SEEJSD

SOUTH EAST EUROPEAN JOURNAL OF SUSTAINABLE DEVELOPMENT

Vol. 7 (2/2023)

Skopje, North Macedonia

SEEJSD

SOUTH EAST EUROPEAN JOURNAL OF SUSTAINABLE DEVELOPMENT

Vol.7 (2/2023)

Publisher: Mother Teresa University in Skopje, Republic of North Macedonia

South East European Journal of Sustainable Development

Managing Editors:
Bekim Fetaji, PhD
Olga Popovska, PhD

ISSN (print) 2545-4463
ISSN (online) 2545-4471
is published twice a year.
Account No. 160016267778815
723019 - 45
Tax No. 4080016561272
Bank: Narodna Banka RM

Technical Editing/Layout: Korab Ballanca
Editorial Office: South East European Journal of Sustainable Development
Mother Teresa University in Skopje, Republic of North Macedonia
Mirce Acev 4, VII floor, Skopje, North Macedonia
Phone: +389 2 3161 004
E-mail: seejsd@unt.edu.mk
Web: www.seejsd.unt.edu.mk

The publication of the Journal is supported by:

Ministry of Culture of Republic of North Macedonia

How Recommendation Algorithms Know What You'll Like

Mirjana Kocaleva Vitanova
Faculty of Computer science, “Goce Delcev” University, Stip, R. N. Macedonia, mirjana.kocaleva@ugd.edu.mk

Marija Miteva
Faculty of Computer science, “Goce Delcev” University, Stip, R. N. Macedonia, marija.miteva@ugd.edu.mk

Elena Karamazova Gelova
Faculty of Computer science, “Goce Delcev” University, Stip, R. N. Macedonia, elena.gelova@ugd.edu.mk

Biljana Zlatanovska
Faculty of Computer science, “Goce Delcev” University, Stip, R. N. Macedonia, biljana.zlatanovska@ugd.edu.mk

ABSTRACT

One of the most used statistical techniques that include machine learning and data miningfor predicting future outcomes with help of data that already exist is known as predictive algorithm. Predictive models are not stable, and they build assumption based on past and present actions. In the paper we are going to introduce Amazon online store and how algorithms know what we like, so they can recommend products to us by their own. One of the biggest innovations in online shopping - first introduced by Amazon - was automatic recommendation generation. The more accurate prediction algorithms are, the more online stores will sell their products. For that reason, prediction algorithms are of great significance for online stores.

CCS CONCEPTS

• Education• Algorithms

KEYWORDS

recommendation, algorithms, Amazon.

1 Introduction

One of the biggest innovations in online shopping - first introduced by Amazon - is automatic recommendation generation. Log in to the site and, right there on the home page, the site will give you suggestions for products you can buy. For example, if you are a JavaScript programmer, you will see recommendations for programming books that use that language, and if you are a mother of young children, you will see how the site mentions toys and children's books.

This homepage personalization is of great benefit to online stores compared to displaying only the top 10 listings or banner ads: page-through traffic and conversion rates are far higher. Customers are more likely to see and buy the products offered.

Prediction algorithms are of great importance to online stores - the more accurate they are, the more online stores will sell. In paper [5] are compared three well known approaches for solving the recommendation problem such as traditional collaborative filtering, cluster models, and search-based methods with their algorithm called item-to-item collaborative filtering. Online calculations based on their algorithm are measured independently of the number of clients and the number of items in the product catalog and produces recommendations in real time and with high quality. Amazon is recognized because his system for personalization and recommendations helps customers to found product that they might not find another way. In this update to their original paper, the authors discuss some of the changes as Amazon has grown [6].

In recent years, Sentimental Analysis was used in all online product firms. Also, many users who are using websites, blogs, online shopping tends to review the products they used. Sentimental Analysis is defined as a concept of data analysis where the collections of reviews are taken into consideration, and those reviews are analyzed, processed, and recommended to the user. In this paper [7], the dataset was collected from the official product sites. First all the reviews had to be pre-processed. After pre-processing is completed, the trained dataset is classified using Naive Bayes and SVM algorithm. These existing algorithms provided the bad accuracy. An ensemble approach will be applied to improve the accuracy of the given scans. An ensemble is an approach to classification by combining two or more algorithms and calculating mode values based on voice references for each algorithm used. In this paper, Naive Bayes, SVM, and Ensemble algorithm were combined. Authors proposed an Ensemble method that helps in providing better accuracy than the current existing algorithm. Once the accuracy is calculated, based on the reviews, the product was recommended for the user.

The study [8] examines how vaccine-related books appear on Amazon. The authors collected vaccine-related books that appeared in the top 10 pages of Amazon search results for seven consecutive days. They also collected Amazon's recommendations for each vaccine book and mapped the recommendation network between these books. Using a network model, they found that books sharing similar views of vaccines were recommended together such that when a user views a vaccine-hesitant book, many other vaccine-hesitant books are further recommended for the user.

Problems that must be solved with such a recommendation algorithm are considered. A big online store like Amazon can have millions of users and millions of items in stock. New customers will have limited information about their preferences, while existing customers may have too much.

The data with which these algorithms work is constantly updated and modified. Customers search the site and prediction algorithms should consider recent item browsing, for example - it does not help if we are looking for a toy for our youngest granddaughter and all we get are jQuery suggestions.The biggest and most important criterion for these systems (apart from accuracy) is the speed. The recommendation algorithm must generate suggestions within a second or more. After all, the customer is in the process of displaying the homepage of the store where the recommendations will appear.

Traditionally, these referral algorithms work by finding similar clients in a database. In other words, they work by finding a set of customers who have purchased or rated the same items. They throw away the items you have already bought or commented on and recommend the rest.For example, if you have already purchased A and B, and a set of such customers also includes C purchases, then C will be recommended for you.

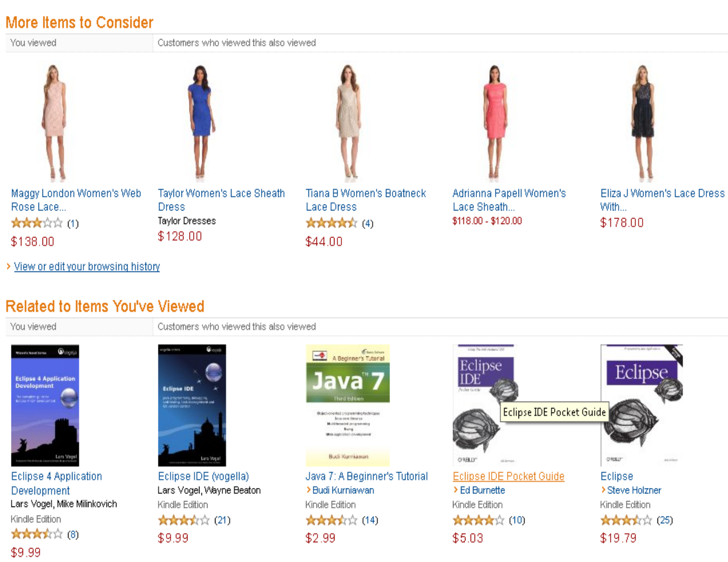

Figure 1 Recommendation process

2 Algorithms

Collaboration

One of the earliest algorithms is known as collaborative filtering. Essentially, the algorithm represents each customer as a vector of all sales elements. Each entry in the vector is positive if the client buys or evaluates an item, negative if the client does not like the item, or empty if the client does not make his or her opinion known.

Most of the records are empty for most of the clients. Some variants of the factors for the popularity of the items may conflict with the importance of the items that are less popular or known. The algorithm then generates its own recommendations by calculating the value of similarity between the current client and everyone else.

The most acceptable way to do this is to calculate the angle between the vectors - the simplest method is to calculate the cosine using a point product divided by the product of the lengths of the vectors. The larger the cosine, the smaller the angle, and therefore more customer-like.

These processes are extremely expensive. There are usually many clients, and many calculations need to be done very quickly. There are techniques for reducing calculations (by taking the customer base or ignoring unpopular items, for example), but in general it is, and always will be, expensive to calculate recommendations this way.

Client clusters

Another traditional algorithm for prediction involves the use of cluster models. Here the goal is to prepare the customer base by dividing them into clusters, and then assigning the current clients to one of the clusters, in theory is choosing the cluster with the most similarity. Since the cluster will be identified, recommendations will come from purchases and ratings from other clients in that cluster.

Although cluster selection works in much the same way as classification algorithm (assuming that we can calculate the characteristic vector that describes the cluster in the same way as the existence of a client vector), the real importance of the algorithm is in creating clusters.

In principle, grouping customer data is done through heuristics: we start with several empty clusters, assign randomly selected clients to each, and then assign other clients to the clusters according to similarity. From the original clusters that are essentially random, sub-algorithms must be used to merge or split the clusters.

The use of cluster models is with less calculations at the point where customer recommendations need to be made quickly. After all, there is less work to be done to find similar clusters than similar clients. Most of the work is done before the clusters themselves are created.

Unfortunately, this method tends to result in low quality referrals from purchases / ratings that are on average within the cluster. A certain customer is no longer the same as most similar customers, but with the average of a large group of customers. Of course, the number of clusters can be increased to adjust the matches, but then there will be opportunities to increase the computation time.

Simple browsing

The next traditional algorithm is a simple search algorithm. For example, if I buy the book Pride and Prejudice by Jane Austen, the search algorithm will search for items in the database for other Jane Austen books, books with same content by other authors, DVDs made by Austen books, and so on. You can see targeted referrals like these in banner ads when you surf the web.

Contents-to-contents

What Amazon did to improve its recommendations was to include collaborative filtering. Instead of trying to find similar customers, he finds the same content. This version is called content-to-content collaborative filtering.

This algorithm matches each of the current buyers who bought an item and ranks the items by similarity, and then builds a list of those same items. A table of similar items must first be built on the website by analyzing the items that customers tend to buy.

Here's how it works: For each item X in the catalogue, we find all the C customers who bought X. For each of these customers, we find all the Y items purchased by C and record that the customer bought both X and Y. Then, for all pairs X and Y, we calculate the similarity between X and Y in the same way as for the collaborative filtering algorithms.

Although this calculation is quite expensive, it can be done in advance. Once the similarity between each pair of items is established, the list of recommendations is easy to obtain.

BELLKOR

Back in 2006, the movie rental company announced a $ 1 million competition to see if anyone could improve on the recommendations made by CineMatch. The goal was to improve the CineMatch score by 10 percent by testing subgroups of the vast Netflix database.

The winner, after almost three years, was a group that called itself BellKor. Competitors who accepted the challenge were given large databases of 100 million ratings, with each rating (between one and five stars) containing the rating date, title, and year of release of the film, and an anonymous user ID. Qualification database were also provided from the selection of 2.8 million ratings with the same information, but without a real rating.

The goal was to develop an algorithm from a large database, apply it to qualifying databases to guess the rating, and then Netflix would check how close the supposed rating was to the actual rating.

It is fascinating to see the strategies that BellKor uses to incorporate its algorithm. It must be emphasized that the BellKor solution is not a single algorithm by itself but can be seen as a series of algorithmic pieces that can be rotated to produce the best response.

Predictions

The first strategy was to create a group of predictors. This describes the average user rating. Suppose the average rating of all movies is 3.5. As an example of a specific movie, Star Wars could have a complete rating of 4, which would be 0.5 better than the average movie.

Our hypothetical user though tends to rate movies below average: we see that his average above all movies is 3.2. Since his average is 0.3 lower than the average, our initial assumption about how he might rate Star Wars would be 4-0.3, or 3.7.

The second strategy used was the realization that time plays a big role in people's ratings. First, the popularity of a movie will change over time. For example, a movie may have a big start and then be forgotten, while another may start with a small start and then become a cult movie. Popularity is also affected by the star or director when they publish a better (or worse) film additionally, and with their appearance, good or bad, in the media.

Time

The overall rating of the user tends to change over time. This could be because the "user" is, in fact, the whole household, also the person making the rating may change, or it could be due to the psychological effect of the user breaking through to a good rating, their next rating may be lower than what would normally be justified (or vice versa: after a series of bad movies, the next good movie may be rated lower than expected).

The following strategy can also be described as part-time: the user can enter a rating for a set of recently viewed movies in one day. The user would like to make ratings unknowingly affect each other (if most movies were good, bad movies would tend to be better than expected ratings, or vice versa), instead of thinking about them all on their own. This strategy is known as frequency.

The development team at BellKor basically describes these strategies mathematically and statistically to provide parameters of the model that can be modified.Taking a large subset of data constantly executes the model, changing the parameters little by little, until they have predicted the ratings of the other smaller subgroups. In this regard, it was possible for them to submit their guesses to the qualifying subgroup.

From all this, we can see that the prediction algorithms are not exactly accurate. Although they provide fast and usually accurate, it does not matter. For Amazon, the prediction engine is a distinguishing factor, and for Netflix it's the main reason for keeping customers in their memberships - after all, once a user has seen Star Wars and his / her collection of extensions, it happens that he / she wants suggestions for others works or will give up searching here.

3 The secret of Amazon recommendation

We wonder what Facebook, Google and Apple know to their users. It is true that Amazon may know more. And massive retailers prove it every day.

"To whom Amazon recommends a product on its website, it will be clear to many that this is not a coincidence."

Basically, the retail giant's recommendation system is based on many simple elements: what the user bought in the past, what items the user has in their virtual carts, the items they rated and liked, and what other clients watched it and bought it. Amazon (AMZN) calls this proprietary mathematical algorithm "content-to-content collaborative filtering," and it's used to greatly customize the browsing experience for returning customers.

The company announced a 29% increase in sales to 12.83 billion dollars during the second fiscal quarter, for a difference of 9.9 billion dollars in the same period last year. Much of that growth probably has to do with the way Amazon has integrated recommendations into virtually every part of the buying process from product discovery to checkout. One can go on Amazon.com and will find multiple panels with product suggestions, then will see areas titled "Most Purchased Together" or other items that customers have also purchased. The company doesn't talk much about how effective recommendations are.

Amazon also assigns recommendations to users via email. Although the website's referral process is more automated, the company provides some employees with numerous software tools to target customers based on shopping and browsing. Actual routing is done by employees, not machines. If an employee oversees promoting a movie for purchase such as Captain America, he should come up with similar movie titles and make sure that customers who have seen other comic book like action movies will get an email encouraging them to watch it Captain America in the future.

Amazon incorporates pretty much all the email marketing standards as any company - but less well known is the fact that the company has a survival-of-the-fittest type of income. This means that if a customer qualifies for both Book Mail and Video Game Mail, the email with the higher average revenue-per-mail sent will win. So, the customer will receive only one email. 41.5 % This tactic prevents email mailboxes from being flooded, at least at Amazon. At the same time, it dramatically increases purchasing opportunities. In fact, the exchange rate and effectiveness of such messages is "very high", significantly more effective than the recommendations on the site. Amazon's conversion rate for referral sales on the site could be 60% higher in some cases if it were based on the performance of other e-commerce sites.

In addition to improving the accuracy of its recommendations, Amazon can also explore more ways to reach consumers. Already, the company has begun selling items previously sold in bulk that were too economical to sell individually, such as a deck of cards or a jar of cinnamon. 62.4% Customers could buy them, but only if they had an order totaling $25 or more. But the company can actively recommend these additional products during checkout when the order will exceed the price threshold, similar to traditional supermarkets that have impulse purchases such as chewing gum and candies.

In that sense, Amazon customers do the same thing they might do in a supermarket, thinking "It's just a few more dollars why not?"

The most important takeaways from personalized recommendations that we see at Amazon include:

1. Increased Sales Revenue

2. Increase Site Traffic
3. Increased user satisfaction
4. Increased customer loyalty
5. Increased engagement of buyers

Recommendation algorithm of Amazon is one of the most complex and efficient in the e-commerce market.

4 Conclusion

E-commerce is a process of selling and buying items via internet. E-stores is available 24/7 every day and there are no country barriers. Everyone can buy something from everywhere in the world with some shipping fees. Recommendation algorithms are best known for their use on web sites used for e-commerce. According to customer input they generate list of recommendation products. To personify the online store for each client recommendation algorithms were used by Amazon. The store is based on customer interest, and changes according customer situation. If Amazon know that we work in education, then the recommendation for us will be books, education tools, notes, pens and so on. Recommendation algorithms are good for e-commerce and their use is indispensable.For almost 20 years the Amazon has been regularly developing its system for recommendation and today is responsible for a large percentage of sales.

REFERENCES

[1] CNNMoney. (2013). Retrieved from http://tech.fortune.cnn.com/2012/07/30/amazon-5/
[2] Industry Report. (2013). Retrieved from http://www.cs.umd.edu/~samir/498/Amazon-Recommendations.pdf
[3] Wordpress. (2013). Retrieved from Software Programming: http://kunuk.wordpress.com/2012/03/04/how-does-the-amazon-recommendation-system-work-analyze-the-algorithm-and-make-a-prototype-that-visualizes-the-algorithm/
[4] Techradar. (2013). Retrieved from http://www.techradar.com/news/internet/how-recommendation-algorithms-know-what-youll-like-1078924
[5] Linden, G., Smith, B., & York, J. (2003). Amazon.com recommendations: item-to-item collaborative filtering. IEEE Internet Computing, 7(1), 76–80. doi:10.1109/mic.2003.1167344
[6] Smith, B., & Linden, G. (2017). Two Decades of Recommender Systems at Amazon.com. IEEE Internet Computing, 21(3), 12–18. doi:10.1109/mic.2017.72
[7] Sadhasivam, J., &Kalivaradhan, R. B. (2019). Sentiment analysis of Amazon products using ensemble machine learning algorithm. International Journal of Mathematical, Engineering and Management Sciences, 4(2), 508.
[8] Shin, J., & Valente, T. (2020). Algorithms and Health Misinformation: A Case Study of Vaccine Books on Amazon. Journal of Health Communication, 1–8. doi:10.1080/10810730.2020.1776423

South East European Journal of Suistanable Development

ISSN (print) 2545-4463

ISSN (online) 2545-4471