www.mecsj.com/
Multi-Knowledge Electronic Comprehensive Journal For Education And Science Publications (MECSJ)
Issue (37),2020
ISSN: 2617-9563

Recommender Systems Principles and methods in Web 2 Applications
An analytical view of the filters used in YouTube

Aminah Aldossary
School of Education, King Faisal University
E-mail: aldossary@kfu.edu.sa

Abstract

Currently, several Web 2 applications encounter many challenges, the most prominent of these challenges is the increasing number of users of the Web also a large amount of data and information that is circulated on Web 2 platforms, which will negatively affect the quality of the output from these applications. Therefore, Web 2 applications have paid great attention to recommendation systems, which greatly contribute to sorting, revising, and inventorying the items that the user may need and presenting it as a recommendation (Isinkaye, et al., 2015). There are many approaches to achieve recommendations like basic techniques of collaborative filtering and content-based approach. These approaches can be done individually or combined depending on the type of recommendations needed by individuals (Nair & Kelkar, 2013). The purpose of these approaches will be demonstrated in this study. An Analytical explanation of how YouTube created its recommenders will be covered in this research.
Keywords: Web 2, Applications, Recommendation Systems, Collaborative Filtering, Content-Based Filtering

الملخص

في الوقت الراهن ، تواجه تطبيقات الويب ٢ العديد من التحديات ، ولعل من أبرز هذه التحديات زيادة اعداد و حاجات مستخدمي الويب و أيضًا كمية البيانات والمعلومات الضخمة التي يتم تداولها على تلك المنصات ،و هذا من شأنه ان يؤثر سلبًا على جودة المخرجات من هذه التطبيقات. لذلك ، أولت تطبيقات الويب ٢ اهتمامًا كبيرًا لأنظمة التوصيات والتي تساهم بشكل فعال في فرز العناصر التي يتطلع المستخدم لها وكذلك مراجعتها و فرزها وتقديمها كتوصية (Isinkaye ، وآخرون ، ٢٠١٥). هناك العديد من التقنيات و الاساليب للوصول لتلك التوصيات التي تتوافق مع احتياجات كل مستخدم على حدا مثل التقنيات الأساسية للتصفية التعاونية والنهج القائم على المحتوى، يمكن تنفيذ هذه الأساليب بشكل فردي أو مجتمعة اعتمادًا على نوع التوصيات التي يحتاجها المستخدمين (Kelkar & Nair، ٢٠١٣). سيتم توضيح هذه التقنيات و الغرض منها في هذه الدراسة كما سيتم تناول شرح تحليلي لكيفية عمل انظمة التوصيات في YouTube.

الكلمات الرئيسية: الويب ٢ ، التطبيقات ، أنظمة التوصية ، التصفية التعاونية ، التصفية القائمة على المحتوى

Multi-Knowledge Electronic Comprehensive Journal For Education And Science Publications (MECSJ)
Issue (37),2020
ISSN: 2617-9563
www.mecsj.com/

1. Introduction

Most of the users around the world use the recommendation systems frequently for purchasing or entertainment such as YouTube, Facebook and Twitter. The principle of the recommendation system is "significant dependencies exist between user- and item-centric activity" (Aggarwal, 2016). Recommendation systems are a web application that recommends several items that are compatible or identical to the user's profile. The dependencies can be found in each item rating by the users. Then will show the outcome in a model for the target user (Aggarwal, 2016). The model for the target user work based on the utility matrix of the user's preferences in the recommendation system (users with similar interest) (Leskovec et al., 2014). However, the system has two classes the first one focuses on the user's preferences items, and the second class is items. The important idea is to know how to find the right recommendation for the target user by analyzing the user's preferences with items (Celdran et al., 2016). Therefore, the goal of recommender systems is to recommend items that are related to the user (Aggarwal, 2016).Additionally, recommender systems suggest a list of top-k items. This list includes items of diverse types to increase the chance that the user would like at least one of these items (ibid). Different methods for building recommendation systems have been developed including collaborative filtering, content-based filtering, and hybrid filtering (Isinkaye, Folajimi, and Ojokoh, 2015).

2. Recommendation filtering techniques

The use of efficient and accurate recommendation techniques is necessary for a system that will provide a useful recommendation to the users. Fig. 1 shows the frame of various recommendation filtering techniques. Collaborative filtering techniques recommend items based on measuring the similarity between users or items (Isinkaye, Folajimi, and Ojokoh, 2015).
CF can be classified into two categories including memory- based and model-based (ibid). On the other hand, content-based filtering techniques recommend items based on target user's information, without looking at other users' opinions (ibid). Furthermore, hybrid filtering techniques combine two or more filtering techniques (ibid).

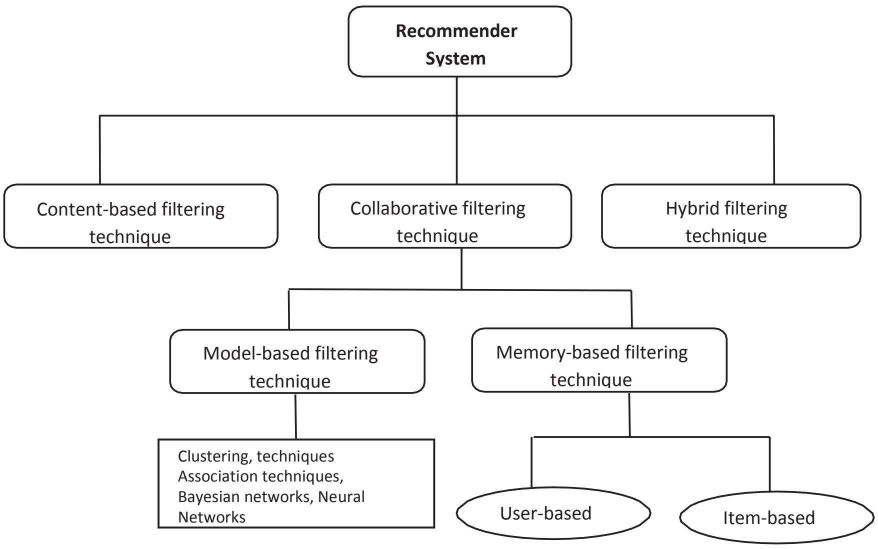

Figure 1 Recommendation techniques

2.1 Memory-Based CF

Memory-based is known as neighborhood-based which is performed by predicting the ratings of user-item based on users whose ratings are similar to the active user or on items that are similar to the target item to generate a recommendation (Aggarwal, 2016).Memory-based can be performed in two ways including user-based and item-based (Isinkaye,Folajimi, and Ojokoh, 2015). **User-based** determines users who are similar to the active user and predicts unspecified ratings of the active user by calculating weighted averages of the rated item by users similar to this user (ibid). On the other hand, **Item-based** predicts the rate for target item of target user by finding other items that are similar to target item (Aggarwal, 2016). According to Aggarwal (2016), item-based CF is more accurate than user-based since item-based methods are more stable with changes to the ratings (ibid). This is because the number of users is larger than the number of items and new users could be added in systems more than new items.

2.2 Model-Based CF

This approach is used to create a model based on extracting some data from the dataset for prediction without using whole dataset. One of the most popular model-based techniques is clustering. CF recommendations compare all dataset to compute the similarity between users/items. Therefore, more time is needed for this computation because of the large size of the dataset (Dakhel and Mahdav, 2011). As a result, in order to compute the similarities between users/items, it is better to aggregate data into the cluster, and each data will be compared with other data in the same cluster (ibid). Similarly, McSherry (2004, cited in Isinkaye, Folajimi and Ojokoh, 2015) explains that clustering techniques work by dividing a dataset into subsets to construct clusters of similar users. Then, recommendations for each user can be made by averaging the rating of other users in a cluster. Any of the distance measures are used it to construct a cluster [2].

2.3 Content-Based Methodology

Unlike the collaborative filtering relied systems that identify users whose preferences are similar to those of the given user and recommend items they have liked, the Content-based recommendation systems (CB recommenders) try to recommend items similar to those a given user has liked in the past (Balabanović & Shoham, 1997) (Lops, et al., 2011). CB systems recommends items based on the target user’s information, without looking at another users’ (i.e. friends or people who have the same taste) likes or opinions.

Multi-Knowledge Electronic Comprehensive Journal For Education And Science Publications (MECSJ)
Issue (37),2020
ISSN: 2617-9563
www.mecsj.com/

3. YouTube

YouTube is one of the largest and most popular online video libraries which has a significant content depends on the users in the creation. As a result of this huge content, YouTube has implemented the recommendation system in order to achieve two main goals. Firstly, on the user level, provides the personalized recommendation to help its users to find high-quality videos link to their interest (Sultana,2015, p.1). Secondly, on the control level, as Davidson et al., (2010, p.293) point out, YouTube recommendations provide direct control over personalized user data. Indeed, there are more than a billion users helped by YouTube's recommendation system to discover more videos that suit their preferences (Covington et al., 2016, p.1). It is clear that YouTube seeks through recommender system to satisfy users by offering what suits their preferences.

3.1 The General Framework

Davidson et al., (2010, p.294) confirm that when applying the recommender system, a number of criteria must be taken into account to achieve the goals such as diversity, connection, recent and understanding by users. Moreover, YouTube's recommendation system includes three main stages: 1. input data 2. related videos 3. ranking recommendation (ibid).

3.1.1Input data

There are two categories of data sources, Firstly, content data such as video metadata as titles, descriptions, and dates (Davidson et al., 2010, p.294). Secondly, user activity data which can be divided into explicit or implicit, explicit activities such as watched, liked, rated and download (Sultana,2015, p.1). Implicit activities create from user behavior during watching videos (Davidson et al., 2010, p.294). A good example is that when a user watched a video for two-minute, this behavior is given a different indication than another user who watched the same video but for 10 minutes. It seems that explicit activities give a high percentage of certainty compared to implicit activities. However, Davidson et al., (2010, p.294) question whether both categories may face difficulties that limit their results. For instance, the first type can be incorrect, incomplete or outdated, in addition, watching a video does not always mean that the user has been impressed with the content.

3.1.2 Related videos

According to Davidson et al., (2010, p.294) one of the most important bases for the building of recommendation system in YouTube is the ability to generate a number of related videos from a seed video. In the same way, Meyer (2012, p.88) explains that YouTube uses association rules of each user for short period usually 24 hours, which work in three steps: firstly, during this period the system collects co-visitation ci, j of

Multi-Knowledge Electronic Comprehensive Journal For Education And Science Publications (MECSJ)
Issue (37),2020
ISSN: 2617-9563
www.mecsj.com/

each pair of videos vi, vj.
Then, the system computes f as a normalization function for vi, vj (ibid). The work of Davidson et al., (2010, p.294) show that" f vi, vj is a normalization function that takes the “global popularity” of both the seed video and the candidate video into account". lastly, the system creates a list containing the seed videos, based on these videos the system searches for videos related to the videos in the list, using the graph of association defined by r vi, vj (Meyer, 2006, p.88). To determine the similarity between reference video vi and another video vj is given by the formula(ibid):

$$r(v_i, v_j) = \frac{c_{i,j}}{f(v_i, v_j)}$$

It is clear that the system uses videos that were viewed or preferred by a user then track all the relevant videos of these videos to find out the closest similarity of vi, which will represent recommendations vj.

3.1.3 Ranking recommendation

According to Meyer (2012, p. 88) after creating the recommendations, the system arranges these recommendations for the user based on three criteria: 1. video quality such as video ratings, comments, view count, sharing and favorites. 2. user specificity which is used to increase videos similar to the user's unique preferences such as view count and time of watching (ibid). 3. diversity in order to add value to the recommendations, the system deletes videos which are very similar to each other (ibid). It seems that this technique is valid because matching videos will not be exciting content for users as well as diversity will allow the user to expand the view of more videos in the same area of interest, therefore, YouTube will maintain its user interaction with these recommendations for the largest possible extent.

Multi-Knowledge Electronic Comprehensive Journal For Education And Science Publications (MECSJ)
Issue (37),2020
ISSN: 2617-9563
www.mecsj.com/

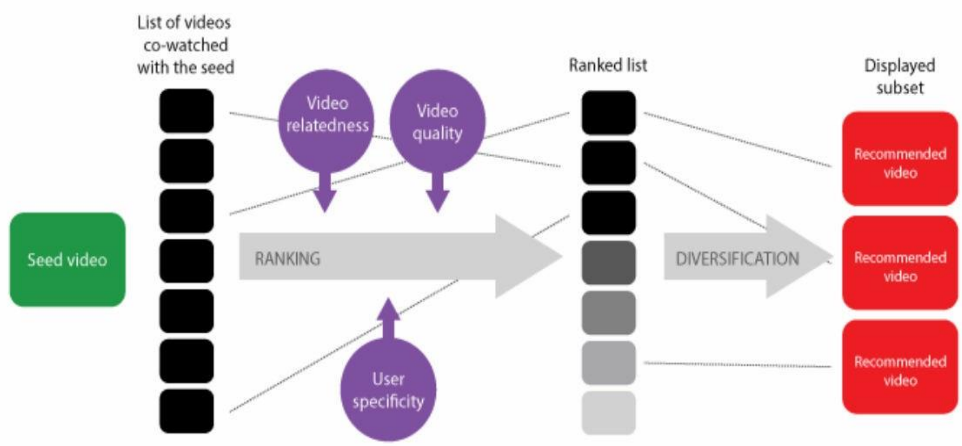

Figure 2: The Recommender System Framework in YouTube.

3.1.4Evaluation

Sultana (2015, p.2) points out that there are a number of metrics to measure the quality of the recommendation system, which is the number of clicks, especially those that result in watching longer portion of the video, furthermore, the extent to which the user follows the recommendations made to him. In addition, Davidson et al., (2010, p.296) conducted an experiment to measure the success rate of the recommendation system in YouTube over the course of 21 days, where the results were:1. 60% of the clicks were for the sections that appeared as a recommendation. 2. The recommended sections acquired a percentage of 207% the clickthrough rate of the higher segments exceeded the view ratio(ibid).

4.conclusion

This paper has focused on recommendation systems as one of the most widely used systems. It has reviewed the definition of recommendation systems and their aspects of use, furthermore, the targets of the recommendation system. Next, it has focused closely on YouTube in terms of how to apply recommendation systems. The recommendation systems are currently witnessing great attention by scientists, researchers, and developers, as a result of the positive results it has achieved in many fields such as electronic commerce, education, service industry, and social networks.

It is worth mentioning that there is recent research focused on the integration between recommendation systems with big data and data mining.

References

- Isinkaye, F., Folajimi, Y. and Ojokoh, B. (2015). Recommendation systems: Principles, methods and evaluation. *Egyptian Informatics Journal*, 16(3), pp.261- 273.
- Nair, R. and Kelkar, K., (2013). Implementation of Item and Content based Collaborative Filtering Techniques based on Ratings Average for Recommender Systems. *International Journal of Computer Applications* , (24), pp. 0975-8887.
- Aggarwal, C. C. (2016). *Recommender systems: The textbook.* New York: Springer.
- Leskovec, J., RAjaraman, A. and Ullman, J. (2014). *Mining of massive datasets*. Cambridge: Cambridge University Press, pp.307-341.
- Celdrán, A., Pérez, M., García Clemente, F. and Pérez, G. (2016). Design of a recommender system based on users' behavior and collaborative location and tracking. *Journal of Computational Science*, 12, pp.83-94.
- Stephen, S. C., Xie, H., and Rai, S. (2017). Measures of Similarity inMemory-Based Collaborative Filtering Recommender System. *Proceedings of the 4th Multidisciplinary International Social Networks Conference,* 32(1).
- Dakhel, G. M., and Mahdavi, M. (2011). A new collaborative filtering algorithm using K-means clustering and neighbors voting. *11th International Conference on Hybrid Intelligent Systems (HIS), Melacca, Malaysia,* pp.179-184.
- Balabanović, M. & Shoham, Y. (1997). Fab: content-based, collaborative recommendation. *Communications of the ACM,* 40(3, March 1997), pp. 66-72.
- Lops, P., Gemmis, M. d. & Semeraro, G., 2011. Content-based Recommender Systems: State of. In: *Recommender systems handbook.* s.l.:Springer, pp. 73-100.
- Sultana, J. (n.d.). *A Survey on YouTube Recommendation Systems*. [online] Academia.edu. Available at: https://www.academia.edu/9175357/A_Survey_on_YouTube_Recommendat

ion_ Systems [Accessed 7 Apr. 2018].

- Davidson, J., Liebald, B., Liu, J., Nandy, P. and Vleet, V., (2010). The YouTube video recommendation systempp. *Proceedings of the fourth ACM Conference on Recommender Systems, Barcelona, Spain,* pp.293-296.
- Covington, P., Adams, J. and Sargin, E. (2016). Deep Neural Networks for YouTube Recommendations. In: *Conference on Recommender Systems*. USA: ACM, pp.191-198.
- Meyer, F. (2012). *Recommender systems in industrial contexts*. PhD. 'UNIVERSITÉ DE GRENOBLE.