

Calò, Daniela Giovanna

Italian contributions on some recent research topics in cluster analysis. (English)

Zbl 1453.62539

Statistica 72, No. 3, 271-286 (2012).

Summary: The paper presents a selective view of the issues that are attracting the interest of Italian statisticians working on clustering methods and applications. It does not aim at providing a comprehensive overview of the wealth of methods developed in Italy on the selected topics: indeed, it focuses on methods dealing with quantitative data and, in this context, only on the most recent literature. The fil rouge is given by the developments which have been inspired in quantitative data clustering by the complex nature of the data nowadays arising in a broad range of applications.

MSC:

62H30 Classification and discrimination; cluster analysis (statistical aspects)

Full Text: [DOI](#)

References:

- [1] G. ADELFIGIO, M. CHIODI, A. D'ALESSANDRO, D. LUZIO, (2010), Clustering of waveforms-data based on FPCA direction, in "Proceedings of COMPSTAT 2010", Physica-Verlag.
- [2] M. ALFÒ, L. NIEDDU, D. VICARI, (2009), Finite mixture models for mapping spatially dependent disease counts, "Biometrical Journal", 51, pp. 84-97. · [Zbl 1442.62232](#)
- [3] A.C. ATKINSON, M. RIANI, (2007), Exploratory tools for clustering multivariate data, "Computational Statistics and Data Analysis", 52, pp. 272-285. · [Zbl 1452.62028](#)
- [4] A.C. ATKINSON, M. RIANI, A. CERIOLI (2010), The Forward Search: theory and data analysis, "Journal of the Korean Statistical Society", 39, pp. 117-134. · [Zbl 1294.62149](#)
- [5] L. AUGUGLIARO, A. MINEO, (2011), Plaid model for microarray data: an enhancement of the pruning step, in B. Fichet et al. (eds.) "Classification and multivariate analysis for complex data structures", pp. 447-456. Springer, Heidelberg.
- [6] J. BAEK, G.J. MCLACHLAN, L. FLACK, (2010), Mixtures of factor analyzers with common factor loadings: applications to the clustering and visualisation of high-dimensional data, "IEEE Transactions on Pattern Analysis and Machine intelligence", 32, pp. 1298-1309.
- [7] S. BALBI, R. MIELE, G. SCEPI, (2010), Clustering of documents from a two-way viewpoint, in "JADT 2010: 10 th international Conference on Statistical Analysis of Textual Data".
- [8] A. BALZANELLA, Y. LECHEVALLIER, R. VERDE, (2011), Clustering multiple data streams, in S. Ingrassia, et al. (eds.) "New Perspectives in Statistical Modeling and Data Analysis", Springer.
- [9] J.D. BANFIELD, A.E. RAFTERY, (1993), Model-based Gaussian and non-Gaussian clustering, "Biometrics", 49, pp. 803-821. · [Zbl 0794.62034](#)
- [10] R. BARAGONA, (2010), Dissimilarity indexes for clustering multivariate time series, available at <http://w3.uniroma1.it/statstsmeh/download/diss>
- [11] R. BARAGONA, F. BATTAGLIA, I. POLI, (2011), Evolutionary Statistical Procedures, Springer-Verlag, Heidelberg. · [Zbl 1378.62005](#)
- [12] F. BARTOLUCCI, (2005), Clustering univariate observations via mixtures of unimodal normal mixtures, "Journal of Classification", 22, pp. 203-219. · [Zbl 1336.62170](#)
- [13] J.P. BAUDRY, A.E. RAFTERY, G. CELEUX, K. LO, R. GOTTARDO, (2010), Combining mixture components for clustering, "Journal of Computational and Graphical Statistics", 19, pp. 332-353.
- [14] D.G. CALÒ, C. VIROLI, (2010), A dimensionally reduced finite mixture model for multilevel data, "Journal of Multivariate Analysis", 101, pp. 2543-2553. · [Zbl 1198.62063](#)
- [15] A. CERIOLI, (2010), Multivariate outlier detection with high-breakdown estimators, "Journal of the American Statistical Association", 105, pp. 147-156. · [Zbl 1397.62167](#)
- [16] R. COPPI, P. D'URSO, P. GIORDANI, (2010), A fuzzy clustering model for multivariate spatial time series, "Journal of Classification", 27, pp. 54-88. · [Zbl 1337.62305](#)
- [17] R. COPPI, P. D'URSO, P. GIORDANI, (2011), Fuzzy and possibilistic clustering for fuzzy data, "Computational Statistics & Data Analysis", doi: 10.1016/j.csda.2010.09.013.
- [18] M. CORDUAS, (2010), Mining time series data: a selective survey, in F. Palumbo et al. (eds.) "Data Analysis and Classification", pp. 355-362. Springer, Heidelberg.

- [19] M. CORDUAS, D. PICCOLO, (2008), Time series clustering and classification by the autoregressive metric, “Computational Statistics & Data Analysis”, 52, pp. 4685-4698. · [Zbl 1452.62624](#)
- [20] P. CORETTO, C. HENNIG, (2010), A simulation study to compare robust clustering methods based on mixtures, “Advances in Data Analysis and Classification”, 4, pp. 111-135. · [Zbl 1284.62366](#)
- [21] P. CORETTO, C. HENNIG, (2011), Maximum likelihood estimation of heterogeneous mixtures of Gaussian and uniform distributions, “Journal of Statistical Planning and inference”, 141, pp. 462-473. · [Zbl 1203.62017](#)
- [22] L. DE ANGELIS, (2011), The multidimensional measurement of poverty: a longitudinal analysis, in “JOCLAD2011 - Book of Abstract”, pp. 49-52.
- [23] A. DE GREGORIO, S.M. IACUS, (2010), Clustering of discretely observed diffusion processes, “Computational Statistics & Data Analysis”, 54, pp. 598-606. · [Zbl 1464.62056](#)
- [24] T. DI BATTISTA, S.A. GATTONE, A. DE SANCTIS, (2011), Dealing with FDA estimation methods, in S. Ingrassia, et al. (eds.) “New Perspectives in Statistical Modeling and Data Analysis”, Springer.
- [25] E. DIDAY, M. NOIRHOMME, (2008), Symbolic Data Analysis, Wiley, New York. · [Zbl 1275.62029](#)
- [26] P. D’URSO, (2000), Dissimilarity measures for time trajectories, “Statistical Methods & Applications”, pp. 53-83. · [Zbl 1454.62258](#)
- [27] P. D’URSO, E.A. MAHARAJ, (2009), Autocorrelation-based fuzzy clustering of time series, “Fuzzy Sets and Systems”, 160, pp. 3565-3589.
- [28] P. D’URSO, E.A. MAHARAJ, Wavelet-based clustering of multivariate time series, “Fuzzy Sets and Systems”, in press.
- [29] A. FARCOMENI, (2009), Robust double clustering, “Journal of Classification”, 26, pp. 77-101. · [Zbl 1276.62040](#)
- [30] G. GALIMBERTI, A. MONTANARI, C. VIROLI, (2008), Penalized factor mixture analysis for variable selection in clustered data, “Computational Statistics & Data Analysis”, 53, pp. 4301-4310. · [Zbl 1453.62094](#)
- [31] G. GALIMBERTI, G. SOFFRITTI, (2009), Discovering multidimensional unobserved heterogeneity through model-based cluster analysis, available at <http://www.statssa.gov.za/isi2009/ScientificProgramme/IPMS/0120.pdf>.
- [32] L.A. GARCÍA-ESCUADERO, A. GORDALIZA, C. MATRÁN, A. MAYO-ISCAR, (2010), A review of robust clustering methods, “Advances in Data Analysis and Classification”, 4, pp. 89-109.
- [33] N. GERSHENFELD, B. SCHONER, F. METOIS, (1999), Cluster-weighted modelling for time-series analysis, “Advances in Data Analysis and Classification”, 397, pp. 329-332.
- [34] F. GIORDANO, M. LA ROCCA, M.L. PARRELLA, (2011), Clustering complex time series databases, in B. Fichet et al. (eds.) “Classification and multivariate analysis for complex data structures”, pp. 417-426. Springer, Heidelberg.
- [35] F. GRESELIN, S. INGRASSIA, (2010), Constrained monotone EM algorithms for mixtures of multivariate t distributions, “Statistics and Computing”, 20, pp. 9-22.
- [36] D.J. HAND, (2009), Modern statistics: the myth and the magic, “Journal of the Royal Statistical Society”, A, 172, pp. 287-306.
- [37] C. HENNIG, (2004), Breakdown points for maximum likelihood-estimators of location-scale mixtures, “Annals of Statistics”, 32, pp. 1313-1340. · [Zbl 1047.62063](#)
- [38] C. HENNIG, (2010), Methods for merging Gaussian mixture components, “Advances in Data Analysis and Classification”, 4, pp. 3-34. · [Zbl 1306.62141](#)
- [39] S. INGRASSIA, R. ROCCI, (2007), Constrained monotone EM algorithms for finite mixture of multivariate Gaussians, “Computational Statistics & Data Analysis”, 51, pp. 5339-5351. · [Zbl 1445.62116](#)
- [40] S. INGRASSIA, R. ROCCI, (2011), Degeneracy of the EM algorithm for the MLE of multivariate Gaussian mixtures and dynamic constraints, “Computational Statistics & Data Analysis”, 55, pp. 1715-1725. · [Zbl 1328.65030](#)
- [41] S. INGRASSIA, C. MINOTTI, G. VITTADINI, (2010), Cluster Weighted Modelling wit Student-t components, available at http://homes.stat.unipd.it/mgri/SIS2010/Program/8-SSVIII_Cladag/881-1507-1-RV.pdf.
- [42] A. IODICE D’ENZA, F. PALUMBO, M. GREENACRE, (2008), Exploratory data analysis leading towards the most interesting simple association rules, “Computational Statistics & Data Analysis”, 52, pp. 3269-3281. · [Zbl 1452.62051](#)
- [43] A. IRPINO, R. VERDE, (2008), Dynamic clustering of interval data using a Wasserstein-based distance, “Pattern Recognition Letters”, 29, pp. 1648-1658. · [Zbl 1147.62054](#)
- [44] T.I. LIN, (2009), Maximum likelihood estimation for multivariate skew normal mixture models, “Journal of Multivariate Analysis”, 100, pp. 257-265. · [Zbl 1152.62034](#)
- [45] E.A. MAHARAJ, P. D’URSO, (2011), Fuzzy clustering of time series in the frequency domain, “Information Sciences”, 181, pp. 1187-1211. · [Zbl 1215.62061](#)
- [46] E.A. MAHARAJ, P. D’URSO, D.U.A. GALAGEDERA, (2010), Wavelet-based fuzzy clustering of time series, “Journal of Classification”, 27, pp. 231-275. · [Zbl 1337.62307](#)
- [47] F. MARTELLA, M. ALFÒ, M. VICHI, (2010), Biclustering of gene expression data by an extension of mixtures of factor analyzers, “The international Journal of Biostatistics”, 4, doi: 10.2202/1557-4679.1078.
- [48] A. MARUOTTI, R. ROCCI, (2010), A semiparametric approach to mixed non-homogeneous hidden Markov models, available at http://homes.stat.unipd.it/mgri/SIS2010/Program/6-SVI_Vicari/851-1532-1-DR.pdf.
- [49] A. MARUOTTI, T. RYDEN, (2009), A semiparametric approach to hidden Markov models under longitudinal observations, “Statistics and Computing”, 19, pp. 381-393.
- [50] C. MAUGIS, G. CELEUX, M.L. MARTIN-MAGNIETTE, (2009), Variable selection for clustering with Gaussian mixture

- models, "Biometrics", 65, pp. 701-709. · [Zbl 1172.62021](#)
- [51] G.J. MCLACHLAN, R.W. BEAN, L. BEN-TOVIM JONES, (2007), Extension of the mixture of factor analyzers model to incorporate the multivariate t-distribution, "Computational Statistics & Data Analysis", 51, pp. 5327-5338. · [Zbl 1445.62053](#)
- [52] A. MONTANARI, C. VIROLI, (2010), Heteroscedastic factor mixture analysis, "Statistical Modelling", 10, pp. 441-460. · [Zbl 07256833](#)
- [53] A. MONTANARI, C. VIROLI, (2010), The independent factor analysis approach to latent variable modeling, "Statistics", 44, pp. 397-416. · [Zbl 1283.62125](#)
- [54] I. MORLINI, (2007), Searching for structure in measurements of air pollutant concentration, "Environmetrics", 18, pp. 823-840.
- [55] I. MORLINI, S. ZANI, (2010), A dissimilarity measure between two hierarchical clusterings, in "CLADAG 2010- Book of Abstract", pp. 219-210.
- [56] E. OTRANTO, (2008), Clustering heteroscedastic time series by model-based procedures, "Computational Statistics & Data Analysis", 52, pp. 4685-4698. · [Zbl 1452.62784](#)
- [57] E. OTRANTO, (2010), Identifying financial time series with similar dynamical conditional correlation, "Computational Statistics & Data Analysis", 54, pp. 1-15. · [Zbl 1284.91593](#)
- [58] F. PALUMBO, D. VISTOCCO, A. MORINEAU, (2008), Huge multidimensional data visualization: back to the virtue of principal coordinates and dendrograms in the new computer age, in C. Chun-Houh et al. (eds.) "Handbook of Data Visualization", pp. 349-387. Springer, Heidelberg. · [Zbl 1147.68464](#)
- [59] D. PEEL, G. MCLACHLAN, (2000), Robust mixture modeling using the t-distribution, "Statistics and Computing", 10, pp. 339-348.
- [60] D. PICCOLO, (1990), A distance measure for classifying ARIMA models, "Journal of Time Series Analysis", 11, pp. 153-164. · [Zbl 0691.62083](#)
- [61] D. PIGOLI, L.M. SANGALLI, (2010), Wavelet smoothing for curves in more than one dimension, available at <http://homes.stat.unipd.it/mgri/SIS2010/1440-1-DR.pdf>.
- [62] A.E. RAFTERY, N. DEAN, (2006), Variable selection for model-based cluster analysis, "Journal of the American Statistical Association", 101, pp. 168-178. · [Zbl 1118.62339](#)
- [63] M. RIANI, A.C. ATKINSON, A. CEROLI, (2009), Finding an unknown number of multivariate outliers, "Journal of the Royal Statistical Society B", B, 71, pp. 447-466. · [Zbl 1248.62091](#)
- [64] R. ROCCI, (2010), Mixing mixtures of Gaussians, GfKI-CLADAG 2010 Book of Abstracts, pp. 27-28.
- [65] R. ROCCI, M. VICHI, (2005), Three-mode component analysis with crisp or fuzzy partition of units, "Psychometrika", 70, pp. 715-736. · [Zbl 1306.62491](#)
- [66] R. ROCCI, M. VICHI, (2010), Two-mode multi-partitioning, "Computational Statistics & Data Analysis", 52, pp. 1984-2003. · [Zbl 1452.62463](#)
- [67] E. ROMANO, A. BALZANELLA, R. VERDE, (2010), A new regionalization method for spatially dependent functional data based on local variogram models: an application on environmental data, available at http://homes.stat.unipd.it/mgri/SIS2010/Program/16-SSXVI_Dibattista/906-1575-1-RV.pdf. · [Zbl 1300.62039](#)
- [68] P.J. ROUSSEEUW, K. VAN DRIESEN, (1999), A fast algorithm for the minimum covariance determinant estimator, "Technometrics", 41, pp. 212-223.
- [69] L.M. SANGALLI, P. SECCHI, S. VATINI, V. VITELLI, (2010), k-mean alignment for curve clustering, "Computational Statistics & Data Analysis", 54, pp. 1219-1233. · [Zbl 1464.62153](#)
- [70] L. SCRUCCA, (2010), Genetic algorithms for subset selection in model-based clustering, available at <http://homes.stat.unipd.it/mgri/SIS2010/Program/1296-1-DR.pdf>.
- [71] L. SCRUCCA, (2010), Dimension reduction for model-based clustering, "Statistics and Computing", 20, pp. 471-484.
- [72] I. VAN MECHELEN, H.-H. BOCK, P. DE BOECK, (2004), Two-mode clustering methods: a structured overview, "Statistical Methods in Medical Research", 13, pp. 363-394. · [Zbl 1053.62078](#)
- [73] R. VERDE, A. IRPINO, (2008), Comparing histogram data using a Mahalanobis Wasserstein distance, in P. Brito (ed.), "COMPSTAT 2008", pp. 77-89. PhysicaVerlag, Berlin. · [Zbl 1147.62054](#)
- [74] J.K. VERMUNT, B. TRAN, J. MAGIDSON, (2008), Latent class models in longitudinal research, in S. Menard (ed.), "Handbook of Longitudinal Research: Design, Measurement, and Analysis", pp. 373-385. Burlington, MA.
- [75] D. VICARI, M. ALFÖ, (2010), Clustering discrete choice data, in Y. LECHEVALLIER, G. SAPORTA (eds.) Proceedings of COMPSTAT2010, pp. 369-378. Physica-Verlag, Heidelberg. · [Zbl 1436.62293](#)
- [76] M. VICHI, (2000), Double k-means clustering for simultaneous classification of objects and variables, in S. Borra et al. (eds.), "Advances in Classification and Data Analysis", pp. 43-52. Springer, Berlin.
- [77] M. VICHI, (2010), Clustering longitudinal multivariate observations, Personal communication, <http://sfc2010.univ-reunion.fr/sfc2010/images/stories/1381-1-DR.pdf>. · [Zbl 1381.62185](#)
- [78] M. VICHI, H.A.L. KIERS, (2001), Factorial k-means analysis for two-way data, "Computational Statistics & Data Analysis", 37, pp. 49-64. · [Zbl 1051.62056](#)
- [79] D. VICARI, M. VICHI, (2009), Structural classification analysis of three-way dissimilarity data, "Journal of Classification", 26, pp. 121-154. · [Zbl 1337.62140](#)
- [80] M. VICHI, G. SAPORTA, (2009), Clustering and disjoint principal component analysis, "Computational Statistics & Data Analysis", 53, pp. 3194-3208. · [Zbl 1453.62230](#)

- [81] C. VIROLI, (2010), Dimensionally reduced model-based clustering through mixtures of factor mixture analyzers, "Journal of Classification", 27, pp. 363-388. · [Zbl 1337.62141](#)
- [82] K. WANG, S.-K. NG, G.J. MCLACHLAN, (2010), Multivariate Skew-t Mixture Models, in "DICTA '09", doi:10.1109/DICTA.2009.88.
- [83] M.S. YANG, K.L. WU, (2006), Unsupervised possibilistic clustering, "Pattern Recognition", 39, pp. 5-21.

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.