Abstract

Huge store of hidden information in text documents is available. Extracting accurate, useful information from this store is very important. Multinomial Naïve Bayes classification algorithm is effective in processing text and extracting accurate information. A new approach of assigning weights to terms based on their positional appearance is proposed. The effectiveness of this approach is demonstrated for two standard text datasets Reuters-21578 and 20-newsgroups. This proposed approach improves average F-measure by 1.0% for Reuters-21578 and by 2% for 20-newsgroups at least.

References


Index Terms

Computer Science
Artificial Intelligence

Keywords

Term Weighting  Document Classification  Multinomial Naïve Bayes Classification  Algorithm