AN OVERVIEW OF CAPABILITIES FOR CLUSTER ANALYSIS OF DATA FOUND IN THE STATISTICA ADVANCED SOFTWARE PACKAGE

I.A. Tikhonov quiteman@mail.ru
SPIN-код: 5941-2719

Bauman Moscow State Technical University, Moscow, Russian Federation

Abstract

The article reviews data clustering capabilities of the STATISTICA software package. We describe the clustering methods found in this product and the specifics of working with them from a practical standpoint. We consider the concept of distance measure between elements of the initial set and certain methods of clustering the initial set of observations, as well as cluster analysis results produced by the algorithms implemented in the STATISTICA Advanced package. There is no doubt that cluster analysis of data is highly relevant and pertinent at present, since data and data analysis results play an increasingly significant role in the information society of today, and clustering provides a better understanding of these data.

Keywords

Data analysis, cluster analysis, clustering, unsupervised classification, STATISTICA, Euclidean space, hierarchical and non-hierarchical clustering methods, joining/tree clustering, k-mean clustering, two-way joining

References

[1] Hartigan J.A. Clustering algorithms. John Wiley & Sons, Inc., 1975. 369 p.

[2] Duran B., Odell P. Cluster analysis. A survey. Springer-Verlag. Berlin – Heidelberg – N.Y. 1974.

[3] Barsegyan, A.A., M.S. Kupriyanov, I.I. Kholod, M.D. Tess, S.I. Elizarov. Analiz dannykh i protsessov [Analysis of data and processes]. St. Petersburg, Peterburg Publ., 2009. 512 p.

[4] Kalinina V.N., Solov'ev V.I. Vvedenie v mnogomernyy statisticheskiy analiz [Introduction to multivariate statistical analysis]. Moscow, GUU Publ., 2003. 66 p.

[5] Ayvazyan A.A., Bukhshtaber V.M., Enyukov I.S., Meshalkin L.D. Prikladnaya statistika: Klassifikatsii i snizhenie razmernosti [Applied statistics: Classification and reduction of dimensionality]. Moscow, Finansy i statistika Publ., 1989. 607 p.

[6] Metody klasternogo analiza. Ierarkhicheskie metody. Available at: http://www.intuit.ru/studies/courses/6/6/lecture/182?page=2 (accessed 10.10.2017).

[7] Obzor STATISTICA. Available at: http://statsoft.ru/products/overview/#advantages (accessed 25.09.2017).

[8] Data clustering: A Review A.K. Jain Michigan State University M.N. Murty. *Indian Institute of Science AND P.J. FLYNN The Ohio State University. ACM Computing Surveys.* Vol. 31, no. 3, September 1999. Available at: http://users.eecs.northwestern.edu/~yingliu/datamining_papers/survey.pdf (accessed 26.09.2017).

[9] Alex Smola and S.V.N. Vishwanathan. Introduction to machine learning. Cambridge University Press, 2008. 234 p.

[10] Bureeva N.N. Mnogomernyy statisticheskiy analiz s ispol'zovaniem PPP “STATISTICA”. Uchebno-metodicheskiy material po programme povysheniya kvalifikatsii “Primenenie programmnykh sredstv v nauchnykh issledovaniyakh i prepodavanii matematiki i mekhaniki”. Nizhniy Novgorod, 2007. 112 p (in Russ.).

Tikhonov I.A. — Master's Degree student, Department of Information Processing and Control Systems, Bauman Moscow State Technical University, Moscow, Russian Federation.

Scientific advisor — M.V. Vinogradova, Cand. Sc. (Eng.), Assoc. Professor, Department of Information Processing and Control Systems, Bauman Moscow State Technical University, Moscow, Russian Federation.