Abstract

It is a big task to provide the accuracy of discovered relevance features in text documents for describing user requirements. Classification of data is biggest issue in more text documents because they have large number of words and data patterns. Most existing popular methods are used by word-based approaches. Still, they have all suffered from the problems of relevance and uncertainty. Over the years, there has been pattern-based methods should perform better result than word-based methods in describing user requirements. But, how to effectively use large scale patterns remains a typical problem in text mining. To overcome this problem, Fuzzy Relevance Feature Discovery Algorithm (FRFDA), classification techniques have been developed for relevance feature discovery. It describes both higher level and low level features based on word patterns. It is also classifies words into categories and updates those word weights based on their relevance and dispensation in patterns. The experimentation result proves that, the proposed FRFDA is better than existing manual and automation methods. The data set Reuters-21578 shows that the proposed model significantly outperforms faster and obtains better extracted features than other methods.


Index Terms

Computer Science

Information Sciences

Keywords

Text mining, fuzzy similarity, feature clustering, text feature extraction, text classification, Fuzzy Relevance Feature Discovery (FRFD), Reuters Corpus Volume (RCV).