Principal Contrast: Evolutionary Algorithm for Clustering

Chi Tim Ng *
Hang Seng University of Hong Kong, China

Huipeng Meng
Hang Seng University of Hong Kong, China

Yan Wu
Hang Seng University of Hong Kong, China

Abstract

This paper proposes a novel high-dimensional clustering method that employs an evolutionary algorithm to obtain linear combinations and labeling configurations optimizing the Wilk statistic in the multivariate analysis of variance of the lower-dimensional transformed data. In particular, we consider the situations where the between group variation in each attribute is diminishing, while such a small between group variation is simultaneously reflected in a large amount of attributes. The performance of the proposed method is tested via simulation studies and real data analysis of labeled genetic data from small round blue cell tumors, gene expression profiles from cancer cells, and return data from hedge fund managers.