International Journal on Recent and Innovation Trends in Computing and Communication
Volume: 3 Issue: 6
ISSN: 2321-8169
3960 - 3961

Personalized Web Recommender System based on Collaborative Filtering

Mr.Mukesh Phadatare
Lecturer
Computer Technology
Babasaheb Gawde Institute of Technology, India

Mr.Sachin More
Lecturer
Computer Engineering
Babasaheb Gawde Institute of Technology, India

Abstract:-It has become much more difficult to access relevant information from the Web with the explosive growth of information available on the World Wide Web. One possible approach to solve this problem is web personalization.

The objective of personalization built on Web usage mining is to recommend a set of objects to the current user, possibly consisting of links, ads, text, products, or services, tailored to the user's perceived preferences as determined by the matching usage patterns.

Keywords:- Web Personalization, Web Access Patterns, *k*-Nearest-Neighbour (*k*NN), Web Usage Mining, Collaborative Filtering, Web Service Recommendation.

*****

I. INTRODUCTION

Collaborative Filtering, firstly proposed by Rich [1], is widely used in commercial recommender systems like Amazon.com [2].The tremendous growth in the number and the complexity of information resources and services on the Web has made Web personalization an indispensable tool for both .

Web based organizations and for the end users.Web personalization has an ability of a site to engage visitors at a deeper level, so that makes the Web experience of a user customized to the user's taste or preferences. Principle fundamentals of Web personalization include modelling of Web objects (such as pages or products) and subjects (such as users or customers), classification of objects and subjects, identical between and across objects and/or subjects, and purpose of the set of activities to be recommended for personalization.

II. PROBLEM DEFINITION

Web usage Mining includes of personalization of websites allows it to recommend a set of objects to the active user, possibly the set of web pages that likely to be accessed in future by users.

The promising Collaborative Filtering approach allows filtering of information or patterns based on collaboration of users, or similarity between items.

III. METHODOLOGY USED

The proposed system forms methodologies by resolving it into two necessary phases which are:-(see fig (1))

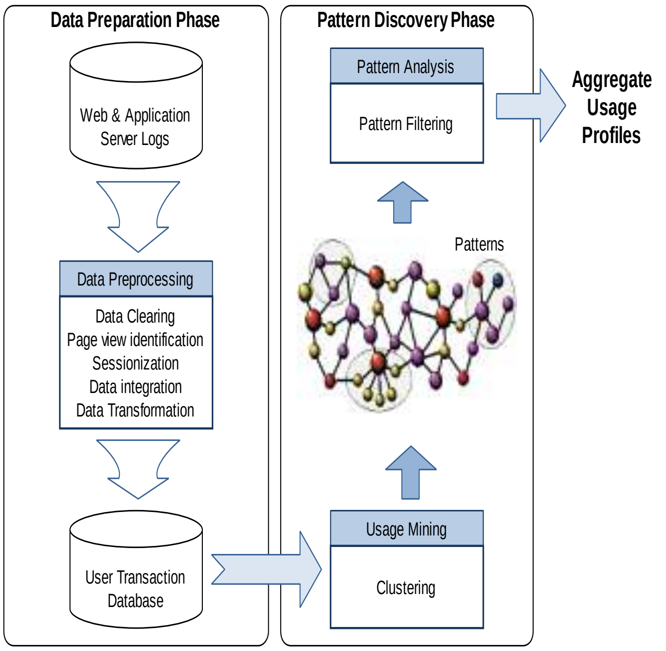

Fig.(1)

1. **Data Preparation Phase**:

This phase considered by first module among the three modules which Data pre-processing.

Data pre-processing module transforms raw Web & Application server logs into User Transaction database by performing data clearing, page view identification, sessionization, data integration and data transformation tasks. This process may involve pre-processing the raw data, assimilating data from multiple sources, and transforming the Integrated data into a form suitable for input into specific data mining processes. Mutually, we refer to this process as data preparation.

2. **Pattern Discovery Phase**:-

This phase includes rest of the two modules Usage Mining & Pattern Analysis.

Usage Mining:-In Usage Mining module, we perform data mining operations that is clustering; result of this is we get different patterns of data. That patterns used in Pattern Analysis.

Pattern Analysis:-This module performs Pattern filtering on given different data patterns so that to obtain Aggregate Usage Profiles. Here we use Collaborative Filtering approach for this task.

Collaborative filtering based on the *k*-Nearest-Neighbour (*k*NN) approach involves comparing the activity record for a target user with the historical records of other users in order to find the top *k* users who have similar tastes or interests. The mapping of a visitor record to its neighbourhood could be based on similarity in ratings of objects, access to related content or pages, or purchase of related items. The known neighbourhood is then used to recommend items not already accessed or purchased by the active user. Thus, there are two primary phases in collaborative filtering:

The neighbourhood formation phase and the recommendation phase. In the context of personalization based on Web usage mining, *k*NN involves measuring the similarity or correlation between the active session $\bar{s}$ and each transaction vector $\bar{t}$ (where $t \in T$). The top *k* most similar transactions to are considered to be the neighbourhood for the session $\bar{s}$, which we denote by *NB*(*s*) (taking the size *k* of the neighbourhood to be implicit):

$$NB(s) = \{\bar{t}s1, \bar{t}s2, \ldots ,\bar{t}sk \}.$$

In order to determine the which items (not already visited by user in active session) are to recommended, a recommendation score is computed for each page view pi $\in$ P based on the neighbourhood for the active session .Two factors are used in determining this recommendation score first is overall similarity of active session to the neighbourhood as a whole, and the average weight of each item in the neighbourhood.

First we compute the mean vector (centroid) of NB(s). The feature value for each page view in the mean vector is computed by finding the ratio of sum of the page view's weights across transactions to the total number of transaction in the neighbourhood. We denote this vector by cent(NB(s)). For each page view p in the neighbourhood centroid, we can now obtain a recommendation score as a function of the similarity of the active session to the centroid vector and the weight of that item in this centroid. In our implementation 'e' have chosen to use the following function, denoted by rec(_s,p):

Rec(_s, p) = _weight(p,NB(s)) X sim(_s, cent(NB(s))), where weight (p,NB(s)) is the mean weight for page view p in the neighbourhood as expressed in the centroid vector. If the page view p is in the current active session, then its recommendation value is set to zero. If a fixed number N of recommendations are desired, then the top N items with the highest recommendation scores are considered to be part of the recommendation set. In our implementation we normalize the recommendation scores for all page views in the neighbourhood.(so that the maximum recommendation score is 1), and return only those which satisfy a threshold test. In this way we can compare the performance of *k*NN across different recommendation thresholds.

IV. SUMMARY

We have proposed an intelligent web recommender system known as web personalization based on sequential Web Access Patterns (WAP).In the proposed system, the sequential pattern mining algorithm CS-mine is used to mine frequent sequential web access patterns. The mined patterns are stored in are stored in the Pattern-tree, which is then used for matching and generating web links for online recommendations. The proposed system has achieved good performance with high satisfaction and applicability.

V. CONCLUSION

We have extended the Collaborative Filtering rule mining by assigning a significant priority to the pages based on time visiting frequency of each page. The proposed prioritizing measure can be used to judge the importance of page to the user, and try to give more useful to user, in order to capture the user's information need more precisely and recommend pages more useful to user.

References

[1] E. Rich, "User modelling via stereotypes," Cognitive Science, Vol.3, No.4, 1979.W.-K. Chen, *Linear Networks and Systems* (Book style). Belmont, CA: Wadsworth, 1993, pp. 123–135.

[2] G. Linden, B. Smith, and J. York, "Amazon.com recommendations: item-to-item collaborative filtering," IEEE Internet Computing, Jan./Feb. 2003.

[3] Web Usage Mining and Personalization -Bamshad Mobasher, DePaul University

[4] Web Mining Based on User Access Patterns for Web Personalization - WANG Xiao-Gang & LI Yue Wuhan University of Science and Engineering, Wuhan City, Hubei Province, China 430073