Abstract

Text classification has become one of the most important techniques in text mining. A number of machine learning algorithms have been introduced to deal with automatic text classification. One of the common classification algorithms is the k-NN algorithm which is known to be one of the best classifiers applied for different languages including Arabic language. However, the k-NN algorithm is of low efficiency because it requires a large amount of computational power. Such a drawback makes it unsuitable to handle a large volume of text documents with high dimensionality and in particular in the Arabic language. This paper introduces a high performance parallel classifier for large-scale Arabic text that achieves the enhanced level of speedup, scalability, and accuracy. The parallel classifier is based on the sequential k-NN algorithm. The classifier has been tested using the OSAC corpus. The performance of the parallel classifier has been studied on a multicomputer cluster. The results indicate that the parallel classifier has very good speedup and scalability and is capable of handling large documents collections with higher classification results.

**Index Terms**

Computer Science  
Artificial Intelligence
Keywords
Arabic text classification  k-NN algorithm  parallel classifier  multicomputer cluster