

Modak, Soumita; Chattopadhyay, Tanuka; Chattopadhyay, Asis Kumar
Unsupervised classification of eclipsing binary light curves through k -medoids clustering.
(English) [Zbl 07481420](#)
J. Appl. Stat. 47, No. 2, 376-392 (2020)

Summary: This paper proposes k -medoids clustering method to reveal the distinct groups of 1318 variable stars in the Galaxy based on their light curves, where each light curve represents the graph of brightness of the star against time. To overcome the deficiencies of subjective traditional classification, we separate the stars more scientifically according to their geometrical configuration and show that our approach outperforms the existing classification schemes in astronomy. It results in two optimum groups of eclipsing binaries corresponding to bright, massive systems and fainter, less massive systems.

MSC:

[62Pxx](#) Applications of statistics

Cited in 1 Document

Keywords:

light curve of variable star; clustering; k -medoids method; complexity invariance distance

Software:

clustervalidation; Silhouettes; AS 136; Binary Maker; clusfind; RobPer; dtw

Full Text: [DOI](#) [arXiv](#)

References:

- [1] Akerlof, C.; Amrose, S.; Balsano, R.; Bloch, J.; Casperson, D.; Fletcher, S.; Gisler, G.; Hills, J.; Kehoe, R.; Lee, B.; Marshall, S.; McKay, T.; Pawl, A.; Schaefer, J.; Szymanski, J.; Wren, J., ROTSE all-sky surveys for variable stars. I. Test Fields, *Astron. J.*, 119, 1901-1913 (2000) · [doi:10.1086/301321](#)
- [2] Batista, G. E.A. P.A.; Keogh, E. J.; Tataw, O. M.; de Souza, V. M.A., CID: an efficient complexity-invariant distance for time series, *Data Mining Knowledge Discovery*, 28, 634-669 (2014) · [Zbl 1294.62188](#) · [doi:10.1007/s10618-013-0312-3](#)
- [3] Bezdek, J. C., *Pattern recognition with fuzzy objective function algorithms* (1981), Plenum Press: Plenum Press, New York · [Zbl 0503.68069](#)
- [4] Bradstreet, D. H.; Steelman, D. P., Binary maker 3.0 - an interactive graphics-based light curve synthesis program written in java, *Am. Astron. Soc. 201st AAS Meeting*, id.75.02; *Bull. Am. Astron. Soc.*, 34, 1224 (2002)
- [5] Caiado, J.; Crato, N., A periodogram-based metric for time series classification, *Comput. Statist. Data Anal.*, 50, 2668-2684 (2006) · [Zbl 1445.62222](#) · [doi:10.1016/j.csda.2005.04.012](#)
- [6] Caiado, J.; Crato, N., Comparison of times series with unequal length in the frequency domain, *Commun. Statist. Simulation Comput.*, 38, 527-540 (2009) · [Zbl 1161.37348](#) · [doi:10.1080/03610910802562716](#)
- [7] Cassisi, C., Montalto, P., Aliotta, M., Cannata, A. and Pulvirenti, A., *Advances in Data Mining Knowledge Discovery and Applications*, Chapter 3: Similarity Measures and Dimensionality Reduction Techniques for Time Series Data Mining, pp. 71-96, Intech, 2012.
- [8] Chattopadhyay, T.; Sinha, A.; Chattopadhyay, A. K., Influence of binary fraction on the fragmentation of young massive clusters – a Monte Carlo simulation, *Astrophys. Space Sci.*, 361, 120-133 (2016) · [doi:10.1007/s10509-016-2705-4](#)
- [9] Dargahi-Noubary, G. R., Discrimination between Gaussian time series based on their spectral differences, *Commun. Statist. Theory Methods*, 21, 2439-2458 (1992) · [Zbl 0800.62527](#) · [doi:10.1080/03610929208830923](#)
- [10] Deb, S.; Singh, H. P., Light curve analysis of variable stars using Fourier decomposition and principal component analysis, *Astron. Astrophys.*, 507, 1729-1737 (2009) · [doi:10.1051/0004-6361/200912851](#)
- [11] Eckner, A., A Framework for the Analysis of Unevenly Spaced Time Series Data. Working Paper. URL: http://eckner.com/papers/unevenly_spaced 2014.
- [12] Eckner, A., Algorithms for Unevenly-spaced time series: Moving averages and other rolling operators. Working Paper. URL: <http://eckner.com/papers/Algorithms%20for%20Unevenly%20Spaced%20Time%20Series.pdf>, 2017.
- [13] Ester, M., Kriegel, H.-P., Sander, J. and Xu, X., A density-based algorithm for discovering clusters in large spatial databases with noise. *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*. AAAI Press, pp. 226-231, 1996.

- [14] Feigelson, E. D.; Babu, G. J., *Statistical Challenges in Modern Astronomy V*, 209 (2013), Springer Science+Business Media: Springer Science+Business Media, New York
- [15] Giorgino, T., Computing and visualizing dynamic time warping alignments in R: The dtw package, *J. Statist. Softw.*, 31, 1-24 (2009) · doi:10.18637/jss.v031.i07
- [16] Graczyk, D.; Soszyński, I.; Poleski, R.; Pietrzyński, G.; Udalski, A.; Szymański, M. K.; Kubiak, M.; Wyrzykowski, L.; Ulaczyk, K., The optical gravitational lensing experiment. The OGLE-III Catalog of Variable Stars. XII. Eclipsing Binary Stars in the Large Magellanic Cloud, *Acta Astronom.*, 61, 103-122 (2011)
- [17] Handl, J.; Knowles, K.; Kell, D., Computational cluster validation in post-genomic data analysis, *Bioinformatics*, 21, 3201-3212 (2005) · doi:10.1093/bioinformatics/bti517
- [18] Hartigan, J. A.; Wong, M. A., A K-means clustering algorithm, *Appl. Statist.*, 28, 100-108 (1979) · Zbl 0447.62062 · doi:10.2307/2346830
- [19] Kaufman, L. and Rousseeuw, P.J., *Finding Groups in Data: An Introduction to Cluster Analysis*, pp. 68-125, John Wiley and Sons, New Jersey, 2005.
- [20] Keogh, E.; Ratanamahatana, C. A., Exact indexing of dynamic time warping, *Knowledge Inform. Syst.*, 7, 358-386 (2005) · doi:10.1007/s10115-004-0154-9
- [21] Kirk, B.; Conroy, K.; Prša, A.; Abdul-Masih, M.; Kochoska, A.; Matijević, G.; Hambleton, K.; Barclay, T.; Bloemen, S.; Boyajian, T.; Doyle, L. R., Kepler eclipsing binary stars. VII. The catalog of eclipsing binaries found in the entire Kepler data set, *Astron. J.*, 151, 68-88 (2016) · doi:10.3847/0004-6256/151/3/68
- [22] Kochoska, A.; Mowlavi, N.; Prša, A.; Lecoœur-Taïbi, I.; Holl, B.; Rimoldini, L.; Süveges, M.; Eyer, L., Gaia eclipsing binary and multiple systems. A study of detectability and classification of eclipsing binaries with Gaia, *Astron. Astrophys.*, 602, A110 (2017) · doi:10.1051/0004-6361/201629957
- [23] Liao, T. W., Clustering of time series data-a survey, *Pattern Recognition*, 38, 1857-1874 (2005) · Zbl 1077.68803 · doi:10.1016/j.patcog.2005.01.025
- [24] Liao, T. W.; Ting, C.; Chang, P.-C., An adaptive genetic clustering method for exploratory minning of feature vector any time series data, *Intl. J. Prod. Res.*, 44, 2731-2748 (2006) · Zbl 1128.62373 · doi:10.1080/00207540600600130
- [25] Lomb, N. R., Least-squares frequency analysis of unequally spaced data, *Astrophys. Space Sci.*, 39, 447-462 (1976) · doi:10.1007/BF00648343
- [26] Maaten, L. V.D., Accelerating t-SNE using tree-based algorithms, *J. Mach. Learning Res.*, 15, 3221-3245 (2014) · Zbl 1319.62134
- [27] Malkov, O.Yu., Oblak, E., Avvakumova, E.A. and Torra, J., Classification of Eclipsing Binaries. Solar and Stellar Physics Through Eclipses, in ASP conference series. Vol. 370. O. Demircan, S. O. Selam and B. Albayrak, eds., 2007.
- [28] Matijević, G.; Prša, A.; Orosz, J. A.; Welsh, W. F.; Bloemen, S.; Barclay, T., Kepler Eclipsing binary stars. III. classification of kepler eclipsing binary light curves with locally linear embedding, *Astron. J.*, 143, 123-128 (2012) · doi:10.1088/0004-6256/143/5/123
- [29] Miller, V. R.; Albrow, M. D.; Afonso, C.; Henning, Th., 1318 new variable stars in a 0.25 square degree region of the Galactic plane, *Astron. Astrophys.*, 519, A12 (2010) · doi:10.1051/0004-6361/200913949
- [30] Modak, S. and Bandyopadhyay, U., A new nonparametric test for two sample multivariate location problem with application to astronomy, *J. Statist. Theory Appl.* 18 (2019), pp.136-146.
- [31] Modak, S.; Chattopadhyay, T.; Chattopadhyay, A. K., Two phase formation of massive elliptical galaxies: study through cross-correlation including spatial effect, *Astrophys. Space Sci.*, 362, 206-215 (2017) · doi:10.1007/s10509-017-3171-3
- [32] Modak, S.; Chattopadhyay, A. K.; Chattopadhyay, T., Clustering of gamma-ray bursts through kernel principal component analysis, *Commun. Statist. Simul. Comput.*, 47, 1088-1102 (2018) · doi:10.1080/03610918.2017.1307393
- [33] Moller-Levet, C. S.; Klawonn, F.; Cho, K.; Wolkenhauer, O., Fuzzy clustering of short time-series and unevenly distributed sampling points, *Adv. Intell. Data Anal. V Lect. Notes Comput. Sci.*, 2810, 330-340 (2003) · doi:10.1007/978-3-540-45231-7_31
- [34] Mowlavi, N.; Lecoœur-Taïbi, I.; Holl, B.; Rimoldini, L.; Barblan, F.; Prsa, A.; Kochoska, A.; Süveges, M.; Eyer, L.; Nienartowicz, K.; Jevardat, G.; Charnas, J.; Guy, L.; Audard, M., Gaia eclipsing binary and multiple systems. Two-Gaussian models applied to OGLE-III eclipsing binary light curves in the large magellanic cloud, *Astron. Astrophys.*, 606, A92 (2017) · doi:10.1051/0004-6361/201730613
- [35] Percy, J. R., *Understanding variable stars* (2007), Cambridge University Press: Cambridge University Press, New York
- [36] Prati, R. C.; Batista, G. E.A. P.A., A complexity-invariant measure based on fractal dimension for time series classification, *Int. J. Natur. Comput. Res.*, 3, 59-73 (2012) · doi:10.4018/jncr.2012070104
- [37] Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, W. T., *Numerical Recipes in C. The Art of Scientific Computing*, 105-128 (1992), Cambridge University Press: Cambridge University Press, Cambridge · Zbl 0845.65001
- [38] Prša, A.; Guinan, E. F.; Devinney, E. J.; DeGeorge, M.; Bradstreet, D. H.; Giammarco, J. M.; Alcock, C. R.; Engle, S. G., Artificial intelligence approach to the determination of physical properties of eclipsing binaries. I. The EBAl project, *Astrophys. J.*, 687, 542-565 (2008) · doi:10.1086/591783
- [39] Rabiner, L.; Juang, B.-H., *Fundamentals of Speech Recognition* (1993), Prentice-Hall, Inc.: Prentice-Hall, Inc., Upper Saddle River, NJ
- [40] Rousseeuw, P. J., Silhouettes: a graphical aid to the interpretation and validation of cluster analysis, *J. Comput. Appl. Math.*, 20, 53-65 (1987) · Zbl 0636.62059 · doi:10.1016/0377-0427(87)90125-7
- [41] Sarro, L. M.; Sánchez-Fernández, C.; Giménez, Á., Automatic classification of eclipsing binaries light curves using neural networks, *Astron. Astrophys.*, 446, 395-402 (2006) · doi:10.1051/0004-6361:20052830
- [42] Scargle, J. D., *Studies in astronomical time series analysis. III - Fourier transforms, autocorrelation functions, and cross-*

- correlation functions of unevenly spaced data, *Astrophys. J.*, 343, 874-887 (1989) · doi:10.1086/167757
- [43] Singh, S.S. and Chauhan, N.C., K-means v/s K-medoids: A comparative study. National Conference on Recent Trends in Engineering And Technology, 2011.
- [44] Soszyński, I.; Udalski, A.; Szymański, M. K.; Wyrzykowski, Ł.; Ulaczyk, K.; Poleski, R.; Pietrukowicz, P.; Kozłowski, S.; Skowron, D. M.; Skowron, J.; Mróz, P.; Pawlak, M., The OGLE collection of variable stars. over 45 000 RR Lyrae stars in the magellanic system, *Acta Astron.*, 66, 131-147 (2016)
- [45] Stefan, A.; Athitsos, V.; Das, G., The Move-Split-Merge metric for time series, *IEEE Trans. Knowledge and Data Eng.*, 25, 1425-1438 (2013) · doi:10.1109/TKDE.2012.88
- [46] Street, R. A.; Christian, D. J.; Clarkson, W. I.; Collier Cameron, A.; Evans, N.; Fitzsimmons, A.; Haswell, C. A.; Hellier, C.; Hodgkin, S. T.; Horne, K.; Kane, S. R.; Keenan, F. P.; Lister, T. A.; Norton, A. J.; Pollacco, D.; Ryans, R.; Skillen, I.; West, R. G.; Wheatley, P. J., Status of superWASP I (La Palma), *Astron. Nachrichten.*, 325, 565-567 (2004) · doi:10.1002/asna.200410281
- [47] Süveges, M.; Barblan, F.; Lecoœur-Taïbi, I.; Prša, A.; Holl, B.; Eyer, L.; Kochoska, A.; Mowlavi, N.; Rimoldini, L., Gaia eclipsing binary and multiple systems. supervised classification and self-organizing maps, *Astron. Astrophys.*, 603, A117 (2017) · doi:10.1051/0004-6361/201629710
- [48] Thieler, A. M.; Backes, M.; Fried, R.; Rhode, W., Periodicity detection in irregularly sampled light curves by robust regression and outlier detection, *Statist. Anal. Data Mining.*, 6, 73-89 (2013) · Zbl 07260353 · doi:10.1002/sam.11178
- [49] Thieler, A. M.; Fried, R.; Rathjens, J., RobPer: An R package to calculate periodograms for light curves based on robust regression, *J. Statist. Softw.*, 69, 1-36 (2016) · doi:10.18637/jss.v069.i09
- [50] Wei, Y., Multi-dimensional time warping based on complexity invariance and its application in sports evaluation. 11th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), IEEE. pp. 677-680, 2014.

This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.