho satisfatório, conforme indicado na Seção 3.

Bibliografias Consultadas

- [1] 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.
- [2] 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.