Rizoiu, Marian-Andrei; **Velcin, Julien**; **Bonnevay, Stéphane**; **Lallich, Stéphane**

ClusPath: a temporal-driven clustering to infer typical evolution paths. (English)

Zbl 1416.62356

Data Min. Knowl. Discov. 30, No. 5, 1324-1349 (2016).

Summary: We propose ClusPath, a novel algorithm for detecting general evolution tendencies in a population of entities. We show how abstract notions, such as the Swedish socio-economical model (in a political dataset) or the companies fiscal optimization (in an economical dataset) can be inferred from low-level descriptive features. Such high-level regularities in the evolution of entities are detected by combining spatial and temporal features into a spatio-temporal dissimilarity measure and using semi-supervised clustering techniques. The relations between the evolution phases are modeled using a graph structure, inferred simultaneously with the partition, by using a "slow changing world" assumption. The idea is to ensure a smooth passage for entities along their evolution paths, which catches the long-term trends in the dataset. Additionally, we also provide a method, based on an evolutionary algorithm, to tune the parameters of ClusPath to new, unseen datasets. This method assesses the fitness of a solution using four opposed quality measures and proposes a balanced compromise.

MSC:

62H30 Classification and discrimination; cluster analysis (statistical aspects)
62M10 Time series, auto-correlation, regression, etc. in statistics (GARCH)
62P20 Applications of statistics to economics
91B84 Economic time series analysis

Keywords:

detection of long-term trends; evolutionary clustering; temporal clustering; temporal cluster graph; semi-supervised clustering; Pareto front estimation

Software:

SPEA2; ClusPath

Full Text: DOI arXiv

References:

[1] Araujo R, Kamel MS (2014) Semi-supervised kernel-based temporal clustering. In: International conference on machine learning and applications, IEEE, ICMLA '14, pp 123-128

[2] Armingeon K., Isler C, Laura Knöpfel DW, Engler S (2011) Comparative political data set 1960-2009. University of Berne

[3] Chakrabarti D, Kumar R, Tomkins A (2006) Evolutionary clustering. In: International conference on knowledge discovery and data mining, ACM, SIGKDD '06, pp 554-560

[4] Chi Y, Song X, Zhou D, Hino K, Tseng BL (2007) Evolutionary Spectral Clustering by Incorporating Temporal Smoothness. In: International Conference on Knowledge Discovery and Data Mining (KDD), San Jose, USA, pp 153-162

[5] De la Torre F, Agell C (2007) Multimodal diaries. In: Multimedia and expo, IEEE, pp 839-842

[6] Smet, Y.; Eppe, S., Multicriteria relational clustering: the case of binary outranking matrices, Evol Multi-Criterion Optim, 5467, 380-392, (2009)· doi:10.1007/978-3-642-01020-0$_3$1

[7] Deb, K.; Pratap, A.; Agarwal, S.; Meyarivan, T., A fast and elitist multiobjective genetic algorithm: NSGA-II, Evol Comput, 6, 182-197, (2002)· doi:10.1109/4235.996017

[8] Dunn, JC, A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters, J Cybern, 3, 32-57, (1973) · Zbl 0291.68033 · doi:10.1080/01969727308546046

[9] Erixon L (2000) A Swedish economic policy: the theory, application and validity of the Rehn-Meidner model. Tech. rep., Department of Economics, Stockholm University

[10] Gaffney S, Smyth P (1999) Trajectory clustering with mixtures of regression models. In: International conference on knowledge discovery and data mining, ACM Press, New York, USA, SIGKDD '99, pp 63-72. doi:10.1145/312129.312198

[11] Halsall-Whitney, H.; Thibault, J., Multi-objective optimization for chemical processes and controller design: approximating and classifying the pareto domain, Comput Chem Eng, 30, 1155-1168, (2006)· doi:10.1016/j.compchemeng.2006.02.010

[12] Kafafy A, Bounekkar A, Bonnevay S (2011) A hybrid evolutionary metaheuristics (HEMH) applied on 0/1 multiobjective knapsack problems. In: Genetic and evolutionary computation, ACM Press, New York, USA, GECCO '11, p 497

[13] Kalnis P, Mamoulis N, Bakiras S (2005) On discovering moving clusters in spatio-temporal data, chap. 21. In: Bauzer Medeiros C, Egenhofer M, Bertino E (eds) Advances in spatial and temporal databases, vol 3633., Lecture notes in computer science, Springer, Berlin, pp 364-381

[14] Liang Z, Tomioka R, Murata H, Asaoka R, Yamanishi K (2013) Quantitative prediction of glaucomatous visual field loss from few measurements. In: International conference on data mining, ICDM '13, pp 1121-1126

[15] Lin WH, Hauptmann A (2006) Structuring continuous video recordings of everyday life using time-constrained clustering. In: Chang EY, Hanjalic A, Sebe N (eds) Multimedia content analysis, management, and retrieval, pp 60730D-60730D-9

[16] MacQueen J (1967) Some methods for classification and analysis of multivariate observations. In: Berkeley symposium on mathematical statistics and probability, vol 1, pp 281-297 · Zbl 0214.46201

[17] Mihăiţă AS, Camargo M, Lhoste P (2014) Optimization of a complex urban intersection using discrete event simulation and evolutionary algorithms. In: International federation of automatic control, IFAC'14, vol 19, pp 8768-8774

[18] Rizoiu MA, Velcin J, Lallich S (2012) Structuring typical evolutions using temporal-driven constrained clustering. In: International conference on tools with artificial intelligence, ICTAI '12, vol 1, IEEE, Athens, Greece, pp 610-617

[19] Rizoiu, MA; Velcin, J.; Lallich, S., How to use temporal-driven constrained clustering to detect typical evolutions, Int J Artif Intell Tools, 23, 1460,013, (2014)· doi:10.1142/S0218213014600136

[20] Rocha, C.; Dias, LC; Dimas, I., Multicriteria classification with unknown categories: a clustering-sorting approach and an application to conflict management, J Multi-Criteria Decis Anal, 20, 13-27, (2013)· doi:10.1002/mcda.1476

[21] Sawaragi Y, Nakayama H, Tanino T (1985) Theory of multiobjective optimization, vol 176. Academic Press, New York · Zbl 0566.90053

[22] Siddiqui ZF, Oliveira M, Gama J, Spiliopoulou M (2012) Where are we going? Predicting the evolution of individuals. In: Hollmén J, Klawonn F, Tucker A (eds) Advances in intelligent data analysis V, vol 7619. Lecture notes in computer science, Springer, Berlin, pp 357-368

[23] Wagstaff K, Cardie C, Rogers S, Schroedl S (2001) Constrained K-means clustering with background knowledge. In: International conference on machine learning, ICML '01, pp 577-584

[24] Xu, T.; Zhang, Z.; Yu, PS; Long, B., Generative models for evolutionary clustering, ACM Trans Knowl Discov Data (TKDD), 6, 7, (2012)

[25] Zitzler E, Laumanns M, Thiele L (2001) SPEA2: improving the strength Pareto evolutionary algorithm. In: Evolutionary methods for design, optimisation and control with applications to industrial problems, EUROGEN '01, pp 95-100