

Algoritmo k-means associado a transformadas na classificação de sinais EEG

Eduardo T. Braun **Cesar R. Rodrigues** **Giovani Baratto** **Alice Kozakevicius**

Universidade Federal de Santa Maria - Grupo de Microeletrônica,

97105-900, Camobi, Santa Maria, RS,

E-mail: {eduardotbraun, cesar.ufsm, Giovani.Baratto, [alice.kozakevicius](mailto:alice.kozakevicius@gmail.com)}@gmail.com,

RESUMO

Neste trabalho estudam-se o algoritmo k-means, suas diferentes variações e sua associação a transformadas de Fourier e wavelets para a classificação de sinais eletroencefalogramas (EEG). O objetivo central dessa classificação é identificar a presença de padrões associados à sonolência em diferentes épocas (intervalos de análise dos sinais EEG).

O algoritmo k-means [6] é amplamente utilizado na identificação de subconjuntos (clusters) com comportamentos similares dentro de um conjunto $\{p_i/p_i \in \mathbb{R}^d, i = 1, 2, \dots, n\}$, com uma quantidade grande (n) de dados. Na verdade, estes subconjuntos (clusters) são formados de acordo com k valores médios arbitrados inicialmente e essas similaridades são obtidas a partir do cálculo das distâncias entre os elementos do conjunto e os k valores médios que representam possíveis centróides para os k clusters.

Uma vez obtida uma configuração para os clusters, os k valores médios são então recalculados, dando início a um processo iterativo no qual as novas distâncias entre os elementos do conjunto inicial e os novos k valores médios de referência são obtidos, permitindo o ajuste dos elementos em cada classe.

A Figura 1 apresenta 4 iterações do algoritmo k-means, considerando o cálculo via distância Euclidiana, sendo o conjunto inicial formado por 3000 pontos aleatoriamente gerados no $[-4, 4] \times [-4, 4] \in \mathbb{R}^2$. Em cada uma das iterações apresentadas na Figura 1 é indicado o novo valor do centróide, recalculado através dos elementos associados a cada um dos clusters, a cada iteração. Diferentes variações para o algoritmo são obtidas cada vez que diferentes métricas são consideradas, como por exemplo a distância de Minkowski considerada em [3]. Além disso, a escolha para o número de clusters k também influencia na dinâmica do algoritmo, permitindo diferentes agrupamentos ao final do processo iterativo.

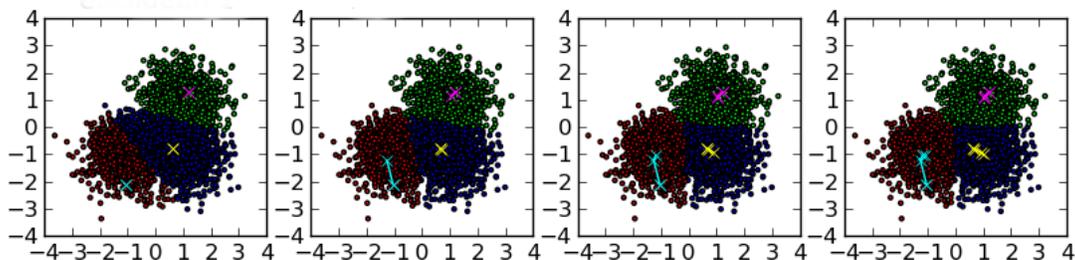


Figura 1: Quatro primeiras iterações do algoritmo k-means, utilizando distância Euclidiana, quando aplicado a uma nuvem de 3000 pontos, $p_i \in \mathbb{R}^2$, distribuídos em 3 clusters ($k = 3$).

Apesar do algoritmo k-means ser amplamente utilizado em diversos tipos de dados [1, 2], ele pode produzir resultados não satisfatórios, quando aplicado a séries temporais [8]. Este baixo desempenho se dá principalmente devido a grande dimensão d de cada elemento do conjunto das séries a serem classificadas, possível alta correlação entre os dados e a grande quantidade de ruído geralmente presente em séries temporais obtidas através de medições associadas a problemas realistas.

Uma maneira de se poder contornar essas situações é aplicar algum tipo de transformada aos dados iniciais, selecionar uma quantidade menor dos coeficientes dessas novas representações e então aplicar o

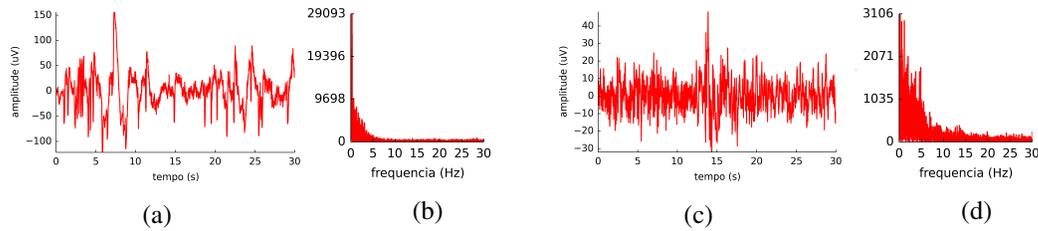


Figura 2: Sinal e espectro de duas épocas retiradas de um mesmo sinal de EEG.

algoritmo k-means a estas novas coleções de coeficientes, como feito em [7, 4, 8].

Neste trabalho, os dados a serem classificados pelo algoritmo k-means são sinais EEG do banco de sinais Physionet[5], sendo considerado apenas o canal F_{pz} para realização dos testes. Na Figura 2 o sinal EEG de um mesmo indivíduo é apresentado em duas diferentes épocas. Em 2(a) está a época associada ao indivíduo acordado. A Figura 2(b) apresenta o espectrograma correspondente. Nas Figuras 2(c) e 2(d) estão os gráficos para o indivíduo dormindo. A transição entre estes dois estados é o foco da análise proposta. Como os sinais EEG possuem uma dimensão extremamente grande, uma primeira abordagem será aplicar transformada de Fourier como em [4], e depois o algoritmo k-means. Neste trabalho a será analisado também o desempenho do algoritmo k-means quando aplicada a transformada wavelet de Haar aos dados iniciais. Como as transformadas wavelets discretas decompõe os sinais em diferentes níveis de resolução e ainda consideram uma informação média de referência, existe aqui um campo fértil de possibilidades de associação desses dados transformados ao algoritmo k-means [7].

Palavras-chave: *K-means, Eletroencefalograma, Sonolência, Transformadas de Fourier e wavelet.*

Referências

- [1] B. Belkhaoui, A. Toumi, A. Khalfallah, A. Khenchaf, and M. S. Bouhlef. Fusion Fourier Descriptors from the E-M, K-Means and Fisher Algorithms for Radar Target Recognition. *International Journal of Computer and Information Technology*, 2(4):603–608, 2013.
- [2] S. Dalmiya, A. Dasgupta, and S. K. Datta. Application of Wavelet based K-means Algorithm in Mammogram Segmentation. *International Journal of Computer Applications*, 52(15):15–19, 2012.
- [3] R. C. de Amorim and B. Mirkin. Minkowski metric, feature weighting and anomalous cluster initializing in k-means clustering. *Pattern Recognition*, 45(3):1061 – 1075, 2012.
- [4] A. Gilbert, S. Muthukrishnan, and M. Strauss. Improved time bounds for near-optimal sparse Fourier representation. *Proc. Wavelets XI SPIE Optics Photonics*, 2005.
- [5] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000 (June 13). <http://circ.ahajournals.org/cgi/content/full/101/23/e215> PMID:1085218;
- [6] J. B. MacQueen. Some methods for classification and analysis of multivariate observations. In L. M. L. Cam and J. Neyman, editors, *Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 281–297. University of California Press, 1967.
- [7] V. V. Shete, A. Charantimath, and D. Bormane. K-complex Detection in Sleep EEG Using Wavelet Transform and Statistical K-means algorithm. *International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering*, 1(4):15–19, 2013.
- [8] M. Vlachos, J. Lin, E. Keogh, and D. Gunopulos. A Wavelet-Based Anytime Algorithm for K-Means Clustering of Time Series. *International Journal of Computer and Information Technology*, 2(4):603–608, 2013.