Abstract

Information retrieval model focuses on the problem of retrieving documents relevant to a user's information need represented as a query. One of the major difficulty of information retrieval is to find the relevance of documents with respect to the user query or information need. The choice of similarity measure is decisive for improving search effectiveness of a IR model. Different similarity measures have been proposed to find most relevant documents with the given query. Vector space model is a popular model and is widely used for information retrieval. The judgment of the relevance between a query and a document is evaluated using cosine similarity between them. However, vector space model does not give reasonable results in terms of precision and recall value. Information retrieval model using Dempster-Shafer theory also known as evidence theory is used in this paper. In this model, each query-document pair is taken as a piece of evidence for the relevance between a document and a query. The evidence is combined using Dempster's rule of combination and the belief committed to the relevance is obtained which then ranked accordingly. To validate the feasibility of this approach, evidences for sample document collection i.e. TREC-9 filtering track i.e. OSHUMED dataset are considered and the results are compared with traditional VSM model in terms of precision and recall measures. It is found that Dempster Shafer Model's performance is better than VSM for information retrieval.
References


Index Terms

Computer Science  Information Sciences

Keywords

Information Retrieval  Dempster Shafer Theory  Evidence Combination  Vector
Space Model

Lucene.