

2nd International Conference on

Big Data Analysis and Data Mining

November 30-December 01, 2015 San Antonio, USA

Selecting training documents for better learning

Abdul Mohsen Algarni
King Khalid University, Saudi Arabia

In general, there are two types of feedback documents: Positive feedback documents and negative feedback documents. Term based approaches can extract many features in text documents, but most include noise. It is clear that all feedback documents contain some noise knowledge that affects the quality of the extracted features. The amount of noise is different from one document to another. Therefore, reducing the noise data in the training documents would help to reduce noise in the extracted features. Moreover, we believe that removing some training documents (documents that contain more noise data than useful data) can help to improve the effectiveness of a classifier. Based on that observation, we found that short documents are more important than long documents. Testing that idea, we found that using the advantages of short training documents to improve the quality of extracted features can give a promising result. Moreover, we found that not all training documents are useful for training the classifier.

Biography

Abdul Mohsen Algarni received a PhD degree in the Faculty of Information Technology at Queensland University of Technology, Brisbane, Australia in 2011. He is currently an Assistant Professor in the Department of Computer Science, King Khalid University. His research interests include web intelligence, data mining, text intelligence, information retrieval, and information systems.

a.algarni@kku.edu.sa

Notes: