

A Biological Driven Method for Correlation Clustering

András Fülöp

Many applications benefit from partitioning data points into clusters. Correlation clustering is a clustering method which originates from document clustering. The problem was introduced by Bansal, Blum and Chawla [1] and was proved to be NP-complete in the same paper. Correlation clustering can be viewed as a graph partitioning problem in which each connection is labelled + (similar) or – (different) depending on the components. In correlation clustering, unlike a large portion of clustering algorithms, the number of clusters is not specified. The goal is to partition the nodes into clusters of similar nodes. There are several existing algorithms for correlation clustering (e.g. [2], [3], [4]).

In this talk we introduce a method with which the correlation clustering problem can be solved. The algorithm is based on a biological optimization method, the Ant System method introduced by Dorigo, Maniezzo, and Coloni in an early report [6], and motivated by the work of Liu and Fu who utilized a similar algorithm for clustering problems [5]. In our talk we discuss our solution and results including comparisons with other existing solutions.

References

- [1] Nikhil Bansal, Avrim Blum, and Shuchi Chawla. *Correlation clustering*. Machine Learning Journal (Special Issue on Theoretical Advances in Data Clustering), 2004.
- [2] Becker, Hila. *A Survey of Correlation Clustering*. COMS E6998: Advanced Topics in Computational Learning Theory, 2005.
- [3] Zimek, Arthur. *Correlation Clustering*. Thesis, Ludwig-Maximilians University, Munich, 2008.
- [4] Achtert, Elke, Böhm, Christian, Kriegel, Hans-Peter, Kröger, Peer, Zimek, Arthur. *Robust, Complete, and Efficient Correlation Clustering*. SIAM Proceedings, 2007.
- [5] Liu, Xiaoyong, Fu, Hui. *An Effective Clustering Algorithm With Ant Colony*. Journal of Computers, vol. 5, no. 4, 2010.
- [6] Dorigo, Maniezzo, Coloni. *Positive feedback as a search strategy*. Technical report, Dipartimento di Elettronica, Politecnico di Milano, 1991.