

Fast and Efficiently Rule Based Recommendations System for Online Shopping

Avneet Kaur

Dept. of Computer Science, Sri Guru Harkrishan College of Management and Technology,
Patiala, Punjab India

Abstract

This work investigates user online browsing and purchasing behaviors, and predicts purchasing actions during a large shopping festival in China. To improve online shopping experience for consumers, increase sales for merchants and achieve effective warehousing and delivery, we first analyse diverse online shopping behaviors based on the 31 million logs generated accompanied with online shopping during a rushed sale event on 11st November, 2016. Based on the obtained user behaviors and massive data, we apply collaborative filtering based method to recommend items for different consumers, and predict whether purchase will happen. We conduct 5-fold cross validation to evaluate the collaborative filtering based recommendation method, and further identify the critical shopping behaviors that determine the precursors of purchases. As online shopping becomes a global phenomenon, findings in this study have implications on both shopping experience and sales enhancement.

Keywords

Filtering Online Shopping, Behavior Analyse, Recommendations, Purchase Prediction, Collaborative Filtering

I. Introduction

In the last years the widespread of several technologies have changed the way people manage, access, and distribute multimedia content. Technologies such as the development and dissemination of P2P networks, and the increase in storage capacity of portable devices had special effect in the worldwide diffusion of multimedia content. Among all the many kinds of multimedia content, music is one of the most popular content now a day. The reason for this is that music is an art and can be shared by many people from different countries, languages, and cultural backgrounds. One point of reference for this affirmation is the number of items sold daily by web-based dealers, or the number of items downloaded and shared via the internet. In its 2007 Digital Music Report, the International Federation of the Phonographic Industry (IFPI), stated that the number of legally downloaded songs in 2006 reach the amount of 795 millions. These facts show us that music is commercially and culturally important.

As the amount of audio content available is increasing several questions arise on how to efficiently access, discover, and present it to the final user. In order to answer all these questions there is the need for new techniques for classifying, searching and retrieving, and recommending audio content. In this thesis we will focus mainly on techniques for recommending audio contents to users. Commercial applications such as content-based music recommendation systems may become increasingly important component of e-commerce applications. One of the advantages of these applications is that they do not need a lot of effort from the user, who is simply presented with potentially relevant items.

Recommendation of music is emerging with force nowadays due to the huge amount of music content and because users normally don't have the time to search through these collections looking for

new items. The main purpose of a recommendation system is to estimate the user's preferences and present him with some items that he doesn't know yet. Currently, most of the audio recommendation systems can be classified in two major kinds. Recommendation systems based on collaborative filtering techniques and content-based recommendation systems. While both kinds of systems have good characteristics, they fail to provide good recommendations in specific situations. Recently a new kind of recommendation systems is emerging, hybrid content-based collaborative filtering recommendation systems. In the next paragraphs we will explain each of these kinds briefly.

The main idea of collaborative filtering (CF) methods is to recommend items to a user by taking into account the rates on those items made by other users. CF systems work by collecting user feedback in the form of ratings and exploit similarities and differences among profiles of several users. According to this feedback can be given explicitly, ratings or annotation, or implicitly such as the time spent in examining the content of the recommendations. Although CF has proved to provide good recommendation it presents some troubles. One of the major disadvantages of collaborative filtering is that it cannot recommend new items since the new items don't have any rate. Another disadvantage of CF is that users need to be involved a lot with the system and provide a lot of rates. This is a disadvantage because it is hard to obtain reliable information from many users.

A. Recommendation Systems

Nowadays, there are a lot of recommendation systems, accessible via internet, which attempt to recommend to users several products such as music, movies, books, etc. In order to understand them first it is necessary to have a description. In a general way, recommendation systems are systems which intend to acquire opinions or preferences about items from a community of users, and use those opinions to present other users with items that are interesting to them. From this general description we can see that recommendation systems need two basic things to work properly: Information about the preferences of the users, and a way to determine if an item is interesting for a user. Normally, the users' information includes external information, such as user profiles, purchases histories, and product ratings. The way to determine whether an item is interesting to a user or not, depends on the kind of recommendation system, and in the techniques used to find similarities among items or users.

The above description is quite general and could be applied even to persons that recommend items to other persons (the salesman in a records' store). A more specific definition of recommendation systems is given in.

"System that produce individualized recommendations as output or have the effect of guiding the user in personalized way to interesting or useful objects in a large space of possible options".

The main keywords in this more formal definition are individualized and personalized. These terms indicate that every user will be presented with different information sources or items depending

on the information the system has about every user. In order to continue or discussion about recommendation systems, how do they work, and which kinds exist currently, we will have to define several terms that we will be using through the rest of this document.

1. Terms and Concepts

The following terms are often used in a recommendation system and the definitions introduced here are based on the work presented in.

i. Item

In the context of recommendation systems, an item represent the information the system posses about any object. An object can be an electronic document, a product, a person, a service or anything that can be represented by information

ii. Recommender

A recommender is any entity that gives personalized recommendations as output to users' preferences. It may be possible that a recommender does not produce a specific output, such as a list, but they might guide somehow the users in an individual way to useful or interesting items. A recommender could be a person or a software system.

iii. Recommendation

This is the output of a recommender; it can be compound by an item or a list of items. The items presented to the users have to be interesting to them, according to the recommender. The criteria used to determine if an item is interesting or not for a user depends exclusively on the technique used by the recommender.

iv. User's Interest

This is an abstract representation of how much a user appreciates an item. This is a subjective concept and it is hard to represent it in an objective way.

v. Prediction

The expected interest of a user in one item. This concept is different to the concept of recommendation. While some systems might present predictions with the actual recommendations, others can produce recommendations only.

vi. Rating

An objective measure representing a user's interest. The possible values of this measure are given according to a scale established by the designer of the recommendation system.

vii. Predicted Rating

An objective measure representing the expected interest of a user in certain item. This measure is estimated by the system and its possible values are elements of a specific scale.

viii. Actual Rating

Objective measure representing the real interest of the user in a specific item. This value is given by the user himself according to the scale of rates of the system.

ix. Prediction Accuracy

A measure that indicates at which extent the predicted rating agrees with the user's actual rating. The more accurate the predictions the better the performance of the recommendation system.

x. Prediction Technique

The specific algorithm that the recommendation system will use in order to calculate the predicted rating of an item.

B. The Recommendation Process

Once we have defined some concepts and terms generally used in recommendation systems now it is necessary to explain the way these systems work and produce recommendations to the users. In a general way every recommendation system follows a specific process in order to create recommendations. Here we will explain this process and each of the steps involved.

If we see the process of recommendation as a black box, as shown in figure, we can identify two sources of information needed as input for the process. These sources of information are the users' profiles and the information about items or products. Ideally the information stored in the profiles is related with the preferences of the users and should be given explicitly by the user itself.

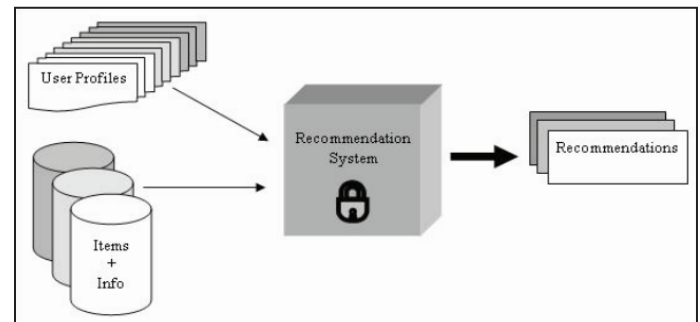


Fig. 1: Recommendation Process as a Black Box

The recommendation process in a more detailed way is showed in figure. It includes the following steps: information recollection, selection, transformation, structuring and presentation. From all the steps presented in figure, the information recollection step is the only one that is not done by the system itself. Below we will give a small description of each of these steps

C. Classification of Recommendation Systems

Since one of the most important components of every recommendation system is the one in charge of making predictions, it is logical to classify them according to the prediction technique they use. Base on the prediction technique they use it is possible to classify the recommendation systems in two main groups: systems that use social-based prediction techniques and systems that use information-based prediction techniques. Additionally to these two first groups there is a special group of recommendation system. This third group is made by systems that use hybrid methods for recommending items.

1. Collaborative Filtering

The growth of the Internet has made it much more difficult to effectively extract useful information from all the available online information. The overwhelming amount of data necessitates mechanisms for efficient information filtering. One of the techniques used for dealing with this problem is called collaborative filtering.

The motivation for collaborative filtering comes from the idea that people often get the best recommendations from someone with similar tastes to themselves. Collaborative filtering explores techniques for matching people with similar interests and making recommendations on this basis.

II. Literature Survey

Benjamin Marlin et. Al (2004) In this paper, Collaborative filtering was initially proposed as a framework for filtering information based on the preferences of users, and has since been refined in many different ways. This thesis is a comprehensive study of rating-based, pure, non-sequential collaborative filtering. We analyze existing methods for the task of rating prediction from a machine learning perspective. We show that many existing methods proposed for this task are simple applications or modifications of one or more standard machine learning methods for classification, regression, clustering, dimensionality reduction, and density estimation. We begin by summarizing the work contained in this thesis, outlining the main developments, and indicating the primary results.

Bamshad Mobasher, Robin Burke et. Al (2006) In this paper, we examine the robustness of model-based recommendation algorithms in the face of profile injection attacks. In particular, we consider two recommendation algorithms, one based on k-means clustering and the other based on Probabilistic Latent Semantic Analysis (PLSA). These algorithms aggregate similar users into user segments that are compared to the profile of an active user to generate recommendations. Traditionally, model-based algorithms have been used to alleviate the scalability problems associated with memory-based recommender systems.

James Salter, Nick Antonopoulos et. Al (2006) In this paper, Directing users to relevant content is increasingly important in today's society with its ever-growing information mass. To this end, recommender systems have become a significant component of e-commerce systems and an interesting application domain for intelligent agent technology. Traditionally, recommender systems employ collaborative filtering—recommending movies, for example, by matching a user to other users with similar tastes and suggesting movies these others have enjoyed.

Abhinandan Das Abhinandan Das, Mayur Datar et. Al (2007) In this paper, Several approaches to collaborative filtering have been studied but seldom have studies been reported for large (several million users and items) and dynamic (the underlying item set is continually changing) settings. In this paper we describe our approach to collaborative filtering for generating personalized recommendations for users of Google News.

Georg Groh, Christian Ehmig et. Al (2007) In this paper, We investigate how social networks can be used in recommendation generation in taste related domains. Social Filtering (using social networks for neighborhood generation) is compared to Collaborative Filtering with respect to prediction accuracy in the domain of rating clubs. After reviewing background and related work, we present an extensive empirical study where over thousand participants from a social networking community where asked to provide ratings for clubs in Munich.

Hugo Siles Del Castillo et. Al (2007) In this paper, The main purpose of a recommendation system is to estimate the user's preferences and present him with some items that he does not know yet. Currently, most of the audio recommendation systems can be classified in two major kinds: recommendation systems based on collaborative filtering techniques and content based recommendation system. While both kinds of systems have good characteristics, they fail to provide good recommendation

is specific situations.

Robert M. Bell, Yehuda Koren et. Al (2007) In this paper, Recommender systems based on collaborative filtering predict user preferences for products or services by learning past user-item relationships. A predominant approach to collaborative filtering is neighborhood based ("k-nearest neighbors"), where a user-item preference rating is interpolated from ratings of similar items and/or users. In this work, we enhance the neighborhood-based approach leading to a substantial improvement of prediction accuracy, without a meaningful increase in running time.

Robert M. Bell, Yehuda Koren et. Al (2007) In this paper, Recommender systems based on collaborative filtering predict user preferences for products or services by learning past user-item relationships. A predominant approach to collaborative filtering is neighborhood based ("k-nearest neighbors"), where a user-item preference rating is interpolated from ratings of similar items and/or users.

Ruslan Salakhutdinov, Andriy Mnih et. Al (2007) In this paper, we show how a class of two-layer undirected graphical models, called Restricted Boltzmann Machines (RBM's), can be used to model tabular data, such as user's ratings of movies. We present efficient learning and inference procedures for this class of models and demonstrate that RBM's can be successfully applied to the Netflix data set, containing over 100 million user/movie ratings. We also show that RBM's slightly outperform carefully-tuned SVD models.

Zan Huang, Daniel Zeng et. Al (2007) In this paper, Collaborative filtering is one of the most widely adopted and successful recommendation approaches. Unlike approaches based on intrinsic consumer and product characteristics, CF characterizes consumers and products implicitly by their previous interactions. The simplest example is to recommend the most popular products to all consumers. Researchers are advancing CF technologies in such areas as algorithm design, human computer interaction design, consumer incentive analysis, and privacy protection.

Zeng, Ming et. al This work investigates user online browsing and purchasing behaviors, and predicts purchasing actions during a large shopping festival in China. To improve online shopping experience for consumers, increase sales for merchants and achieve effective warehousing and delivery, we first analyse diverse online shopping behaviors based on the 31 million logs generated accompanied with online shopping during a rushed sale event on 11st November, 2016. Based on the obtained user behaviors and massive data, we apply collaborative filtering based method to recommend items for different consumers, and predict whether purchase will happen. We conduct 5-fold cross validation to evaluate the collaborative filtering based recommendation method, and further identify the critical shopping behaviors that determine the precursors of purchases. As online shopping becomes a global phenomenon, findings in this study have implications on both shopping experience and sales enhancement. In this article, we investigate how the big discounts and promotions offered on November 11st influence the sales of e-commerce websites, consumers' online shopping behaviors and the popularity of items based on the logs cleaned from the DPI dataset. The sales of e-commerce are stimulated sharply by discounts and promotions. We find that the sales per half an hour reach the peak at 00:30 and 10:30 am on November 11st that are half an hour later than the time

that big discounts and promotions start. Customers are more likely to shop around on the website before they make orders. The last conclusion is that the sales of the most popular items on November 11st could be increased several times. We also apply a collaborative filtering based approach to recommend items to users, and five-fold cross validation is conducted to evaluate the effectiveness of the proposed method. Moreover, we test the efficacy of several feasible precursors of purchasing actions (e.g., the effect of total browsing time, the number of clicks, product categories, and time of day in future purchases) by examining whether visiting the shopping site prior to the sale event or browsing a coupon page is indicative of future purchases. The effect of holidays and shopping season to retails is critical. By far being the busiest shopping season of the year and this period can determine the difference between profit and loss for the year for many retailers. Mobile clearly is the big story. Despite accounting for a smaller overall percentage of spending, mobiles have an outsize impact on retail sales growth. In terms of traffic, mobile outpaced desktop retail traffic by a factor of 2 and was higher also on the Cyber Monday, when online retailers promote exceptional bargains immediately following the Thanksgiving holiday weekend in the US.

Sandeep K. Raghuvanshi et. al Recommendation systems are information filtering systems that urge to predict preferences that user might have for an item over other. Recommendation systems are very popular in applications like movies, books, research articles, search queries, social tags, product, financial services, restaurants, twitter pages, job, university, friends and what not. To increase product sales is the primary goal of recommendation system by bringing a relevant item to the user and thus increasing the overall profit, which covers the functional goal of recommendation system such as [1]—relevancy, serendipity and diversity. Most popular recommender systems of today are Group Lens recommender system, Amazon.com recommender system, Netflix Movie recommender system, Google News personalization system, Face book friend recommendations; link prediction recommender system Recommendation system serves as a useful tool for users in expanding their interest and their experience over the Internet. Recommendation accelerates profits for developer and business person by knowing their customers well serving them best. Along with mobiles and computers, they open new security doors for the automobile industry and devices used on daily basis. Among several solutions and facilities, there are some issues related to available recommendation system that needs to be addressed specifically to take most out of them

III. Proposed Work

With the recent advances in technology, there is an emerging presence of social media and social networking systems. From the point of view of a sociologist, social media can be characterized as “collective goods produced through computer-mediated collective action”. In the case of multimedia enriched social network systems, such as last.fm, the collective goods are musical tracks and the collective action is the process of crafting individual profiles of musical preference and linking them either explicitly, via bonds of friendship, or implicitly, through collaborative annotation.

This collective action leads to the creation of an implicit social networking structure, which we aim to further explore. In particular given the success of item recommendation systems in commercial websites, such as Amazon.com and Net ix, it is considered worthwhile to revisit the recommendation problem through the novel perspective of social networking. In general, recommendation

systems aim to provide personalized recommendations of items to users based on their previous behavior as well as on other information gathered by item descriptions and user profiles. However, no emphasis has been placed yet on personalization based explicitly on social networks.

The reason is that despite there is an increasing interest in the exploration of social networks, there does not exist a concrete dataset that includes both explicit bonds of friendships among users and free-form collaborative annotation of items. This is due to that most social media systems do not allow for free access to all user profiles or lists of friends.

Given the incentives of the widespread adoption of social networks and of the lack of some previous study that directly addresses the problem of efficiently integrating the added value knowledge provided by those networks in the field of collaborative recommendation, we propose a new methodology that tackles the aforementioned issues.

A. Online Recommendation

Based on the result of offline data, recommend online which will reduce waiting time for users greatly, at the same time increases the accuracy of recommendation through the user’s credibility clustering method. The user’s satisfaction will be enhanced all-round. The detailed steps are as follows:

1. Calculate the similarity of target user u_a with each cluster center W_k according to similarity formula(1), distribute the target user to the cluster with the closest similarity;
2. Take all user’s information from the cluster u_a belongs to in data base;
3. Calculate the modified similarity $\text{sim}'(u_a, u_i)$ of target user u_a with each user in this cluster according to formula, choose the N neighbors that have the closest modifying similarity as the closest neighbor set U' ;
4. Calculate the Prediction rating P'_{u_a, m_x} given by target user u_s to all projects that haven’t been rated according to formula, choose the few resources that have the highest rating as recommendation.

IV. Conclusion and Future Scope

In this article, we investigate how the big discounts and promotions offered on November 11st influence the sales of e-commerce websites, consumers’ online shopping behaviours and the popularity of items based on the logs cleaned from the DPI dataset. The sales of e-commerce are stimulated sharply by discounts and promotions. We find that the sales per half an hour reach the peak at 00:30 and 10:30 am on November 11st that are half an hour later than the time that big discounts and promotions start. Customers are more likely to shop around on the website before they make orders. The last conclusion is that the sales of the most popular items on November 11st could be increased several times. We also apply a collaborative filtering based approach to recommend items to users, and five-fold cross validation is conducted to evaluate the effectiveness of the proposed method. Moreover, we test the efficacy of several feasible precursors of purchasing actions (e.g., the effect of total browsing time, the number of clicks, product categories, and time of day in future purchases) by examining whether visiting the shopping site prior to the sale event or browsing a coupon page is indicative of future purchases. The effect of holidays and shopping season to retails is critical. By far being the busiest shopping season of the year and this period can determine the difference between profit and loss for the year for many retailers. Mobile clearly is the big story. In

addition, generalizing the analysis and proposed algorithm in this paper on dataset in different years and different websites can be very convinced, which is also left as our future work till we obtain the dataset in other years and other websites. We wonder if the key shopping behaviors observed from this dataset would be similar to that seen in other shopping websites during the shopping festival. Last but not least, we did not look into such cases nor proximity of information search behavior, where two connected individuals may search for similar items on the JD.com shopping site. To further improve the accuracy of recommendation and prediction of purchase actions, advanced techniques such as deep learning and transfer learning can be applied into these tasks. We leave these issues as future work

References

- [1] Huang, Zan, Daniel Zeng, Hsinchun Chen, "A comparison of collaborative-filtering recommendation algorithms for e-commerce." *IEEE Intelligent Systems* 5, pp. 68-78, 2007.
- [2] Cacheda, Fidel, Víctor Carneiro, Diego Fernández, VreixoFormoso, "Comparison of collaborative filtering algorithms: Limitations of current techniques and proposals for scalable, high-performance recommender systems." *ACM Transactions on the Web (TWEB)* 5, no. 1: 2, 2011.
- [3] Tso-Sutter, Karen HL, Leandro BalbyMarinho, Lars Schmidt-Thieme, "Tag-aware recommender systems by fusion of collaborative filtering algorithms." In *Proceedings of the 2008 ACM symposium on Applied computing*, pp. 1995-1999. ACM, 2008.
- [4] Takács, Gábor, IstvánPilászy, BotyánNémeth, DomonkosTikk, "Scalable collaborative filtering approaches for large recommender systems." *The Journal of Machine Learning Research* 10, pp. 623-656, 2009.
- [5] Das, Abhinandan S., Mayur Datar, Ashutosh Garg, Shyam Rajaram, "Google news personalization: scalable online collaborative filtering", In *Proceedings of the 16th International Conference on World Wide Web*, pp. 271-280. ACM, 2007.
- [6]. Groh, Georg, Christian Ehmig, "Recommendations in taste related domains: collaborative filtering vs. social filtering", In *Proceedings of the 2007 international ACM conference on Supporting group work*, pp. 127-136. ACM, 2007.
- [7]. Hu, Yifan, Yehuda Koren, Chris Volinsky, "Collaborative filtering for implicit feedback datasets." In *Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on*, pp. 263-272. IEEE, 2008.
- [8] Candillier, Laurent, Frank Meyer, Marc Boullé. "Comparing state-of-the-art collaborative filtering systems." In *Machine Learning and Data Mining in Pattern Recognition*, pp. 548-562. Springer Berlin Heidelberg, 2007.
- [9] Bell, Robert M., Yehuda Koren, "Improved neighborhood-based collaborative filtering." In *KDD cup and workshop at the 13th ACM SIGKDD International Conference on knowledge discovery and data mining*, pp. 7-14. sn, 2007.
- [10] Koren, Yehuda, "Factor in the neighbors: Scalable and accurate collaborative filtering." *ACM Transactions on Knowledge Discovery from Data (TKDD)* 4, No. 1, 2010.
- [11] Li, Bin, Qiang Yang, XiangyangXue, "Transfer learning for collaborative filtering via a rating-matrix generative model", In *Proceedings of the 26th Annual International Conference on Machine Learning*, pp. 617-624. ACM, 2009.
- [12] Ekstrand, Michael D., John T. Riedl, Joseph A. Konstan, "Collaborative filtering recommender systems", *Foundations and Trends in Human-Computer Interaction* 4, No. 2, pp. 81-173, 2011.
- [13] Bach, Francis, Jacob Abernethy, Jean-Philippe Vert, and TheodorosEvgeniou. "A New Approach to Collaborative Filtering: Operator Estimation with Spectral Regularization." (2008).
- [14] Lee, Joonseok, Mingxuan Sun, Guy Lebanon, "A comparative study of collaborative filtering algorithms." *arXiv preprint arXiv:1205.3193*, 2012.
- [15] Bell, Robert M., Yehuda Koren, "Scalable collaborative filtering with jointly derived neighborhood interpolation weights", In *Data Mining, 2007. ICDM 2007. Seventh IEEE International Conference on*, pp. 43-52. IEEE, 2007.
- [16] Liu, Fengkun, and Hong Joo Lee. "Use of social network information to enhance collaborative filtering performance." *Expert systems with applications* 37, No. 7, 4772-4778, 2010.
- [17] Mobasher, Bamshad, Robin Burke, Jeff J. Sandvig. "Model-based collaborative filtering as a defense against profile injection attacks." In *AAAI*, Vol. 6, pp. 1388. 2006.
- [18] Marlin, Benjamin, "Collaborative filtering: A machine learning perspective." PhD diss., University of Toronto, 2004.
- [19] Raghuwanshi, Sandeep K., R. K. Pateriya, "Collaborative Filtering Techniques in Recommendation Systems", In *Data, Engineering and Applications*, pp. 11-21. Springer, Singapore, 2019.
- [20] Zeng, Ming, Hancheng Cao, Min Chen, Yong Li, "User behaviour modeling, recommendations, and purchase prediction during shopping festivals", *Electronic Markets* 29, No. 2, pp. 263-274, 2019.