
Clustering Algorithms for Storage of Tick Data

Gabor I. Nagy* and Krisztian Buza

Budapest University of Technology and Economics
Magyar tudósok körútja 2, H-1117 Budapest, Hungary
gnagy@tmit.bme.hu, buza@cs.bme.hu

Abstract. Tick data is one of the most prominent types of temporal data, as it can be used to represent data in various domains such as geophysics or finance. Storage of tick data is a challenging problem because two criteria have to be fulfilled simultaneously: the storage structure should allow fast execution of queries and the data should not occupy too much space on the hard disk or in the main memory. We present two clustering-based solutions, in particular, our recently-developed clustering algorithms, SOHAC and SOPAC. These algorithms are designed to support the storage of tick data and are under publication (see References). We evaluate our algorithms both on publicly available real-world datasets, as well as real-world tick data from the financial domain provided by one of the world-wide most renowned investment bank. In our experiments, we compare our approaches, SOHAC and SOPAC, against a large collection of conventional clustering algorithms from the literature. The experiments show that our algorithm substantially outperforms – both in terms of statistical significance and practical relevance – the examined clustering algorithms for the tick data storage problem. Additionally, we present our most recent research directions related to clustering algorithms for tick data storage.

References

- NAGY, G.I. and BUZA, K. (2012): Partitional Clustering of Tick Data to Reduce Storage Space. *IEEE 16th International Conference on Intelligent Engineering Systems*, to appear.
- NAGY, G.I. and BUZA, K. (2012): Efficient Storage of Tick Data That Supports Search and Analysis *12th Industrial Conference on Data Mining, LNCS*, Springer, to appear.

Keywords

TICK DATA, CLUSTERING, STORAGE, APPLICATION, FINANCE

* The first author is PhD-student