Survey on Personalized Web Recommender System

Santosh Kumar
Department of Information Technology
ABES Engineering College, Ghaziabad 201009, Uttar Pradesh, India
Email: santoshg25@gmail.com

Varsha
Department of Computer Science and Engineering
Krishna Institute of Engineering & Technology Ghaziabad 201206, Uttar Pradesh, India
Email: varsha327@gmail.com

Received: 24 November 2017; Accepted: 19 January 2018; Published: 08 July 2018

Abstract—Recommendation system plays an essential role in searching any information from World Wide Web. Recommender system handles Information straining problem and improve customer correlation by providing best services. It suggests items or services to users according to their interest, navigation behavior or demographic information. This paper performs a survey on different approaches available for recommender system and performs a comparative analysis of different algorithms. Further, a discussion about various application areas has been done. At the end, issues and challenges in recommender systems have been discussed.

Index Terms—Recommendation System, Navigation Behavior, demographic information.

I. INTRODUCTION

Extracting information from internet is like finding sapphire from the sea. There is huge volume of information in the Internet and finding the relevant and required information is tedious task. Users have many options and choices to get answer to their queries. Users face difficulty in effectively making decisions from too much information in response to their query. This problem is called information overload [33]. The solution to information overloading is the recommender system. Recommender systems handle this problem by providing the users a personalized, exclusive content and service. It uses data mining methods and countless recommendation algorithms to find the relevant information and services from the huge information source. It suggests items to users according to their likes, navigation behavior or demographic information [24]. In earlier days, recommendation was done through analysis of print media like newspapers, magazines, books, novels etc. Individuals make their query to the librarian about the books to suggest; this was time taking process and it took days or months to make any recommendation. But in this fast growing era, different business organizations and individuals have to take decisions instantly. The progress of social webs like Facebook, Twitter and Internet capable mobile devices like smart phones and tablets has further increased the volume of online dynamically growing information. These mobile devices empower users to freely make, upload and share data and add media objects such as text, images, videos etc. Recommendation system is being used to provide personalized information from this massive amount of generated information. Personalized recommendation system is a Data Filtering System which recommends items to the users based on consumers rating, user’s past navigation pattern, demographic knowledge etc. Recommendation system suggests the items like music, movies, news, books, images, web pages etc. Using recommender system information or items are produced to users in which they have comfort [1]. Recommender system uses several recommendation algorithms in providing personalized, exclusive content and services to the user. These Recommendation Algorithms can be categorized into collaborative filtering (CF) recommendation methodologies, content-based recommendation methodologies, and hybrid approaches [21]. There are many websites which use recommender system to recommend data or item to customers like Netflix, Rotten Tomatoes, Movielense, IMDB, Amazon,
eBay, Google News and social networking sites like Facebook, twitter, google+, LinkedIn, Instagram etc. to recommend friends or followers. In Fig. 1 recommender system offer recommendations of item to user based upon their past history of likes, purchases, or demographic information.

II. RELATED WORK

Jianxun Liu, et al. [8] proposed personalized recommendation system for various web services. In this competitive era there is large amount of web services available and finding the best web service is tough task. A Location- Aware Personalized Collaborative Filtering for Web Service Recommendation has been used. This technique uses Pearson Correlation Coefficient (PCC) to calculate the similarity between the customer and item and it improves QoS estimation. X. Chen, et al. [9] proposed a unique collaborative filtering algorithm used in large scale web services. A region-based CF algorithm which is working in two phases has been used. In first phase, the customers are divided into different areas based on their physical location. In the second phase, to offer web services, it discovers similar customers for the new customer and creates estimation of web services. Zhang J., et al. [10] proposed personalized recommendation to support “information overload”. Author used user-created labels of social labeling system to find customer choice for recommendation but it is tough to find customer with the parallel taste due to the customer lived in different countries and communicates in different languages. Author introduces the model that uses item features to make customer performance models through collaborative filtering. This model not only adds area features into a personalized recommendation but also spotting the unspecified relationship between customers. This method provides better result and makes recommendation easy for customers who lived in different domains. Feng H., et al. [11] proposed new recommendation plan called TOTAR (Temporal Overlapping Community detection using Time-weighted Association Rules). They proposed an approach based on Time weighted overlapping community detection and association rule mining. A number of modern plans to integrate, time effects have been proposed in TOTAR to improve the prediction of user dynamic interests over time. Author used datasets from Movie Lens and Netflix to compare TOTAR with multiple traditional plans. The trial results displays that TOTAR exceed other present algorithm in correctness and variety. They continue this study in a number of ways like in community detection, the selection strategy of the start nodes i.e., the seeding strategy. Shen J., et al. [12] had proposed Travel attraction recommendation system collects the knowledge from social media. Author describes the personalized attraction similarity (PAS) model to combine travel information and customer feedbacks to recommend attraction to appropriate customers. Author used Classification technique to search parallel attraction based on customer liking then the similarity between all attractions are classified into favorite and un-favorite groups with multi-modality fusion. This method exceeds several approaches with the welfare of combined intelligence and increases customer communication. The main motive of this method is to combine customer feedback and collective intelligence. Mould, D. R. [13] proposed recommendation by using trust clusters. Trust clusters make recommendation more effective and accurate. Method of recommendation by similarity factor is slower. Here Trust clusters are merged with memory based Algorithm and uses item based collaborative filtering and customer based collaborative filtering. To test approach [13] used Pearson correlation algorithm and Automated Collaborative Filtering. These algorithms improve the quality and accuracy of recommendation. Liu, Ning-Han [14] Music recommendation is important and progressive research area. Author used better similarity evaluation which improves the quality of music recommendation. Similarity evaluation is done by three methods First one is Minimum Likelihood Estimation (MLE), second is Genetic algorithm (GA) and third is Genetic Programing (GP). In [14] authors used Personalize distance function by using Euclidean distance function to calculate the similarity. These approaches are used to improve the quality and music selection mechanism. Polatidis, Nikolaos, and Christos K. Georgiadis [15] proposed a multilevel recommendation method to facilitate the decision making by suggesting recommendation of high quality and accuracy. Author used collaborative filtering to achieve accuracy in recommendation and for similarity factor, Pearson co-relation coefficient (PCC) and Weighted Pearson Correlation coefficient (WPC) has been used. To calculate the accuracy of the recommendation system Collaborative Filtering Delivers the Mean Absolute Error (MAE) metrics. These methods increase the quality and accuracy of recommendation and eliminate noisy users. Kang, Guosheng, et al [16] had proposed Active Web Services recommendation method which recommends Web Services to a customer based on the customer’s past history. This method calculates the similarity between the customer’s functional benefits and web services, based on the likeness in the functional and nonfunctional features. Active Web Service Recommendation System, organizes the web services in sequence of topmost recommendations for the customer on the basis of similarity matrix. Xu, Wei, et al. [17] proposed recommendation system to provide customized information services for business purpose. Author used two level recommendation methods to provide personalized R&D (Research and Development) project information for customers. Author used AHP technique as the rank aggregation method. In future AHP technique can be used with other data fusion techniques such as Condorcet fusion and other techniques that model score distribution to make more efficient and accurate recommendation. Shahabi, Cyrus, and Yi-Shin Chen [19] proposed Yoda structure to reduce the complexity in hybrid recommendation approach. Yoda is structured as a
tunable model recommendation for off-line and employed for real-time recommendation for on-line. Author used prolonged Yoda to contain more recommendation sets. The recommendation sets can be achieved from various sources, such as social experts, web navigation patterns, and collections of customer evaluations. Moreover, the extended Yoda can learn the confidence values routinely by using implicit customers’ relevance response through web navigations using genetic algorithms (GA). Yoda reduce the complexity of system as the amount of customers or items and accuracy is significantly increased by using the GA-based learning mechanism. Naruchitparames, Jeff, et al. [20] proposed recommendation system for social networks web services which has complex and heterogeneous large data that is difficult to handle. Author used method to associate recommendation systems via complex network system, rational theory and a Pareto optimal genetic algorithm in a two-step approach to provide quality, colleague recommendations while concurrently determining an entity’s perception of relationship [19]. Merging network topology and genetic algorithms produce better recommendations. Network topology and genetic algorithm represents strong potential for developing link recommendation systems using this combined approach of personal interests and the underlying network. Lopes, Prajyoti, and Bidisha Roy [23] provide real time dynamic recommendation to registered or unregistered users of website. Authors proposed an action based rational recommendation technique which makes use of lexical patterns to generate recommendation of items. Further showed the effectiveness of proposed approach by evaluating it on real time e-commerce data and comparing the system with user based and product based techniques. Authors shown that this system results into good quality and accuracy. Also it minimizes limitations of traditional recommendation system. system. Abderrahim, Naziha et al. in [31] proposed a social trust-aware system for recommendation of web services. This approach has been compared with conventional similarity-based and trust-based service recommendation methods and it performed better. Authors have shown experimentally that the proposed service recommendation method significantly outperforms conventional similarity-based and trust-based service recommendation methods. Fridi, Asmaa et al. in [32] have used semantic information from Linked Open Data to enhance the effectiveness of traditional collaborative filtering to improve the results.

III. RECOMMENDATION SYSTEM

The area of recommendation system is becoming widespread because at the present time information is weighed down everywhere and in this state recommendation comes for the assistance [1]. In today’s world Internet is key part of our life whether it is in Education, Hospitals, Music, Online Books, Online Shopping, e-Commerce and movies sites. Users of World Wide Web are suffering from information overload. The solution to this problem is recommendation system. Recommendation systems are data filtering system that recommends products or services based on users rating, past activities or navigation patterns, demographic information etc [24]. Recommender system makes decision making easy and fast for the customer and help customer to find essential data for customer from the large sea of information there are various types of recommendation systems like book recommendation system, Friend recommendation system, Music recommendation system, News recommendation system etc. uses different approaches to make convenience for customer to find relevant items Recommender system is broadly categorized its techniques as Content-based, Knowledge-based, Collaborative Filtering and demographic-based [2, 3]. Classification of recommender system techniques is shown in Fig. 2. and is illustrated in next section.

A. Recommendation System Techniques

A.1 Collaborative Filtering

Collaborative filtering based reference methods benefits people to make decision based on the choices of other individuals who share parallel interests [29]. In this customers with similar interest are discovered. Customers giving identical rating to any product are considered as customers of similar interests. Such customers form a cluster or a zone. A customer rates those items that he has not rated earlier [5, 6]. As an example consider there are two users one is user1 and another is User2 if User1 likes item {A, B, C} and User2 likes item {A, B, C, D} then according to the similar choices of {A, B, C} between two users item D is recommended to User1 because both User1 and User2 has maximum similarity of item {A, B, C} so both users consider as user with similar taste. Collaborative Filtering algorithms separated into two wide classes: memory based and model based algorithms [1].

A.1.1 Model Based Collaborative Filtering

Model based collaborative filtering is based on offline mode. This technique compresses the complicated data set into lower dimensional set. It uses the Bayesian Network to capture relationship between user and item relation. Matrix factorization is another important technique that is first to point the QOS (Quality-of-service) prediction problem [22].

A.1.2 Memory Based Collaborative Filtering

Memory based collaborative filtering depends on an item to item or customer to customer similarity to mark estimation for customer on the items which customer has not used yet. If we put additional dataset into memory and added new data into existing set it reduces the performance of big dataset. This difficulty can be overcome by Recalculating correlations like Pearson correlation and vector correlation. Memory based collaborating filtering method are separated into two classes one is customer based collaborative filtering and another is item based Collaborative filtering [27].

Copyright © 2018 MECS

I.J. Information Engineering and Electronic Business, 2018, 4, 33-40
i) Item based collaborative filtering

Item based Collaborative Filtering recommends items based on comparison between item-item similarities. This method first discovers the correlation between the customers and then forecasts the rating of item for the customers [27].

Drawback- Data sparsity problem, Gray sheep problem, Cold-start problem.

ii) Customer based collaborative filtering

Customer based collaborative filtering suggests items, products or services based on customer-customer similarity. Calculation of rating to items is completed depending upon the similar item is rated by different or similar type of customers [27].

Drawback- Data sparsity problem, Gray sheep problem, Cold-start problem.

A.2 Content based filtering

Content based filtering recommendation method searches items with the help of previous history or navigation pattern of customer [7]. Content-based recommendation technique is based on the data about item content and ratings of customers have given to items. This system combines these ratings to profile of the customer’s likes based on the characteristics of the rated items for example if user liked shapes of green square and green circle then system start building customer profile like {green, square, circle} and match this with other similar items like {circle, pentagon, green, cube} and recommend these to user with the help of user past likes and navigation pattern. Content based recommender systems perform well with fresh items. Content based filtering uses two types of methods, first regular data retrieval that is “cosine similarity measure” and second is “machine learning method” [23].

A.3 Hybrid Recommendation Systems

Hybrid recommendation system is association of two or more techniques to gain more powerful methods. Most of the users like Amazon, Facebook, Pandora and Netflix that use hybrid recommender system. The most general hybrids are those of content-based and collaborative filtering. It benefits to achieve better performance and overcome the drawbacks of regular recommendation methods [25]. Hybrid of Content-based and collaborative filtering system help to avoid cold start problem and is able to notice user’s interests in the form of products or services.

A.4 Demographic-based Recommendation Systems

Demographic recommendation systems are based on the demographic summary of the customer. Demographic-based recommendation suggests the product by considering the user’s demographic data like sex, age, date of birth, scholastic qualification and other individual structures. Demographic recommender assumes that customers with same demographic data will rate the products similarly and recommender obtains cluster of similar demographic data. For example if two users user1 and user2 share most probably the same demographic data and give rating to same product then demographic recommender system will obtain cluster of similar demographic data [30]. Here demographic data is needed to forecast best rating for recommendation of items to the customers [28].

A.5 Knowledge-based Recommendation Systems

Knowledge-based Recommendation Systems suggests items by considering the customer past history, then it forecasts the customer’s liking based on the implications. Knowledge based recommender system does not consider customer rating rather it uses case based reasoning. Here items are suggested in the form of cases and it suggests cases to most similar customer’s profile. [26]
B. Pros and Cons of Recommendation System

<table>
<thead>
<tr>
<th>Techniques / Algorithms</th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collaborative Filtering</td>
<td>• Knowledge about the item features is not needed.</td>
<td>• Get worse by grey sheep problem, cold-start problem, shilling attack, and malicious customers’ rating.</td>
</tr>
<tr>
<td></td>
<td>• Collaborative filtering works independently because with CF technique one can look the preferences of individual users from outside the box. Quality of items can be easily evaluated.</td>
<td>• Here the quality depends on historical rating of user.</td>
</tr>
<tr>
<td>Content based Filtering</td>
<td>• Capable to endorse fresh and unpopular items.</td>
<td>• Unable to notice users interest in the form of products or services.</td>
</tr>
<tr>
<td></td>
<td>• Not any cold start and sparsity problem.</td>
<td>• Difficult to find product value.</td>
</tr>
<tr>
<td>Hybrid Recommendation Systems</td>
<td>• Overcomes the difficulties of content based filtering and collaborative filtering.</td>
<td>• Unsuitable grouping of recommender method leads to poor recommendation or decline precision outcome.</td>
</tr>
<tr>
<td>Demographic-based Recommendation</td>
<td>• It provides the recommendation before customer gives the rating to item. It is not centered on customer item rating.</td>
<td>• There is a Probability of improper personal classification.</td>
</tr>
<tr>
<td>Knowledge-based</td>
<td>• It is based on unambiguous knowledge so no need to provide rating to items.</td>
<td>• Customer profile conservation is one of the challenging tasks.</td>
</tr>
</tbody>
</table>

IV. APPLICATION AREAS OF RECOMMENDATION SYSTEM

- **Entertainment** – Entertainment comprises movies and music type of recommendation system to provide relevant items to customers without information overload e.g. - YouTube, Netflix, Rotten Tomatoes, Movielense, IMDB, Gnoosic, Shazam etc.
- **E-learning** - E-learning sites use recommendation system to indigently recommend actions based on actions of previous learners. Learners can read relevant information by their choice e.g. - E-learning for kids, Kindle, Google News, Amazon, eBay etc. The recommendations for e-learning could be an online activity such as doing an exercise, reading posted messages on a conferencing system, etc [34]
- **E-Commerce** - E-Commerce sites comprises online book purchase, online shopping, e-banking, Health, Education, government, Taxes, business, tourism etc. e.g. - Amazon, Levies, eBay, Snapdeal, Paytm etc.

![Fig.3. Area of applications](image-url)
V. CHALLENGES AND ISSUES OF RECOMMENDATION SYSTEM

A. Cold start

It is challenging to make recommendations for new customers whose profile is almost empty also customer has not liked or rated any item so likelihood in this case is unknown to the system. This is called cold start problem. If any item is fresh then that can also suffer from cold start problem. Both of these new customer and new item problems can be resolved by hybrid approaches.

B. Trust

Trust relationships between users can useful in the design of social recommender systems. These systems use the idea that users in the same social network tend to share similar interests. Existing recommender approaches based on social trust relationships do not fully utilize such relationships and thus have low prediction accuracy or slow convergence speed [35].

C. Scalability

As the increase of numbers of customers and items, the system needs more resources for processing the information and making recommendations. Majority of resources are used with the purpose of finding users with similar tastes and items with similar descriptions. This problem can be solved by using different types of filters.

D. Sparsity

Many online shopping websites have a large number of customers and several items. There is always a possibility that customers can rate just limited items or like inappropriate items. So by using collaborative and other methods recommendation system generate areas of similar type of customer profiles. If customers liked just few items then it is hard to determine customers of parallel taste because of lack of data.

E. Privacy

Recommender systems are dependent on the user’s feedback. This feedback reveals the information about the user’s interests like their political opinions, sexual orientations, and personal preferences. Sometimes such information can be highly sensitive, which leads to privacy concerns. This requires advanced algorithms to be used by recommender system.

F. Ramp-up problem

It is similar to cold-start problem. In this fresh items cannot be suggested to any customer till they acquire some kind of rating.

G. Synonym problem-

Most recommender systems are unable to differentiate between closely related items. Collaborative filtering is unable to compute the similarity between such items. Different methods, such as automatic term expansion, the construction of a thesaurus, and Singular Value Decomposition (SVD) can solve synonymy problem [33].

H. Shilling attack

Attackers can introduce preferred profiles into recommender systems so that the recommendation results are biased according to them that may lead to a noteworthy negative impact on the robustness of the systems.

I. Over Specialization Problem

The learning methods applied on content-based filtering try to find the most relevant documents on the basis of user’s past navigational pattern. This approach recommends only those type of documents which have been seen in the past ignoring the exploration of new documents. This problem is called over-specialization problem. As user’s interest changes with time, recommender system does not recommend the documents according the current interest of users. Instead of adapting to the user’s interests after the system has received feedback one could try to predict a user’s interests in the future and recommend documents that contain information that is entirely new to the user.

A recommender system provides recommendations to the user after deciding between two types of information:

- Exploitation: The system chooses documents similar to which user has shown a preference.
- Exploration: The system chooses documents where the user profile does not provide evidence to predict the user’s reaction.

VI. CONCLUSION

In this paper various Recommendation System Techniques like Collaborative Filtering, Model Based Collaborative Filtering, Memory Based Collaborative Filtering, customer based collaborative and item based collaborative filtering have been discussed. Further recommender systems have been categorized into hybrid, demographic-based, and knowledge - based recommendation systems and their pros and cons have been pointed out. Then papers discusses about various application areas of recommendation system. At last different challenges and issues of recommendation system like Cold start, Trust, Scalability, Sparsity, Privacy, Ramp-up problem, Synonym problem, Shilling attack, Over Specialization Problem are discussed. Collaborative filtering has cold start problem and not suited for new users. In content based filtering there is some outlier navigation pattern due to which recommender system is unable to find clear interest of users thereby providing undesired recommendation. While using hybrid approach these problems are somehow minimized.

REFERENCES


Authors’ Profiles

**Dr. Santosh Kumar** is Professor at ABES Engineering College, Ghaziabad and has completed his PhD from the School of Computer and Systems Sciences, Jawaharlal Nehru University, New Delhi, India. He completed his MTech from Allahabad Agricultural Institute, Allahabad, Uttar Pradesh, India after completing his BE in Computer Science and Engineering from Madan Mohan Malviya Engineering College, Gorakhpur, Uttar Pradesh, India. His research interests are Web Mining, Databases and Data Warehousing.

**Ms. Varsha** is M. Tech. Scholar in Computer Science and Engineering department at Krishna Institute of Engineering & Technology Ghaziabad, India. She has completed her B. Tech in Information Technology from Dr. K. N. Modi Institute of Engineering and Technology, Ghaziabad, India. Her research interests are Web Mining and Data Mining.

How to cite this paper: Santosh Kumar, Varsha," Survey on Personalized Web Recommender System", International Journal of Information Engineering and Electronic Business(IJIEEB), Vol.10, No.4, pp. 33-40, 2018. DOI: 10.5815/ijieeb.2018.04.05