Integrated Semantic Web Usage Mining with Fuzzy Technique to Improve Accuracy of Recommendation System

ABSTRACT

of

THESIS

SUBMITTED TO

BABASAHEB BHIMRAO AMBEDKAR UNIVERSITY

LUCKNOW

BABASAHEB
BHIMRAO
AMBEDKAR
UNIVERSITY

LUCKNOW

प्रज्ञा शील करुणा

ESTABLISHED 1996

FOR THE DEGREE OF

Doctor of Philosophy

IN

COMPUTER SCIENCE

Submitted by

Bhupesh Rawat

Enrolment No. 966/14

Supervisor

Prof. Sanjay Kumar Dwivedi

DEPARTMENT OF COMPUTER SCIENCE
SCHOOL FOR INFORMATION SCIENCE & TECHNOLOGY
BABASAHEB BHIMRAO AMBEDKAR UNIVERSITY
(A CENTRAL UNIVERSITY; NAAC- 'A' GRADE)
VIDYA VIHAR, RAEBARELI ROAD, LUCKNOW-226 025 (U.P.), INDIA

2018

Abstract

With the abundance of information available on the web it is becoming increasingly difficult for users to find the items which best match with their preferences. The items to be recommended could be such as 'what course to choose', 'what movie to watch', and 'what book to read' among others. This is commonly referred to as 'information overload' problem. In this context, recommender systems help user to find the most appropriate items based on their preferences. Recommender systems have been used successfully in a wide variety of domains such as e-commerce, e-government and e-resources among others. There are several recommendation techniques which are found to be useful in making recommendations in several domains such as book, movie, and news article among others. The choice of a particular technique depends on the type of data available with the system. For example, if the 'user item rating matrix' is available then the 'collaborative filtering technique' is most appropriate and if the items' features are available then 'content filtering approach' is most suitable. Majority of existing recommender systems are based on content filtering technique, collaborative filtering technique and their combinations. Content filtering approach recommends item (learning objects/learning actions) based on the profile of a user which consists of learners' preferences for a set of items. On the other hand, collaborative filtering systems take into account the opinion of other users in order to find similarity of a group of users with an active user.

In addition to 'information overload' issue, the existing recommender systems are also suffering from the issue of *'one size fits all'* which means that the same course/learning resources are being offered to all the learners without taking into account their differences in terms of their level of knowledge, skill, and interest. This problem could be addressed by building the profile of a learner which consists of learner's preferences. The profile is then used by a recommender system to make recommendations to users.

Furthermore, we have also found that the existing clustering algorithms do not employ '*cluster validation mechanism'* in order to evaluate the quality of clusters produced by these algorithms. This leads to the formation of clusters which do not represent the actual profile of a user and are not consistent with the actual number of clusters present in the given dataset. Moreover, the applications which use these clusters for recommendation of items might end up recommending items which do not match with the learners' profile. In order to deal with this issue, we have suggested cluster evaluation methods which ensure that, the clusters created are of high quality.

In addition, existing recommender systems based on collaborative filtering technique suffer from *sparsity issue*. The sparisty is caused by insufficient ratings in the user item rating matrix which prevent a recommender system from making good quality of recommendations. Furthermore, traditional recommender approaches were not able to fully explore the semantic tools which could be exploited in order to learn more about learners' preferences. With the emergence of semantic web, a large number of semantic tools and techniques have become available which can be exploited in order to elicit additional preferences of learners. In this context, we aim to enrich the user item rating matrix by utilizing resource description framework and Apache Jena rules in order to improve the accuracy of recommendations.

The thesis proposes a course recommender framework which recommends different data mining courses to learners based on their profile. Experiments were conducted on the real world dataset in order to evaluate the performance and accuracy of the proposed framework which shows that the framework is able to improve the accuracy of recommendations significantly. The proposed approach based on k-means and collaborative filtering algorithm provided a lowest RMSE of 39%, a highest precision of 90%, a highest recall of 72% and an F1 score of 79%. Furthermore, the enhanced framework which is based on resource description framework, Apache Jena rules and RDF factbase yielded lowest RMSE of 25%, a highest precision of 98%, a highest recall of 90% and an F1 score of 93%. We not only compared our approach (M2) which is based on enhanced framework with (M1) previous framework which is not based on resource description framework but with other similar approaches in order to see the improvement. The significant improvement obtained by the enhanced approach is due to enriched user 'item rating matrix'. The evaluation of precision and recall is done @ 10 which means that the accuracy and relevancy of courses is measure out of 10 courses. For instance a recall of 90% suggests that, out of 10 courses 9 are relevant for a learner which helps a learner in improving his subject performance and overall academic performance as well.