

Research on precision marketing of banking based on improved collaborative filtering algorithms

Yuning Bian¹, Yeli Li², Qingtao Zeng³, Yanxiong Sun⁴, Linxuan Yu⁵ and Wei He⁶

¹Beijing Institute Of Graphic Communication, Beijing, 102600, China,
774200483@qq.com

²Beijing Institute Of Graphic Communication, Beijing, 102600, China,
1605872754@qq.com

³Beijing Institute Of Graphic Communication, Beijing, 102600, China,
954276545@qq.com

⁴Beijing Institute Of Graphic Communication, Beijing, 102600, China,
tongxin_yxsun@163.com

⁵Beijing Institute Of Graphic Communication, Beijing, 102600, China,
951101614@qq.com

⁶Beijing Institute Of Graphic Communication, Beijing, 102600, China,
871613503@qq.com

Abstract. With the rapid development of the Internet, the banking industry is facing great challenges. How to attract more customers under the huge impact of the Internet is the most important issue for banks to consider. In the context of big data, banks are using a variety of means to collect data and build their own customer portraits to adjust their marketing strategies. This article provides an improved collaborative filtering algorithm --- time-related similarity calculation method. This improved collaborative filtering algorithm can provide a new reference for the bank's precision marketing.

1. The concept of big data

1.1. Features of the Age of Big Data

After entering the 21st century, the technological revolution has driven the whole society forward again and again." Big data" is also the word slowly entering people's eyes. I believe that we are no stranger to this word, next, let's talk about the characteristics of big data in this day. "Big data" is, as the name suggests, a large amount of data [1]. In fact, the basic characteristics of big data can be summed up in four aspects: large capacity, diversity, low value density, and speed. Next, let's address these four characteristics.

Firstly, the capacity is large. The general amount of data can reach the level of TB to PB leap. Secondly, diversity. There are many types of data. For example, the types of data are generally web logs, audio, video, pictures, geographic location information, and so on. Thirdly, low value density, in a large amount of data is not a lot of data, such as in a very long video data, the generally useful data may be only two or three seconds. Fourthly, big data on the processing speed of data has its own requirements, generally in line with the "1 second law." That is, the processing results are generally



Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](#). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

given in the second time range, the time is too long to lose value. The speed requirement of this point is also the most essential difference between large data processing technology and traditional data mining technology.

1.2. Big data and banking integration is imperative

Big data is one of the most potential technologies in the 21st century, from the current payment methods more and more intelligent, we will be more and more intelligent consumption and financial management model. The development prospects of big data technology in commercial banks are also very broad. As the hub of the whole society's currency circulation, commercial banks can rely on retrieval technology to collect a large amount of customer behavior information [2]. These customer behavior information is an important reference data for our bank product recommendations to help banks improve their business insight and trend forecasting capabilities, increase the profitability of banks, respond quickly to market changes, enhance risk warning capabilities, meet customer purchase needs, enhance core competitiveness.

2. The concept of precision marketing

2.1. The concept of enterprise marketing and precision marketing

Marketing strategy refers to the enterprise in the course of business to the actual needs of customers as the landing point, according to the previous accumulation of purchasing experience to obtain the customer's purchasing power and demand information, market expectations and other data information. Thus there is a planned arrangement to organize various production and business activities. The marketing strategy mainly studies the various situations faced by the marketing of enterprises under the current market conditions. In other words, the marketing strategy is mainly studied in the market environment and marketing environment, how to accurately understand, analyze, select and seize market opportunities to meet the consumer's shopping needs, so that enterprises to maximize wealth, so that its long-term survival and development.

Precision marketing, as the name implies, belongs to a kind of marketing strategy, compared with the traditional marketing strategy, the focus of precision marketing is more clear. The first step of precision marketing is to position precision, using modern information technology to establish personalized customer communication system so that enterprises can achieve capital expansion in a low-cost way. Precision marketing has become the backbone of network marketing, but also has an attitude in the network marketing concept of one of the core views [3].

2.2. Steps to Precision Marketing

Precision marketing has three main meanings:

First, the ultimate goal of precision marketing is marketing without marketing. However, to achieve this goal, it is necessary to go through step-by-step marketing.

Second, precision marketing is a precise system assurance and means, and this means is also measurable.

Third, the purpose of precision marketing is to achieve low-cost sustainable business objectives. Precision marketing is a popular marketing term, in a comprehensive sense is to use the emerging media to accurately give enterprise decision makers the corresponding marketing program [4].

3. Products in the banking industry

3.1. Introduction to banking products

With the deepening of Internet technology, more and more banks have their own mobile phone clients. Customers can learn about different wealth management products through the mobile clients of major banks [5]. Industry experts said that in the future, traditional financial banking and the Internet to strengthen cooperation and complementary advantages will become the main trend, Such as Soufan

and other P2P financial management will be the traditional bank's strict credit review mechanism and strong data management mechanism and Internet credit audit technology, not only can make financing services to cover more small and micro enterprises, It can also help reduce the cost of financing for small micro-enterprises.

In the market of bank financial products, each bank pays great attention to the promotion of its own product brand, such as ICBC's "steady profits", Everbright Bank's "sunshine financial management", Minsheng Bank's "extraordinary financial management", China Merchants Bank's "Bank Jinbao" and so on, in the market have a certain brand awareness [6].

3.2. Elements of Banking Financial Products

(1) Publisher

That is, the seller of wealth management products, generally is the development of financial products financial institutions. Investors should generally pay attention to the issuer's research and development, investment management strength. Financial products issued by powerful financial institutions are more reliable. In addition, some investment channels are eligible, small financial institutions may not be eligible to participate in these investments, which creates a restriction on the direction of investment for issuers, and ultimately affects the return on wealth management products, so that the credit of powerful institutions is more reliable.

(2) Subscribers

That is, the investor of banking products. Some wealth management products are not intended for the general public, but are available for targeted subscription groups.

(3) Term

Any financial product is issued with a deadline. Most of the financial products issued by banks have a relatively short term, but there are also foreign banks to launch a term of 5 to 6 years of wealth management products. Therefore, investors should be clear about the adequacy of their own funds and the investment period of the possible liquidity needs, to avoid the inconvenience. When investing in long-term wealth management products, investors also need to pay attention to macroeconomic trends, interest rates and other indicators have a general judgment, to avoid interest rates and other fluctuations caused by losses or liquidity difficulties.

(4) Price and gain

Price is the core element of financial products. The purpose of the fund-raiser's sale of a financial product is to earn the equivalent of the price of the product, and the investor's investment is exactly the same as the price of the financial product he buys. The yield represents the percentage of the product's income to investors as a percentage of the investment. It is the rate of return calculated on the terms of the product after the end of the investment management period.

(5) Risk

In the effective financial market, risk and benefit are always equal, only to assume the corresponding risks can obtain the corresponding return. In practice, financial markets are not always effective or not always effective. Due to the existence of information asymmetry and other factors, there may be low risk high-yield, high risk low-yield potential in the market. Therefore, investors should learn more about the risk structure of wealth management products, so as to make a judgment and evaluation of them to see if they match the income.

(6) Liquidity

Liquidity refers to the liquidity of an asset, which is a pair of contradictions with yield, which is why some economists define interest as the "price of liquidity". Under the same conditions, the better liquidity, the lower the yield, so investors need to make a trade-off between the two.

(7) Other rights nested in wealth management products

Financial derivatives such as options are often embedded in wealth management products, especially in some structured wealth management products. For example, an early redemption clause for an investor is a right (although not necessarily the best option).

Therefore, investors should fully explore the information when choosing wealth management products, and make full use of this right [7].

3.3. Build of user portraits

(1) The concept of user portraits

The user portrait is mainly intended to depict a user's characteristics. The user's characteristics are abstracted into the tagged user model. Labels are characteristic identifications that are highly condensed into the user's information. Distinguish different users from computers by labeling them [8]. Labels are characterized by two different representations, one of which is semantic: it makes it easy to understand the meaning of each label [9]. Another is short text: each label represents only one meaning to make it easier for the machine to extract standardized information.

The build of the user portrait begins with the initial label, followed by manual fine-graining and finally repeated iterations. Label selection should have attenuation and weight. With user portraits, users can be classified, the standard of classification is the similarity between labels.

(2) Classification of user portraits

User portraits can be divided into: qualitative user portraits, qualitative user portraits and quantitative verification, quantitative user portraits. Quantitative is mainly quantity, qualitative is mainly a characteristic.

Qualitative user portraits: qualitative interviews, user type segmentation, building user portraits. Advantages: Fast and easy, can dig deep into the use of scenes and motivations. Cons: Missing data validation.

Qualitative user portraits and quantitative verification: qualitative interviews, user type segmentation and quantitative data verification, building user portraits. Pros: Data and qualitative combination verification. Disadvantages: Heavy workload, high cost.

Quantitative user portraits: user group segmentation assumptions, data collection and clustering analysis, and building user portraits. Advantages: Fully supported by data, the user's characteristics and proportions can be obtained through statistical analysis. Disadvantages: High statistical requirements, difficult to understand the use of the scene, difficult to tap the user's emotional tendencies and behavior behind the reasons and deep motivation.

4. The combination of big data and banking products

4.1. Traditional recommendation algorithms

The traditional recommended algorithm is the collaborative filtering algorithm. Co-filtering algorithms are the most classic and easy-to-operate recommended algorithms. It is divided into two main types, one is based on user recommendations, the other is based on item recommendations.

The main focus of user-based recommendations is on the "users" themselves, with the main goal being to find users who are similar to themselves, and then recommending what these "similar users" like to their customers. What's common in life is "what else do people like like you?" [10].

Item-based recommendations focus primarily on the "items" themselves, and the main goal is to find similar items that are similar to those of their customers and then recommend them to customers. What's common in life is "what else you might like".

For example: in Figure 1, both User A and User B like Item A and Item B. Then you can think that user A and user B are more similar to the user, that is, each other as "neighbors." Then you can user B like item D recommended to user A.

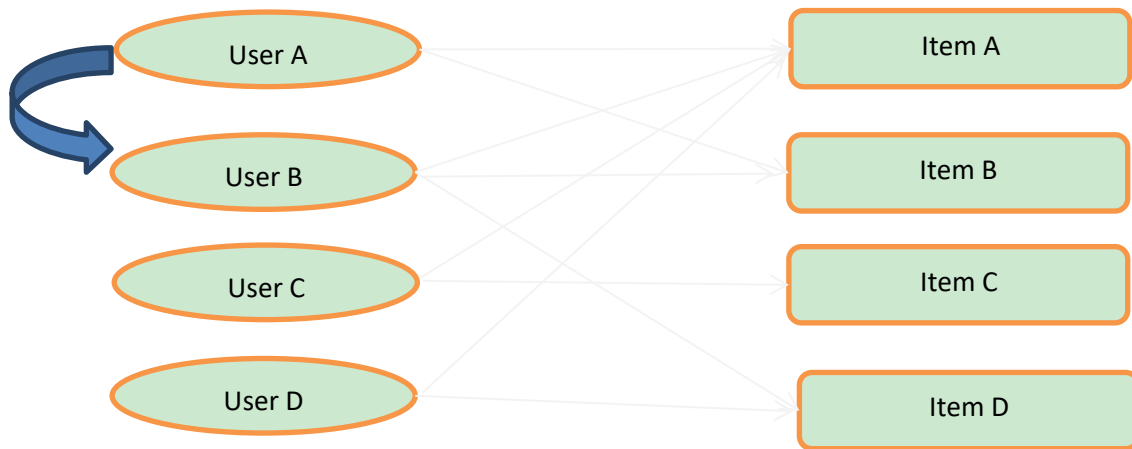


Figure 1. Based on user recommendations

In Figure 2, the item combination A-D is also the most liked by the user at the same time, so you can think of item A and item D the most similarity. Therefore, you can recommend item B. to users who like item A.

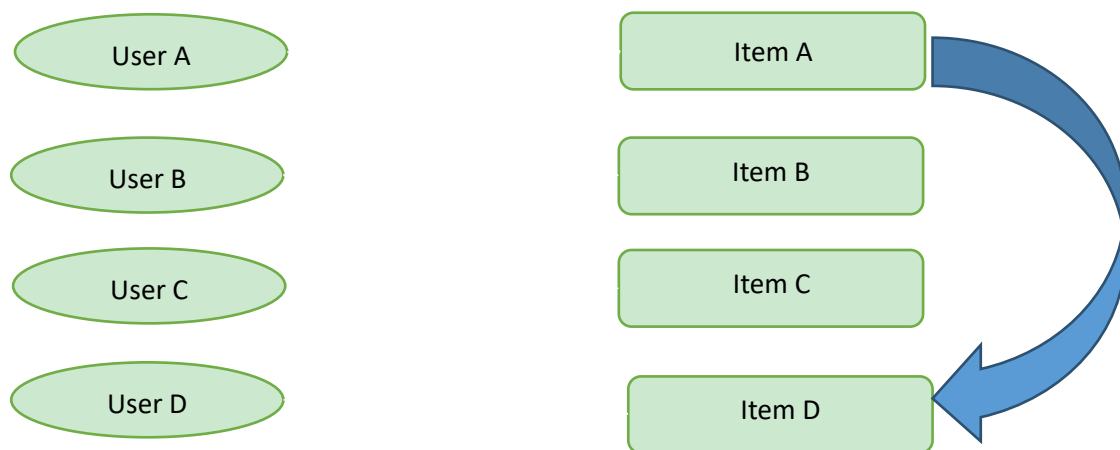


Figure 2. Recommendations based on items

4.2. Improvement of the recommended algorithm for collaborative filtering

Based on the traditional collaborative filtering algorithm, there are many disadvantages, so this paper presents two aspects of improvement:

(1) Consider the popularity of items and user activity

User-based: We believe that if two users do the same thing with cold-calling items, they have the same interest. It is proposed to calculate the similarity between users using the following formula:

$$W(u,v) = \frac{\sum_{i \in N(u)} \frac{1}{\log(1 + |N(i)|)}}{\sqrt{|N(u)| |N(v)|}}$$

Where, u, v represent the different two users. N(i) represents the number of users who have had behavior toward si. This formula is a good solution to the effect of user u and user v common interest list of popular items on their similarity.

Item-based: We believe that active users contribute less to the similarity of items than inactive users. Propose to calculate the similarity between items using the following formula:

$$W(i, j) = \frac{\sum_{u \in N(i) \cap N(j)} \frac{1}{\log(1 + |N(u)|)}}{\sqrt{|N(i)| |N(j)|}}$$

Among them, i, j represent $s(i)$ represent the different two items. $N(u)$ represents the number of items that user u has produced behavior.

(2) Consider the impact of time

None of the above algorithm improvements take into account the time factor, in fact, the time factor also has an impact on the accuracy of the recommended algorithm. User-based: If two users are interested in an item during the Same Time Period, they are more similar. Add the time factor to the calculation and you can use the following formula:

$$W(u, v) = \frac{\sum_{i \in N(u) \cap N(v)} \frac{1}{1 + \partial^* |Tui - Tvi|}}{\sqrt{|N(u)| |N(v)|}}$$

Where Tui indicates when the user u behaves on the item i , $\frac{1}{1 + \partial^* |Tui - Tvi|}$ Available at any $|Tui - Tvi|$ is replaced by a regressive and larger than 0 function. After finding a user with similar interests to the current user u , this group of users' recent interests are closer to the current interest of the user u than before, which can be calculated using the following formula:

$$P(u, i) = \sum_{v \in S(u, k) \cap N(i)} W(u, v) * r(v, i) * \frac{1}{1 + \partial (T0 - Tvi)}$$

(Where $T0$ represents the current time.)

Item-based: Usually the user likes the item more similar, in calculating the similarity of the item with the time factor, use the following formula:

$$W(i, j) = \frac{\sum_{u \in N(i) \cap N(j)} \frac{1}{1 + \partial^* |Tui - Tuj|}}{\sqrt{|N(i)| |N(j)|}}$$

The user's recent behavior reflects the user's current interest more than the previous behavior, and $P(u, i)$ can be calculated using the following formula:

$$P(u, i) = \sum_{j \in S(i, k) \cap N(u)} W(i, j) * r(u, j) * \frac{1}{1 + \partial (T0 - Tuj)}$$

4.3. Time-related calculation of recent interests combined with precision marketing

In response to the improvements to the collaborative filtering algorithm proposed in the previous section, we use the calculation method of time-related similarity to recent interests (Time-generation calculation of nearest interest similarity) to achieve the experimental validation of this article. The idea of verification is: according to a user's evaluation of financial products to achieve the improvement of user similarity based on the collaborative filtering algorithm recommendation, similarity calculation using time-related recent interest similarity calculation [11].

(1) Find user A's (user_id_1) interests.

(2) Find the user group set that has the same financial product interests and hobbies as user A in the same time period.

(3) Find the collection of financial products that this group likes "recently" <products_id>.

(4) Recommend these financial products to user A.

But sometimes we come across data because of data expansion between two users, large data on one side and small on one side, but both are called obvious linear relationships. We introduced the Pearson correlation coefficient to measure the linear correlation between the two variables.

The interval range of Pearson values is -1 to 1. Where -1: full negative correlation, 1: full positive correlation, 0: not related. Correlation coefficients in the range of 0.8 to 1.0 are very strong correlation, strong correlation in the range of 0.6 to 0.8, medium correlation in the range of 0.4 to 0.6, weak correlation in the range of 0.2 to 0.4, and very weak or non-related in the range of 0.0 to 0.2.

5. Summary and outlook

The personalized recommendation of bank products proposed in this paper is not complicated enough and can only be applied in standardized fund products and wealth management products. Especially suitable for the current bank outlets, mobile banking and other electronic screen normal, mainstream personalized recommendation products. But in real-world precision marketing, personalized recommendations for more complex portfolios require more machine learning and financial knowledge to support them [12]. Therefore, the study of recommended products meet customer expectations, reduce customer disappointment rate is the future we should pay attention to and research topics.

Acknowledgments

Beijing science and technology innovation service capability construction project (PXM2016_014223_000025) and Beijing Institute of Graphic Communication 2018 R&D Project (Ec201802).

References

- [1] Li Guojie, Cheng Xueqi. Big Data Research: A Major Strategic Field of Future Science and Technology and Economic and Social Development - The Status of Big Data Research and Scientific Thinking. *Journal of the Chinese Academy of Sciences*, 2012 (06): 647-657.
- [2] Zhang Fang. Research on Mobile Banking Strategy of China Construction Bank A Branch. Guangxi: Guangxi Normal University, 2019.
- [3] Lin Qingpeng. Precision Marketing Strategy Research based on Big Data Mining.
- [4] The Impact and Countermeasures of Big Data Applications on The Business Development of Commercial Banks.
- [5] Party." Big Data" Trend Commercial Bank Response Strategy Research *New Finance*, 2012 (12): 25-28. 8.html.
- [6] Wang. Research on the Precision Marketing Path of Big Data-based E-Commerce Enterprises. *J/OL. Technology Economy Market*, 2019 (08): 156-158 .2019-10-17. <http://kns.cnki.net/kcms/detail/36.1122.n.201909.1423.144>.
- [7] Li Xiaoyang. The property of commercial data in the context of big data. *J/OL. Jiangsu Social Sciences*: 1-10.2019-10-18. <https://doi.org/10.13858/j.cnki.cn32-312/c.20190918.013>.
- [8] Deng Elegant. Design and Application of Precision Marketing of Commercial Banks in the Background of Big Data.
- [9] Zhang Zhuang, Feng Xiaonian, Qian Tieyun. User Portrait Method based on multi-modal fusion technology *J/OL. Journal of Peking University (Natural Science Edition)*: 1-8 (2019-10-17) .<https://doi.org/10.13209/j.0479-8023.2019.097>.
- [10] Jiang Zongli, Wang Yihe. Recommendation Algorithm based on series excavation's point of interest .J., *Journal of Beijing University of Technology*, 2019,45 (09): 853-858.

- [11] Li Xiaofeng, Li Dong. High sparse array combination recommendation algorithm based on deep learning.
- [12] Zhou Dongqi. Precision Marketing Research on J Commercial Bank Financial Products in the Background of Big Data.