Saša Bošnjak
Mirjana Marić
Zita Bošnjak

Choosing a Collaborative Filtering Algorithm for e-Commerce

Article Info:

Management Information Systems,
Vol. 3 (2008), No. 1,
pp. 011-015

Received 12 Januar 2008
Accepted 24 April 2008

UDC 004.421.2:339

Summary

This paper aims to give an overview of recommender systems as one of key factors of e-commerce development. Also it describes different types of recommender systems, and methods they employ in order to produce personalized recommendations. Furthermore, the role of these systems in the Internet sphere is investigated. The paper covers some problems in implementations as well as some still open issues in recommender system field.

Key words

recommender systems, data mining, web personalization

Introduction

The biggest e-commerce organizations offer millions of products belonging to thousands of different groups, aiming at an equal share of profits by selling the products that are available only online. Internet has proven itself an excellent and by far the cheapest distribution channel for businesses that base their profits on long tail distribution of popularity and availability. One of the biggest challenges in this environment is choosing/discovering a specific product. That is why, according to Chris Anderson (Anderson, 2006), one of the three key elements – the driving forces of new Internet (social) revolution - is creating connections between supply and demand. This element has a goal of easing the costumer's discovery of those items (products), potentially interesting for him, which are not among the most popular ones. Practically, it aims to cut the search costs. Internet as a platform enables us to cut these costs by utilizing the *wisdom of the crowds* effect (Surowecki, 2005). Google Page Rank, Amazon.com personalized picks, Netflix recommender, are the biggest applications that utilize this effect through the automatic tracking of user behavior. It is, thanks to the development of IT technology, fairly easy to precisely measure and find patterns of user behavior, creating user tastes and preferences in real time. Users have now become the tastemakers themselves. Internet is out of the information age and has entered the recommending age. Thanks to the Web 2.0, the Tim Berners-Lee idea of worldwide information sharing (Berners-Lee, 1999) has been brought to a large number of users. In Long tail economy information is abounded and different mechanisms are allowing us to gather and use them for decision making by providing us with some recommendations. This, often called *navigation* layer of Long tail, provide us with enough information to surf (buy, read, listen) along the whole length of incredible long supply side on the Internet. One of the most important mechanisms that enable this is called a recommender system. The most frequently used recommender systems are so called collaborative filtering systems.

This paper aims to explain the role of recommender systems on the internet, as well as their major types. It also examines the way recommender systems find personalized recommendations.

1. Long Tail and Personalization

In his renewed book titled *The Long Tail* (Anderson, 2006) Chris Anderson systematized, based on some previous work done by the MIT group, the relationship between product sales and popularity. The obtained graph, named the same as the book, explains the way e-commerce business make equal profits from products that can be bought in brick and mortar shops as well as from the products that cannot be bought in offline shops. Due to the limited space in offline shops each product has a shelf rent. In order to cut these costs the major offline stores can only offer the products that are most popular. Being among the most popular products, a certain product is more likely to be sold. In this way the risk associated with the shelf rent is minimized. This means that most of the available products (items) cannot be found in major supermarket chains (theaters, bookstores, newspapers, television…).

Nevertheless, according to the research of Rapshody.com, more than 98% of their items are accessed by at least one user. The Long tail distribution shows these kinds of relations. This proportion has questioned the classical CRM Pareto optimum rule, showing that it does not apply to internet tailing. The Pareto optimum rule, taken from the economical wealth distribution, also known as the 80/20 rule, states that 80% of the profits come from 20 % of the costumers and thus most efforts in CRM should be averted to these costumers. Anderson showed that equal profits can be made by selling less of more products. He divided sellers into 3 groups: physical, hybrid and digital retailers. The first ones offer only products that can be found in classical distribution channels. The second group consists of hybrid retailers that sell products that are available online utilizing the digital catalogs to broaden their supply and save on huge warehouses. The third group is made of sellers that sell pure digital products (like iTunes), that make extra savings by eliminating the shipment. As an extra, an entry to a database and few megabytes for storing are almost free, and there are no economic reasons for these sellers to have each single item in their supply. This new approach to a large number of market niches is enabled by (Anderson, 2006):

1. democratization of tools of production
2. cutting the costs of consumption by democratizing distribution,
3. connecting the supply and demand.

Due to the major drop in computer costs and the development of new areas for PC usage, costs of production have been dramatically cut. Today individuals can offer different digital products or services, and even factories can be built for a reasonable price. This is most obvious in the media world, because everyone with a PC can become a producer, director, musician, writer, etc. Through blogs every Internet user can very easily share his views with a world, thus influencing other people. By the democratization of tools of production by mass PC usage, unbounded supply side is created driving Long tail along the horizontal axis. If PCs enabled everyone to become a producer, than the Internet has made everyone a potential distributor by cutting down the distribution costs.

In the situation where everybody is a producer and a distributor, and there is a large number of customers, there must be a way for them to cope with this new mass of products (services). That role is given to the third driving force of the Long tail economy, connecting supply and demand. This connection is done using different types of filters. Filters include blogs, playlists, recommendations, etc., in one word, everything that is considered being an opinion making factor. The opinion of the Internet users is not formed based on the most frequent influences, such as media or politicians, but through the number of interactions with other users that share their views. This way of interaction Anderson calls postfilters. The difference between them and the classical influences (prefilters) is in the fact that postfilters do not apriori make a decision but try to find the best decision among the alternatives which are already on the market. This means that every option should be marketed and left open to the market evaluation. That is exactly what postfilters do by channeling and not predicting user behavior making it the most important factor in filtering the options. The amplifying instead of predicting is the main difference between the post- and pre- filters. The limited shelf space makes the offline markets extremely discriminatory. The players on the demand side therefore use different methods for forecasting when deciding which products are to be offered. Research (Atlee & Por, 2006) has shown that the percentage of the expert's accuracy in forecasting future events is very low (the most correct is weather forecasting with 30% accuracy). Because of this, a large number of potentially successful products are discarded before entering the market. On the other side, in the Long tail markets where self rent is low, all products can at some point be available. That changes the role of filters, as Anderson put it, from market gatekeeper to market advisor. Instead of using demographic or psychological date collaborative filtering approaches each user and uses his preference date for filtering.

These kinds of filters are based on the principles of wisdom of the crowds. This principle states that the more diverse the decision group is the better the result of the decision process will be. The term is coined by James Surowecki in his book (Surowecki, 2005). He identified four major factors in creating this phenomenon: diversity of opinion, independence, decentralization, and aggregation.

Making decision in this manner is in the long tail economy depicted by the new kind of systems called recommender systems. The role of these systems is to find within the countless items the one which is most suited for the given user knowing his preferences, as well as preferences of other users.

2. Recomender Systems Types

Wikipeida.com defines recommender systems as *programs which attempt to predict items (music, movies, books, news, web pages…) that a user may be interested in, given some information about the user`s profile.*

Content based recommenders make recommendations by analyzing content of the objects ranked by the user and objects that are to be recommended. Many algorithms are used to analyze mainly textual documents and find some patterns in the content of these documents. These patterns are used when making a recommendation. The content based recommenders encounter two major problems. The first one is the representation of objects, and a second one is the creation of user profiles. All algorithms use keywords to represent object. In this way, an object is represented by some keywords (broad classes) that describe it, and recommendation is performed by discovering objects with keywords similar to objects previously preferred by the user. This way of recommendation has shown the best results in recommending text documents such as web pages or news.

An advantage of content based recommenders is their high relevance, but they are computationally expensive when used for music or video content search. Apart from high relevance, the quality of recommended item is difficult to measure in content based recommendations.

From all different approaches for obtaining recommendations, the most popular one is by far the collaborative filtering (CF). This approach makes recommendations by discovering correlation between users of the recommender system. It is a unified approach for discovering the list of potentially interesting items (items not been accessed by the active user) and predicting its relevancy to the active user. The idea is to reach the final recommendation only on the basis of similar users and their actions. CF systems distinguish two kinds of decisions: the prediction and the recommendation. Prediction is the most probable mark user would give to an item had he accessed it previously, and the recommendation is a set of n items that would be liked by the user. In order to discover these decisions multiple approaches are used. The most frequently used approaches are user based CF, item based CF, dimensionality-reduction algorithms and link analysis (Wang, deVries, & Reinders, 2005).

The first group of models uses information about users and items like explicit user ratings or transaction data (access, request logs, page view time, etc.) It is clear that user based approach does not use only item or user date but the date about their interactions. User item system discovers the users with the highest correlation with the active user (nearest neighbor) by comparing ratings for the same items. The recommendation is provided as the highest rated item of the similar users. Advantage of this way of recommending is the elimination of item content analysis which makes it suitable for any kind of items. Also, user based CF provides both the relevant and quality recommendations. The quality is assured through the collaboration of users, and self-improvement of these kinds of systems which makes them one of the corner stones of Web 2.0.

The major problem which accrues when using a CF system is the so called *cold start* and resembles the absence of ratings given by a new system user. If the system is to make recommendation for a new user and he has not previously rated any items, user based CF cannot be done. This is why these systems usually demand several user ratings before making any recommendations. The same problem applies to items as well. To overcome this, different kinds of CF systems, called *item based CF,* had been developed. These systems do not search for correlations between users, but try to find correlations between items. Recommendation is made by trying to find items with high correlation to items rated by the user. Advantages of item based CF are similar to user based, high relevancy and quality. Also, it is more scalable because of the slower change in number of items than the number of users. The vector of item ratings is much less sparse than a vector of user ratings. These features make item based CF one of the most popular recommender systems and the Amazon.com personalized picks is developed using this method.

In general, CF systems encounter some constraints (Sarwar, Karypis, Konstan, & Riedl, 2001):

- scalability: nearest neighbor algorithms demand computations that are more time consuming as the number of users and items rises. It is imperative that the algorithms used can handle several million users and items.
- scarcity: in real life recommender systems use a large data set. In these systems users have ranked less than 1% of items (amazon.com had 2 million books and 1% of that is 20000). Because of this, CF systems cannot make a good recommendation all the time.

To overcome the above mentioned problems, *dimension reduction algorithms* for CF are used. These algorithms try to solve the scarcity problem by reducing the user-item rating matrix. To do this, they employ different algorithms that use singular value decomposition of latent classes.

Link analysis algorithms, such as Google`s Page Rank, are also used for making recommendations. Research on three different datasets showed (Wang, deVries, & Reinders, 2005) that link analysis algorithms slightly outperform other algorithms. However, the research does not single out the overall best algorithm, but it does not treat all datasets as a part of a system. The next section examines some of the factors for choosing the right algorithm.

3. Choice Factors

In order to choose the appropriate recommender system for a specific e-commerce application, several issues must be considered. The best algorithm is not always the one that gives the most correct overall recommendations. In a dataset containing data that belongs to many different groups (classes), such as product groups or different genres, the precision of recommendations can be different for different types of items. This means that some users will get very precise recommendation, while others, that prefer some other type of items which are hard to predict, will get a significantly lower precision. As major concerns in making a decision about the choice of the algorithm for collaborative filtering we identify four different factors: rating schemes, number of users, computational time and explaining ability.

As in all data analysis problems, data is the most important issue. The correct interpretation of input data in great deal influences the success of the analysis process itself. Since the data used as input to the recommender system is basically user-item rating matrix, the way it is constructed is a critical issue. Rating schemes that are used most often are the 1-to-5 scale, or the binary system. Each of these schemes has its flaws. The most obvious one is the rating semantics of different systems. Most of them use star representation and it is unclear what does, for example 3.5 stars mean. Some systems use ratings only to show the positive ratings (1-5), while some others use it to show negative ratings as well – in that case, 1 star denotes a not interesting content. The problem then arises for different algorithms are tested using the data that is formed on different rating schemes. Also the rating schemes in general have only one rating that is supposed to show the average user opinion about the item. If we take the example of apparel, does the rating mean quality, user taste, price or all together aggregated into a single rating? And if it is aggregated, was there some weight assigned to different aspects of the items? Another problem with scale rating schemes is the rating variance. It means that, if the group of users has an average rating for item A 3 and item B also 3, depending on the variance of the ratings the prediction of the item to the active user should not be the same (Martin, 2006).

One of the recommender systems major advantages is at the same time one of its drawbacks. The self-improvements can lead to the creation of feedback loop when the most popular item is highly related with almost every other item (MyStrands call this the Coldplay effect).

To overcome some of above mentioned problems, some systems use binary (love/hate) ratings. Some issues are resolved using this scheme, but there are still some unsolved issues. The major one is the interpretation of a rating. What do 10 negative marks mean? Some systems do not use rating schemes at all, but use information from user tracking, such as playlists, blogroles, links or visits. The major issue about choosing rating schemes is still undergoing serious research. The Grouplens research showed that users wanted to rate as many features as they can, but research of Burson (Burson, 2007) actually show that despite the user preference of more control over personalization, they make poor job evaluating their skills. As a result of this, if we allow a user to rate different kinds of item features, they can end up with items that do not suit them.

The research on rating scheme influence is still rather new and it goes into two directions: trying to explain the influence of a single rating on the overall recommendation for the group, and trying to assess ratings with respect to user moods.

The number of users must be carefully examined when implementing the collaborative filtering system. The largest systems such as Netflix or Yahoo! radio have million of new users and millions of new ratings each day. That makes it extremely hard to employ some of the algorithms and the companies combine collaborative filtering with some other kind of systems, mostly expert (Yahoo) or content-based. This has led to development of meta recommender systems (Schafer, Konstan, & Riedl, 2004) which use different recommendation systems to get an overall prediction.

Another interesting issue is the reuse of recommendations through web services, which can ease an initial recommendation system implementation. A large number of users demand a scalable system, and that there is some tradeoff between accuracy and computational time. The research in (Wang, deVries, & Reinders, 2005) shows the computational time comparison of the most popular algorithms.

Making recommendations should not be a black box, but should rather be able to explain the reason for recommending some item. Depending on the type of an algorithm, different levels of explanations can be derived. The more data is used for making the recommendation, the explanation is more easily made. The tradeoff here is between the accuracy and explanation power of a model.

Conclusion

Most of the recommender system evaluation research has been focused on finding the algorithm that has the best precision in terms of the statistical measures like RMSE or measures used in classification such as precision, recall, or AOC. These measures can express the mathematical side of the recommendation, but have a problem of different inputs. To find the best suited collaborative filtering, research must shift to other issues related to a complete recommender system, and not just its algorithm part. Therefore, future research should go in a holistic direction, trying to assess the complete system with all its aggregated parts.

References

Anderson, C. (2006). Long Tail: How Endless Choice Is Creating Unlimited Demand. London: Random House Business Books.

Atlee, T., & Por, G. (2006). Intelligence as a Field of Multi-Disciplinary Study and Practice. Retrieved October 18, 2007, from Blog of Collective Intelligence: http://www.community-intelligence.com/blogs/public/2007/01/a_source_document_for_collecti.html

Berners-Lee, T. (1999). Weaving the Web. New York: Orion Business Books.

Berners-Lee, T., Hall, W., Hendler, J., Shadbolt, N., & Weitzner, D. (2006). Creating a Science of the Web. Science Vol. 313. no. 5788 , 769-761.

Burson, K. (2007). Interpersonal Miscalibration and its Impact on Product Choice. Journal of Consumer Research , 34 , 104-110.

Golden, S., & Huberman, B. (2006). Usage Patterns of Collaborative Tagging Systems. Journal of Information Sciences , 32(2) , 198-208.

Huang, Z., Zeng, D., & Chen, H. (2007). A Comparison of Collaborative-Filtering Recommendation Algorithms for E-Commerce. IEEE Intelligent Systems , 22 (5) , 68-78.

Martin, J. (2006, October). Ratings Variance in Recommender Systems. Retrieved March 5, 2007, from ACM Recommender Systems Conference: http://blog.recommender06.com/wp-content/uploads/2006/09/joliemartin.pdf

Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001). Item-Based Collaborative Filtrering Recommendation Algorithms. Proceedings of the 10th International WWW Conference. Hong Kong.

Schafer, J., Konstan, J., & Riedl, J. (2004). The View through MetaLens: Usage Patterns for a Meta-Recommendation System. IEE Proceedings Software, 151 (6), (pp. 267-279).

Surowecki, J. (2005). The Wisdom of the Crowd. New York: Anchor.

Wang, J., deVries, A., & Reinders, M. (2005, December). A User-Item Relevance Model for Log-Based Collaborative Filtering. Retrieved April 11, 2008, from Information and Communication Theory Group: http://ict.ewi.tudelft.nl/pub/jun/ecir06.pdf

Bošnjak Saša

University of Novi Sad
Faculty of Economics Subotica
Segedinski put 9-11
24 000 Subotica
Serbia

Email: bsale@ef.uns.ac.rs

Marić Mirjana

University of Novi Sad
Faculty of Economics Subotica
Segedinski put 9-11
24 000 Subotica
Serbia

Email: mprokic@ef.uns.ac.rs

Bošnjak Zita

University of Novi Sad
Faculty of Economics Subotica
Segedinski put 9-11
24 000 Subotica
Serbia

Email: bzita@ef.uns.ac.rs