Scholars Journal of Engineering and Technology (SJET) **ISSN 2321-435X (Online)**
Sch. J. Eng. Tech., 2014; 2(1):54-59 **ISSN 2347-9523 (Print)**

(An International Publisher for Academic and Scientific Resources)
www.saspublisher.com

Research Article

Recommendation of News Groups to the Users Based on Cobweb Clustering

R. Vidya*, S. Akila Devi, A. Askarunisa
Department of computer Science and Engineering, Vickram College of Engineering, Enathi, India

***Corresponding author**
R. Vidya
Email: rvidyakanna@gmail.com

Abstract: Internet provides drastic access to the news articles from different information sources around the world. The main approach is used to find out the users preference from both news content and user information. Incremental clustering is done on the web news document in order to group the documents for recommendation. The idea of conceptual clustering is used. It finds the similarity between them which is called as correlation measures. Here the data is collected from data set through various web sites of news group.
Keywords: recommendation, news groups, cobweb clustering, stemmer

INTRODUCTION

Recommender systems are those are likely said to be as the information filtering (IF) technique that which presents the information items such as news article, music, politics *etc* to the users on the basis of their interest. Today web based news reading services, like google news and yahoo provides a wide range of information about the news to the people. They provide plenty of information source and current affairs and happenings around different part of the world. Hence different news sites information are collected and the information are collected from various news groups. Clustering of the news groups are done, which means grouping of the data into similar groups. Each group called as the cluster which contains objects that are similar between themselves and dissimilar to objects of other groups. The most general setting in which recommender system are studied is presented in figure1.Known user are represented as a matrix of n users and m items, which each cell are u and I corresponds to the ratings given to the item *i* by the user u is a user ratings matrix typically sparse as most users do not rate most items.

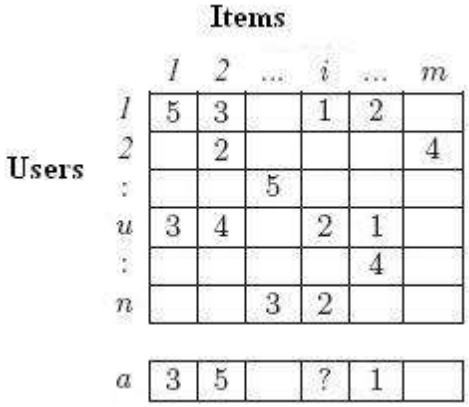

Fig. 1: General setting in recommender systems

News Recommendation system is inevitable in the news website to recommend the updates news to the interested users. However the system consumes high resources for recording users news navigation history, extracting creating profiles from history and maintaining user interest matrix.

Our proposed Conceptual and User Recommendation through Cobweb Clustering presents accurate news page to user with low computational process and storage requirements.

The two basic entities which appear in any Recommender System are the user (sometimes also referred to as customer) and the item (also referred to as product in the bibliography). A user is a person who utilizes the recommender system providing his opinion about various items and receives recommendations about new items from the system. The goal of Recommender Systems is to generate suggestions about new items or to predict the utility of a specific item for a particular user. In both cases the process is based on

the input provided, which is related to the preferences of that user. Hao Wen *et al.* [1] reported a hybrid method for personalized recommendation of Web news to users has been presented. They proposed an approach which classifies Web pages by calculating the respective weights of terms. A user's interest and preference models are generated by analyzing the user's navigational history. Kim *et al.* [2] proposed collaborative filtering principle for a network consists of group of customers. His study discuss about users contribution by uploading multimedia content, writing wiki pages, and posting blog articles. Adomavicius [3] in his work he explored a number of item ranking techniques that can generate recommendations that have substantially higher aggregate diversity across all users while maintaining comparable levels of recommendation accuracy.

In Hao Ma *et al.* [4] paper, they aimed at providing a general framework on mining Web graphs for recommendations, they first propose a novel diffusion method which propagates similarities between different nodes and generates recommendations; then we illustrate how to generalize different recommendation problems into our graph diffusion framework. McFee *et al.* [5] in their paper, they propose a method for optimizing content-based similarity by learning from a sample of collaborative filter data. Oliveira *et al.* [6] presented an online recommendation system that eases the matching of a user with the most relevant products and services. Their paper discuss the results gathered on experimental data analysis and the statistical hypotheses tests that were performed, which allowed concluding in which circumstances trust-based recommendation is advantageous. Cheng *et al.* [7] proposes an adaptive recommendation mechanism that rests on a congestion-aware scheduling method for multi-group travelers on multi destination travels. Tao *et al.* [8] reported a personalized ontology model for web information gathering. As a model for knowledge description and formalization, ontologies are widely used to represent user profiles in personalized web information gathering. Zorzo *et al.* [9] reported an adaptive automaton in recommendation systems. The recommendation systems look for to offer customized products to their users. An accurate user profile can greatly improve a search engine's performance by identifying the information needs for individual users Kenneth Wai-Ting Leung and Dik Lun Lee [10]. In this paper, they proposed and evaluated several user profiling strategies. Dimitrios Pierrakos and Georgios Paliouras presented [11] a knowledge discovery framework for the construction of Community Web Directories, a concept that they introduced in their recent work, applying personalization to Web directories. In this context, the Web directory was viewed as a thematic hierarchy and personalization was realized by constructing user community models on the basis of usage data.

Problem Definition

In this work, a user's preference model scores a Web site based on the degree to which the user prefers to retrieve information from that Web site. The recommendation rating process of the proposed system can be divided into two steps. First, a content-based algorithm is utilized to determine the probability of recommending Web content to a user, considering the factors of the user's interest and preference models, the Web content, and the time limitation. Second, the method of collaborative filtering is used to modify the probability of recommending Web content. The system will distribute some test Web content, which has been well classified and identified by users. The users who send back positive responses are considered as the trusted users.

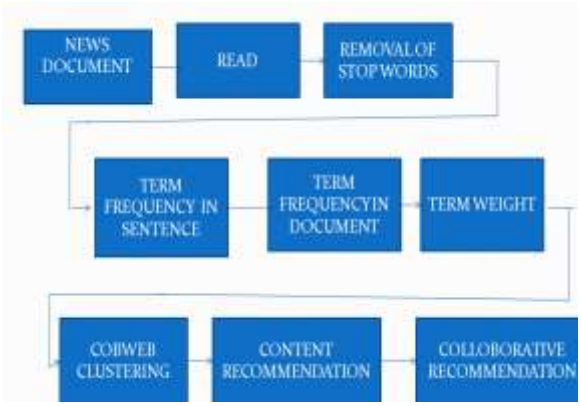

Fig. 2: Block diagram

A New Database is that which stores all the repository of the document information. Then Read Document it reads the texual content from the information which has been stored. Now the process starts by the removal of the stop words. After the stop words it means Prepositions and the conjunctions. It is called as the stemmer process. Term sentence Frequency is that it checks the number of occurrence of the term in the

sentence. Term document Frequency is that it checks occurrence of the word in the overall document.

Recommendation System

Determining the content of a Web page using the Web page classification method; and utilizing the Nave Bayes model for updating the user's interest and preference models. In this work, a user's preference model scores a Web site based on the degree to which the user prefers to retrieve information from that Web site. The recommendation rating process of the proposed system can be divided into two steps. First, a content-based algorithm is utilized to determine the probability of recommending Web content to a user, considering the factors of the user's interest and preference models, the Web content, and the time limitation. Second, the method of collaborative filtering is used to modify the probability of recommending Web content.

Clustering

Cluster analysis or clustering is the task of assigning a set of objects into groups (called clusters) so that the objects in the same cluster are more similar (in some sense or another) to each other than to those in other clusters. In proposed system, we include incremental clustering techniques for clustering user's history of news pages which we used are COBWEB Clustering.

Techniques

TF-IDF TF - IDF, term frequency-inverse document frequency, is a numerical statistic which reflects how important a word is to a document in a collection or corpus. It is often used as a weighting factor in information retrieval and text mining. The tf-idf value increases proportionally to the number of times a word appears in the document, but is offset by the frequency of the word in the corpus, which helps to control for the fact that some words are generally more common than others.

Stemmer

Affix removal conflation techniques are referred to as stemming algorithms and can be implemented in a variety of different methods. All remove suffices and/or prefixes in an attempt to reduce a word to its stem. The algorithms that are discussed in the following sections, and those that will be implemented in this project, are all suffix removal stemmers. A context free algorithm removes endings with no restrictions placed on the circumstances of the removal.

RESULTS

Incremental Conceptual Clustering

Clustering can be considered the most important unsupervised learning problem; so, as every other problem of this kind, it deals with finding a structure in a collection of unlabelled data. A loose definition of clustering could be "the process of organizing objects into groups whose members are similar in some way". A cluster is therefore a collection of objects which are "similar" between them and are "dissimilar" to the objects belonging to other clusters. The results are shown by various screen shots from the fig. 3-10.

Fig. 3: Displays the text from the document

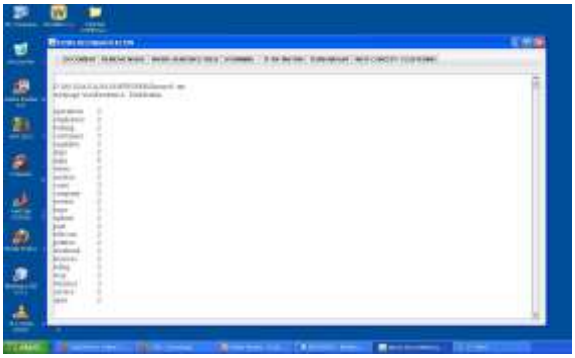

Fig. 4: Number of occurrences of words in the sentence

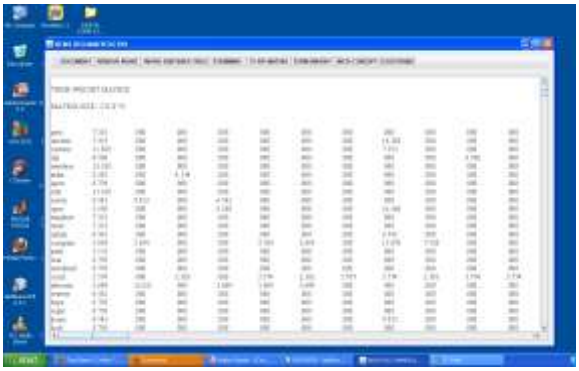

Fig. 5: Term weight calculation

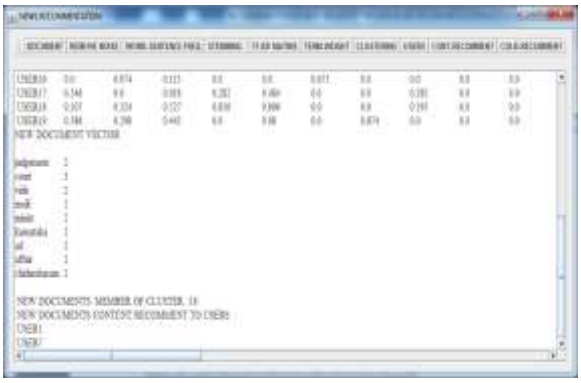

Fig. 6: Conceptual Recommendation

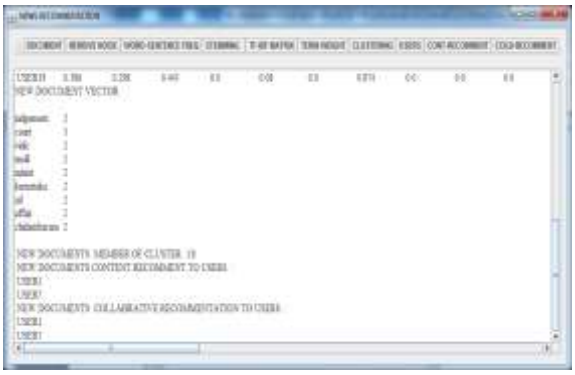

Fig. 7: Colloborative Recommendation

In this project the improved Recommendation of the documents is plotted with the measurement of the similarity between the documents centroids and the Cluster Formation are shown in Fig. 8.

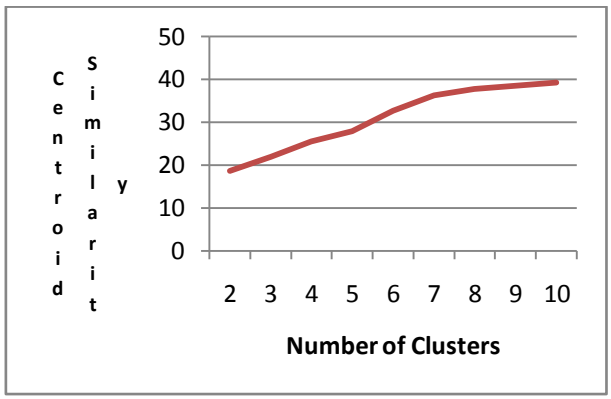

Fig. 8: Improved recommendation Rate

The Efficiency of the accuracy is measured with the improved Precision rate by means of the Cobweb Clustering that gives an improved precision rate when compared with K-means used in the existing system as shown in Fig 9 and the diminishing value of the Recall points when compared to the existing system.

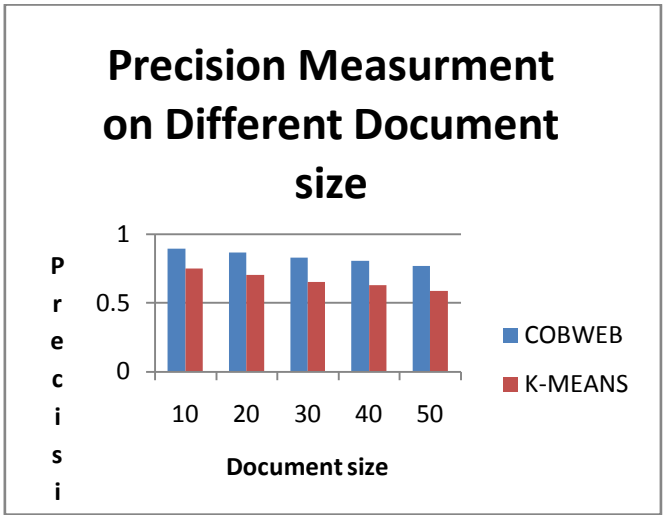

Fig. 9: Improved Precision Rate

The diminishing value of the Recall points when compared to the existing system is shown in Figure 6 that indicates that the technique decreases the error rate.

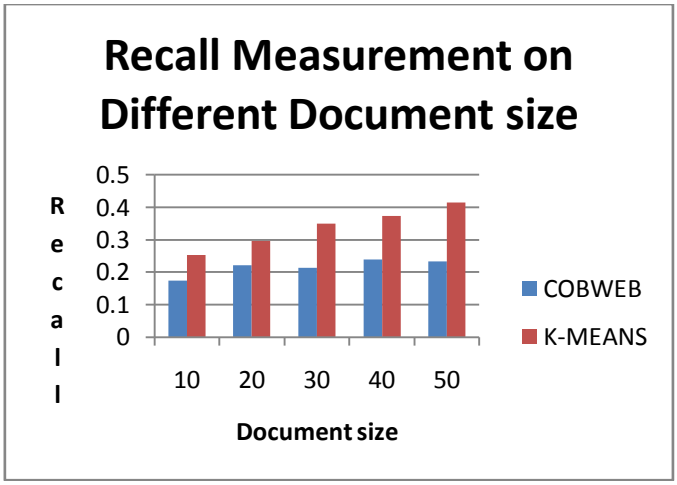

Fig. 10: Diminishing Recall Value

CONCLUSION

In this project Web News Recommendation is performed to the user by means of content and collaborative means. Web News is collected from various news sites and it explores the intra relation towards the news articles. The similarity measurements are applied and the maximum dimensionality of the document is increased. The System supports an efficient incremental conceptual clustering. The content and collaborative based recommendation system is tested with data collecting from various news sites. These sites produce news from an events happened in the society. Various news sources used the news content almost are same or little variations. So that many number of redundant news documents are processed by recommendation system per every event generating documents. The user navigation history documents data have great relevance when successive news generated from the event.

REFERENCES

1. Hao Wen, Fang L, Guan L; A Hybrid Approach for Personalized Recommendation of News on the Web. Expert Systems with Applications, 2012; 39: 5806–5814.
2. Kim HK, Ryu YU, Cho Y, Kim JK; Customer-Driven Content Recommendation over a Network of Customers. IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, 2012; 42(1): 48-56.

3. Adomavicius G; Improving Aggregate Recommendation Diversity Using Ranking-Based Techniques. IEEE Transactions on Knowledge and Data Engineering, 2012; 24(5): 1-15.
4. Ma H, King I, Rung-Tsong Lyu M; Mining Web Graphs for Recommendations. IEEE Transactions on Knowledge and Data Engineering, 2012; 24: 1051-1064.
5. McFee B, Barrington L, Lanckriet G; Learning Content Similarity for Music Recommendation. IEEE Transactions on Audio, Speech, and Language Processing, 2012; 20: 2207-2218.
6. Oliveira ADR, Bessa LN, Andrade TR, Filgueirase LVL, Sichman JS; Trust-based recommendation for the social Web. IEEE (Revisit IEEE America Latina Latin America Transactions), 2012; 10: 1661 – 1666.
7. Sheng-Tzong Cheng, Gwo-Jiun Horng, Chih-Lun Chou; The Adaptive Recommendation Mechanism for Distributed Group in Mobile Environments. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, 2012; 42: 1081-1092.
8. Tao X, Li Y, Zhong N; A Personalized Ontology Model for Web Information Gathering. IEEE Transactions on Knowledge and Data Engineering, 2011; 23(4): 496-511.
9. Zorzo SD, Cereda PRM, Gotardo RA; Adaptive Automata in Recommendation Systems. IEEE (Revista IEEE America Latina) Latin America Transactions, 2011; 9: 152-159.
10. Wai-Ting Leung K, Lee DL; Deriving Concept-Based User Profiles from Search Engine Logs. IEEE Transactions on Knowledge and Data Engineering, 2010; 22(7): 969-982.
11. Pierrakos D, Paliouras G; Personalizing Web Directories with the Aid of Web Usage Data. IEEE Transactions on Knowledge and Data Engineering, 2010; 22(9): 1331-1344.