World Academy of Science, Engineering and Technology
International Journal of Educational and Pedagogical Sciences
Vol:9, No:01, 2015

A Similarity Measure for Classification and Clustering in Image Based Medical and Text Based Banking Applications

Authors : K. P. Sandesh, M. H. Suman

Abstract : Text processing plays an important role in information retrieval, data-mining, and web search. Measuring the similarity between the documents is an important operation in the text processing field. In this project, a new similarity measure is proposed. To compute the similarity between two documents with respect to a feature the proposed measure takes the following three cases into account: (1) The feature appears in both documents; (2) The feature appears in only one document and; (3) The feature appears in none of the documents. The proposed measure is extended to gauge the similarity between two sets of documents. The effectiveness of our measure is evaluated on several real-world data sets for text classification and clustering problems, especially in banking and health sectors. The results show that the performance obtained by the proposed measure is better than that achieved by the other measures.

Keywords : document classification, document clustering, entropy, accuracy, classifiers, clustering algorithms
Conference Title : ICEE 2015 : International Conference on Engineering Education
Conference Location : Singapore, Singapore
Conference Dates : January 08-09, 2015