Abstract

Information filtering (IF) systems usually filter data items by correlating a set of terms representing the user’s interest with similar sets of terms representing the data items. Many techniques have been employed for constructing user profiles automatically, but they usually yield large sets of data. Various dimensionality-reduction techniques can be applied in order to reduce the number of terms in a user query. A new framework is described to classify large scale documents and retrieve the documents related to the user’s query based on the application of trained artificial neural network (ANN) model. Its novel feature is the identification of an optimal set of documents that are relevant to the user. As a case study the government orders issued by Tamil Nadu state government, a state in India are classified according to their semantic similarity. Various neural architectures such as back propagation neural network (BPN), radial basis function (RBF), Learning Vector Quantization (LVQ) and Support vector machines (SVM) are used and their performance evaluation is analyzed.

References
- Kohonen, T., Kaski, S., Lagus, K. and Honkela, T. 1996. Very Large Two-Level SOM for Browsing of Newsgroups”. In Proc. of ICANN'96 International Conference on Artificial Neural Networks, 269-274.

Index Terms

Computer Science
Information Sciences

Keywords
Document clustering  Artificial Neural Networks (ANN)  Learning form examples