Multiple-Resolution Stream Clustering Using Graph Maintenance

Marwan Hassani, Pascal Spaus, and Thomas Seidl

Data Management and Data Exploration Group
RWTH Aachen University, Germany
`{hassani,spaus,seidl}@cs.rwth-aachen.de`

Abstract. Challenges for clustering streaming data are getting continuously more sophisticated. They are driven by stream properties such as the continuous data arrival, the time-critical processing of objects, the evolution of the data streams, the presence of outliers and the varying densities of the data. Due to the continuously evolving nature of the stream, it is crucial that stream clustering algorithms autonomously detect clusters whose number, shapes and densities vary as the stream flows. We present the first hierarchical density-based stream clustering algorithm based on cluster stability, called *HASTREAM* [2] which is able to meet the above mentioned requirements.

We show additionally, that HASTREAM inherited efficiency issues as the main drawback of density-based hierarchical clustering algorithms, as these were not the scope of its contribution. We present then *I-HASTREAM* [1], a first density-based hierarchical clustering algorithm that has considerably less computational time compared to the first presented algorithm. I-HASTREAM utilizes and introduces techniques from the graph theory domain to devise an incremental update of the underlying model instead of repeatedly performing the expensive calculations of the huge graph. Specifically the Prim's algorithm for constructing the minimal spanning tree is adopted by introducing novel, incremental maintenance of the tree by vertex and edge insertion and deletion. The extensive experimental evaluation study on real world datasets shows that I-HASTREAM is considerably faster than HASTREAM while delivering almost the same clustering quality.

References

1. Marwan Hassani, Pascal Spaus, Alfredo Cuzzocrea, and Thomas Seidl. Adaptive stream clustering using incremental graph maintenance. In *BigMine 2015 at KDD'15*, pages 49–64, 2015.
2. Marwan Hassani, Pascal Spaus, and Thomas Seidl. Adaptive multiple-resolution stream clustering. In *MLDM '14*, pages 134–148, 2014.