

CLUSTERING OF EUROPEAN COUNTRIES BY AN INFLATION RATE AND CLUSTERS RESEARCH

ALEKSEJS LOZKINS

Faculty of Applied Mathematics and Control Processes, Saint Petersburg State University
Universitetskii prospekt 35, 198504 Petergof, Saint-Petersburg, Russia

Supervisor: D. Sc. Bure V. M., Professor of Department of Mathematical Game Theory and Statistical Decisions

E-mail: hippy92@yandex.ru

In analysis of socioeconomic information there are often situations when initial numerical data contain various kinds of inaccuracies (measurement error, intentional misrepresentation, error of calculation, error of mathematical model and other possible sources of error). In this regard, it is important to choose those classifications (classification methods) of study objects which have the stability property with respect to random noise.

The current work represents clustering analysis of European countries by an inflation rate as a demonstration of study on clusters stability. The following methods are being considered: hierarchical, k-means and c-means clustering. A stability criterion is introduced and a study on stability of clustering methods with different parameters is carried out.

The work is aimed to get a reliable and accurate clustering which is stable with respect to random perturbations. There are two cases of random perturbations being examined: perturbations generated by truncated normal distribution and perturbations generated by uniform distribution. Consideration of the two different distributions allows more objective assessment of the stability of clustering methods.

REFERENCES

- [1] G. James et al. *Unsupervised Learning // An Introduction to Statistical Learning*. Springer New York, 2013, 373–418.
- [2] M. Blaug. *Phillips curve // Economic Theory in Retrospect*. Cambridge University Press, 1997, 678–686.
- [3] V. M. Bure, E. M. Parilina. *Theory of Probability and Mathematical Statistics*. Lan Saint-Petesburg, 2013, 416. (in Russian)
- [4] Trading Economics [e-resource]: URL: <http://tradingeconomics.com/country-list/inflation-rate-mom> (date of treatment: 08.03.2015).
- [5] L. L. McQuitty. Similarity analysis by reciprocal pairs for discrete and continuous data. *Educational and Psychological measurement*, **26** (4):825–831, 1996.
- [6] J. A. Hartigan, M. A. Wong. Algorithm AS 136: A k-means clustering algorithm. *Applied statistics*, 100–108, 1979.
- [7] N. R. Pal, J. C. Bezdek, R. J. Hathaway. Sequential competitive learning and the fuzzy c-means clustering algorithms. *Neural Networks*, **9** (5):787–796, 1996.
- [8] C. Garcia-Osorio, C. Fyfe. Visualization of high-dimensional data via orthogonal curves. *Journal of Universal Computer Science*, **11** (11):1806–1819, 2005.