www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 2 Issue 12 December, 2013 Page No. 3556-3560

Semantically Driven Personalized Recommendations on Sparse Data

Kaushik, Tridev [1], Srivastava, Kavita [2]

[1]MCA Student, IITM, Janakpuri, Delhi
3devkaushik@gmail.com

[2]Associate Professor, IITM, Janakpuri, Delhi
srivastava.kavita1507@gmail.com

Abstract: *Recommendation techniques play significant role in the design of online applications especially in e-commerce applications. Generally recommendation techniques use filtering methods. Filtering methods fall in the category of content filtering and collaborative filtering. Content filtering requires the matching of user's profile with product features. Content filtering doesn't take into account the similarity among users' profiles. Another filtering method is called collaborative filtering. Recommender systems which use collaborative filtering also consider the similarity among other users' profiles also for providing recommendations. Such kind of recommender systems compute the ratings about a product feature by considering ratings specified by users with similar profiles.*

Recommender systems often suffer from the problem of sparse data. If the products to be sold online have several features and all features must be rated by the users and if the product is promoted online and survey is presented to thousands of online users, it may happen that not all users participate in the survey. Even if all users participate in the survey they do not provide ratings on all features of the product. It results in several missing values in user-item matrix. This matrix is sparse in nature. If matrix with sparse data is presented as input to the recommender system, the recommender system may not work correctly on it. Therefore the missing values must be filled before the data is fed to the recommender system.

In this paper we propose an approach to handle the problem of sparse data by using user profile similarities in a social net work. Each user's profile is augmented with an additional attribute called trust. The value of trust represents the degree of trust of social network users on the given user. When a user completes a given survey for a product and he/she skips one or more ratings, then the trust value from his/her profile is retrieved to fill this value. Next this user's friends' list is retrieved and the rating specified by these users are also retrieved. On the basis of these rating values and the trust value, missing rating value is computed.

Experimental results show that the rating values are computed with reasonably good accuracy.

Keywords: Cold Start, Ontology, RDF, Recommender System, Semantics, Sparsity, Trust.

1. Introduction

The use of Internet is an integral part in our daily life and we choose online applications to carry out different tasks whether it is viewing the balance in our bank account, getting information on a product or learning online. As we surf Internet and browse web pages we look for relevant information. Our browsing behaviour also indicates what we are looking for and web page developers can take advantage of this information [1].

Traditional businesses are also expending their market by making their products available online and let customers view and purchase the products online. Recent years have seen manifold emergence of e-commerce websites. Now it has become hard for users to get the best product they are looking for at best price. It is a challenge for an e-commerce website to provide right information to the appropriate customer. When a user navigates a website, he/she is presented with a list of recommended items after spending certain amount of time in browsing. When user makes a purchase, then also he is provided with a list of recommended items. The more relevant the items are to a user, more satisfying his/her experience with the website and ultimately it will lead to a returning customer.

Hence, recommender system is a significant element for an e-commerce website. A recommender system is a software component that analyses the user's current and past browsing behaviour and integrates it with the profile of the user and makes use of this whole information for matching with the product profiles.

In order to generate recommendation list, recommender systems rely on the user profile information, users' browsing behaviour information, users' buying history and ratings provided by users on product surveys.

Recommender systems that use collaborative filtering techniques, also take into consideration the information obtained from several users. Collaborative filters recommend items purchased by users with similar profiles as the current user.

A schematic diagram of the recommender systems is given in figure 1.

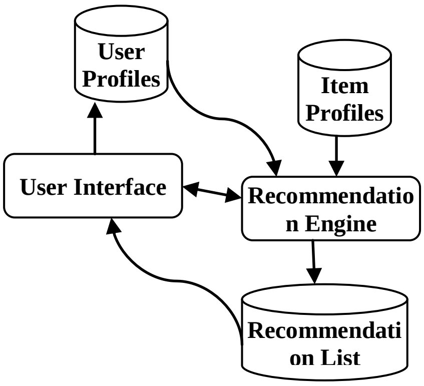

Figure I. A Recommender System

Item preferences of users are represented in rating matrix. Figure 2 shows a rating matrix.

U / P	U1	U2	…………	Un-1	Un
P1			………		
P2			………		
….			………		
Pn			………		

Figure II. Rating Matrix

Here U1, U2,….Un represent users to whom product is presented with an aim to get rating on all features P1, P2, …, Pn. In order to provide accurate predictions, the recommender system utilizes the ratings specified by a large number of users. There are several issues in obtaining sufficient information in rating matrix. A user can rate only very small proportion of the total items; there are many missing entries in the rating matrix. The nature of this rating matrix is sparse. Recommender system requires the values filled in place of missing entries for providing reasonably accurate estimate of users' choice.

The Cold Start Problem

Another problem faced by recommender systems is the cold start problem [2], [3]. This problem arises when a new item is being promoted. The ratings matrix does not contain any rating value for new products. It means recommender system would assume that no user has preferred this item. Therefore its value can never be upgraded from the initially assigned value until online users find out that item by themselves and then they choose to rate these items. So for a long time new and recent items hardly get their place in recommendation lists. There are several ways to alleviate the cold start problem. The new and recently added products should be given to a higher initial rating value and these items must be included in the recommendation list of the users who have well past buying history. Later when these users do not prefer to purchase these items, their ratings can be decremented appropriately. If some of the users prefer these items their profiles as well as rating matrix can be updated appropriately.

The Sparsity Problem

Sparsity problem is more severe because of two reasons[4]. First, it is difficult for a user to rate a large number of items. Second, a product may have a large number of features and users don't rate all of these features. Another issue is that users' preference may change over time and if the recommender system does not update its recommendation list with changing user preferences, its predictions may be wrong and result will be more annoying users.

In this paper we present a novel dynamic approach to alleviate the sparsity problem in rating matrix. Our approach is based on the assumption that preferences of user and his/her likelihood to purchase an item is similar to the users with similar profiles as the profile of the given user. Our approach takes advantage of the social network of a given user. If a user does not provide the rating about a product or product feature, these values can be computed with the help of ratings provided by users who are included in the friends' list of given user. So if the rating matrix is initially sparse, the missing values can be computed over time. After some time in operation, the rating matrix will retrieve the missing values. If there are certain features for which sufficient information cannot be obtained, these product features may be regarded as non-significant for users while making purchase decisions. These features can be removed from the rating matrix.

While adding a user in the friends' list, a trust value is also computed and added in that user's profile. The missing rating values can be computed using the trust value stored in user's profile and the trust values stored in profiles of other users who are included in the friends' list.

Main contributions of this research paper include (1) presenting a theoretical framework of semantically enhanced recommender system, and (2) presenting a trust based scheme to obtain missing rating values.

Rest of the paper is organized as follows. A literature survey of related work is given next. Then the proposed system is described. Experiments are done with the collected data and results are shown. The paper concludes with the conclusion and future scope of improvements.

2. Literature Survey

There are numerous methods which have been proposed to deal with the problem of missing data [7], [8], [9]. Majority of research focussed on methods specific to applications. The recommendations provided by these methods may be better, but they are not generalized methods. There are also approaches proposed for dimensionality reduction of rating matrix [6]. Other techniques include application of associative retrieval techniques in the bipartite graphs of products and customers, similarity based on product features and content boosted collaborative filtering.

The dimensionality reduction approaches to deal with sparse data determine the least rated product features and non-responding users to remove rows and columns of rating matrix. The techniques for dimensionality reduction include Principle Component Analysis (PCA) [7] and Latent Semantic Indexing (LSI) [8], [9]. These approaches require information that is both difficult and expensive to obtain.

However, the objective of addressing sparsity problem is to find missing data and not losing potentially useful information. The proposed approach aims to fill missing values in rating matrix.

3. Proposed System

Generally the item preference of a user does not vary to a great extent. Users tend to like an item which is also liked by other users in their peer group. Here we introduce the procedure of obtaining missing rating values in the rating matrix.

The rating matrix, R(i, j) is generated for each product that is being promoted online. Here i represents the product feature $\boldsymbol{p}_i$ and j represents the user $\boldsymbol{u}_j$. The format of rating matrix is given figure 2.

To obtain initial rating values for a product, it is promoted on an online social network (OSN). When a user register on an OSN, his/her friends' list is created through the contact list stored in the email account which is used in registering on OSN. The user is asked to rate the item. Some of the users provide rating values to all features of the item whereas some users may skip certain ratings. The proposed method for obtaining missing values works as follows.

1. The rating matrix is scanned to find all the missing values and indexes of all the missing values R(i, j) are collected.
2. Then, the missing values are updated implicitly through trust values in the friends' list of this user and their corresponding rating values about that product feature or they are obtained explicitly from all users in the friends' list.
3. While working implicitly, the user which has not specified the rating value for a particular product feature $\boldsymbol{p}_i$, his friends' list is searched and trust values are retrieved.
   a. Retrieve the $\boldsymbol{p}_i$ value in the same row in all columns which belong to users included in the friends' list.
   b. Compute the combined trust value using (I).
   c. Compute the resulting pi value using (II).
   d. Write the computed $\boldsymbol{p}_i$ value in place of the missing rating value in rating matrix.
4. If there are not sufficient number of users in the friends' list who have rated this product feature, the rating is obtained explicitly from all these users by sending them a message.

5. Then the missing rating value is obtained as given in step 3.
6. If it is not possible to obtain the rating value about a product feature either implicitly or explicitly, it is assumed that this product feature is not relevant from users' point of view and can be discarded. The corresponding row from the rating matrix is removed.
7. The procedure stops when the rating matrix gets values in all cells and undesired rows are removed.

4. Experiments

For applying our proposed method, a survey is conducted about a latest mobile phone model. The survey is presented to the users of Online Social Network (OSN). The users who have participated in the survey were not included in the same friends list. The questionnaire used in the survey consists of number of mobile phone features. These features include ***battery life, design, and ease of use, durability, interface, connectivity, display, sound, camera and processor***. The partial dataset generated is given in table I.

Table I. Dataset

U (USER) / SPECIFICATIO	**U1**	**U2**	**U3**	**U4**	**U5**	**U6**	**U7**	**U8**	**U9**	**U10**
Battery life	2	5			2		2	2		3
Design	4				5					
Ease of use	4	5	4		6	5	6	2	3	5
Durability	4						3			
Interface	1				4		6		4	
Connectivity	4	3			3		5	2		6
Display	4						5		2	
Sound	1	7			4		4			3
Camera	5	5			6		2	3		4
Processor	4				2		7		4	

Users labelled U1, U2, U5, U7, and U8 are in the same friends' list. Now in order to fill missing rating values, the rating matrix is scanned for all empty cells.

Let us choose to fill the values not entered by the user U5. We have chosen this user because the column corresponding to this user has nearly all filled values except for the attributes Durability and Display. Therefore this user is an active participant in our survey, therefore considered as a significant user.

Now we must calculate the trust value that user U5 has on other users (U1, U2, U7 and U8) in the friends' list.

This trust value is computed by using profile similarity of these users. The profile similarity can be computed by using pearson's correlation coefficient r, given in expression (I) below [6].

$$r = \frac{\sum (X - \mu_X)(Y - \mu_Y)}{N \sigma_X \sigma_Y} \qquad \textbf{(I)}$$

Here X and Y are trait values in two profiles. Consider the profiles of user U1 and U5; let the trait values are given as follows.

<u>X (U1)</u>	<u>y(U5)</u>
.3	.8
.5	1
-.1	.4
.6	1.1
-.3	.2

By using (I), we have computed r=1.0. If trust calculated on a scale from 1 to 10, trust of U5 on user U1 is considered as 10.

Similarly we have computed, trust value of U5 on users U2, U7 and U10 also.

Their trust value is given in table II.

Table II. Trust Values

USERS	**TRUST VALUE (t)**
U 1	**10**
U2	**8**
U7	**7**
U8	**9**

Now, using these trust values, the missing rating values are computed by expression II, given below.

R(i, j) =

$$\frac{\sum_{k=1}^{n}}{n}$$

[k varies for all users in the friends' list of user j] **(II)**

5. Results

The rating matrix is filled with all values by using (I) and (II). The column belonging to U4 is discarded as it is not a significant user. Likewise the row belonging to attribute

Design is also discarded, as it is found to be non-significant attribute. The resulting values are given in Table III.

Table III. Rating matrix with all entries.

U (USER) SPECIFICATIO	U1	U2	U3	U5	U6	U7	U8	U9	U10
Battery life	2	5	**3**	2		2	2		3
Ease of use	4	5	4	6	5	6	2	3	5
Durability	4	3	3	**3**	3	3	4	4	
Interface	1	3		4		6		4	
Connectivity	4	3		3		5	2		6
Display	4		5	4		5		2	
Sound	1	7	6	4		4			3
Camera	5	5	4	6		2	3		4
Processor	4		5	2		7		4	

The values computed above are compared with values explicitly taken by users. The result is found to be satisfactory.

6. Conclusion

This paper describes how the accuracy of recommendation systems can be improved when semantically-enhanced methods are applied. In our approach, we make use of semantics by applying trust ontology in user profiles. This is a domain-based method that makes inferences about user's interests and a taxonomy-based similarity method is used to refine quality of rating matrix thereby improving overall results.

References

[1] A RECOMMENDATION SYSTEM FOR THE SEMANTIC WEB
Victor Codina and Luigi Ceccaroni

[2] Berners-Lee T., J. Hendler, and O. Lassila. 2001. The Semantic Web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American* 284 (5), 34-43.

[3] Blanco-Fernández, Y. et al. 2008. A flexible semantic inference methodology to reason about user preferences in knowledge-based recommender systems. *Knowledge-Based Systems* 21 (4), 305-320.

[4]. Cantador, I., Bellogín, A., and Castells, P. 2008. Ontology-based personalised and con-text-aware recommendations of news items. *Proc. of IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, 562–565.

[5] Codina, V. 2009 Design, development and deployment of an intelligent, personalized recommendation system. Master Thesis. Departament de Llenguatges i Sistemes In-formàtics, Universitat Politècnica de Catalunya. 101 pp.

[6] Fink, J. and Kobsa, A. 2002. User Modeling for Personalized City Tours. *Artificial In-telligence Review* 18 (1), 33-74.

[7] Gawinecki, M., Vetulani, Z., Gordon, M., Paprzycki, M. 2005. Representing users in a travel support system. *Proceedings - 5th International Conference on Intelligent Sys-tems Design and Applications 2005, ISDA '05*, art. no. 1578817, 393-398.

[8] Middleton, S.E., De Roure, D. C., and Shadbolt, N.R. 2001. Capturing Knowledge of User Preferences: ontologies on recommender systems. *In Proceedings of the First International Conference on Knowledge Capture (K-CAP 2001), Oct 2001, Victoria, B.C. Canada*.

[9] Sieg, A., Mobasher, B., Burke, R. 2007. Ontological user profiles for personalized Web search. *AAAI Workshop - Technical Report WS-07-08*, 84-91.