

## Adaptação de métodos para o agrupamento de formas tridimensionais

Inácio R. A. do Nascimento<sup>1</sup>

CCEN/UFPE, Recife, PE

Getúlio J. A. do Amaral<sup>2</sup>

CCEN/UFPE, Recife, PE

Representações geométricas e imagens de objetos são agentes de estudos e pesquisas em diversos campos e aplicações, tais como Biologia, Medicina, Neurociência, Arqueologia e Reconhecimento Facial. A Análise Estatística de Formas (*Statistical Shape Analysis*) é um ramo da estatística, utilizado para trabalhar com representações geométricas e formas de objetos. A ideia central da abordagem geométrica da análise de formas é a utilização da representação do próprio objeto. Assim, um modo de caracterizar uma forma é detectar um conjunto finito de pontos em torno da silhueta desse objeto. A esse conjunto de pontos é dado o nome de marcos e a posição desses pontos relaciona-se com as coordenadas cartesianas. Temos que uma configuração é uma coleção de marcos em um determinado objeto, sendo representada matematicamente por uma matriz  $\mathbf{X}$  de dimensão  $k \times m$  de coordenadas cartesianas de  $k$  marcos em  $m$  dimensões. Forma é toda a informação que resta quando são removidos os efeitos de locação, escala e rotação de um determinado objeto [2].

Em diversas situações do cotidiano é relevante a classificação dos indivíduos de um conjunto de dados em grupos, seja para auxiliar no entendimento do fenômeno estudado ou simplesmente para organizar os dados. Análise de Agrupamento é um conjunto de técnicas que tem como objetivo criar grupos, de tal maneira que os elementos de cada grupo sejam similares entre si e os grupos sejam diferentes.

Em várias situações, faz-se necessário o agrupamento de objetos pertencentes ao espaço de formas, os quais são espaços não-euclidianos. Neste sentido, diversos autores adaptaram algoritmos para o agrupamento de formas. Considerando formas bidimensionais, pode-se citar o trabalho feito por [1] que realiza a adaptação da versão de Hartigan-Wong do K-médias. Porém, no cenário de formas tridimensionais, não existem muitos trabalhos envolvendo adaptações de métodos para o agrupamento de formas 3D. Um dos trabalhos mais significativos nesse quesito foi desenvolvido por [6], o qual apresenta a adaptação das versões de Lloyd do K-médias e do K-médias aparado.

O presente trabalho tem como principal objetivo adaptar e analisar os seguintes algoritmos no contexto de formas tridimensionais: *CLARANS* [5] e *Hill Climbing* [3]. As qualidades dos grupos formados foram avaliadas usando técnicas de validação de agrupamento. As medidas utilizadas foram os índices de Rand Ajustado [4] e Silhueta [7]. No contexto de formas, as observações são representadas por matrizes de configuração e a formação dos grupos é feita de acordo com a distância Riemmaniana, uma medida de dissimilaridade adequada para o espaço de formas. Os métodos propostos foram implementados na linguagem de programação R. Os desempenhos dos algoritmos foram analisados por intermédio de experimentos de simulação e aplicações em conjuntos de dados reais pertencentes a literatura da área. Nos experimentos, simulou-se  $n_1$  objetos correspondentes a um grupo e  $n_2$  objetos correspondentes a outro grupo. Os grupos foram

---

<sup>1</sup>inacio.nascimento@ufpe.br

<sup>2</sup>gjaa@de.ufpe.br

definidos por uma distribuição normal multivariada com vetor médio tridimensional de dimensão  $3k$  e uma matriz de covariância  $\Sigma$  de dimensão  $3k \times 3k$ . Considerou-se dois tipos de dispersão entre os marcos: isotropia (os marcos possuem aproximadamente a mesma variabilidade) e anisotropia (os marcos não possuem aproximadamente a mesma variabilidade).

Quanto aos resultados, as análises foram interpretadas com base nos valores obtidos pelos índices de Rand Ajustado e Silhueta. Quanto mais próximo os valores desses índices for de 1, melhor o agrupamento gerado. Para as simulações, os resultados apresentaram indícios de que para os cenários de isotropia e anisotropia com baixa dispersão, os algoritmos *CLARANS* e *Hill Climbing* adaptados para o agrupamento de formas 3D, apresentaram bons desempenhos. Enquanto que para os casos de anisotropia com alta dispersão, os valores medidos pelos índices foram baixos. Nas aplicações em dados reais, os desempenhos dos métodos apresentaram um panorama bem semelhante com os resultados obtidos para os dados simulados.

## Agradecimentos

Os autores agradecem a agência brasileira CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) pelo suporte financeiro.

## Referências

- [1] Amaral, G. J. A., Dore, L. H., Lessa, R. P. and Stosic, B. K-means algorithm in statistical shape analysis, *Communications in Statistics-Simulation and Computation*, 11:1016–1026, 2010. DOI: 10.1080/03610911003765777.
- [2] Dryden, I. L. and Mardia, K. V. S. *Statistical shape analysis: with applications in R, 2a. edição*. John Wiley & Sons, 2016. ISBN:9781119072492.
- [3] Friedman, H. P. and Rubin, J. On some invariant criteria for grouping data, *Journal of the American Statistical Association*, 62:1159–1178, 1967. DOI: 10.1080/01621459.1967.10500923.
- [4] Hubert, L. and Arabie, P. Comparing partitions, *Journal of Classification*, 2:193–218, 1985. DOI: 10.1007/BF01908075.
- [5] Ng, R. T. and Han, J. Clarans: A method for clustering objects for spatial data mining, *IEEE transactions on knowledge and data engineering*, 14:1003–1016, 2002. DOI: 10.1109/TKDE.2002.1033770.
- [6] Vinué, G., Simó, A. and Alemany, S. The k-means algorithm for 3d shapes with an application to apparel design, *Advances in Data Analysis and Classification*, 10:103–132, 2014. DOI: 10.1007/s11634-014-0187-1.
- [7] Rousseeuw, P. J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis, *Journal of Computational and Applied Mathematics*, 20:53–65, 1987. DOI: 10.1016/0377-0427(87)90125-7.