A Survey on Popular Recommender Systems

Ashish Kumar, Yudhvir Singh
Computer Science and Engineering Department
UIET, Maharishi Dayanand University
Rohtak, India

Abstract- Recommender system is one of the most common research topic now a days. Social networking and ecommerce websites are highly dependent on recommender system. It help users to discover the hidden items of their choice without much effort. It is a software tool which uses a user's preference to discover new items from a large pool of available items in real time. In this paper we mainly focused on various types of recommender system and how the different online platforms making use of recommender system to give personalized service to its users.

Keywords- Recommender system, collaborative filtering, amazon recommender system, Netflix recommender system, linkedin-skills, youtube recommender system, tapestory, yahoo questions.

I. Introduction

Recent advances in the area of IOT and electrical devices in the past few years have the changed the way of people used to live their lives [1]. These advancements in technologies have generated a large amount and variety of data. Every click on the smart devices generating the data. Solving this enormous data is new the challenge for data scientists and data architects. This data may not be useful for the users directly but it contains huge value for the service providers. To mine the information from the overloaded data, there is continuous need to upgrade the software tools and technologies [2]. Recommender system provides the solution of this problem in the domain of ecommerce business and social networking sites. Recommender systems act as a salesmen for the users when they visit on ecommerce websites. It helps the service provider to increase their revenue and helps the user to recommend items or products of their preferences.

Goldberg, Nichols, Oki and Terry [4] developed the first recommender system i.e. Tapestry in 1992. It was an electronic messaging system that allowed its users to rate messages wither 'good' or bad. One of the simplest types of recommender system is non personalized recommender system [3]. Recommendations produced by these systems are independent of the user. Each user will get the same recommendation list because user's preference will not be taken into care. Generally this recommended item's list contains the most popular items, items from best sellers or may be the latest added item. It all depends on the business plan by the company. But this type of recommendation does not add much value to business, so personalized recommender system is widely used in the industry.

In the personalized recommender system user's preference is taken in to the care. The newly recommended item list purely depends on the past interaction of the user with the system.

II. Background

Recommender systems are classified in the following categories-

1. Collaborative filtering based recommender system: it first time comes into the picture in early 1990 for dealing with the overload in online information space. Tabestry [4] was a messaging system based on collaborative filtering developed at Xerox Palo to deal with the large amount of incoming documents via electronic emails. Collaborative filtering based recommender systems involves two stages; in the first stage it computes the similarity between the target user (for which we want recommendations) and all other users in the system and in the second stage it calculate the list of predicted items based on the calculations of the first stage [5].

Collaborative filtering based recommender system is further classified in two types. First is user-user based collaborative filtering. GroupLens [6], Ringo [7] and BellCore [8] uses the user-user based filtering technique for recommending item *I* to the user *U* based on the rating pattern from the other users of the system for the same item. Second type of collaborative filtering based recommender system is item-item based collaborative filtering. Idea behind it is if two items have same users' like or dislike then they are considered to be of same category. Sarwar et al. [9] and Karypis [10] described it first time.

2. Content based recommender system- idea behind it is it takes user's interest as an input and generate a list of recommended products/items as an output. It doesn't depends on the information provided by other users in the form of ratings [11]. It is popularly used for recommending the text based products from which we can find the textual description like messages, web pages, news etc.

Hybrid recommender system- it is combination of several types of recommender system. Reason behind combining the several recommender systems is that one type of recommender system may not perform well in all type of situations. E.g. collaborative filtering based recommender system has problem of cold start where content based recommender system doesn't have this type of issue [12]. In most of the hybrid recommender systems collaborative filtering based recommender system are combined with content based recommender system [13].

III. Popular Recommender System

In this section, we will present a few online businesses that uses recommender system to provide their users customized service.

1. Amazon.com – amazon uses recommender system to personalize their online store as per customers. They recommend sports items to an athlete and toys to a new mother. They don't follow traditional user-user based recommender system instead they use item-item based recommender system. Generally it is assumed that number of users on an ecommerce website is much higher than the number of items. While computing the similarity among the users, in traditional user-user collaboration approach, the time complexity of computation would be higher than the time complexity in item-item based recommender system. This approach provide the advantage over the traditional collaborative filtering approach in the manner that it is suitable for large data set without reducing the dimensionality, sampling or partition [14].

2. Netflix recommender system – it consists of a number of recommender systems, collectively all of them increase the user's experience. Homepage of Netflix generally contains 40 rows and 75 videos per row. Each row displays the results based on one type of recommender system. Main recommenders that Netflix uses on its platform are-

- Personalized video ranker (PVR), top-N video ranker – purpose of this algorithm is to provide the list of popular recommendation to each user in the catalog [16].
- Trending ranker – purpose of this algorithm is to recommend the viral videos to the users.
- Continue watching ranker – it sorts the recently watched videos to make prediction whether the user is intended to resume or rewatch the video or user is not interested in the video any more. This prediction is based on several factors like time elapsed since viewing, the point of abandonment, whether any other video is watched after that.
- Video-video similarity algorithm – this is basically an impersonalized algorithm, it find similar videos from catalog and add them in the same category.
- Page generation algorithm – each account of Netflix is generally used by several members of home or friends. Each one has different mood or taste. Netflix shows multiple rows of recommendations to deals with this problem. It uses the output generated by all other algorithms and generate one row of recommendatios. Each of these algorithm is based on statistical and machine learning technique (supervised and unsupervised approaches) [15].

3. Linkedin – while adding the skill set on linkedin. They have provided the functionality of type-ahead. For this purpose they have evaluated several public listing of skills but none of them is complete. Linkedin is one of the world largest professional body that allows professionals to connect each other. It contains millions of professionals. Each have different skill set. Creating a skill set list for covering all types of skill is quite difficult.

Linkedin come up with the solution of this problem by creating a list of skills directly from the user profile. Now fetching the keyword from the user profile may or may not be skill. Beside this may skills are duplicate as well e.g. c programming and c coding both are same [17]. For this they have extracted the comma separated keywords from the skill set section of the profile. Now from these keywords some entities are not skills like company name, these have to be removed. After this ambiguity has to be removed from the keywords. Some keywords has multiple meanings. All these are removed from the skill set list. Clustering was used for this purpose. Next step is removing the duplicity. Multiple name of same skill also adding unnecessary item in the list and make it large. Crowdsourcing is the solution for this problem. Linkedin has used the Amazon's mechanical turk for building the crowdsourcing. By using all these methods they have reduced the 150000 skill topics to 50000 skill topics. Once the skill set is ready they created the skill inference algorithm to recommend the new skill to the users [18, 19].

4. Youtube - it is the world's largest and most popular video platform which allows its users to upload the videos, watch the videos and sharing of the videos. It also allows users to interact with the videos by liking/disliking the video or by commenting on the video. Everyday billions of the videos are played on the youtube by millions of the users. Recommending the best video to a particular user based on his preference is the challenge that youtube has solved using the statistical methods. In order to keep the users engaged and entertained, it is important that it regularly update its recommendations as per the user's interaction with the system. Youtube focuses on the recommending the recent, fresh and diversified videos to the users. It should not happen that user thinks why this video is recommended to him.

For recommending the videos to user, its system take video content i.e. metadata of the video as well as user's activity data. By user activity data means that user's interaction with the system by any mean it can be watching a video, liking/disliking a video or it can be commenting on a particular video.

Youtube, first of all find out the all related videos [20]. It basically creates the groups of related videos. All videos in a particular group is similar somehow. In order to like a video to the user, youtube combine the related videos association rules with the personal activity of the user on the system. Once the list of possible recommended videos is prepared then these videos are ranked based on several parameters:

- Video quality
- User specification
- Diversification

Instead of recommending the just most relevant videos to the user, youtube consider the relevance and diversification across the categories. Youtube uses the batch oriented pre computed approach instead of on demand computation [21].

5. Tapestry – it was developed at Xero Palo Alto research center. Motivation behind developing this mailing system was increasing use of emails as a result user was getting overloaded with the huge data [22]. They suggested solution for dealing with this huge data is to filter the incoming data, no matter from where they are coming in just filter the data based on the interest. Tapestry was one of the first system that uses the collaborative based filtering. It not just simply compute the similarity between the documents neither it filter the document when they arrived on the system. Instead they run the systematic queries on the new document arrived and on the old documents on the database.

Tapestry is more than just a simple mailing service because it handles all the incoming documents as well. Despite this it also store all the old mails. To find the document of interest it runs the ad hoc queries to filter out the keywords. On verifying the result of query, it sore the results in the database and in future when new document is arrived it uses the results of already filtered document to filter this new document [4].

6. **Yahoo: Answers –** in today's date it is one of most famous question answering system. Recommending the right question to the right user is highly suggestible for effective use of system. It get over 30 million questions and answers [22]. Unlike other platforms of similar kind it broadcast questions to all the user on the site. Its main focus is on satisfying the asker's need [23, 24]. Instead of just answering the question, user can give thumbs up or thumbs down to any question.

Thread of question start once a user posted a question and assign it to a category. Question will remain in open state till it is not resolved. With the help to textual attributes, question can be described. And using the collaborative filtering it is recommended to users who has shown interest in similar attributes [26].

To find the attributes of question, yahoo uses the textual description of question. It may be title of question or may be its body. Then it tokenize the question in to tokens. User has to select the category of question to which it belongs. It may be "internet", "youtube" or "dogs" etc. A user ID is created by the system. Users can be of several types: asker, answerer, best answerer, question voters, answer voters. Goal of the system is to predict the user who will answer the question. For this purpose interaction features are used by the classifier to evaluate the match between the user and the question. Once data set is ready yahoo implement its classifiers to find the right user for right question [25].

IV. Conclusion and Future Scope

In this paper, we presented the problems faced by some big giant of the internet and we briefly explained the solution provided by them. Handling the large number of users, recommender systems plays an important role. Feature extraction is most crucial part of the recommender system. Effectiveness of recommender system is highly dependent on the quality of features. Recommender system can be applied on large number of problems. By combining filtering techniques with image processing, it can be used to automate the traffic signals. In highly competitive environment, some threats are also there which effects the performance of the recommender system. Attackers may inject fake profiles in the system to affect the results of the recommender system. That is the issue to be dealt with.

References

[1]. Jayavardhana Gubbi, Rajkumar Buyya, Slaven Marusic, and Marimuthu Palaniswami. Internet of things (iot): A vision, architectural elements, and future directions. Future generation computer systems, 29(7):1645-1660, 2013.
[2]. Aristomenis S Lampropoulos and George A Tsihrintzis. Machine Learning Paradigms: Applications in Recommender Systems, volume 92. Springer, 2015.
[3]. Gediminas Adomavicius and Alexander Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE transactions on knowledge and data engineering, 17(6):734{749, 2005.
[4]. Goldberg, D., Nichols, D., Oki, B.M. and Terry, D. (1992) Using Collaborative Filtering to Weave an Information Tapestry. Communications of the ACM, 35, 61-70.
[5]. Dakhel, Gilda Moradi, and Mehregan Mahdavi. "A new collaborative filtering algorithm using K-means clustering and neighbors' voting." In *Hybrid Intelligent Systems (HIS), 2011 11th International Conference on*, pp. 179-184. IEEE, 2011.
[6]. Konstan, Joseph A., Bradley N. Miller, David Maltz, Jonathan L. Herlocker, Lee R. Gordon, and John Riedl. "GroupLens: applying collaborative filtering to Usenet news." *Communications of the ACM* 40, no. 3 (1997): 77-87.
[7]. Shardanand, Upendra, and Pattie Maes. "Social information filtering: algorithms for automating "word of mouth"." In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 210-217. ACM Press/Addison-Wesley Publishing Co., 1995.
[8]. Hill, Will, Larry Stead, Mark Rosenstein, and George Furnas. "Recommending and evaluating choices in a virtual community of use." In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 194-201. ACM Press/Addison-Wesley Publishing Co., 1995.
[9]. Linden, Greg, Brent Smith, and Jeremy York. "Amazon. com recommendations: Item-to-item collaborative filtering." *Internet Computing, IEEE* 7, no. 1 (2003): 76-80.
[10]. Karypis, George. "Evaluation of item-based top-n recommendation algorithms." In *Proceedings of the tenth international conference on Information and knowledge management*, pp. 247-254. ACM, 2001.

[11]. Mladenic, Dunja. "Text-learning and related intelligent agents: a survey." *IEEE Intelligent Systems* 4 (1999): 44-54.

[12]. Kumar, Ashish, Deepak Garg, and Prashant Singh Rana. "Ensemble approach to detect profile injection attack in recommender system." In *2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 1734-1740. IEEE, 2015.

[13]. Burke, Robin. "Hybrid recommender systems: Survey and experiments." *User modeling and user-adapted interaction* 12, no. 4 (2002): 331-370.

[14]. Linden, Greg, Brent Smith, and Jeremy York. "Amazon. com recommendations: Item-to-item collaborative filtering." *IEEE Internet computing* 1 (2003): 76-80.

[15]. Gomez-Uribe, C.A. and Hunt, N., 2016. The netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems (TMIS)*, *6*(4), p.13.

[16]. Amatriain, X. and Basilico, J., 2016, September. Past, present, and future of recommender systems: An industry perspective. In *Proceedings of the 10th ACM Conference on Recommender Systems* (pp. 211-214). ACM.

[17]. A. Mathes. Folksonomies-cooperative classi_cation and communication through shared metadata. Computer Mediated Communication, 47(10):1{13, 2004.

[18]. F. Ricci, L. Rokach, and B. Shapira. Introduction torecommender systems handbook. Springer, 2011.

[19]. Bastian, M., Hayes, M., Vaughan, W., Shah, S., Skomoroch, P., Kim, H., Uryasev, S. and Lloyd, C., 2014, October. LinkedIn skills: large-scale topic extraction and inference. In *Proceedings of the 8th ACM Conference on Recommender systems* (pp. 1-8). ACM.

[20]. R. Agrawal, T. Imieli´nski, and A. Swami. Mining association rules between sets of items in large databases. SIGMOD Rec., 22(2):207–216, 1993.

[21]. Davidson, J., Liebald, B., Liu, J., Nandy, P., Van Vleet, T., Gargi, U., Gupta, S., He, Y., Lambert, M., Livingston, B. and Sampath, D., 2010, September. The YouTube video recommendation system. In *Proceedings of the fourth ACM conference on Recommender systems* (pp. 293-296). ACM.

[22]. L. Rao. Yahoo mail and im users update their status 800 million times a month. TechCrunch, Oct 28 2009. http://techcrunch.com/2009/10/28/yahoo-mail-and-im-usersupdate- their-status-800-million-times-a-month/.

[23]. E. Agichtein, C. Castillo, D. Donato, A. Gionis, and G. Mishne. Finding high quality content in social media, with an application to community-based question answering. In Proc. international conference on Web search and web data mining (WSDM'10), pages 183–194, 2008.

[24]. C. Shah and J. Pomerantz. Evaluating and predicting answer quality in community qa. In Proc. conference on Research and development in information retrieval (SIGIR'10), pages 411–418, 2010.

[25]. Dror, G., Koren, Y., Maarek, Y. and Szpektor, I., 2011, August. I want to answer; who has a question?: Yahoo! answers recommender system. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 1109-1117). ACM.

[26]. D. Horowitz and S. Kamvar. The anatomy of a large-scale social search engine. In Proc. International World Wide Web Conference (WWW'10), 2010.