Survey on Recommendation System

Pratibha Chole, Prof.Pradeep Patil

M.E Student, Dept. of Computer Engineering, VPCOE, Baramati, Maharashtra, India

Assistant Professor, Dept. of Computer Engineering (PG), VPCOE, Baramati, Maharashtra, India

ABSTRACT: Recommender system help users in finding items of interest. It can be defined as the subclass of information filtering system that presents list of items which are based on user interest. Big data is all about handling vast amount of data. In modern years, the amount of customers, services and online information is increasing rapidly, so the big data analysis problem is generated for service recommender systems. As numbers of alternative services are growing, recommending services that users preferred effectively have become an important research issue. Service recommender systems have been revealed as expensive tools to help users deal with services overload and provide appropriate recommendations to them. In this paper survey of various recommendation approaches like content based, collaborative, Hybrid approach and its related research in the recommender system are described.

KEYWORDS: Recommender system; Preferences; Scalability; Big data; Hadoop

I.   INTRODUCTION

Every day people take decisions like which movie should I see? , which city should I visit? , what should I eat? etc. there are many choices and a little time to explore them all. Recommender system helps people to take decisions in these complex information spaces. Recommender systems are a type of information filtering that presents lists of items (films, songs, books, images, videos, products.) which are based on user interest. Examples are Netflix, Pandora etc. The amount of data in real world increases rapidly that cause's the big data problem. Big data means large amount of data. Big data refers to a dataset which is difficult to capture, store, manage, process and examine with the available present technology within the endurable speed and time. Volume, Variety and Velocity are the factors of big data. Volume means the quantity of data that is generated. Varity may be in the form of structure, unstructured, semi structure data. Velocity means speed of generation of data. Big data management is a serious problem to IT companies. Also it poses heavy impacts on service recommender systems. In today's world everybody needs to purchase the best thing not only in sensible cost but also in short time. In such situation recommender system plays an important role.

II.   RECOMMENDATION SYSTEM CONCEPT

According to the definition of recommender system in [1], it was defined as system that produces individualized recommendations as output or has the effect of guiding the user in a personalized way to interesting or useful services in a large space of possible options.
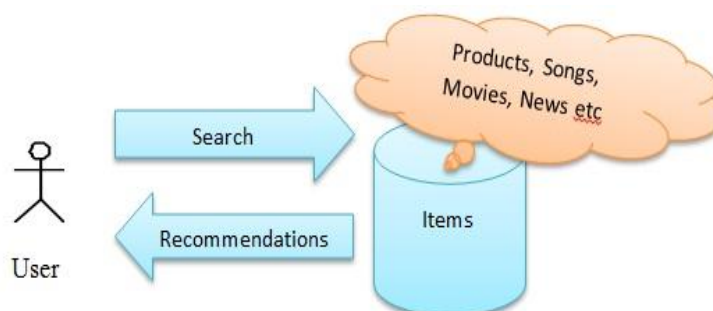

Fig 1: Basic Recommendation System Concept

As shown in Fig.1. User interacts with items stored in system which may be product, movie, news and gives recommendations to the user. Recommendation system consists of three parts: user, recommendation algorithm, recommend object. The most unfavorable is recommendation algorithm, it is the center part. Current recommendation methods usually can be arranged into three main categories: content-based, collaborative, and hybrid recommendation approaches [1].

A. Content-based approaches

In Content-based recommendation approach an item is recommended based on the item description as well as user profile of interest .It recommend services similar to those the user liked in the past[1]. The content information is used to form a profile of user's interest and these profiles utilized to recommend unseen items or information to user. User profile is generated by observing the description of content rated by a user before. Content based filtering analyzes item description to classify items that are of scrupulous interests to user. There are two major perceptions in Content based approach Item Profile and User Profile. Item corresponds to items that can be recommended for user and frequently stored in the database. User Profiles help to recommend items based on interest, the interest can be determined based on the previously viewed items. Content based approach utilizes a variety of classification learning techniques like Decision tree, Nearest Neighbor method, Probabilistic methods. The importance of document for a specific user is decided by term weighting technique like TF-IDF (Term Frequency-Inverse Document Frequency) [2].

B. Collaborative filtering approaches

Collaborative filtering means recommend services to the user that users with similar tastes preferred in the past. [3] CF gives either predictions or top N recommendations for user. CF is basically about the collecting and examining the large amount of information on user's behavior, preferences, browsing activities and predicting what the user will like based on their similarity to other users. There are different methods for evaluating similarity among users like Cosine Based Similarity, Pearson Correlation Coefficient (PCC). It can be classified into following types:

1. User based collaborative filtering

In this type of Collaborative filtering the prediction of the rating of an item for a user rely on the ratings of the same item rated by similar users.

2. Item based collaborative filtering

In this type of Collaborative filtering the prediction of the rating of an item for a user depends upon the ratings of other similar items by the same user.

Depending on the algorithmic technique Collaborative Filtering recommendation approach can be separated into two classes that is memory based and model based Collaborative Filtering. In Memory Based approach all the collected data reside in the memory and utilized for building predictions. Here the ratings of users are directly utilized for making predictions. Basically, it may use all users or only user's who are in the neighborhood to the active user. Over again, there are two variations in Memory Based Approach like User Centric and Item Centric. In model based Collaborative Filtering the models are constructed and trained based on existing data and used for making predictions [3].

C. Hybrid approaches

Hybrid approaches merge content-based and CF methods in several different ways. Various Hybridization techniques are shown in table 1.Various hybrids have been examined, including the hybridization techniques specifically weighted, switching, feature Combination, Cascading, meta-level, feature augmentation and mixed [4]. They are described as follows:

- Weighted: Collaborative and Content-based methods are implementing separately and then combining their predictions.
- Switching: A certain switching criterion is utilized by the system to exchange between two recommendation systems operating on the same object.
- Feature Combination: In a single Recommendation algorithm Features from different recommendation systems data sources are placed.
- Cascading: For this category, one recommendation system processed the results given by another
- Meta Level: In this technique, a feature such as a model learned by one recommendation is used as input to other.
- Feature Augmentation: The output of single system is utilized as an input attribute to other.
- Mixed: It merges two or more approaches at the same time for example: Content-based and Collaborative Filtering.

International Journal of Innovative Research in Computer and Communication Engineering

TABLE 1:

HYBRIDIZATION TECHNIQUES

Hybridization Techniques	Description
Weighted	The scores (or votes) of Many recommendation techniques are combined together to produce a single recommendation.
Switching	The system switches between recommendation techniques depending on the current situation.
Feature combination	Features from different recommendation data sources are placed together into a single recommendation algorithm.
Cascade	One recommender refines the recommendations given by another.
Meta-level	The model learned by one recommender is used as input to another.
Feature augmentation	Output from one technique is used as an input feature to other.
Mixed	Recommendations from several different recommenders are presented at the same time.

Recommendation techniques with their pros and cons are given in table no 2[4].

TABLE 2:

TRADEOFFS BETWEEN RECOMMENDATION TECHNIQUES

Techniques	Pluses	Minuses
Content based	There is no need for data on other users. No cold start and sparsity problem. It is able to recommend users with not only unique taste but also new popular items.	Data should be in structured format. Inadequate to use quality judgments from other users. .
Collaborative Filtering(CF)	Knowledge about an item feature is not needed. It has better scalability. Also it reduces the sparsity problem.	There is new user as well as new item cold start problem.
Hybrid	It overcomes the difficulties of Collaborative filtering or Content-based approaches.	It is very complicated to implementation. It depends on external data.

III.   CHALLENGES OF RECOMMENDATION SYSTEM

Various challenges of Recommender system are cold start problem, scalability and sparsity problem [5]. Cold start problem is a situation when a new user or item just enters in the system. It's difficult to give recommendations to new users as his profile is nearly empty and he hasn't rated any items yet so his taste is unknown to the system. Scalability is the property of system specifies its ability to handle growing amount of information in a graceful manner. With enormous growth in information over internet, it is obvious that the recommender systems are having an explosion of data and thus it is a great challenge to handle with continuously growing demand. Sparsity is the problem of lack of information.

IV. RELATED WORK

There have been many recommender system developed not only in academia but also in industry. In [1], Adomavicius and Tuzhilin give an analysis of the field of recommender systems and describe three current recommendation approaches: content-based, collaborative, and hybrid recommendation approaches as well as they describe different limitations of present service recommendation methods, and discuss feasible expansions that can raise recommendation abilities and prepare recommender systems applicable to an even wide range of applications. In [6], authors proposed a KASR(Keyword Aware Service Recommendation) method for presenting a personalized service recommendation list and effectively recommending the most appropriate services to the users. In this method keywords indicates users preferences, and a user-based Collaborative Filtering algorithm is used to generate appropriate recommendations and it is implemented using Hadoop in real world dataset thus the result show that KASR improves accuracy and scalability of service recommender system. In [7] Personalized Travel Recommendation System is given by considering automatically mined knowledge from travel photo logs specific user profiles or attribute, group type in photo contents. An information theoretic measure is used to demonstrate that user profiles are informative and effective for travel recommendation. In [8] this paper Recommender system for sport videos is given, which is based on the audiovisual consumption. System can produce recommendations for live as well as recorded events. In [9], to recommend Web services actively authors proposed a novel Web service recommendation approach and framework based on user's interests as well as preferences according to usage history. AWSR is implemented and deployed on the Web. AWSR (Active Web Service Recommendation) system can recommend Web services which are uniform with user's interests and preferences from usage history as well as it satisfy their functional and non-functional requirements. Jin et al. [10] propose a large-scale video recommendation system. In this system an item-based CF algorithm is used. Proposed approach implemented in Qizmt, that is a .Net Map-Reduce framework. The system can be utilized for large-scale video sites. In[11] "Big Data" environment designing and implementation of scalable recommender Systems are possible due to the development of cloud computing software tools like Apache Hadoop, Map-Reduce.CF algorithm is executed on Hadoop and scalability problem solved by dividing data set. If the amount of data grows the proposed method doesn't have favorable scalability and efficiency. In [12], spam emails means where one's email storage inbox is attacked with emails that make no sense, they are on the rise. The Bayesian spam detection scheme with context matching is developed by implementing the keyword stripping using the Porter Stemmer algorithm. As the root or stem word is only considered that makes keyword search more efficient.

V. CONCLUSION

In this paper survey of various recommendation approaches like content based, collaborative filtering and hybrid have been discussed which are used to build the recommender system. As the Big data management is main issue in recommender system. Big data is managed by using distributed computing platform known as Hadoop.

REFERENCES

[1]   G. Adomavicius and A. Tuzhilin, "Toward the Next Generation of Recommender Systems: A Survey of the State of- the-Art and Possible Extensions," IEEE Trans. Knowledge and Data Eng., vol. 17, no. 6, pp. 734-749, June 2005.
[2]   Michael J. Pazzani and Daniel Billsus, "Content-based recommendation systems," Springer Berlin Heidelberg. pp. 325-341,2007
[3]   Badrul Sarwar, George Karypis, Joseph Konstan, and John Ried, "Item-based Collaborative Filtering Recommendation Algorithms," May 1-5, 2001, Hong Kong.
[4]   R. Burke, "Hybrid Recommender Systems: Survey and Experiments," User Modeling and User-Adapted Interaction, vol. 12, no. 4, pp. 331-370, 2002.
[5]   L. Sharma and A. Gera, "A Survey of Recommendation System: Research Challenges," International Journal of Engineering Trends and Technology (IJETT), vol. 4, May 2013.
[6]   Shunmei Meng, Wanchun Dou, Xuyun Zhang,"KASR: A Keyword-Aware Service Recommendation Method on Mapreduce for Big Data Applications," IEEE Trans. on Parallel and Distributing systems, vol.25, no.12, December 2014.
[7]   Y. Chen, A. Cheng, and W. Hsu, "Travel Recommendation by Mining People Attributes and Travel Group Types from Community-Contributed Photos," IEEE Trans. Multimedia, vol. 25, no. 6, pp. 1283-1295, Oct. 2013
[8]   M. Alduan, F. Alvarez, J. Menendez, and O. Baez, "Recommender System for Sport Videos Based on User Audiovisual Consumption," IEEE Trans. Multimedia, vol. 14, no. 6, pp. 1546-1557,Dec. 2012
[9]   G. Kang, J. Liu, M. Tang, X. Liu, and B. cao, "AWSR: Active Web Service Recommendation Based on Usage History," Proc. IEEE 19th Int'l Conf. Web Services (ICWS), pp. 186-193, 2012
[10]   Y. Jin, M. Hu, H. Singh, D. Rule, M. Berlyant, and Z. Xie, "MySpace Video Recommendation with Map-Reduce on Qizmt," Proc. IEEE Fourth Int'l Conf. Semantic Computing, pp. 126-133, 2010.
[11]   Z.D. Zhao and M.S. Shang, "User-Based Collaborative-Filtering Recommendation Algorithms on Hadoop," Proc. Third Int'l Workshop Knowledge Discovery and Data Mining, pp. 478-481, 2010
[12]   B. Issac and W.J. Jap, "Implementing Spam Detection Using Bayesian and Porter Stemmer Keyword Stripping Approaches," Proc. IEEE Region 10 Conf. (TENCON '09), pp. 1-5, 2009.

BIOGRAPHY

Pratibha R. Chole received Bachelor of Engineering Degree in IT from Savitribai Phule Pune University. She is now pursuing M.E. degree.

Pradeep M. Patil has received his Bachelor of Engineering Degree from Dr. Babasaheb Ambedkar Marathawada University, Aurangabad in 1995 and M.E. Degree from Shivaji University Kolhapur in 2005. He is now pursuing his Ph.D. from Sathyabhama University, Chennai.