

13ª Mostra da Produção Universitária

Rio Grande/RS, Brasil, 14 a 17 de outubro de 2014.

UTILIZAÇÃO DE ALGORITMOS DE AGRUPAMENTO PARA CLASSIFICAÇÃO DE GALÁXIAS

GIL, Vanessa de Oliveira
EMMENDORFER, Leonardo
FERRARI, Fabricio
vanessa.gil@furg.br

Evento: Encontro de Pós-Graduação
Área do conhecimento: Astrofísica Extragaláctica

Palavras-chave: galáxias; agrupamentos; mineração de dados.

1 INTRODUÇÃO

No contexto da Astronomia, a classificação de galáxias mostra-se como uma importante etapa conceitual para compreender suas características físicas e seus processos de formação e evolução. Com o advento de grandes *surveys* astronômicos contendo milhares de objetos, a classificação dessas galáxias torna-se problemática, pois até o momento a prática mais comum de classificação de galáxias é a visual, guiada pelo diagrama de Hubble, que não é capaz de atender a demanda atual.

A necessidade de automação e de um método de classificação conciso é evidente, portanto a utilização de aprendizagem de máquina não-supervisionada é de grande importância. A abordagem utilizada será a de análise de agrupamentos, onde procura-se encontrar grupos de observações fortemente relacionadas de modo que os objetos que pertençam ao mesmo grupo sejam mais similares entre si do que com os que pertençam a outros grupos. A evolução das estruturas das galáxias é um dos passos para compreendermos como a matéria do universo se agregou e formou as estruturas perceptíveis atualmente.

2 REFERENCIAL TEÓRICO

Edwin Hubble organizou e sintetizou padrões representativos da morfologia das galáxias e propôs uma sequência baseada na complexidade das estruturas. As galáxias elípticas são visualmente mais simples que as lenticulares, enquanto as galáxias espirais com braços mais abertos são as mais complexas. Não são todas as galáxias que se encaixam perfeitamente no esquema de Hubble, portanto isso torna difícil sua classificação em termos morfológicos devido à assimetria, núcleos descentralizados, dentre outros. Por isso, são utilizadas medidas não paramétricas da morfologia das galáxias.

O CAS é um método não paramétrico simples, onde Abraham et al.(1996) introduziu o índice de concentração (C) e Shade et al.(1995) apresentou a assimetria rotacional (A) como uma maneira de distinguir automaticamente as galáxias de tipo precoce e tardia classificadas por Hubble. Além disso, a suavidade (S) também é um parâmetro utilizado para quantificar a fração de luz que está contida em estruturas de pequena escala. Há também outras duas medidas não paramétricas para quantificar a morfologia de uma galáxia que suavizam os problemas do sistemas CAS. O primeiro é o coeficiente de Gini (G) que quantifica a distribuição relativa do fluxo dentro de um pixel associado com a imagem da galáxia e o segundo é o

momento de luz (M).

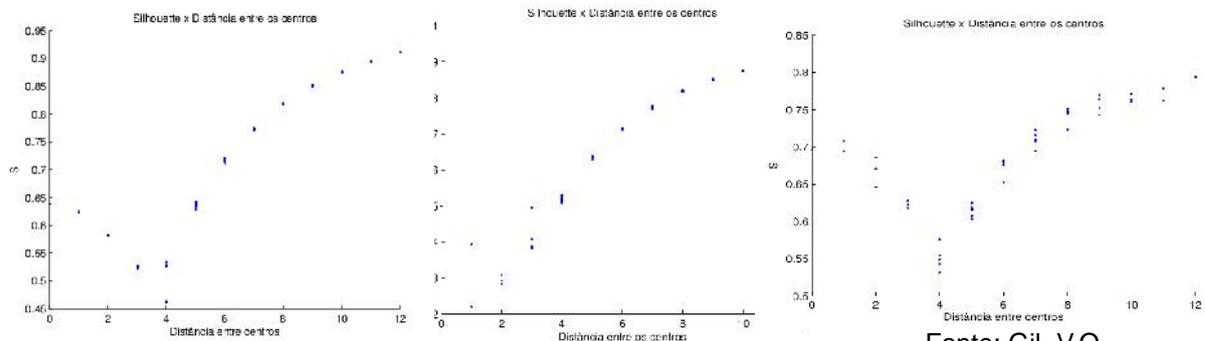
3 PROCEDIMENTO METODOLÓGICO

A metodologia escolhida para este trabalho utiliza a análise exploratória de dados através de técnicas de agrupamento objetivando analisar os resultados para detectar classes de galáxias por meio de parâmetros morfométricos. Fazem parte do conjunto de dados usados, os dados reais e sintéticos que contenham medidas morfométricas de galáxias, onde os parâmetros utilizados nos algoritmos de agrupamentos EM, K-means e DBSCAN são escolhidos através de Análise de Componentes Principais e Seleção de Atributos. Após será aplicado um algoritmo que valide os resultados obtidos, o *Silhouette de Russeeuw, P.J. (1987)*.

4 RESULTADOS e DISCUSSÃO

Foram utilizados dados sintéticos bidimensionais para a obtenção de resultados preliminares, onde aproximamos duas gaussianas até que seus centros coincidissem e aplicamos os algoritmos de agrupamento. Após, o *Silhouette* foi aplicado para inferir a qualidade dos agrupamentos gerados pelos três algoritmos. A figura 1 ilustra os resultados obtidos que se mostram satisfatórios. Conforme os centros das gaussianas se aproximaram, tornou-se mais difícil para os algoritmos definir claramente os grupos. O EM se mostrou mais adequado à aplicação.

Figura 1 – Resultados obtidos com o K-means, EM e DBSCAN.



Fonte: Gil, V.O.

5 CONSIDERAÇÕES FINAIS

Espera-se encontrar um número correto de grupos tal que separe os diferentes tipos morfológicos de galáxias de acordo com suas propriedades morfométricas presentes nos *surveys* utilizados. Nesse espaço de parâmetros não surgem grupos de objetos com diferenças muito nítidas. Esse resultado reforça a existência de uma continuidade morfométricas nas populações de galáxias elípticas e espirais.

REFERÊNCIAS

Abraham, R. G., et al. 1996, ApJ, 471, 694.

SHADE, D., et al., 1995, ApJ, 451, L1.

ROUSSEEUW, P.J.; Silhouette: a graphical aid to the interpretation and validation of clusters analysis, Journal of Computational and Applied Mathematics, 1987.