

A Review on Current Recommendation Systems

Shivani Sharma
Research Fellow
CDAC, Mohali

Manish Dhir
Project Asstt.
CDAC, Mohali

ABSTRACT

In recent years, E-commerce, web service and web information system have been used explosively. Massive explosion of the world-wide-web and the emergence of e-commerce have encouraged designers to develop recommendation systems. Web users demonstrate a variety of navigational patterns by clicking series of web pages. These patterns can be understood by mining user logs using WUM. One of the widely used applications of Web Usage Mining is Online Recommendation and prediction. Generally, all the recommendation systems follow a framework for generating efficient recommendations. Various recommendation systems use different approaches based on the sources of information they utilize. The accessible sources are user information (demographics), the product information (keywords, genres) and the user-item ratings. This paper gives introductive information about recommendation system, their techniques, and algorithms and also describes some existing works.

General Terms

Recommendation System, Pattern detection.

Keywords

Recommendation System, Knowledge based systems, Web Usage Mining.

1. INTRODUCTION

The rise in quantity of data of knowledge over web in previous couple of years has caused the growing risk of data flooding that successively has created the matter of accessing relevant data to the users. Conjointly with the hike in variety of internet sites and sites, webmasters notice it difficult to formulate the content in accordance with the user's want. The data demand of internet users are often worked out by evaluating user's web navigation behaviour. Web Usage Mining (WUM) is employed to extract data from net user's access logs by using data processing techniques. One in all the applications of WUM is Recommendation system that is personalised info filtering technique accustomed either verify whether or not a definite user can approve a given item or to spot a listing of things which might be of great importance to the user.

Recommender frameworks help clients discover and evaluate things of investment. Recommendation Systems can utilize data mining strategies for making suggestions utilizing learning gained from the activity and qualities of the clients. The architecture of an online web recommendation system based on web usage mining basically consists of three phases: Data Pre-processing, Pattern detection and generating recommendations. Data pre-processing and Pattern detection phases are performed offline and the recommendations are generated online. Data pre-processing involves transforming the web access logs and user profiles into format appropriate for the system. Pattern detection involves using data mining techniques like clustering, sequential pattern mining or association rule mining. Current recommendation systems exhibit certain limitations such as intelligence, adaptability, flexibility, limited accuracy. The disadvantage of these

systems is that the new user obtains the recommendations only on the basis of his current navigation.

There are many definitions of recommender systems. One of the first was presented by Paul Resnick and Hal R. Varian in 1997. They claim that "in a typical recommender system, people provide recommendations as inputs, which the system then aggregates and directs to appropriate recipients" [4]. These systems are usually defined in terms of their functionality as the systems or agents that suggest the products to the users who purchase products on e-commerce sites [3]. The recommender systems help the consumer to make the decision what to buy. The recommender systems can be classified because of the level of personalization into non-personalized and personalized methods (Figure1). The former methods do not consider the characteristics and preferences of the customers, whereas the latter tightly depends on the user profile.

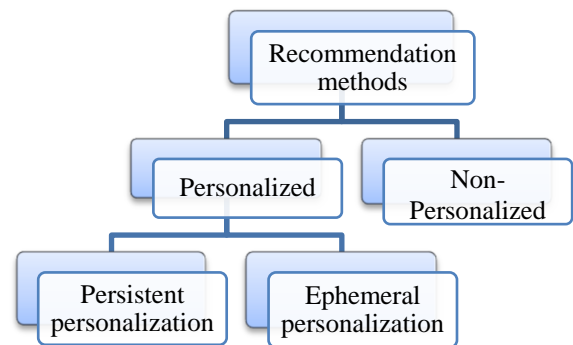


Fig 1: Level of Personalization in Recommender system

The example of the non-personalized method is the recommendation that suggests the products which were best rated in the past by all the customers in average ("best rated") or the number of their copies, which were sold, is the greatest ("best buy"). In order to create this kind of recommendation the statistical methods are used commonly. Moreover, in another variant of non-personalised approach even new items can be recommended, e.g. the books published in the last month. This kind of recommendation depends on the policy of the e-commerce and belongs to the techniques where not much calculation is required. The main feature of those suggestions is that they are the same for all customers. Usually, it is easy and quite convenient for a user to find one item from the list of the most popular ones. This is similar to the situation, when someone goes to the bookstore and finds there the shelf with the bestsellers. However, some research claims that the recommender systems are only those ones, which produce personalized recommendations [5]. In other words, the output of these systems is the individualized recommendation that helps to guide the single user to products or services that fulfil their particular needs. As a result, they cope with information overload better than the non-personalized methods and enable to find and purchase the

right items from the large amount of possible choices. The personalized recommendation is based either on the demographic information about users or on the analysis of the past behaviour of the user in order to predict their future behaviour (collaborative and content-based filtering) [6]. Moreover, the personalization can be either persistent or ephemeral [7]. Persistent personalization, based on the previous users' behaviours, enables to create unique list of products for each user. The requirement that ought to be fulfilled in this situation is that customers must log into the system in order to create user profile for each of them. In a persistent personalized recommendation each person on the Web site sees different recommendations because they depend directly on the users' personal data. The recommendations rely on the information derived from the survey responses, purchasing history, products ratings, etc. The user profile is not necessary in the ephemeral personalization. In this case the recommendations are created according to the users' behaviours during a current session, their navigation and selection [1]. In this technique the recommendations are the same for all users.

The main goals of the recommender systems are as follow:

- To cope with information overload
- To help all customers (new, frequent, and infrequent) to make decisions what products to buy, which news to read next, which movie is worth watching, etc.
- To convert observers to buyers
- To build credibility through community and maintain the loyalty of the customers
- To inviting customers to come back
- To enhance e-commerce sales and cross-sell

2. RECOMMENDER SYSTEM

Recommender Systems collect and exploit various types of information about users, products, and interactions between users and products to generate a personalized list of items that fits user's current needs. This information may be gathered while the user is interacting with the system or can be acquired from other systems the user has interacted with. The most commonly used information about the interaction between users and products is the set of subjective ratings assigned by the users to previously experienced products. The system then uses these ratings for predicting the ratings for products not yet experienced. The products with the highest estimated ratings are then recommended to a user.

Recommender systems techniques can be classified into four main categories [17]: Collaborative Filtering, Content Based Filtering, Knowledge Based systems and Hybrid systems.

Collaborative Filtering (CF) is a recommendation technique which emulates a simple and effective social strategy called "word-of-mouth" [16]. Here, personalized recommendations for a target user are generated using opinions of users having similar tastes to those of the target user. This is a domain independent approach, which computes recommendations by leveraging historical log data of users' online behaviour. Usually for this type of recommendations, the user must provide explicit opinion about an experienced item.

Content Based (CB) Filtering is another major technique that generates a recommendation for a user analyzing only the

ratings of the target user. Content Based techniques analyze items' descriptions to identify items that are of a particular interest to the target user. Typically, these CB systems learn the user profiles while analyzing the features of the items that were rated by the user in the past. Various machine learning techniques are exploited in order to learn these predictive models. Content Based Filtering plays a very specific role in RS literature, as it does not suffer from the new-item problem, i.e., the impossibility to recommend an item for which no rating was previously collected. Therefore, it is typically preferred to CF in those domains where the main need is to recommend recent items or data sparsity is very high [18].

Knowledge-based (KB) systems use a decision model to make inferences about the user needs. Usually a KB system computes or simply stores some recommendation rules designed by a domain expert. It can also use previous user interactions with the system and later exploit similar observed cases to generate a prediction [14]. The main advantage of KB systems is their flexibility to incorporate various types of rule, e.g., business rules, into the recommendation process. Moreover, these systems have no cold start problem, as no user input is required to bootstrap the system. However, such systems are very expensive to build as they require accurate input from a domain expert. Because of that, pure knowledge based systems are not very common nowadays and usually they are combined with the previously mentioned techniques []. Knowledge based system could be easily extended to use contextual information.

Hybrid Systems combines several RS techniques to overcome the limitations of pure techniques []. Hybrid systems often provide better accuracy than pure techniques [12]. Moreover, CF does not provide a direct way to integrate context into the prediction model. Therefore, a natural way to integrate such additional information is to use a hybrid system. In fact, most contextual post-filtering approaches would fall into the group of cascading hybrid methods. Here rating prediction is done by CF and adapted by different predictive system that uses context. The classical pre-filtering CACF method [13] can be also seen as a hybrid system.

3. ALGORITHMS TO BUILD RECOMMENDATION SYSTEM

There are three important criteria that determine how useful an algorithm is and that are: (i) Quality of Predictions, (ii) Scalability/Speed, (iii) Easily Updated. There are three types of algorithms used to build recommendation system:

3.1 Memory primarily based Algorithms

Memory-based algorithms approach the cooperative filtering downside by exploitation the complete information. As delineate by Breese et. al [1], it tries to seek out users that area unit almost like the active user (i.e. the users we would like to form predictions for), and uses their preferences to predict ratings for the active user.

Advantages:

- The quality of predictions is quite sensible.
- This may be a comparatively easy rule to implement for any scenario.
- It is incredibly simple to update the information, since it uses the complete information when it makes a prediction.

Disadvantages:

- It uses the complete information when it makes a prediction, thus it must be in memory it's terribly, very slow.
- Even once in memory, it uses the complete information when it makes a prediction, thus it's terribly slow.
- It will generally not build a prediction sure as shooting active users/items. This will occur if the active user has no things in common with all people that have rated the target item.
- Over fits the information. It takes all random variability in people's ratings as feat, which might be a true downside. In different words, memory-based algorithms don't generalize the information in the least.

3.2 Model primarily based Algorithms:

Memory-based recommendation systems don't seem to be perpetually as quick and ascendable as we'd like them to be, particularly within the context of actual systems that generate period recommendations on the idea of terribly massive datasets. to attain these goals, model-based recommendation systems area unit used. Model-based recommendation systems involve building a model supported the dataset of ratings. In different words, we tend to extract some info from the dataset, and use that as a "model" to form recommendations while not having to use the whole dataset when. This approach doubtless offers the advantages of each speed and quantifiability.

Advantages:

- Scalability: Most models ensuing from model-based algorithms area unit abundant smaller than the particular dataset, so even for terribly massive datasets, the model lands up being sufficiently small to be used expeditiously. This imparts quantifiability to the system.
- Prediction speed: Model-based systems are possible to be quicker, a minimum of as compared to memory-based systems as a result of, the time needed to question the model (as hostile the total dataset) is typically abundant smaller than that needed to question the total dataset.
- Avoidance of over fitting: If the knowledge set over that we tend to build our model is representative enough of real-world data, it's easier to do to avoid over fitting with model-based systems.

Disadvantages:

- Inflexibility: as a result of building a model is usually a time- and resource-consuming method, it's sometimes harder to feature knowledge to model-based systems, creating them inflexible.
- Quality of predictions: the actual fact that we tend to don't seem to be exploitation all the data (the whole dataset) obtainable to United States of America, it's doable that with model-based systems, we do not get predictions as correct like model-based systems. It ought to be noted; however, that the standard of predictions depends on the means the model is constructed. In fact, as is seen from the results page, a model-based system performed the simplest among all the algorithms we tend to tried.

3.3 Association Rules

Association rules presents a novel algorithmic rule that doesn't perform like every others we tend to worked with. Association rules try and connect the causative relationships between things. Associate in Nursing association rule primarily is of the shape $A_1, A_2, A_3, \dots, B_1, B_2, B_3$, It makes an attempt to indicate however a series of things will confirm another series of things. For a additional concrete example, if we tend to same $A \rightarrow B, C$, that will mean that the looks of item own someone's history would imply that B and C would be there moreover. it is not simply the things that matter, however; another necessary issue is that the confidence of a rule. Confidence is that the intuitive plan of however applicable a rule is. It will vary from zero to one. If the arrogance is one, then we all know that the rule continuously applies - that's, anytime we tend to see A, we tend to conjointly see B and C. However, if the arrogance is zero, it means that it's ne'er correct - A doesn't imply B and C.

Advantages:

- It is improbably quick. Building the matrix takes a awfully short quantity of your time, and so all recommendations afterward area unit instant.
- It generates credible results - what you would possibly consider recommending someone given what they need seen.
- It works well with distributed information sets, particularly if we tend to implement a multi-level association rule index that has higher levels of generalization (in case the lower ones do not have enough information). we tend to did not feel the necessity to implement it, however the choice was there.

Disadvantages:

- It generates obvious results. as an example, as a result of everybody rated and enjoyed Star Wars, it nearly always advocate that you just go watch it if you haven't.
- It cannot predict ratings, solely order preferences. If you wish to truly predict ratings you would like to use a special algorithmic rule.

4. RELATED WORK

Hiral and Meera [3] proposed a hybrid architecture that integrates product information with user's access log data and then generates a set of recommendations for that particular user. In this work, an architecture that integrates product information with user's access log data and then generates a set of recommendations for that particular user. This architecture uses more than one data mining algorithms like clustering and pattern matching algorithms. For Clustering they use K-mean Clustering & other algorithm for pattern matching is Boyer Moore Pattern Matching Algorithm. To enhance this work, an optimization algorithm will be used. Here Particle Swarm optimization algorithm has been used which is a soft computing algorithm of Artificial Intelligence. It will help to improve the results on the basis of intelligence.

Sneha Y.S, G. Mahadevan, Madhura, [1][2] proposed an structural engineering which incorporates semantic data with web utilization mining. Longest Common Subsequence is utilized to create a rundown of suggested items. This system improved the performance of the existing Recommender systems by overcoming the new item problem. The system

comprised of online and offline phase. RDF model is used for incorporating Semantic information. The System does not involve clustering of user profiles which leads to searching of patterns by traversing all usage logs thus leading to more time consumption and thus reducing performance of the overall system.

Hadi Khosravi et al [4] executed a Semantic Recommendation Procedure for Electronic item Catalog. The key perspectives incorporates displaying and arrangement of the web articles, matching among items and deciding the set of the activity to be suggested for personalization. This system implemented the significant elements of web recommendation and personalization like Modelling of web pages or web products and customers, Mapping between appropriate customers and products and determining the set of recommendations. This system made use of Ontology and OWL (Web Ontology Language) for web product classification. This system avoided false positive recommendations i.e. the product will be recommended even though it would not be relevant to the customer.

Mehradad Jalali et al [5] proposed an online recommendation System utilizing LCS Algorithm. The system involved two phases that work in conjunction with each other i.e. the online and offline phase. Data pre-treatment and navigation pattern mining is carried out in offline phase while predictions are generated in the online phase. The accuracy of expectation is 73%.

Suleyman Salin and Pinar Senkul [6] have utilized an architecture which uses semantic data in pattern generation phase for Web Usage Mining. Spade, is utilized to create regular access successions. Here sequential association rule mining is used.

Xin Sui, Suozhu Wang, Zhaowei Li [7] have researched and proposed a model which integrates Semantic Web and Agent Personalized Recommendation System (SWARPS) in E-Commerce. It involves implementation of AI-Multi Agent technique. The architecture is divided into various agents (monitor agent, web analysis agent, semantic conversion agent, data mining agent, preference analysis agent, semantic recommendation agent and recommendation engine agent) that carry out specific functionalities. The agents work jointly with each other to provide recommendations. The system exhibits limited degree of intelligence, autonomy and flexibility.

Himangni Verma and Hemant Verma [10] have put forth a customer end based system where at customer end web action of the customer framework is assessed. This assessment helps in developing the transactional patterns that helps in personalization and behavioural analysis. A frequent access pattern algorithm is used to divide the data into time slices. This data is formulated using Hidden Markov Model (HMM) for discovering the pattern in data which aids in generating many accurate and efficient recommendations.

5. CONCLUSION

With the growing demand of information over internet in last few years has caused the growing risk of information flooding which in turn has created the problem of accessing relevant data to the users. Also with the hike in number of websites and web pages, webmasters find it challenging to formulate the content in accordance with the user's need. The information demand of the online users can be figured out by evaluating user's web navigation behaviour. Web Usage Mining (WUM) is used to extract knowledge from Web user's access logs by

employing Data Mining Techniques. One of the applications of WUM is Recommendation system. Recommender frameworks help clients discover and evaluate things of investment. Recommendation Systems can utilize data mining strategies for making suggestions utilizing learning gained from the activity and qualities of the clients. A number of recommendation systems have been introduced but still current recommendation system exhibits certain limitations such as intelligence, adaptability, flexibility, limited accuracy.

6. REFERENCES

- [1] Sneha Y.S, G. Mahadevan, Madhura, :An Online Recommendation System based on web usage mining and semantic web using LCS Algorithm", IEEE 2011
- [2] Sneha Y.S, G. Mahadevan, Madhura, : A Personalized Product Based Recommendation System Using Web Usage Mining and Semantic Web", International Journal of Computer Theory and Engineering (IJCTE) Vol. 4, No. 2, April 2012
- [3] Mehrdad Jalali1, Norwati Mustapha, Md. Nasir B Sulaiman, Ali Mamat, "A Web Usage Mining Approach Based on LCS Algorithm in Online Predicting Recommendation Systems" 12th International Conference Information Visualisation IEEE 2008.
- [4] SuleymanSalim et al "Using Semantic Information for web usage mining based recommendation" International Conference IEEE 2009.
- [5] Xin Sui, Suozhu Wang, Zhaowei Li "Research on the Model of Integration with Semantic Web and Agent Personalized Recommendation System "Proceedings of the 2009 13th International Conference on Computer Supported Cooperative Work in Design.
- [6] A. Apostolico, "String editing and longest common subsequences", Handbook of Formal Languages, Vol. 2. Linear Modeling: Background and Application, Springer Verlag, Berlin, chapter 8, 1997, pp. 361–398.
- [7] RVVSV Prasad, V Valli Kumari "A Categorical Review of Recommender Systems" , International Journal of Distributed and Parallel Systems (IJDPS) Vol.3, No.5, September 2012 Hadi Khosravi Farsani, and Mohammadali Nematbakhsh "A Semantic Recommendation Procedure for Electronic Product Catalog", World Academy of Science, Engineering and Technology 22 2006.
- [8] Suguna, D, Sharmila, "Clustering Web log Files - A Review", International Journal of Engineering Research & Technology (IJERT) Vol. 2 Issue 4, April – 2013 ISSN: 2278- 0181
- [9] Himangni Rathore, Hemant Verma, "Analysis on Recommended System for Web Information Retrieval Using HMM", International Journal of Engineering Research and Applications ISSN : 2248-9622, Vol. 4, November 2014.
- [10] R.Baraglia, F.Silvestri," An online recommender system for large Web sites", Web Intelligence, IEEE/WIC/ACM, 2004, pp. 20–24.
- [11] Sneha Y S, G. Mahadevan, Madhura, "A Personalized Product Based Recommendation System Using Web Usage Mining and Semantic Web", 2011 International Conference on Information and Computer Application (ICICA 2011).

- [12] A. Singh and S. Sharma, "Role of Page ranking algorithm in Searching the Web: A Survey", *International Journal of Engineering & Technology, Management and Applied Sciences*, Vol. 1, Issue 1, June 2014.
- [13] P. Bari and P. M. Chawan. "Web Usage Mining", *Journal of Engineering Computers & Applied Science*, Vol. 2, No. 6, pp. 34-38, 2013
- [14] R. Sharma and K. Kaur, "Review of Web Structure Mining Techniques using Clustering and Ranking Algorithms", *International Journal of Research in Computer and Communication, IJRCCT*, Vol. 3, No. 6, pp. 663-668, 2014.
- [15] C. Tsai, C. Lai and M. Chiang, "Data mining for internet of things: A survey," *IEEE Communication Surveys & Tutorials*, Vol. 16, No. 1, 2014.
- [16] F. S. Gharehchopogh and Z. A. Khalifelu, "Analysis and evaluation of unstructured data: text mining versus natural language processing" *Application of Information and Communication Technologies (AICT)*, 2011 5th International Conference on. IEEE, pp. 1-4, 2011.
- [17] C. Kaur, and R. R. Aggarwal, "Reference Scan Algorithm for Path Traversal Patterns.", *International Journal of Computer Applications*, Vol. 48. Pp no. 20-25, 2012
- [18] V. Losarwar, M. Joshi, "Data Preprocessing in Web Usage Mining", *International Conference on Artificial Intelligence and Embedded Systems (ICAIES'2012)*, pp. 15-16, 2012.