IJITKM Volume 6 • Number 2 • July-December 2013 pp. 138-142

A Technical Analysis of Recommender Systems for Web Personalization based on Data Mining Methods

Renu Tilwani[1],Deepak Dembla[2] Dr. Surendra Sharma[3]
[1]Department of CSE, Arya Institute of Engg. & Technology, Jaipur, INDIA
[2,3]Associate Professor, Deptt. of CSE, Arya Institute of Engg. & Technology, Jaipur, INDIA
renutilwani@gmail.com, deepak_dembla@yahoo.com,surendrasharma286@gmail.com

Abstract: The rapid progress of internet technologies in last few years has imposed a heavy information burden on users. This raises the requirement of personalized website design. The benefit of present approach for personalized website design is its cost-effective features, but on the other hand current approach cannot easily provide a more refined personalized service due to lack of a web usage data. In this paper, the design recommender system is analyzed as a technique for website design personalization. We analyzed current recommender systems, architecture and design framework, issues and challenges and also discuss various recommendation techniques. We also identify recommendation techniques as a technology for designing Website personalization system. Here, techniques and methodologies of Data Mining (DM) for information extraction are compared for recommendation systems. After this investigation it is concluded that Sparsity and Cold-start are the two problems of present recommender system and some other architectural framework of recommender system should be implemented to tackle these problems.

Keywords: Web Design, Recommender System, Personalization.

1. INTRODUCTION:

With the rapid development of internet technologies in recent decades, the requirement for personalized service on website has been increased [1]. But present Website personalization system simply gathers design elements in accordance with the user's choice. Website personalization system should be enhanced to reduce user's trouble in the process of personalization. These studies inspect various personalization services, which are used effectively in other fields of website development and then propose a design recommender system.

Recommender Systems (RSs) are software agents which gives the suggestion for items to the user according to his/her preference [2] [3] [4]. These suggestions help user in decision-making processes such as what items to buy or what online news to read or what music to listen. A Recommender System usually focuses on a specific type of item (e.g., books, or news, movies) and accordingly its design, its graphical user interface, and the core recommendation technique is used to generate the recommendations are all tailored to provide helpful and valuable suggestions for that specific type of item. Recommender Systems are very helpful for those users who are lacking sufficient personal experience to evaluate the number of alternative items present on website. For example, a book recommender system that assists users to select a book to read. In the popular website such as Amazon.com, this site employs a RS to personalize the online store for each customer [4].

Commercial websites use Recommender systems to suggest products to their customers [2]. Items are recommended to the customers on the basis of their demographic information and by analyzing their past buying behavior. Analyzing the past buying behavior of customers and their demographic information helps the site to adapt itself according to customer.

2. COMPONENT OF RECOMMENDER SYSTEM

A recommender system is composed of three elements as shown in Figure 1. First, a Recommendation content which is presented to users has to be prepared. Second, User's preferences or behavioural data on these contents must be collected. Third, then recommendation technique is choose from number of alternatives technique to analysis these user data and select the optimal content to each user [5].

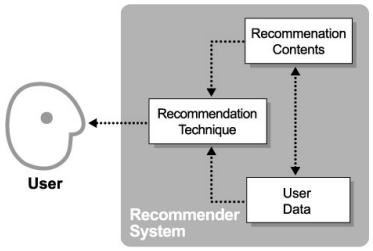

Figure1. Components of Recommender System

3. CLASSIFICATION OF RECOMMENDATION TECHNIQUE

There are six different classes of recommendation techniques:

Content based Recommendation: In this approach, given set of documents are examined on the basis of ratings given by the previous users. User interests are based on the features of the objects rated by that user.

Collaborative-based Recommendation: These approaches suggest the items to active clusters with similar tastes of items for other users liked in the past. The similarity in taste of two users is computed based on the similarity in the rating history of the users. Therefore, this approach is also known as people-to-people correlation.

Demographic-based Recommendation: In this technique, system suggests items on the basis of demographic profile of the user. These approaches are famous in the marketing literature.

Knowledge-based Recommendation: In this technique, knowledge-based system recommends the items based on specific domain knowledge.

Community-based Recommendation: In this technique, system recommends items on the basis of user's friend preferences. It has been observed that people tend to depend more on suggestions from their friends rather than unknown individuals. This study generates interest in community-based systems or, as or as they typically referred to, social recommender systems.

Context-based Recommender system:

In this technique, system recommend item to customer on the basis of its contextual information such as time, place and company of other peoples (e.g., for watching movies or dining out).

Hybrid Recommender Systems: These Recommender Systems are based on the combination of the above mentioned techniques. A hybrid system combining techniques A and B tries to use the advantages of A to fix the disadvantages of B.

4. DESIGN RECOMMENDER SYSTEM FOR WEBSITE PERSONALIZATION SERVICE

4.1 BASIC FRAMEWORK The basic design of recommender system consists of three components:

- **Recommendation Content** A website is a composed of many web pages and a webpage contain various elements such as text, graphics, color, layout, etc. Thus we can use Webpage design elements as recommendation content for designing website personalization system [5].In present personalization design service, users choose their desired elements out of many alternatives specified by designers, and then the personalization system gather these preferences into a page design.

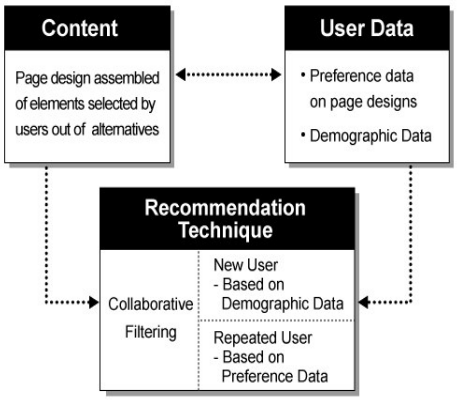

Figure3: Main issues of the Design Recommender System

- **User Data**

We can divide the user data in two groups: explicit user data and implicit user data. The data that are collected by asking users to give their own preferences by filling out a form that requests particular information are the explicit user data. On the other hand, the data that are collected by analysing past buying behaviour of user are implicit user data [5].

- **Recommendation Technique**

Website design is composed of many multimedia elements that are interactive and have complicated features. Since website design has a subjective character, its assessment varies with the personal tastes of users. Therefore, it is quite difficult to analyze and classify website design. Even if it could be done, the result would be undesirable in most cases. Therefore, content based approach is not appropriate for analyzing and dividing page designs into classes and recommending page designs according to the result. In collaborative filtering technique, we find user group that has most similar taste for a given user by analysing by analyzing user's preference data on a particular page designs [3][5][6]. Thus, we can say that collaborative filtering approach is better than content based approach for recommendation of subjective contents like page design.

4.2 STRUCTURE AND PROCEDURE

The recommender system for the website personalization is consists of four main modules based on a personalization database (Figure 4).

- **User Registration Module** Those customer who use the services of recommender system for website design personalization , first register themselves as new users by submitting their general information such as demographic and background data. All the gathered user data are stored in server log files as a user profile database of the recommender system.
- **Design Elements Registration Module** By using this module, designers can classify page design into suitable constituent elements and register their various design alternatives for each element in a design element database. Registered design alternatives are the basic ingredients with which users can make their own page designs [5].

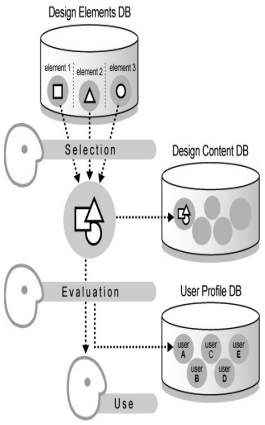

Figure4. Recommender System Structure for Website

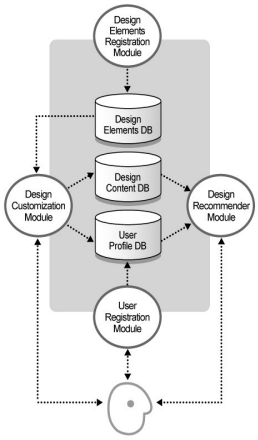

Figure5 Procedure of the Design Customization Module.

- **Design customization module**

The design customization module and present personalization service follow the same procedure in website design. However, it is improved in two aspects as follows: (i), a page design selected by a user is registered in a design content database, which can be reviewed and used by another user as personalization content; second, the page design are always rated by the user whenever they register, review or use the page design. Users can state preferences by rating page design presented to them. Therefore, the recommender system can have users' preference data on each page design, and these Preference data plays a big role in recommending items to user Figure 5 summarise the procedure of design customization module

- **Design Recommender Module**

In this module, collaborative filtering approach is implemented to identify users with similar taste and figure out the nearest neighbours of each user by using their preference data. To do so, all the users who design their own page in the design customization module or use a page design suggested by the system should offer their preferences to the page design by filling out a rating form, and then the results are stored in the user profile database. The design recommender system than identify users group who have the most similar preference pattern to the given user and then select the page designs that they give a high rate of preference to in the design contents database. Instead of preference data, demographic data is used by recommender system for newly registered user to solve the new user problem as stated above[5][6]. Therefore, in case of new user, a page design that demographically similar users have preferred will be suggested. By following the above procedure, a page design that matches the demands of each user predicted and recommended.

6. ANALYSIS OF DATA MINING METHODS FOR RECOMMENDER SYSTEMS

Recommender Systems (RS) implement techniques and methodologies of Data Mining (DM) for information extraction such as Similarity measures, Sampling, Dimensionality Reduction, Classification, Association-Rule-Mining (ARM) and Clustering[4]. Similarity Measures is one of the preferred approaches to collaborative filtering (CF) recommenders is to use the kNN classifier. This classification approach is based on defining an appropriate similarity or distance measures. Processing the entire data set is computationally too expensive. This problem can be solved using one another technique of data mining, sampling. Sampling includes selecting a subset of relevant data from a large data set. It is used for both pre-processing and final data interpretation steps. It can also be used to create training and testing datasets. Dimension Reduction is a another data mining technique which shows the notions of density and distance between points, which are critical for clustering and outlier detection, become less meaningful in highly dimensional spaces. This is known as the Curse of Dimensionality. Dimensionality reduction transforms the original high-dimensional space into a lower-dimensionality. There are two most relevant dimensionality reduction algorithms in the context of Recommender System such as Principal Component Analysis (PCA) and Singular Value Decomposition (SVD). Principal Component Analysis (PCA) is a classical statistical method to find patterns in high dimensionality datasets [7]. Singular Value Decomposition (SVD) is a powerful technique for dimensionality reduction [8].

Another technique of data mining is classification, it is a mapping between a feature space and a label space, where the features represent characteristics of the elements to classify and the labels represent the classes [4]. A restaurant Recommender System, for example, can be implemented by a classifier that classifies restaurants into one of two categories (good, bad) based on a number of features that describe it. There are various types of classifiers, but we discuss about either supervised or unsupervised classification. In supervised classification, a set of labels or categories is known in advance and we have a set of labelled examples which form a training set. In unsupervised classification, the labels or categories are unknown in advance and the task is to suitably arrange the elements at hand. One another technique of data mining for recommender systems is clustering. It is a process that assigns items to groups so that the items in the same groups are more similar than items in different groups. The goal is to discover meaningful groups that exist in the data [9]. Similarity is estimated using a distance measure. The objective of a clustering algorithm is to reduce intra-cluster distances while maximizing inter-cluster distances. There are two main categories of clustering algorithms such as, hierarchical and partitioned. Partitioned clustering algorithms divide data items into non overlapping clusters such that each data item is in exactly one cluster while Hierarchical clustering algorithms successively cluster the items within found clusters, producing a set of nested cluster organized as a hierarchical tree. The k-means clustering and DBSCAN algorithm is a partitioned clustering. There is another semantic document clustering technique, in which annotated pages will be created semantically and these are grouped into clusters. Such type of categorization is achieved by clustering the web document based on semantic similarity between the ontology terms that characterize them [10].

8. CHALLENGES AND LIMITATIONS

In this section, we present some of the common hurdles in deploying Recommender Systems:

Sparsity: Most users do not rate most items and hence the user ratings matrix is typically very sparse. In Collaborative Filtering systems sparsity decreases the probability of finding a set of users with similar ratings.

The Cold-start Problem: New items and new users problems in recommender systems are collectively known as the coldstart problem [3][6]. In Collaborative Filtering systems, an item cannot be recommended unless some user has rated it before. This problem applies to both new items and obscure items. The new-item problem is also often referred to as the first-rater problem. On the other hand content based approaches do not rely on rating of user. In Content based approach, the content-based predictions of similar users can also be used to further

improve predictions for the active user [10]. But, the new-user problem is difficult to tackle, since without any previous knowledge of user choice it is not possible to find similar users or to build a content-based profile.

Fraud: Commercial websites are using Recommender Systems widely for personalizing their site according to customers adopted by commercial websites that affects the overall profitability of sellers. This has provoked many unsocial vendors engaging in different forms of fraud to use recommender systems for their benefits. Either, they attempt to increase the average desirability of their own products (push attacks) or decrease the ratings of their competitors (nuke attacks). These types of attack are often referred as shilling attacks [12] or profile injection attacks [7].

9. CONCLUSION & FUTURE SCOPE

In this technical review, recommender system is examined as a technique for website personalization. We discuss current recommender systems and identified recommendation techniques, as a technology for designing Website personalization system, which recommends items to customer by identifying the user's with similar preference. Then we discuss recommender system design and structure of recommender system. We also study data mining techniques that are used to mine the web data for personalization. After this investigation we can conclude that sparsity and Cold-start are the two problem of present recommender system and some other architectural framework of recommender system should be implemented to tackle these problems.

REFERENCES:

[1] Xiaosheng Yu, Shan Sun " Research on Personalized Recommendation System Based on Web Mining" In Proceedings of the 2010 International Conference on E-Business and E-Government , 2010, pp. 346-349,ACM Press.

[2] J. Ben Schafer, Joseph Konstan, John Riedl "Recommender Systems in E-Commerce"Proceedings EC'99 Proceedings of the 1st ACM conference on Electronic Commerce, pp. 158-166, New York, USA1999.

[3] Prem Melville & Vikas Sindhwani "Recommender System" In Encyclopaedia of machine learning, chapter 705,2010,pp. 829-838,Boston ,MA ,2010.

[4] Sanjeev Kumar Sharma, Dr. Ugrasen Suman "Design and Implementation of Architectural Framework of Recommender System for e-Commerce" In Proceedings of International Journal of Computer Science and Information Technology & Security (IJCSITS), Vol. 1, No. 2, December 2011

pp. 153-162.

[5] Jong-Hwan Seo, Kun-Pyo Lee "Development of Website Design Personalization Service Using Design Recommender System" In: Proceedings of the 6th Asian Design Conference, October 2003.

[6] Prem Melville and Raymond J. Mooney and Ramadass Nagarajan "Content-Boosted Collaborative Filtering for Improved Recommendations" Proceedings of the Eighteenth National Conference on Artificial Intelligence (AAAI-2002), pp. 187-192, Edmonton, Canada, July 2002.

[7] Jolliffe, I.T. Principal Component Analysis. Springer, 2002.

[8] Hartigan, J.A. 1975. Clustering Algorithms (Probability & Mathematical Statistics). John Wiley & Sons Inc.

[9] Paterek, A. "Improving regularized singular value decomposition for collaborative filtering." In: Proceedings. of the KDD Cup and Workshop,2007, pp. 39-42, ACM , New York.

[10] Sanjeev Kumar Sharma, Ugrasen Suman, "A Semantic enhanced data mining framework for web Personalization," In: Proceedings of International conference on data analysis, data quality and metadata management (DAMD-2010), Singapore, 2010, pp. 49-57.

[11] Eirinaki, M., Vazirgiannis, M., Varlamis, I. "SEWeP: Using Site Semantics and a Taxonomy to enhance the Web Personalization Process." In: Proceedings. of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2003, pp. 99–108.

[12] Mukherjee, R., Jonsdottir, G., Sen, S., Sarathi, P. "MOVIES2GO: an Online Voting based Movie Recommender System." In: Proceedings of the Fifth International Conference on Autonomous Agents, 2001, pp. 114–115., ACM Press.