Abstract

This system proposes an efficient text classification approach which is based on multi-layer SVM-NN text classification and two-level representation model. Automated text classification is attractive because it frees organizations from the need of manually organizing document bases, which can be too expensive. This system proposes two-level representation model to represent text data, one is for representing syntactic information using tf-idf value and the other is for semantic information using Wikipedia. Further, a multi-layer text classification framework is designed to make use of the semantic and syntactic information. The proposed framework contains three SVM-NN classifiers in which two classifiers are applied on syntactic level and semantic level in parallel. The outputs of these two classifiers will be combined and given as input to the third classifier, so that the final results can be obtained. Experimental results on benchmark data sets like 20Newsgroups and Reuters-21578 have shown that the proposed model improves the text classification performance.

Index Terms

Computer Science  Artificial Intelligence

Keywords

Wikipedia  Semantics  Text classification  Text representation  Multi-layer
classification  SVM