Abstract

The exponential growth of the internet has led to a great deal of interest in developing useful and efficient tools and software to assist users in searching the Web. Document retrieval, categorization, routing and filtering can all be formulated as classification problems. However, the complexity of natural languages and the extremely high dimensionality of the feature space of documents have made this classification problem very difficult. We have different methods for text classification: the Naive Bayes classifier, the nearest neighbor classifier, SVM (Support Vector Machine), Feature Selection, Feature Extraction Algorithms, decision trees and a subspace method. Each method involved has its own advantage and disadvantage. In order to avoid these ambiguities and redundancies, some of these methods can be combined.
together to produce highly accurate results. In addition to this, the produced algorithm will help to enhance the performance of the overall text classification system.

References

- A, Basu, C. Watters and M. Shepherd faculty of computer science Support Vector machine for text classification, Delhousie University, Canada
- "Feature selection for text classification". Inoshika, Dilruksh

Index Terms

Computer Science

Algorithms

Keywords

Text Classification  Support Vectors  Training Set  Feature Set