A Survey on Recommendation System

Debashis Das
M.Tech, School of Computer Engineering KIIT University, Bhubaneswar, India

Laxman Sahoo
Professor, School of Computer Engineering KIIT University, Bhubaneswar, India

Sujoy Datta
Professor, School of Computer Engineering KIIT University, Bhubaneswar, India

ABSTRACT
Recommendation systems have become extremely common in recent years. It helps the customer to discover information and settle on choices where they do not have the required learning to judge a specific item. It can be utilized as a part of different diverse approaches to encourage its customer with effective information sorting. It is a software tool and techniques that provide suggestion based on the customer's taste to discover new appropriate thing for them by filtering personalized information based on the user's preferences from a large volume of information. Users taste and preferences should be constructed accurately in order to provide most relevant suggestions. This survey paper compare's and details the various type of recommender system and popular recommendation algorithms and its uses.

Keywords
Recommendation system, Types of the recommendation system, Feedback techniques

1. INTRODUCTION
Data become the key factor for everything, but nowadays the size of data is increasing exponentially. In June 2015, India had a web client of around 354 million and is likely to reach 500 million in 2016, in spite of being the 2nd biggest client hub on the planet. The infiltration of e-commerce of web based business is low appeared differently in relation to business areas like the United States (266 million, 84%), or France (54 million, 81%), yet is creating at an unprecedented rate, including around 6 million new members consistently every month[1]. This amount of datasets cannot be managed efficiently by the common databases management system. The datasets in the form of semi-structured data and unstructured data like image, audio, video, JSON documents, wet log and search patterns etc. cannot be stored and handled by traditional databases, so the concept of Big Data came into the picture.

According to IBM," Everyday, internet user generates 2.5 quintillion bytes of information- so much that 90% of the information on the planet today has been generated In the most recent 2 years alone [2]. This information originates from everywhere, for example, social media posts, images, videos, transition records of both e-commerce and no e-commerce, satellite data etc. This data is called Big Data. Tech America Foundation describes big data as " Big data is a term that defines huge volumes of high velocity, complex and variable data that require advanced techniques and technologies to enable the storage, capture, distribution, management, and analysis of the information"(Tech America Foundation's Federal Big Data Commission, 2012)[3].

Initially, Big Data describe by 3 V's (Variety, Volume and Velocity). Volume is described as the quantity of information produced by individuals or organizations. The sources may be internal or external. Velocity is defined as the rate at which data is generated. Variety is represented as various sort of information extracted from various sources like Facebook, twitter in addition to various feedback websites. In addition to the 3 V's, other V's have also been mentioned (veracity, variability, value). IBM mentioned Veracity as the fourth V, which define the uncertainty of information. Variability: SAS introduced Variability and Complexity are two more dimensions of big data. Value: Oracle presented Value as a defining attribute of big data.

In our day to day life, indeed, even to settle on requirements like which motion picture to watch, which novel to peruse, where to eat, we rely on our associates, news on the daily papers, what's more, common reviews, and so forward to help us find what is great for us. This support from our surroundings gives us an easy way to find out the best alternative without having much effort to filtering through the different choices available in the market. In this era of technology, the Recommendation system is an application that filtered personalized information and gives the way to understand a user's taste and to suggest appropriate things to them by considering the patterns among their likes and ratings of various things.

In the area of both e-commerce and no e-commerce, recommendation system is broadly studied and used to achieve maximum profit and to fulfill the precision marketing goal. Schafer et al. researched how RS support e-commerce websites to maximize profits and analysis the recommendations stems at numerous market-leading websites [4]. Amazon [5] got a hike up 20%-30% on sales by using recommender system [6].

2. RECOMMENDATION SYSTEMS
Have you ever amused how the "People you may know" feature on Facebook or LinkedIn? This feature suggests a list of people whom you might know, who are similar to you based on your friends, friends of friends in your friend circle, current location or may be past location, skill sets, groups, liked pages, and so on. These recommendations are specific to you and differ from user to user.

The First RS [7] was created by Goldberg, Nichols, and Oki&Terry in 1992**.** A recommendation system is an approach to the issue like to provide suitable things to the customer despite of searching lots of items. Although People's tastes vary from one to another but they also follows some pattern. RS are software tools and techniques that provide suggestion based on the individual's taste to discover new required content for them like useful products on e-commerce sites like amazon.in, videos on YouTube, posts on the wall of the social media like Facebook, News recommendation on online news websites automatically. RS perceive suggestion's consequently to the customers by analyzing previous browsing history, the feedback assigned to the products and different user's behavior.

Recommendation systems usually produce a number of suggestions in one of the given techniques (Figure. 1)

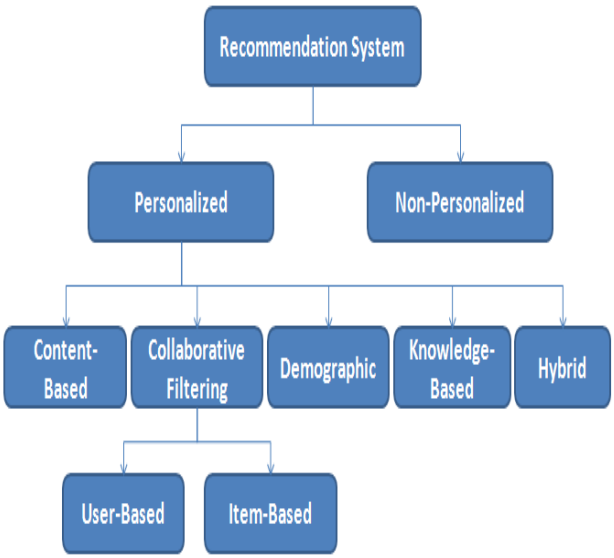

Figure. 1: Types of Recommendation System

2.1 Personalized Recommendation System

In this type of recommender system, the system goals at recommending users desired items based on their past behavior also interpersonal relationship of social networks by considering three perspectives: 1) Interpersonal impact, which implies whom you would believe, 2) Interest circle derivation, which implies whose interest is similar to yours, 3) User individual interest which has the effect on what items you would enthusiasm for [8].

Personalized recommendation systems are categorized into five different kinds be influenced by their approach to recommendation [9].

2.1.1 Content-based filtering

This type of filtering techniques filtering built on the customer preferences and description about the item. Basically, these algorithms aim to suggest items or product which are alike to that items that user enjoyed in past or is looking at in the present-day.

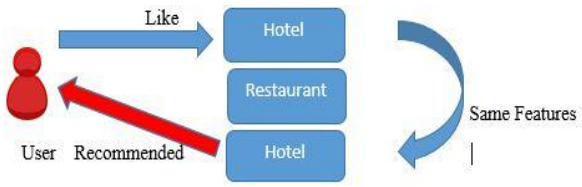

Fig. 2: Content-Based Recommendation

Merits

1. Other user's data not required.

2. No data sparsity as well as cold start.

Demerits

1. Content analysis is essential to define the item features.

2. The excellence of the product can't be estimated. The likeness calculation is incomplete to the product description.

2.1.2 Collaborative filtering

Collaborating filtering approaches build the system by considering the user's past behavior (rating is given to those items, previously parched or chosen items) and an additionally similar decision made by different users, then use the system to calculate the item or else rating that the user may perhaps interested in.

User-based CF algorithms makes suggestions by considering users having similar interest. It relates user as per the rating is assign to the product.

In the Figure. 3, in the 1st place user identified with the 3rd user rather second because the rating given by the third user is very alike to the 1st one. That is the reason item 3 is suggested to the user.

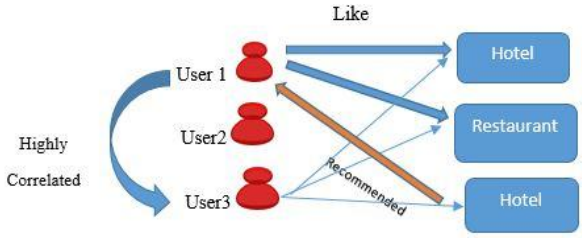

Figure. 3: User-Based Collaborative Filtering

Item-based CF algorithms depends on the items as the user rated items comparably are probably similar. From Figure. 4, user2 and user3 rated item-1 and item-3 so, it assumes that item 1 and 3 become similar. As user1 like item 1, item 3 is suggested.

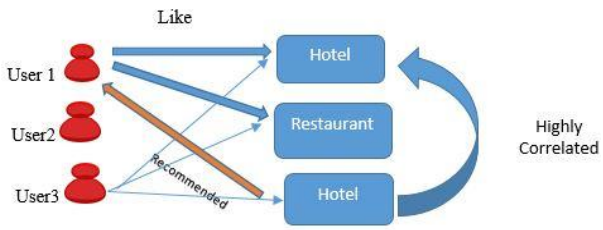

Figure. 4: Item-Based Collaborative Filtering

Merits

1 .The excellence of the item can be estimated through user ratings.

Demerits

1. Cold start problem for different users and new products.

2. Stability vs. plasticity issue.

2.1.3 Demographic filtering

Demographic recommendation technique only considers the data of the user like age, gender, employment status about the user only home possession and even location also. The recommends is made by considering demographic similarities to the user.

Merits

The technique is domain independent because Item feature is not needed.

Demerits

1. Collection of demographic information give rise to privacy issue.

2. Plasticity vs. stability issue.

2.1.4 Knowledge-Based recommender system

Consider a scenario like we don't buy house, car etc. frequently so in these scenarios the rating about the items

doesn't play a great role, in these type of situation knowledge-based recommender system came into the picture. This type of system deployed in a specific domain where the perches history is small. In these type of system, the algorithm considers the knowledge about the item and its feature, user preference (asked explicitly), recommendation criteria before giving the recommendation. The accuracy of the model is judged in view of how helpful the prescribed thing is to the client.

Before building these sorts of recommended frameworks, we anticipate the accompanying inquiries:

- What kind of information about the items is taken into the model?
- How are user preferences captured explicitly?

There are two basic type of knowledge-based recommender system.

- Constraint-based
- Case-based

Case-based recommenders center on the retrieval of comparable things on the premise of various sorts of likeness measures, whereas Constraint-based recommenders depend on with respect to an unequivocally characterized set of recommendation principles.

2.1.5 Hybrid recommender system

Hybrid recommender systems are built by joining different recommender systems to build a more robust framework. By combining several recommender systems, we can reduce the demerits of one method through the merits of one more system and accordingly construct a more robust system. For instance, through combining collaborative filtering techniques, where the model fails after new items don't have ratings, through content-based methods, wherever feature info around the items is accessible, new items can be recommended more precisely and powerfully.

Previously constructing a hybrid model, we study the subsequent queries:

- What methods must be combined to accomplish the business solution?
- How should we combine a number of systems and their outcomes for good predictions?

3. RECOMMENDER SYSTEM FEEDBACK TECHNIQUES

The basic process of an RS is information feedback, as it gives the data that RS needs with a specific end goal to give suitable suggestions to the clients in view of their preferences. Basically, feedback techniques are divided into three types [10]:

3.1 Implicit Feedback Technique (IFT)

By using this technique information is obtained without user's consciousness but obtained based on the user's action during the process. User's taste and interest are measured without seeking for user's consent. An IFT captures and interprets user's feedback by using application domain-dependent tools and some methodologies. This type of IFT can be found in various applications such as browsing history, web consumption history, and mouse movements or even search pattern.

Merits

1. IFT can be collected at much lower cost.
2. IFT is unproblematic, it does not put a load on the user of the recovery system.
3. It is less accurate as compare to EFT but large information can be collected at a lower cost.

Demerits

1. IFT is vulnerable to noise.
2. IFT is less accurate compared to EFT.
3. IFT is difficult to interpret.

3.2 Explicit Feedback Technique (EFT)

This approach involves the users for assigning either numeric or score rating for evaluating the product. The common scenario of explicit ratings is completed on an arranged discrete scale (example-Mark out ten). Ratings given on these measures permit these judgments to be handled statistically to give distributions, average and so on. The EFT helps users in expressing their interest and taste on the particular object [10, 11].

There are three core methodologies to get clear relevance feedback [12, 13]:

1. Like/dislike – items or goods are categorized as relevant/ irrelevant using a binary rating scale.
2. Ratings – judgments of items or goods are made using a numerical scale.
3. Text comments- comments about the item is obtained.

Merits

1. EFT is simple to use [14].
2. The feedback can be done either be positive or negative i.e. helps the user to specify what they like and what they don't like [14].
3. The accuracy of EFT seems to be higher than IFT.

Demerits

1. EFT is absolute. For example, a user listening to music many times, the users can still express is interested if the user listens to the music once without listening to it several times [14].
2. The problem of intrusiveness is one of the challenges affecting EFT [14].
3. Using numerical scale can be confusing as the user might not be consistent in giving their rating.
4. User's rating might not show the true opinion of users.
5. EFT is susceptible to noise.
6. It is sensitive to user context.

3.3 Hybrid Feedback Technique (HFT)

HFT is the combination of both IFT and EFT. This approach utilizes both combinations of numerical rating scores and human behavior in predicting items of interest and taste to the users.

Merits

1. HFT helps to recover the forecast rating accuracy.

2. HFT is the mixture of both Implicit and explicit methods.

Demerits
1. HFT is not cheap.

2. It is computationally intensive.

4. LITERATURE REVIEW

In [15], author implements user-based Collaborative Filtering algorithm on a distributed cloud computing platform that is Hadoop is used, to solve the scalability problem of Collaborative Filtering method. Merits are 1. Better for finding interests for similar items. 2. Personalized recommendation. Demerits are 1.Doesnot consider similar users interests.

In [16], the author concentrates on personalized travel recommendation and show promising applications by utilizing the freely accessible community-contributed photos. They propose to lead customized travel recommendation by further considering specific user profiles or attributes as also travel group types (e.g., family, couple, and friends). Merits are 1. Community contribution is a good attribute to consider for providing recommendations. 2. A lot of information can be extracted using community-contributed photos. Demerits are 1. Privacy issues can arise when community-contributed photos are processed. 2. Very complex to execute.

In [17] author proposed an approach which includes item-to-item collaborative filtering to discover meaningful interesting videos among the large scale of the videos and this methodology is executed in Qizmt which is a.NET MapReduce framework. Merits are 1. Provides better recommendation for same item using interests of similar users. Demerits are 1. Does not consider similar Interests, 2. Complex to implement.

In [18] author has put forth a KASR method for the personalized recommendation. Here user based collaborative filtering algorithm is used. To make the technique more effective and scalable it is executed on Hadoop. Jaccard coefficient and Cosine similarity measure are used for evaluation. They show that the proposed recommendation method is better than the current traditional methods. The major advantages are 1. Scalable 2. More efficient than traditional methods. There are also some demerits associated with this like 1. Jaccard Coefficient method is not so accurate. 2. User's positive and negative reviews are not separated. Sentiments in the text are not considered for calculation.

In [19], the author suggested a novel clustering technique built on Latent Class Regression model (LCRM), which is basically ready to consider both the general ratings and feature-level opinion values (as extracted from textual reviews) to perceive reviewers' inclination homogeneity. In the examination, they tried the proposed recommender algorithm with two true datasets. More notably, they compared it with different related methodologies, including the non-review based technique and not-LCRM based variations.

In [20], propose a system that considers the location as an attribute for giving the recommendation to users. Merits are 1. Better for location specific services. 2. Reduces transmission cost overhead. Demerits are 1. Not suitable where a location is not an attribute of concern. 2. Lack of similarity calculation thus not suitable for bigger datasets.

In [21], they suggested recommendation technique that examines the distinction among the feedbacks of the customer to recognize the customer's predilections. These techniques consider clear ratings, an activity that can report the data sparseness issue. In view of these techniques, they also lead an experimental investigation of online-restaurant client feedbacks to make a restaurant RS and exhibit the efficiency of the suggested technique.

In [22], they suggested a collaborative filtering method for creating the suggestion for various items using ratings and feedbacks accessible on twitter. They have evaluated feedbacks given by blipper (a feedback website) for four unique products using CF method.

5. CONCLUSION

In this paper, we try to briefly describe the various type of recommendation techniques and its type. We also discuss the feedback techniques for recommender system.

In future, various other attributes and techniques can be developed and evaluated for efficient implementation of recommendation systems.Also by combining recommendation systems with machine learning (ML) and natural language processing (NLP), we can develop powerful and efficient recommendation systems which will consider various aspects. Using machine learning, we can train the system to provide best recommendations based on its past experiences. This will result in a very efficient recommendation system with has its own intelligence to predict the best interest of the user and hence provide recommendations with high accuracy.

6. ACKNOWLEDGMENTS

We are thankful to the faculty members of School Of Computer Science and Engineering Department of KIIT University, Bhubaneswar for their cooperation and suggestions.

7. REFERENCES

[1] Website link https://en.wikipedia.org/wiki/E-commerce_in_India

[2] Website link https://www-01.ibm.com/software/data/bigdata/what-is-big-data.html

[3] TechAmerica Foundation's Federal Big Data Commission. (2012). Demystifying big data: A practical guide to transforming the business of Government. retrieved from http://www.techamerica.org/Docs/fileManager.cfm?f=techamerica-bigdatareport-final.pdf.

[4] Schafer, J. Ben, Joseph A. Konstan, and John Riedl. "E-commerce recommendation applications." *Applications of Data Mining to Electronic Commerce*. Springer US, 2001. 115-153.

[5] Fernández, Alberto, et al. "An overview on the structure and applications for business intelligence and data mining in cloud computing." *7th International Conference on Knowledge Management in Organizations: Service and Cloud Computing*. Springer Berlin Heidelberg, 2013.

[6] Xiang, Liang. "Recommendation system practice." *BeiJing Youdian Publication House, BeiJing* (2012): 6.

[7] Witten I. H. and Frank I. Data Mining, Morgan Kaufman Publishers, San Francisco, 2000.

[8] GK, Kiran Kumar Bathi, and N. Sheela. "Personalized Recommender System." *PARIPEX-Indian Journal of Research* 4.8 (2016).

[9] Asabere, Nana Yaw. "Review of Recommender Systems for Learners in Mobile Social/Collaborative Learning." *International Journal of Information* 2.5 (2012).

[10] F Ricci, L. Rokach, B. S Paul. KantorRecommender Systems Handbook", pg.77.

[11] Núñez-Valdéz, Edward Rolando, et al. "Implicit feedback techniques on recommender systems applied to electronic books." *Computers in Human Behavior* 28.4 (2012): 1186-1193.

[12] Jawaheer, Gawesh, Martin Szomszor, and Patty Kostkova. "Comparison of implicit and explicit feedback from an online music recommendation service." *proceedings of the 1st international workshop on information heterogeneity and fusion in recommender systems*. ACM, 2010.

[13] Lops, Pasquale, Marco De Gemmis, and Giovanni Semeraro. "Content-based recommender systems: State of the art and trends." *Recommender systems handbook*. Springer US, 2011. 73-105.

[14] Amatriain, Xavier, Josep M. Pujol, and Nuria Oliver. "I like it... i like it not: Evaluating user ratings noise in recommender systems." *International Conference on User Modeling, Adaptation, and Personalization*. Springer Berlin Heidelberg, 2009.

[15] Zhao, Zhi-Dan, and Ming-Sheng Shang. "User-based collaborative-filtering recommendation algorithms on hadoop." *Knowledge Discovery and Data Mining, 2010. WKDD'10. Third International Conference on*. IEEE, 2010.

[16] Chen, Yan-Ying, An-Jung Cheng, and Winston H. Hsu. "Travel recommendation by mining people attributes and travel group types from community-contributed photos." *IEEE Transactions on Multimedia* 15.6 (2013): 1283-1295.

[17] Jin, Yohan, et al. "MySpace video recommendation with map-reduce on qizmt." *Semantic Computing (ICSC), 2010 IEEE Fourth International Conference on*. IEEE, 2010.

[18] Meng, Shunmei, et al. "KASR: A Keyword-Aware Service Recommendation method on MapReduce for big data applications." *IEEE Transactions on Parallel and Distributed Systems* 25.12 (2014): 3221-3231.

[19] Chen, Li, and Feng Wang. "Preference-based clustering reviews for augmenting e-commerce recommendation." *Knowledge-Based Systems* 50 (2013): 44-59.

[20] Luo, Zhiling, Ying Li, and Jianwei Yin. "Location: a feature for service selection in the era of big data." *Web Services (ICWS), 2013 IEEE 20th International Conference on*. IEEE, 2013.

[21] Liu, Hongyan, et al. "Combining user preferences and user opinions for accurate recommendation." *Electronic Commerce Research and Applications* 12.1 (2013): 14-23.

[22] Esparza, Sandra Garcia, Michael P. O'Mahony, and Barry Smyth. "Mining the real-time web: a novel approach to product recommendation." *Knowledge-Based Systems* 29 (2012): 3-11