

Tree-based mining contrast subspace

ABSTRACT

All existing mining contrast subspace methods employ density-based likelihood contrast scoring function to measure the likelihood of a query object to a target class against other class in a subspace. However, the density tends to decrease when the dimensionality of subspaces increases causes its bounds to identify inaccurate contrast subspaces for the given query object. This paper proposes a novel contrast subspace mining method that employs tree-based likelihood contrast scoring function which is not affected by the dimensionality of subspaces. The tree-based scoring measure recursively binary partitions the subspace space in the way that objects belong to the target class are grouped together and separated from objects belonging to other class. In contrast subspace, the query object should be in a group having a higher number of objects of the target class than other class. It incorporates the feature selection approach to find a subset of onedimensional subspaces with high likelihood contrast score with respect to the query object. Therefore, the contrast subspaces are then searched through the selected subset of one-dimensional subspaces. An experiment is conducted to evaluate the effectiveness of the tree-based method in terms of classification accuracy. The experiment results show that the proposed method has higher classification accuracy and outperform the existing method on several real-world data sets.