



A web recommendation system considering sequential information



Rajhans Mishra^{a,*}, Pradeep Kumar^b, Bharat Bhasker^b

^a Indian Institute of Management, Indore, Information Systems Department, Indore, India

^b Indian Institute of Management, Lucknow, IT and Systems Department, Lucknow, India

ARTICLE INFO

Article history:

Received 23 March 2013

Received in revised form 29 January 2015

Accepted 6 April 2015

Available online 14 April 2015

Keywords:

Web recommendation system

Soft clusters

Sequential information

Singular value decomposition (SVD)

ABSTRACT

With the rapid growth of information technology, the current era is witnessing an exponential increase in the generation and collection of web data. Projecting the right information to the right person is becoming more difficult day by day, which in turn adds complexity to the decision making process. *Recommendation systems* are intelligent systems that address this issue. They are widely used in e-commerce websites to recommend products to users. Most of the popular recommendation systems consider only the content information of users and ignore sequential information. Sequential information also provides useful insights about the behavior of users. We have developed a novel system that considers sequential information present in web navigation patterns, along with content information. We also consider soft clusters during clustering, which helps in capturing the multiple interests of users. The proposed system has utilized similarity upper approximation and singular value decomposition (SVD) for the generation of recommendations for users. We tested our approach on three datasets, the MSNBC benchmark dataset, simulated dataset and CTI dataset. We compared our approach with the first order Markov model as well as random prediction model. The results validate the viability of our approach.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Massive digitization of data and affordable processing capabilities have motivated organizations to shift from the traditional world of mass production to the new world of customization with respect to their offerings [49]. The development of e-commerce platforms has allowed companies to provide more options to the customers. Decision support systems are required to collect the huge amount of data, process it and project it to the managers in support of customization. Recommendation system is a decision support system which can provide the desirable information to the customers as per their needs.

Recommendation systems are used by e-commerce organizations to suggest products to their customers. The products can be recommended, based on the top sellers on a site, demographics of the customer, analysis of the past buying behavior of the customer, etc. Recommendation systems generate recommendations for the users by exploring their requirements and likes. They generate

different recommendations to suit different users, thus providing customized web interface to the users. Thus, the web is personalized for each user using a recommendation system. Currently most of the e-commerce organizations have enabled recommendation systems at the back end, offering web recommendations to the users.

Recommender systems have been developed using data mining techniques [16,6,7,14,22,35,42,57], heuristics [39,11,46,50] and finding association patterns among the items [24,31]. Examples of popular recommender systems include Amazon.com [33] for books, CDs and various other products, Movie Lens [36] for movies, VERSIFI [5] for news, PHOAKS system for relevant information to users on web [55] and Jester system for jokes [16].

1.1. Motivation and problem definition

With the rapid growth of technology and the web, generation and accumulation of digital data have become easier. Advancement in techniques from diverse domains like machine learning, pattern recognition and statistics have made it possible to mine and unfold interesting as well as unknown patterns from the data.

Web data exists in various formats like url visits, web page content and incoming and outgoing hyperlinks to the page. Based on the data being analyzed or mined, web mining can be classified into three

* Corresponding author at: Faculty Block B, First Floor, IIM Indore, Indore, (M.P.), India 453331. Tel.: +91 731 2439 550; fax: +91 731 2439661.

E-mail addresses: rajhansm@iimidr.ac.in (R. Mishra), pradeepkumar@iiml.ac.in (P. Kumar), bhasker@iiml.ac.in (B. Bhasker).

different categories, namely, web usage mining, web content mining and web structure mining. Web content mining is the application of data mining techniques to the content published on the web. Web structure mining operates on the web's hyperlink structure [56,13]. Web usage mining derives novel, implicit and useful patterns from the usage data of users [10,54,58].

Web recommendation systems are an important and popular tool to analyze users' behavior over the web and to generate recommendations as per their preferences. It supports organizations in intelligent decision making with respect to their customers' needs by automating the recommendations as per their preferences. Hence, it works as a decision support system for organizations.

Presenting a web user with his most probable next page visit is an interesting and challenging problem. Consider a web user who has registered with an online megastore website like ebay.com. In his session visit, he has traversed the pages like entertainment, books, electronics, footwear and so on. As a store manager, it will be an interesting problem to provide in advance to the web user the set of few web pages he/she might likely visit in his/her current session. Thus, any system providing a recommendation for the next one or two pages can be helpful in projecting the desired product or category to the user. Desired product/ category will have more probability to be purchased, which, in turn, can improve expected profits for any online e-commerce firm.

While building recommender systems, the sequentiality aspect of the user session is ignored. Sequential aspect of web user sessions has been considered by probabilistic models like Markov model while designing a web recommendation system. However, problem with the probabilistic model is that switching probability among web categories should be known a priori, and may require the knowledge and experience of a domain expert. Even if a domain expert is available, the estimation of exact probability among states (web categories/pages) is an open problem and cannot be easily addressed.

Recommendation system is a type of decision support system designed to discover user preferences, and to study them in order to anticipate their needs. They provide recommendations to customers as per their taste within a given domain. Formally, in a recommendation framework, there exists a large number (n) of items or products $P = \{P_1, P_2, P_3, \dots, P_n\}$, which are described by a set of k attributes or features, $F = \{F_1, F_2, F_3, \dots, F_k\}$. Each product is defined by one or more features from the feature set. There is also a large set of m users, $U = \{U_1, U_2, U_3, \dots, U_m\}$ and for each user, a set of ratings about the quality of observed products is maintained in the database. Now, we formally define the problem as follows: For a new user p , the task of the system is to generate the set of next web page visit, based on the web page visits of similar profile available in database U . While predicting, the system should also consider the web page visit order.

Design of recommender system may be viewed as a combination of clustering and classification tasks. In this paper we have proposed the framework for the design of a recommender system using a combination of similarity upper approximation technique (for clustering web user sessions) and singular value decomposition (for predicting the next web page visit) algorithm [17,37]. To capture the sequentiality property of the data, we have used S^3M similarity measure [28] while performing clustering task.

1.2. Contribution and paper organization

In this paper, we designed a recommendation system for web users considering the sequential aspect of a web user session. The proposed recommendation system is different from sequential pattern mining algorithms. Sequential mining algorithms provide the patterns that exist in the sequences. In our work we have proposed a system which generates the recommendations to the users, considering the sequential information that exist in their usage patterns of web pages.

In our proposed model, rough set based similarity upper approximation clustering technique has been used that generates overlapping

clusters. Overlapping clusters contain common elements, hence the boundary of these clusters become soft. Soft clusters are desired since they capture multiple interests of the users. They allow any user to be placed in multiple categories.

We have performed experiments to validate the results of our recommendation system. We have utilized three datasets for the experiments, the MSNBC dataset, simulated dataset and CTI dataset. We have evaluated the performance of our recommendation system on these three datasets and validated our results with the first order Markov model as well as a random prediction model.

The rest of the paper is organized as follows; related work has been discussed in Section 2. The architecture of the proposed system has been discussed in Section 3. Section 4 reports the experimental results and discussions while the conclusion and future work are given in Section 5.

2. Related work

The journey of recommenders system started with research papers on collaborative filtering by Resnick et al. [46], Shardanand and Maes [50] and Hill et al. [21]. Recommendation systems are designed using various techniques including k-NN, decision tree, clustering, regression, heuristic methods, neural networks and association rule mining [41]. Based on the type of techniques used, recommendation systems can be classified as content based and collaborative based systems [1].

The content based approach has originated from the information retrieval [46,48] and information filtering domain [36]. Content based recommender systems generate recommendations based on users' past preferences. The rating for any item for any user is calculated based on ratings of similar items given by the user. Many researchers treat it as a classification problem where the goal is to learn a function that predicts which class a document belongs to (i.e., either liked or not-liked). Others view it as a regression problem in which the goal is to learn a function that predicts a numeric value (i.e., the rating of the document) [4,30,38,43,32].

Collaborative systems are different from content based systems in the sense that they first find similar users for target users and then generate recommendations based on preferences of similar users [1]. In this approach, recommendations are made by finding correlations among the users. The main objective of collaborative filtering is to find rating of the items, not seen by the current user, using the ratings of similar users.

GroupLens [26], Video Recommender [21] and Ringo [50] are examples of recommender systems that use a collaborative filtering algorithm for automatic prediction. Collaboration based recommender systems can be further classified into two classes, memory based (heuristic based) and model based collaborative systems [1].

Memory based systems compute the similarity between users based on users' ratings. The algorithms of memory based systems are heuristics that make recommendations based on an entire collection of items pre-rated by the users [7,39,11,46,50]. Model based collaborative recommender systems generate the descriptive model of the system, based on the users' preferences, using various data mining and machine learning techniques. The techniques that are used include Bayesian models, clustering models, latent semantic models as singular value decomposition, probabilistic latent semantic analysis, multiple multiplicative factor, latent Dirichlet allocation and Markov decision process based models [59]. The predictions for a new user are made based on the constructed model. Kumar et al. [29] have used a simple probabilistic model for collaborative filtering. There are various other probabilistic modeling techniques used for building recommender systems, available in literature.

Zang et al. [60] have proposed a knowledge based recommender system that utilizes opinion mining and rough set association rule mining to find out the associations between product attributes from user data. Castellano et al. [8] have proposed a new system for web recommendation using fuzzy sets and neural networks. Other web

Download English Version:

<https://daneshyari.com/en/article/552426>

Download Persian Version:

<https://daneshyari.com/article/552426>

[Daneshyari.com](https://daneshyari.com)