

A Rough Hypercuboid Approach for Feature Selection in Approximation Spaces

Abstract:

The selection of relevant and significant features is an important problem particularly for **data** sets with large number of features. In this regard, a new feature selection algorithm is presented based on a rough hypercuboid approach. It selects a set of features from a **data** set by maximizing the relevance, dependency, and significance of the selected features. By introducing the concept of the hypercuboid equivalence partition matrix, a novel representation of degree of dependency of sample categories on features is proposed to measure the relevance, dependency, and significance of features in approximation spaces. The equivalence partition matrix also offers an efficient way to calculate many more quantitative measures to describe the inexactness of approximate classification. Several quantitative indices are introduced based on the rough hypercuboid approach for evaluating the performance of the proposed method. The superiority of the proposed method over other feature selection methods, in terms of computational complexity and classification accuracy, is established extensively on various real-life **data** sets of different sizes and dimensions.