Choosing a Collaborative Filtering Algorithm for E-Commerce

OPEN ACCESS

Volume : 6

Special Issue : 1

Month : September

Year: 2018

ISSN: 2321-788X

Impact Factor: 3.025

Citation:
Muthulakshi, A. (2018). Choosing a Collaborative Filtering Algorithm for E-Commerce. *Shanlax International Journal of Arts, Science and Humanities*, 6(S1), pp.55–57.

DOI:
https://doi.org/10.5281/zenodo.1410969

A.Muthulakshi, M.Phil.,
Research Scholar, Morappur Kongu College of Arts and Science
Morappur

Abstract

This paper aims to give an overview of recommender systems as one of the key factors e-commerce development. Also, of e-commerce it describes different types of recommender systems, and the methods they employ 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.

Introduction

Collaborative filtering, according to the website Techopedia, is a 'technique commonly used to build personalized recommendations on the Web.' The process matches the purchase patterns of one shopper to similar shoppers, or it might make recommendations based on your prior online purchases or browsing history, or the purchases of individuals who have shopped for similar items. Collaborative filtering is an example of a new wave of customer segmentation - that is, using buyer behaviors to divide consumers into groups that share same purchasing or browsing habits.

The large 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. The 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 large challenges in this environment is choosing/discovering a specific product. That is why, according to Chris Anderson (Anderson, 2006), one of the three crucial 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 particular these costs by utilizing the wisdom of the crowds effect (Surowecki, 2005).

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 purchase in the offline shops, Due to the limited space in offline shops each product has a shelf rent. To cut these costs, the major offline stores can only offer the products that are most popular. Being among the most popular products, a determined product is more likely to be sold. In this way the risk associated with the shelf rent is minimized. This means that the 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, show that it does not apply to internet tailing. The Pareto optimum rule, taken from the economic 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 advantage could be made by the selling less of more products. He divided sellers into three groups: physical, hybrid and digital retailers. The first ones offer only products that can be found in the 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 the 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 every 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

Recommender Systems Type

Content-based recommenders make recommendations by analyzing a content of the Objects ranked by the user and objects that are to be the 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 primary problems. The first one is the representation of the 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 the recommendation is performed by the discovering objects with keywords similar to objects previously preferred by the user. This way of endorsement has shown the best results in recommending text documents such as webpages or news.

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 Object

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.

Choice Factors

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 correct overall recommendations. In a dataset containing data that belong 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 a very precise recommendation, while others, that prefer some other type of items which are hard to predict will get a significantly lower precision. As primary concerns in deciding on 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 primary issue. The correct interpretation of input data in a great deal influences the success of the analysis process itself. Since the data used as input to the recommender by system basically user-item rating matrix, the way it is constructed is a critical issue. Rating schemes that are used by most often are the 1-to -5 scales, 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, stars mean. Some systems use ratings only to show the practical grade (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 the 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.

Conclusion

Most of the recommender system evaluation research has been focused on the finding the algorithm that has the best precision regarding 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 absolute system with all its aggregated part

References

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