Accurate Marketing Research Based on Big Data Analysis of Online Education User Behavior

Weilin Li

Xi'an Eurasia University, Xi'an, China

Keywords: marketing; big data; online behavior learning; social network analysis; metadata

Abstract: In modern marketing, online learning based on big data is used to achieve accurate marketing. At present, with the rapid development of mobile information technology, online learning based on mobile network has become the mainstream. By using metadata and social network analysis, new indicators are created to identify customers most likely to be converted into mobile Internet users. The main indicators are discretionary income, time and social learning. Through the use of historical data, machine learning prediction model is trained and validated, and used to select the experimental group, select online customer behavior data to validate the model. The results show that the model also shows good performance in a longer period of time. 98% of the customers in the experimental group are transformed into final target customers after the end of the activity.

1. Introduction

At present, mobile phones are the main way to connect to the Internet. Mobile Network Operators (MNOs) usually use short messages to improve customers'awareness and feedback on new products and services. In the domestic market, MNOs usually carries out thousands of text publicity activities every year, and receives several promotional messages every month. In order to ensure that customers will not be regarded as spammers, but provide useful information, which is the main focus of MNOs in the marketing process; for this particular marketer, precise marketing is achieved through the strategy of not sending the same text information every 14 days.

Deciding which service to send to which customer often depends on the marketing team's "intuition", that is, what is the right audience in the sales process, which is called target positioning. In a recent IBM study, 80% of marketers reported that their target selection was based on their "feelings" and that a data-driven approach might improve the efficiency of text-based activities because it was closer to providing "the right offer" to the right customers. For example, previous studies have shown that data-driven methods can reliably predict the personality of mobile and Facebook users.

This paper evaluates MNO's best practices in large-scale "Internet Data" experiments based on data-driven, validates the online learning behavior analysis model, and proves that using machine learning and social network analysis will lead to higher conversation rate, rather than the best practice marketing method. At the same time, when the motion response data is not available, the historical natural adoption rate data is used to train the model.

2. Analysis of Online Learning Behavior

2.1 Practice Model

Current MNOs best practices rely on the experience of marketing teams to determine which customers should receive information about a particular activity. The marketing team usually chooses customers to calculate directly using metadata of some simple indicators (such as sending, receiving time, average recharge, etc.). For this special "Internet Data" activity, the number of texts recommended by the marketing team in the literature [9] survey is based on the average revenue per user of the user. (ARPU), equal to the average monthly income, and prepaid customers. As shown in

  DOI: 10.25236/ismeem.2019.070

Table 1, the variables of the experimental group in the model are selected (according to practice, these customers are most likely to become target customers).

The experimental group consisted of 50,000 randomly selected customers.

Table 1. Variables used to select experimental groups

Serial number	Experimental group variable
1	Send at least four text messages per month
2	Receive at least four text messages per month
3	Mobile phone with “data support”
4	“Unexpected data usage”
5	Medium or higher ARPU segment

2.2 Data Driven Method

2.2.1 Characteristic

For each subscriber in the experiment, more than 350 features were obtained from the use of metadata, subscription data and value-added services. According to the relevant research, we can infer the social graph between customers to calculate the new features. This paper only considers the customer interaction more than three times a month. In this social chart, the strength of the connection is the weighted sum of two months'phone calls and text messages. Using this social chart, we calculate about 40 social characteristics.

2.2.2 Model

This article uses six months of metadata for development and model training. In this paper, natural user data acquisition is used to train the model. Then, by comparing these natural users with those who did not use the mobile Internet in the same period according to the people who just started using the mobile internet, the goal of this method is to complete the following confirmation: (1) the user's behavior may be interested in using the internet; (2) who will continue to use the mobile internet. Then, 50,000 natural users and 100,000 non-Internet users were randomly selected. Note that the natural converter is only one way to extract the features of the customers that may be converted. Constructing a training set for learning behavior analysis as shown in Table 2.

Table.2 Training Set

Size	Classifier	Description
50K	Natural user	From December to March, there are 50K natural adopters with less than 50KB of data per month (unexpected data usage). More than 1MB of data per month in April and May
100K	Non-internet user	Do not use the internet

Several modeling algorithms, such as support vector machine and neural network, are tested according to the above training set to classify natural converters. The implementation of the model is a trade-off between accuracy and stability. According to past experience, stability is taken as the evaluation criterion.

In this paper, the cross validation model is used to execute the training set, which only depends on several key variables. In the initial sample selection, several centralized features are selected as the final model evaluation index.

2.2.3 Training Set Verification

Figure 1 shows the performance of historical data measured using lift curves, with about three of the top 20% customers improving. This means that if you choose 20% of the highest scoring customers, the model will be three times more than the number of random choices from the sample.

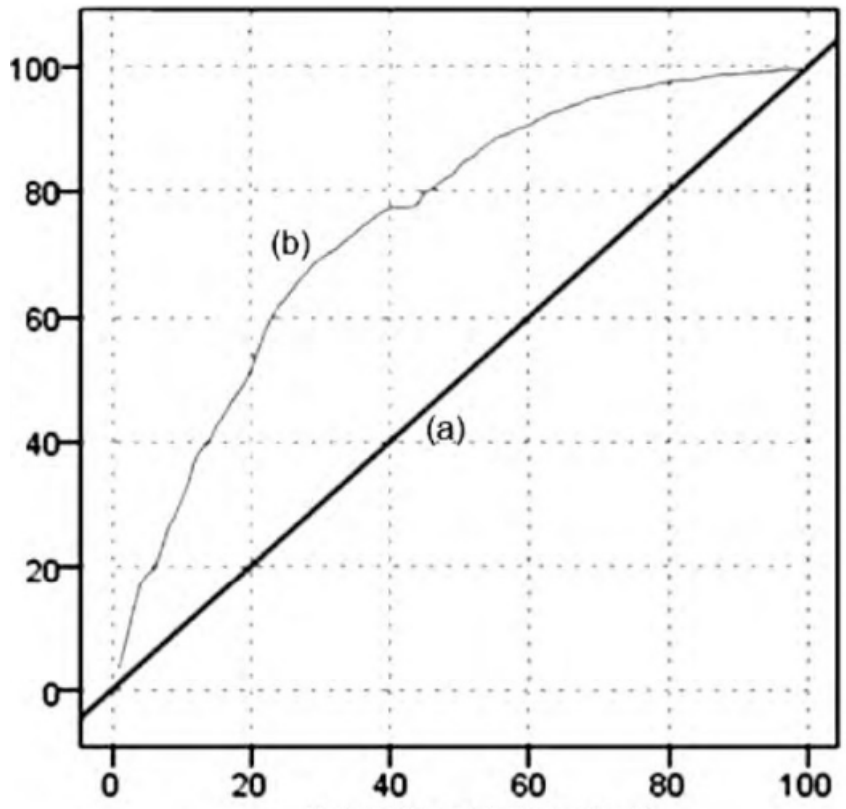

Figure.1 Out-of-sample verification of the model using the lift curve

(a) in the case of random selection (b) display the result of the selection by using the model

The model was used to select the experimental group, and the marketing team selected the appropriate control group. It was stipulated that the selected control group should be explicitly excluded when choosing the experimental group.

3. Experiments

Through large-scale experiments, data-driven methods are compared with current MNOs best practices. The conversion rates of the control group and the experimental group were compared.

The experimental process is described as follows: the selected customers receive a short message informing them that they can activate 15MB traffic data usage at half the market price, and the validity period of 15MB traffic is 15 days. Customers can activate the package by sending text messages with codes to a short number, which is a common marketing activity for operators. The text message received by the client contains information and instructions on how to activate it.

The conversion rate between the experimental group and the control group was astonishing. The conversion rate of the experimental group selected by the model was 642%, while that of the control group selected by the best practice method was only 0.5%, as shown in Figure 2 (a). The difference is very significant (p value is less than 2.2e-16).

Figure 2 (a) Conversion rates (b) of the control group (best practice method) and the experimental group (data-driven method) used the percentage of people updating the data plan after using the number included in the campaign, with an error condition of using the blaker method and a confidence interval of 95%.

The goal of this experiment is not only to get customers to accept the offer, but also to get them to start again after the trial period. Therefore, the update rate is compared in Figure 2 (b). After using half-price packages, the customer buys another package service between the two groups. The results show that 98% of the reformed people in the experimental group bought the second complete price package, while only 37% of the control group updated their package; this means that 6.29% of the experimental group was converted in the second month, while the control group was 0.19%.

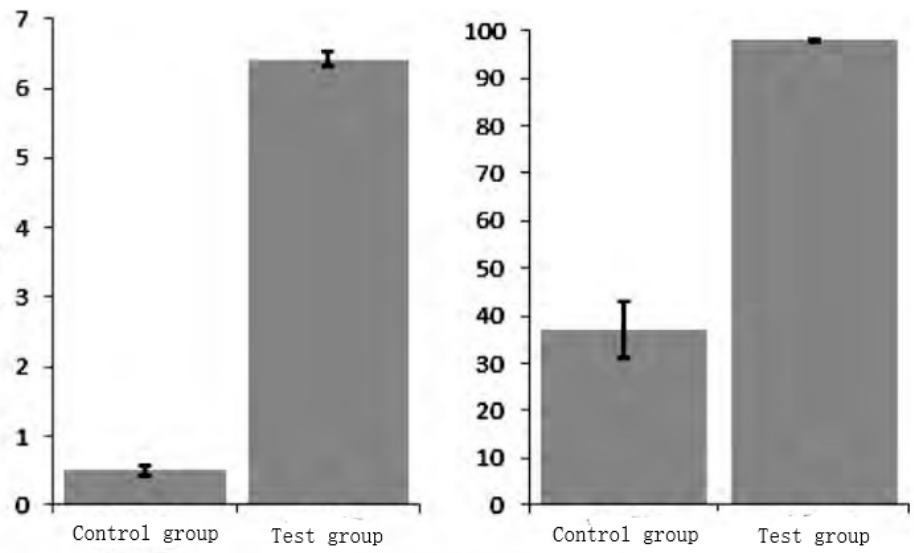

Fig.2 the update rate camparison

Although this is not the ultimate goal of the data-driven approach adopted in this paper, the posterior classification of the features selected by the model will bring some interesting qualitative analysis. In fact, most of the features have three types: disposable income, time and social learning. Marketing teams believe that discretionary income is generally important and assume that customers with high ARPUs are more likely to switch. However, the model does not select total expenditure as an important variable to help predict conversion. In fact, looking at the ARPU of people who receive text messages, you will find that the performance of the low ARPU part is slightly higher.

At the same time, the use of social functions in customer selection may increase retention rates. It can be inferred that when the user's neighbors are mobile data users, the value they get from mobile data will also increase. In other words, we hope that network externalities exist in mobile Internet data. This means that choosing users whose nearest neighbors are already using mobile data may use this network effect locally to create observed persistent effects, and the retention rate of users will be very high in the second month. The marketing team is impressed by the power of this method, and is now studying how to implement this function in the operation of MNO and how to use it more systematically in future activities. Data-driven marketing and customer understanding have opened up a better research approach, using behavioral patterns, compared with current best practices, the conversion rate of Internet data activities has been significantly improved. It is expected that such a method can greatly reduce the sending of spam information by providing more relevant services to customers.

4. Conclusion

This paper mainly uses the online marketing of telecom operators as the research case background. Through the method of big data analysis, according to the process of feature selection, model, training set, etc., the discretionary income, time and social learning are the main research features. The control group constructed by the marketer is the reference selection experiment group, which proves that the data-driven method has better experimental results in the target customer selection.

References

[1] Li Yang. Research on online learning behavior analysis model in big data environment [D]. Harbin: Harbin University of Science and Technology, 2017.

[2] Fang Haiguang, Luo Jinping, Chen Junda, et al. Quantitative self-MOOC adaptive learning

system based on educational big data[J]. Journal of Electrotechnical Education, 2016, 37 (11): 38-42+92.

[3] Feng Xiaoying, Zheng Qinhua, Chen Pengyu. Research on Evaluation Model of Online Cognitive Level from the Perspective of Learning Analysis[J]. Journal of Distance Education, 2016, 34 (06): 39-45.

[4] SUN Xiaowei. Analysis of Online Learning Behavior Based on SPOC Platform Log Data and Its Influencing Factors[J]. Journal of Shenyang Normal University(Natural Science), 2017, 35 (01): 103-107.

[5] Zhang Hongling, Feng Bozhen, Li Hui, et al. Research and Countermeasures of Network Education Teaching Based on Online Learning Behavior Data Analysis[J].Information Science, 2017, 35 (09) :74-78.

[6] Yuan Zicong. Mathematical classroom teaching strategy based on learning behavior analysis[J]. Educational Information Technology, 2017, (04): 18-20.

[7] Li Shuang, Wang Zengxian, Yu Wei, et al. Research on Online Learning Behavior Input Analysis Framework and Measurement Index--Based on LMS Data Learning Analysis[J]. Open Education Research, 2016, 22 (02) : 77-88.

[8] Yang Dingzhong. Research on Digital Learning Behavior Based on Educational Big Data[J]. Educational Teaching Forum, 2017, (22): 77-79.

[9] Wang Hongjiang, Mu Su, Huang Jie, et al. Real-time analysis method and application research of independent learning input[J]. Journal of Electrotechnical Education, 2017, 38 (10): 44-50.

[10] Xiao Yumin, Meng Bingwen, Tang Tingting, et al. Big Data Research and Practice for Wisdom Education: Value Discovery and Path Exploration[J]. Journal of Electrotechnical Education, 2017, (12): 5-12.

[11] Chen Ganguo. An Analysis and Prediction Strategy of Cadre Online Cheating Learning Behavior[J]. Computer Engineering, 2017, 43 (09) : 17-22+28.

[12] Cao Jianfang, Hao Yaojun. Evaluation of large-scale online learning behavior based on parallel Adaboost-BP network[J].Computer Applications and Software, 2017, 34 (07):267-272.

[13] Lü Haiyan, Zhou Lijun, Zhang Jie.Application of Educational Data Mining in the Analysis of Students' Online Learning Behavior under the Background of Big Data[J]. Computing Technology and Automation, 2017, 36 (01): 136-140.

[14] Xing Wei. Research on the Factors Affecting the Learning Effect of Online Learning Behavior in Mixed Learning[J]. Curriculum Education Research, 2017, (25): 37-38.

[15] Zhang Shuai, Zhao Yuying. Research on the Performance Model of “Leading” Online Learning Behavior[J]. Times Education, 2017, (08) :6.

[16] Liu Yan. Characteristics Analysis of Learners' Participation in Online Learning Behavior in Distance Education[J]. Times Education, 2016, (18): 51.