

Abstract: A new method is proposed that combines dimension reduction and cluster analysis for categorical data. A least-squares objective function is formulated that approximates the cluster by variables cross-tabulation. Consequently, individual observations are assigned to clusters in such a way that the distributions over the categorical variables for the different clusters are optimally separated. In addition, we consider existing methods with similar objectives in a unified framework and derive a new algorithm for the GROUPALS method. A simulation study is used to appraise the methods in a structural fashion and compare the performance with respect to each other. Moreover, with respect to the clustering, we compare the results to k-means cluster analysis based directly on the full dimensional data. Our results show that, in general, all methods perform similarly. However, the joint dimension reduction and clustering methods tend to outperform, with respect to the retrieval of the true underlying cluster structure, the full dimensional clustering when noise is added to the data.