

COMPARAÇÃO DOS MÉTODOS DE SEGMENTAÇÃO DE IMAGENS OTSU, KMEANS E CRESCIMENTO DE REGIÕES NA SEGMENTAÇÃO DE PLACAS AUTOMOTIVAS

Leonardo Meneguzzi¹; Marcelo Massoco Cendron²; Manassés Ribeiro³

INTRODUÇÃO

Análise de clusters ou *clusterização* é a tarefa de segmentar e agrupar um conjunto de objetos. Esse conceito de agrupamento vem auxiliando o ser humano nas suas mais variadas tarefas. Sua aplicabilidade se estende por diversas áreas como medicina, engenharia, agricultura, meio ambiente entre outras. O projeto inicialmente tinha foco em segmentar placas automotivas, o projeto estava relacionado a engenharia de tráfego e segurança pública, e posteriormente para a clusterização de galáxias.

PROCEDIMENTOS METODOLÓGICOS

O projeto inicial tinha como objetivo comparar os métodos de segmentação de imagens *Otsu*, *K-means* e *Crescimento de Regiões*. No primeiro momento foram estudados os três métodos, visto a viabilidade de cada um e possíveis formas de aplicar. O método melhor estudado e que determinado através de pesquisas como de melhor desempenho para essa aplicação foi o K-means. Foram testados várias

¹Aluno do Instituto Federal Catarinense - Campus Videira - Técnico em informática - E-mail: meneguzzileo@gmail.com

²Professor Orientador. Instituto Federal Catarinense - Campus Videira - Técnico em informática - Email: marcelo.cendron@ifc-videira.edu.br

³Professor Orientador. Instituto Federal Catarinense - Campus Videira - Técnico em informática - Email: manasses@ifc-videira.edu.br

linguagens para se implementar, e também se usáramos dados bidimensionais ou tridimensionais.

O projeto posteriormente mudou seu foco, para segmentar galáxias utilizando os métodos de segmentação já estudados. Os dados utilizados para testes e aplicação já estavam disponíveis, eles foram obtidos de um outro projeto paralelo. Sendo assim o trabalho foi subdividido em duas etapas.

Na primeira etapa foi realizada a revisão bibliográfica sobre o algoritmo de mineração de dados "K-means", e foi estudado a melhor forma de se aplicar. Também foi definido a linguagem que iríamos usar, e que o método trabalharia com dados bidimensionais, e quais os procedimentos necessários para de chegar ao resultado esperado. A linguagem utilizada foi Java. Também definimos os testes, para isso foram escolhidos cinco conjuntos de dados que foram convertidos em uma matriz de duas colunas com a posição dos pontos para ser possível aplicar o algoritmo.

Na segunda etapa foram executados os testes, sendo que todos os conjuntos por pré-definição do projeto foram separados em dois agrupamentos. Não foi necessário uma grande bateria de testes, nem muitos ajustes para melhor funcionamento, uma vez que o método se comportou bem em todos os testes desenvolvidos.

RESULTADOS E DISCUSSÕES

Análise de clusters ou clusterização é a tarefa de agrupar um conjuntos de objetos. Esses objetos são agrupados seguindo algum padrão. Esse método é uma técnica da mineração de dados(área que tem como objetivo filtrar e analisar informações e resumi-las), que tem aplicações em análise de dados estatísticos, reconhecimentos de padrões, análise de imagens, entre outros. Clusterização poderia ser então entendido como um processo de organização de dados/objetos em grupos cujos membros de alguma forma são semelhantes. Esse agrupamento pode ser feito com vários tipos de algoritmos.

Para tal tarefa foi usado o K-means, um dos mais simples métodos de algoritmos de aprendizagem não supervisionada de Clusterização. Esse método usa um procedimento simples para classificar agrupamentos sem a necessidade de especificar muitos parâmetros. A necessidade do projeto consiste em separar constelações em dois agrupamentos. Foi utilizado esse algoritmo implementado em Java, com dados de tabelas bidimensionais.

Nos seguintes agrupamentos de dados do projeto WISE os resultados das separações foram (os gráficos a seguir foram gerados apenas para visualização dos resultados):

Figura 1 - Gráfico $w1-w2 \times g-i$ não clusterizado e clusterizado

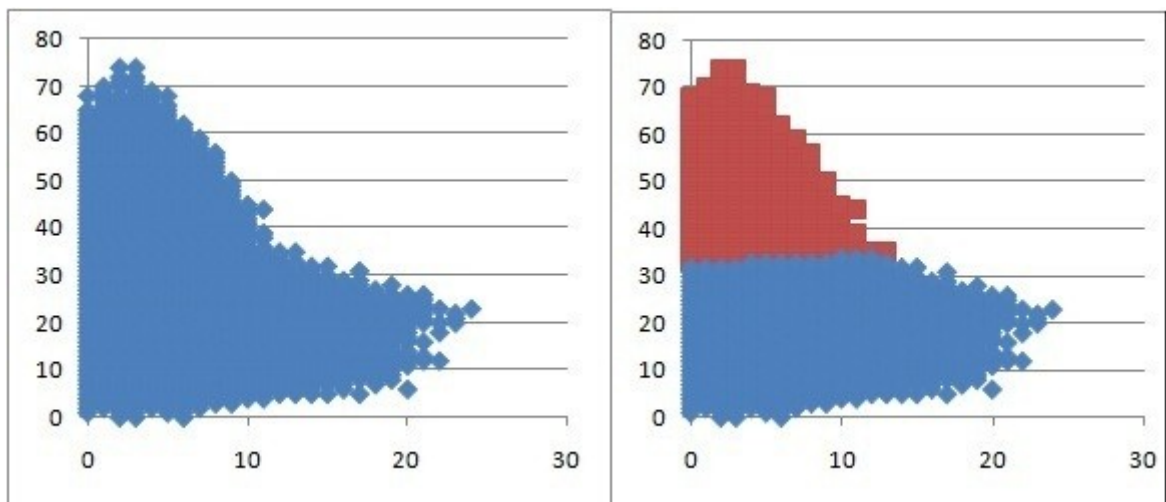


Figura 2 - Gráfico w_1-w_2 x $i-z$ não clusterizado e clusterizado;

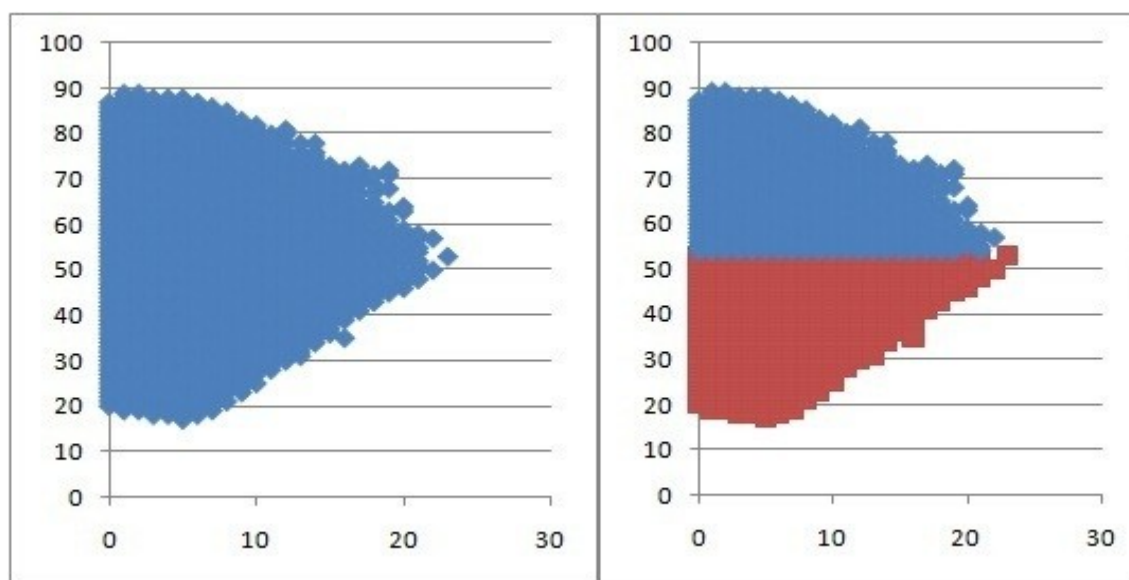
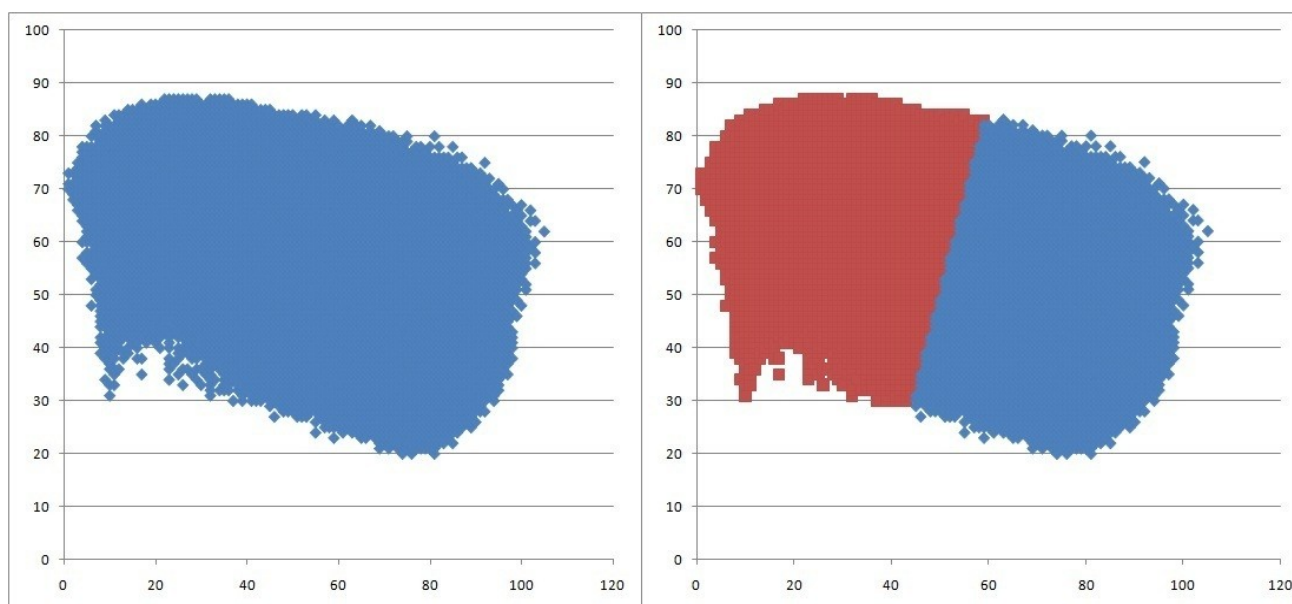


Figura 3 - Gráfico w_1-w_3 x $g-r$ não clusterizado e clusterizado;



Todos os agrupamentos foram separados em dois clusters.

CONSIDERAÇÕES FINAIS

Os resultados não foram apenas importantes para o escopo desse projeto, sua maior relevância se deu no contexto geral do mapeamento de galáxias, pois permitiu que através da linguagem de programação Java, fosse possível realizar a classificação com uma linguagem independente de plataforma e com fácil entendimento. Através de um simples algoritmo de Mineração de dados generalizado pode-se obter resultados satisfatórios.

Além disso, a base desse projeto foi utilizada em outro projeto de pesquisa do mesmo coordenador, em que foi melhorada a velocidade através do uso da placa de processamento gráfico (placa de vídeo) o qual não poderia ter sido realizado se não tivesse sido implementado na linguagem apresentada.

REFERÊNCIAS

DUNHAM, M.H. – Data Mining Introductory and advanced topics [S.l.]:Prentice Hall, 2003

Wright, E. L. et al. The Wide-field Infrared Survey Explorer (WISE): Mission Description and Initial On-orbit Performance. v. 140, p. 1868–1881, dez. 2010.