A Study of Recommender System Techniques

Article *in* International Journal of Computer Applications · June 2012

DOI: 10.5120/7269-0078

2 authors:

Reena Pagare
MIT ART,DESIGN AND TECHNOLOGY UNIVERSITY

Anita Shinde
Marathwada Mitra Mandal's College of Engineering

A Study of Recommender System Techniques

Reena Pagare
Assistant Professor
Department Of Computer Engineering
M. I. T. College Of Engineering Pune

Anita Shinde
Department Of Computer Engineering
M. I. T. College Of Engineering Pune

ABSTRACT
Many clients like to use the Web to discover product details in the form of online reviews. These reviews are given by other clients and specialists. User-given reviews are becoming more prevalent. Recommender systems provide an important response to the information overload problem as it presents users more practical and personalized information services. Collaborative filtering techniques play vital component in recommender systems as they generate high-quality recommendations by influencing the likings of society of similar users.

General Terms
Recommender Systems, Algorithms.

Keywords
Collaborative Filtering, Sparsity problem, Trust network.

1. INTRODUCTION
Recommender systems give advice about products, information or services users might be interested in. Recommendation systems generate a ranked list of items on which a user might be interested. Recommendation systems are constructed for movies, books, communities, news, articles etc. They are intelligent applications to assist users in a decision-making process where they want to choose one item amongst a potentially overwhelming set of alternative products or services. Recommender systems are personalized information filtering technology used to either predict whether a particular user will like a particular item or to identify a set of N items that will be of interest to a certain user. It is not necessary that a review is equally useful to all users. The review system allows users to evaluate a review's support by giving a score that ranges from "not helpful" to "most helpful". If a particular review is read by all users & found helpful then it can be assumed that new user might appreciate it. Controversial reviews are the reviews that have a variety of conflicting rating (ranking). Controversial review has both passionate followers and motivated enemy without clear majority in either group. The Recommender System uses information from user profile & interaction to tell possible items of concern. It is useful to approximate the degree to which specific user will like a specific product. The Recommender systems are useful in predicting the helpfulness of controversial reviews [1].

Recommender systems are a powerful new technology for extracting additional value for a business from its user databases. These systems help users to find items they want to buy from a business. Recommender systems benefit users by enabling them to discover items they like. They help the business by generating more sales. Recommender systems are rapidly becoming a crucial tool in E-commerce on the Web.

This paper is organized as follows. Section 2 explains related work in recommendation systems categories. Section 3 introduces commonly used recommendation strategies. Section 4 concludes the paper.

2. RELATED WORK
2.1 Recommender System Categories
2.1.1 Ontological-Based Recommender Systems
Decentralized architectures, like peer-to-peer (P2P) networks, have inspired the progress of ontological-based recommender systems [2]**.** The distributed neighborhood-based recommender System is introduced which contains an epidemic-style protocol that preserves areas of like-minded users, and distributes information in a robust. This is done without any central involvement & in a dynamically changing large-scale environment. In [3], a multilayer semantic social network model is introduced. This model defines a system from different viewpoints. This recognizes a set of users having similar interest that correlate at different semantic levels. In [4] the concept of user contexts is used which corresponds to the different ranks of specificity to ontology. It creates a recommendation from the set of items most charged by the user and which can adjust to the level of specificity of the information presented to the user.

2.1.2 Recommenders in Collaborative Tagging Systems
In [5], the construction of collaborative tagging is evaluated. The collaborative tagging allows anyone particularly consumers to freely connect keywords or tags to data or content. It has also determined consistency in user activity, tag frequencies, kinds of tags used etc. It describes dynamical model of collaborative tagging that calculate stable patterns and narrates them to replication and shared knowledge.
[6] It introduces a generic model of collaborative tagging to recognize the dynamics behind it. It has observed the distribution of frequency of use of tags. The generic model uses power law distribution of tags. It combines model of tagging with feedback cycles & information value to generate stable distribution of tags. The collaborative tag suggestions algorithm uses score for each user. It actually defines a set of criteria for good tagging system. The tag suggestion uses this criterion to find high quality tags. They have eliminated noise & spam.

3. RECOMMENDATION STRATEGIES
The methods used for recommendations can be content based, collaborative filtering and trust based.

3.1 Content Based Methods
In content based method, items similar to those that user has previously purchased or reviewed are suggested. Here the scope of this recommendation is limited to the direct region of the users' previous purchase history or score. Content based system does not use any preference data and provides

recommendation directly based on similarity of items. Similarity is computed based on item attributes using appropriate distance measures. Content-Based (CB) Recommender Systems mean that the recommendations to a specified user based on the descriptions of the items. First, domain knowledge professionals are required to examine the items. Then categories of these items are listed. Finally, the system will use these categories of items to match the characters of a specific user.

Content-based filtering chooses documents based on the contents of documents & each user's preference. In content-based filtering, users can obtain suitable documents that match with their interests.

3.2 Collaborative Filtering Methods

Collaborative filtering creates personalized recommendations by combining the knowledge of similar users in the system. In collaborative Filtering (CF) technique, the recommendation process is automated by building on users' opinions of items in a community. Collaborative Filtering (CF) is based on the principle that the finest recommendations for an individual are given by people who have similar flavor. Collaborative filtering identifies users with choice similar to the target user and then computes predictions based on the score of the neighbors. Collaborative filtering considerably progresses recommendation system. The recommendation for a target item is based on other users' ranking of item instead of study contents. The job in collaborative filtering is to guess the usefulness of product to a particular user which is based on a database of user votes.

Collaborative filtering algorithms guess ranking of a target item for target user with help of grouping of the ranking of the neighbors (similar users) that are known to item under consideration. The six algorithms of collaborative filtering are evaluated. The input to algorithms is taken as interaction matrix A of order M x N = (aij) where M is number of consumers (c1,c2,c3,….cM) & N is number of products (p1,p2,p3,……..pN) . The recommendations are based on transactions. The value of aij can be either 0 or 1 where 1 means transaction between ci & pi (ci has brought pi) & 0 means absence of transaction. The output of algorithm is probable scores of product for each consumer. The recommendations consist of a ranked list of K products [7].

3.2.1 The user-based algorithm

This algorithm is used to predict target consumer's future transactions by combining the observed transactions of similar consumers. First the algorithm calculates a consumer similarity matrix WC = (wc_{st}) which determines the similarity score based on row vector of A. A high value of wc_{st} shows that consumers s & t have similar liking as they have already brought many similar products. WC A will give the products' probable score for each consumer. Resulting matrix will be containing element at c^{th} row & p^{th} column combine s the scores of the similarities between consumer c and other consumers who have purchased product p[8].
User based algorithms compute the recommendation of item for particular user in three steps. In the first step, it searches n users in database which are similar to active user. In second step, it calculates union of the items purchased by these users and link a weight with every item based on its significance in the set. In the third step, from the union it chooses and recommends the N items which have the highest weight and which have not already been bought by the active user.

3.2.2 The item-based algorithm

This algorithm is same as user based algorithm except it determines product similarity instead of consumer similarity. It calculates a product similarity matrix WP = (wp_{st}) which is based on the column vectors of A. A high wp_{st} shows that products s and t are similar as many consumers have brought both of them. A WP will give the products' probable scores for each consumer. Resulting matrix will be containing the element at the c^{th} row and p^{th} column combines the scores of the similarities between product *p* and other products that consumer *c* has purchased. This algorithm provides higher efficiency and comparable or better recommendation quality than the user-based algorithm for many data sets [9].
The primary motivation behind item based algorithm is the truth that the customer is more likely to buy items which are related (similar) to the items he has already bought in past. Means by analyzing the historical purchasing information, we can directly find the similar items.

3.2.3 The dimensionality-reduction algorithm

This algorithm compresses original interaction matrix & produce recommendations which are based on compressed, less-sparse matrix to simplify the sparsity problem. It applies standard singular-vector decomposition (SVD is a matrix factorization technique that factors an $m \times n$ matrix R into three matrices) to decompose the interaction matrix A into U· Z · V' where U and V are two orthogonal matrices of size M x R and N x R respectively, and R is the rank of matrix A. Z is diagonal matrix of size R x R which has all singular values at its diagonal values. SVD can be used in recommender systems & has two features. It can be used to capture hidden association between customers & products which indicates prediction of likeliness of specific product by customer. SVD can be used to construct a low-dimensional image of customer-product space & calculates region in reduced space[10].
The dimensionality-reduction algorithm requires the longest runtime because after reduction computing consumer similarity needs considerable CPU cycles.

3.2.4 The generative-model algorithm

This algorithm makes use of hidden class variables to clarify the patterns interaction between consumers & products. This algorithm approximates appropriate possibility & conditional probability. Based on estimated probability it creates score of product p for consumer c. This algorithm groups together consumers & product in the order of the latent (hidden) classes so simplifies the problem of data sparsity [11]. The approach used in this is generalization of a statistical technique called as probabilistic Latent Semantic Analysis which was initially extended in the context of information retrieval. Actually the probabilistic latent semantic models are closely related to dimension reduction methods & matrix decomposition techniques such as singular value decomposition. The method accomplishes competitive recommendation and calculation accuracies, is highly scalable, and extremely flexible.

3.2.5 The spreading-activation algorithm

This algorithm focuses on sparsity problem by discovering transitive associations between consumers & products by using bipartite consumer-product graph. The algorithm traverses through the graph to find out transitive connections. It describes transitive properties between users with help of social network to access extra details for recommendation

reasons. The algorithm uses a method based on trust interfaces which is transitive association between users that are member of social network [12]. The main objective here is to develop an effective technique that offers high quality recommendations when sufficient data is not available.
The spreading activation algorithm consists of a series of nodes both user and item nodes. These nodes are then connected by edges where each edge has a weighting representing the ratings the item has received from the users. The higher the weight of the edge the higher the rating that item has received. The item nodes then send back "pulses" to the active user and their neighbors thus spreading the activation to the other nodes in the neighborhood of the active user.

3.2.6 The link-analysis algorithm

This algorithm is used in web page ranking & social network analysis. In this consumer-product graph, the global graph structure is used to help collaborative filtering under sparse data [13]. In this graph first set of nodes consists of products, services & information items for probable utilization. The second set consists of consumers or users. The feedback and transaction are represented as links connecting nodes between these two sets. This graph is referred as consumer-product graph. The link analysis algorithms such as HITS (Hypertext Induced Topic Selection) & PageRank are used for identifying essential web pages in a Web graph. The idea behind link analysis recommendation algorithm is to able to extract helpful link structure details from the consumer-product graph & make more effective recommendation with sparse data.
The World Wide Web is a graph that contains linked hypertext documents i.e a link p->q means that webpage p gives recommendation to surfers who are visiting p that they should visit q. Kleinberg [14] distinguished between two types of Web pages which represents a certain topic. The first are authoritative pages that include vital content information of the topic. The second types are hub pages. Hubs are mainly resource lists, connecting many authorities on the topic without directly containing the authoritative information. In this model hubs and authorities share a mutual supporting relationship which forms communities where the hubs are closely connected to the authorities.
PageRank [15] is another link analysis approach to webpage ranking, which is used by the Google Search engine. This algorithm decides ranking of pages by allocating each page p a rank based on its significance called PageRank. In particular, PageRank is the probability of visiting p in a random walk of the entire web.

3.2.7 Evaluation of algorithms

These algorithms show varied performance under different training sets. The dimensionality-reduction algorithm requires the highest runtime. The item-based algorithm works slowly for the large number of products. The spreading-activation and link-analysis algorithms require less number of iterations to achieve acceptable recommendation quality. The generative model algorithm is very efficient as it needs a small number of hidden classes for quality recommendations. The spreading-activation algorithm is especially fast. The link-analysis algorithm usually performs the best.

Table 1. Comparative Analysis of Collaborative Filtering Recommendation Algorithm

Collaborative Filtering Recommendation Algorithm	Advantages	Disadvantages
User-based algorithm	Simple to implement	Suffer serious scalability problems Due to sparsity accuracy is low.
Item-based algorithm	Better performance & quality than user based algorithm. Less computation. Provides higher efficiency. Faster than user based algorithm	Slow for large number of items.
Dimensionality-reduction algorithm	Simplifies the sparsity problem.	Requires the highest runtime.
Generative-model algorithm	Scalable Flexible	Expensive due model building One can lose useful information due to reduction models.
Spreading-activation algorithm	Relaxes the sparsity & cold start problem Fast as it computes recommendation only for target consumers	Works only when sufficient data is not available
Link-analysis algorithm	Better performance than user based & Item based algorithm. Useful when sparse data is available.	Works only when sparse data is available

3.3 Trust-Based Methods

People generally like recommendations from their friends who they know & trust. Trust is bet about future dependent actions of others. In Trust Based Recommendation systems, trust network is used in which users are joined by trust scores which indicate how much faith they have in each other. The information from this trust network is used in Trust-Based Methods. In [16] the choice between recommendations from friends & recommender systems is given. If the quality and usefulness is taken into account, friends' recommendations are preferred even if the recommendations given by the recommender systems have high originality factor. Friends are treated as more experienced to create good and valuable recommendations if compared with recommender systems.

Trust-aware recommender systems take input a matrix consists of ratings about objects from users where the users are represented as rows & the objects are represented as columns & the value in the cell represents the rating given by user to particular object [17]. Along with this, one more matrix is considered as input to system in which user can state their trust on other users which forms a matrix of trust ratings about users.

The user's trust network is constructed for generating predictions [18]. It has three steps. The first step is direct trust. The direct has two methods: explicitly or implicitly. In explicit method, the user himself or herself decides how much they trust others. In implicit method, the system decides the level of trust from particular observed user features. The second step is propagation of trust. It is possible to propagate the trust i.e. create new relations among users. The third step is predicting ratings. From the trust network, we can predict what ratings the particular user would give for items.

Table 2. Comparative Analysis of Recommendation Strategies

Recommendation Strategy	Advantages	Disadvantages/Limitations
Content Based Recommendation	User independence Transparency	Limited content analysis Over-specialization
Collaborative Filtering	Easy to create and use Explain-ability of the results New data can be added easily More applicable in the reality	It totally depends on human ratings Sparsity(Insufficient data) Scalability Ignore the social relationships. Cold start problem (Low Accuracy)
Trust Based Recommendations	Avoid cold start problem. Alleviate the sparsity problem	Difficult to develop a trust network

4. CONCLUSION

The huge volume of information flowing on the web has given rise to the need for information filtering techniques. Recommendation systems are effectively used to filter out excess information and to provide personalized services to users by employing sophisticated, well thought-out prediction algorithms. The Content-Based (CB) systems require explicit domain knowledge and heavy knowledge engineering. The CF only needs the ratings made by users to the items. Trust methods solve cold start problem, data sparsity problem. Amongst the six algorithms of collaborative filtering discussed here, link-analysis algorithm works better in terms of precision, recall & F-measure but it works efficiently only when sparse data is available. This limitation can open a new era of research. More research can be carried out on how link-analysis will work better when the data available is not sparse data. We can combine two or more collaborative filtering algorithms to overcome sparsity problem.

REFERENCES

[1] Patricia Victor and Chris Cornelis , Martine De Cock and Ankur M. Teredesai Trust- and Distrust-Based Recommendations for Controversial Reviews 1541 1672/11/$26.00 © 2011 IEEE INTELLIGENT SYSTEMS

[2] V. Diaz-Aviles, "Semantic peer-to-peer recommender systems," M.S.thesis, Comput.-Based New Media Group, Inst. Comput. Sci., Albert Ludwigs Univ. Freiburg, Freiburg, Germany, 2005.

[3] I. Cantador and P. Castells, "Multilayered semantic social network modeling by ontology-based user profiles clustering: Application to collaborative filtering," in Proc. Manag. Knowl. World Netw., 2006, pp. 334–349.

[4] S. Kim and J. Kwon, "Effective context-aware recommendation on the semantic Web," Int. J. Comput. Sci. Netw. Security, vol. 7, no. 8, pp. 154–159, Aug. 2007.

[5] S. Golder and B. Huberman,"The structure of collaborative tagging systems" in Proc. CoRR, 2005, pp.1–8.

[6] Z. Xu, Y. Fu, J. Mao, and D. Su, "Towards the semantic Web: Collaborative tag suggestions," in Proc. CollaborativeWeb Tagging Workshop WWW, Edinburgh, U.K., 2006

[7] Zan Huang, Daniel Zeng and Hsinchun Chen A Comparison of Collaborative-Filtering Recommendation Algorithms for E-commerce 1541-1672/07/$25.00 © 2007 IEEE

[8] J.S. Breese, D. Heckerman, and C. Kadie, "Empirical Analysis of Predictive Algorithms for Collaborative Filtering," Proc.14th Conf. Uncertainty in Artificial Intelligence (UAI 98), Morgan Kaufmann, 1998, pp. 43–52.

[9] M. Deshpande and G. Karypis, "Item-Based Top-N Recommendation Algorithms," ACM Trans. Information Systems, vol.22, no.1, 2004, pp. 143–177.

[10] B. Sarwar et al., "Application of Dimensionality Reduction in Recommender Systems: A Case Study," Proc. WebKDD Workshop at the ACM SIGKDD, 2000; http://glaros.dtc.umn.edu/ gkhome/node/122

[11] T. Hofmann, "Latent Semantic Models for Collaborative Filtering," ACM Trans. Information Systems, vol. 22, no. 1,2004,pp.89–115.

[12] Z. Huang, H. Chen, and D. Zeng, "Applying Associative Retrieval Techniques to Alleviate the Sparsity Problem in Collaborative Filtering," ACM Trans. Information Systems, vol. 22, no. 1, 2004, pp. 116–142

[13] Z. Huang, D. Zeng, and H. Chen, "A Link Analysis Approach to Recommendation under Sparse Data," Proc. 2004 Americas Conf. Information Systems, 2004

[14] Kleinberg, J. Authoritative sources in a hyperlinked environment. in Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (1998).

[15] Brin, S., and Page, L. The anatomy of a large-scale hyper textual web search engine. in Proceedings of the 7th International World Wide Web Conference (Brisbane, Australia, 1998).

[16] Rashmi Sinha and Kirsten Swearingen. Comparing Recommendation made by Online Systems and Friends. In the Proc. of the DELOS-NSF Workshop on Personalization and Recommender Systems in Digital Libraries, Ireland.

[17] Zhili Wu, Xueli Yu and Jingyu "An Improved Trust Metric for Trust-aware Recommender Systems" 2009 IEEE.

[18] Leonardo Zanette, Claudia L.R. Motta, Flávia Maria Santoro, Marcos Elia "A Trust-based Recommender System for Collaborative Networks" 2009 IEEE.