

Abstract

Yasser El Sonbaty

a feature ion algorithm with redundancy reduction for text classification

document classification involves the act of classifying documents according to their content to predefined categories. one of the main problems of document classification is the large dimensionality of the data. to overcome this problem, feature ion is required which reduces the number of ed featuresthus improves the classification accuracy. in this paper, a new algorithm for multi-label document classification is presented. this algorithm focuses on the reduction of redundant features using the concept of minimal redundancy maximal relevance which is based on the mutual information measure. the features ed by the proposed algorithm are then input to one of two classifiers, the multinomial naive bayes classifiertthe linear kernel support vector machines. the experimental results on the reuters dataset show that the proposed algorithm is superior to some recent algorithms presented in the literature in many respects like the f1-measurethe break-even point.