A novel cluster combination algorithm for document clustering.

Xu, Sen; Wang, Zhenggang; Li, Xianfeng; Cao, Rui

Summary: Ensemble techniques have been successfully applied in the supervised machine learning area to increase the accuracy and stability of base learner. Recently, analogous techniques have been investigated in unsupervised machine learning area. Research has showed that, by combining an ensemble of multiple clusterings, a superior solution can be attained. In this paper, we solve the cluster combination problem in term of finding a “best” subspace and formulate it as an optimization problem. Then, we get the solution according to basic concept and theorem in linear algebra whereupon a novel cluster combination algorithm is proposed. We compare our algorithm with other common cluster ensemble algorithms on real-world datasets. Experimental results demonstrate the effectiveness of our algorithm.

Keywords: clustering analysis; cluster combination; documents clustering
doi:10.1007/978-3-642-36669-7_24