Research on the Behavioral Classification of Consumer Brand Transformation Based on Big Data of Cigarette Consumption

Lili Zhu[a] , Yuhua Mo*[b]

[a]zhulili-lily@qq.com,[*b] Corresponding author: moyuhua95@gmail.com

Internet Research Centre, China Tobacco Guangxi Industrial Co., LTD., Nanning, Guangxi, China

Abstract: With the emergence of big data technology, it provides new solutions for consumer behavior insights and innovative marketing strategies. This paper constructed a consumer behavior classification model based on the perspective of brand transformer from two dimensions: brand switch probability and price trend tendency, which was called CBST model. The NMF model was used to calculate the probability of consumer brand switch, and the ARIMA model was used to predict the trend of price change when consumers converted brands. Furthermore, the model effect was verified by taking the data of cigarette consumers' scan code brand record in a certain city for one year as an example. Specifically, consumer behavior can be divided into four types: high switch probability and high price tendency; low switch probability and high price tendency; low switch probability and low price tendency; high switch probability and low price tendency. For each type of consumer behavior, targeted marketing strategies could be formulated from three aspects: consumption trend, value of conversion, and attention degree, which could provide suggestions for effective customer relationship management and efficient allocation of marketing resources for enterprises.

Keywords: brand switch; NMF model; consumer behavior analysis; consumer classification; big data

1. INTRODUCTION

In the era of digital economy, with the intelligence of mobile devices, consumers have more diverse channels to understand product information. The range of products that consumers can choose gradually expands, accompanied by a decrease in brand and product loyalty among consumers. Consumer brand switching behavior not only intensifies competition between enterprises, increases advertising and promotion costs, and makes it difficult for them to acquire customers; but also motivates enterprises to continuously innovate products, timely observe changes in consumer behavior and preferences, and predict consumption trends.

The emergence of big data technology provides new solutions for consumer behavior insights and innovative marketing strategies [1]. Firstly, big data technology can analyze historical purchasing data of consumers to help enterprises better understand consumption habits and insights into consumer psychology. Secondly, big data technology can help enterprises predict future consumer demand [2], fully utilizing the economic value of data as a production factor, and finding opportunities from data to cope with market uncertainty challenges. Finally, big

ICEMBDA 2023, October 27-29, Tianjin, People's Republic of China

DOI 10.4108/eai.27-10-2023.2341920

data technology can help enterprises optimize marketing strategies. Based on consumer evaluations, enterprises can efficiently allocate limited resources through effective marketing channels and methods, accurately positioning brands and enhancing brand image [3]. This paper analyzes consumer behavior from the perspective of consumer brand switching, using big data technology, and illustrates the application of the fast-moving consumer goods attention model in innovative marketing strategies by taking cigarette products as an example, providing some suggestions and inspiration for enterprises.

2 LITERATURE REVIEW

Brand is an important competitiveness of an enterprise, especially for fast-moving consumer goods. A good brand image in the minds of consumers can help reduce consumer choice costs and improve decision-making efficiency. Researchers have mainly studied the factors that influence consumer brand switching behavior from two aspects: brand switching intention and influencing factors. Regarding brand switching intention, Yan Lin analyzed the effect of product involvement degree on consumer brand switching intention based on the interaction between switching cost and perceived risk. Her research found that product involvement degree has a negative impact on consumer brand switching intention. When the degree of product involvement is high, consumers perceive the switching cost to be relatively high as well, and they are not easily willing to switch brands. Conversely, it means the switching intention is high [4]. Lin Bingkun and Lü Qinghua also analyzed the impact of consumer confusion on brand switching behavior based on switching costs and provided suggestions for companies to alleviate consumer confusion by releasing product information [5].

Regarding the analysis of factors influencing brand switching behavior, Gönül and Srinivasan analyzed the impact of direct price reduction and coupon giving on consumer brand switching behavior, taking into account two sources of variance in consumer heterogeneity and brand loyalty tendencies. Their research provides recommendations for customer relationship management, emphasizing the perceived value of coupons compared to direct price reductions and the relationship between brand loyalty tendencies and household income [6]. Liu Jianxin and Fan Xiucheng analyzed the impact of celebrity endorsement scandals on consumer brand switching behavior and analyzed its mediating mechanism [7]. Compared to national brands, recommendations based on private labels have a higher conversion success rate because private labels benefit more from quality signals [8].

With the development of digital technology, big data has had an impact on traditional consumer analysis theories, methods, and operations management [9]. Erevelles and Fukawa et al. built an analytical framework using three types of company resources: physical capital, human capital, and organizational capital to collect, extract, and apply consumer big data according to resource theory [10]. Erkayman and Erdem et al. compared the accuracy of machine learning and deep learning techniques in predicting consumer switching decisions and found that deep learning with k-fold cross-validation technique performed better when analyzing consumer data [11]. Furthermore, Rahim and Mushafiq et al. used three machine learning methods: multi-layer perceptron (MLP), support vector machine (SVM), and decision tree classification (DTC) to analyze repeat purchase behavior classifications [12].

After applying big data technology to analyze consumer purchasing records, it is common to stratify and classify consumers in order to provide targeted marketing strategies. One commonly used model is the RFM model [13], which categorizes customers into superstar, golden customer, typical customer, occasional customer, everyday shoppers, dormant customers, etc. Monalisa and Nadya et al. then used fuzzy C-means algorithm to cluster consumers from LWC Company into three categories, providing three combinations of customer lifetime value strategy [14]. Song and Zhao et al. improved the time series data used in the RFM model to ensure accuracy [15]. However, the RFM model also has some drawbacks. For example, it may not accurately classify customers with small spending amounts because it primarily focuses on the amount of purchase made by customers. Therefore, research needs to be conducted to explore different dimensions for classifying consumers.

Applying the above brand switching research theories and big data technology, models to product analysis can help identify new consumer conclusions. Caracciolo and Furno et al. used big data to study the relationship between consumer groups and wine types, finding that brand switchers prefer medium-quality wines while loyal consumers purchase basic wines [16]. Taking cigarette products as an example, Shiffman and Goldenson used ADJUSST research method and Tobacco Dependence Index (TDI) to analyze changes in US adults who smoked over 1 year or completely converted to JUUL-brand electronic nicotine delivery system (ENDS) dependency [17]. Cowie and Swift et al. analyzed the relationship between brand switching behavior in Australian cigarette consumers and smoking cessation activities, finding that brand switching does not seem to affect smoking cessation behavior as it neither inhibits nor promotes smoking cessation [18].

This paper aims to understand consumer behavior from the perspective of consumer brand switching, utilizing big data technology to build a consumer behavior classification model, providing a new perspective for researching consumer classifications.

3 Model

The consumer behavior switch tendency (CBST) model for individual consumers, constructed from a consumer brand switching perspective, has two dimensions: the probability of a consumer making a brand switch, and the trend of changes in the price of switching brands.

$$\mathrm{CBST} = \mathrm{F}(\mathrm{switch}, \mathrm{tendency}) \tag{1}$$

The probability of a consumer making a brand switch is measured using the Non-negative Matrix Factorization (NMF) model. This model assumes that within a time period, there are m consumers $i = 1,2,\cdots,m$, choosing from n brands $j = 1,2,\cdots,n$, where n is a positive integer representing the number of mutually exclusive choices available to consumers. By establishing a random utility matrix U, the random utility derived from choosing a brand can be described.

The dimensionality of the random utility matrix U is $m \times n$, where m represents the number of consumers and n represents the number of brands. Each element u_{ij} represents the random utility derived by consumer i choosing brand j.

To model and predict random utilities, the NMF model is used to decompose the U matrix into two non-negative matrices W and H, satisfying:

$$U \approx W_{m*k} * H_{k*n} \tag{2}$$

The Non-negative Matrix W represents the consumer brand purchase features, with dimensions $m \times k$, where k is the number of features. Each element w_{ik} represents the relationship strength between the ith consumer and the kth feature. The Non-negative Matrix H represents the brand features, with dimensions $k \times n$. Each element h_{kj} represents the relationship strength between the kth feature and the jth brand.

By decomposing the NMF model, we obtain the matrices W and H such that the error between the product of the two matrices and the original matrix U is minimized. The objective function F(W,H) is defined as:

$$\mathrm{Min}_{(W,H)}\ F(W,H) = \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n}(U - (WH)^T)^2 + \frac{1}{2}(|W|_F^2 + |H|_F^2) \tag{3}$$

$$W = [\, w_{ik} \,]\ ,\ \ |W|_F^2 H = [\, h_{kj} \,]$$

$$\text{s.t. } w_{ik} \geq 0\, , h_{kj} \geq 0, \forall i, k, j$$

To compute the gradients of F with respect to w_i and h_j, we have:

$$\frac{\partial F}{\partial w_i} = -\sum_{i=1}^{m}(u_{ij} - w_i h_j^T)\, h_j + \lambda w_i \tag{4}$$

$$\frac{\partial F}{\partial h_j} = -\sum_{j=1}^{n}(u_{ij} - w_i h_j^T)\, w_i + \lambda h_j \tag{5}$$

Using these gradients, we obtain the final matrices W and H. The matrix P is then calculated using the formula:

$$P = [\, p_{ij}] = W_{m*k} * H_{k*n} \tag{6}$$

Where p_{ij} represents the probability of a consumer i switching brands from j to another brand. The rows of matrix P represent consumers and the columns represent brands.

To estimate the consumer's price trend when switching brands, we use an ARIMA model of order p and q. For a given set of historical prices $\{p_1, p_2, \cdots, p_T\}$, we predict the price trend of consumers at time t as follows:

$$P_t = \beta_1 p_{t-1} + \beta_2 p_{t-2} + \cdots + \beta_p p_{t-p} + \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2} + \cdots + +\alpha_q \varepsilon_{t-q} \tag{7}$$

Where$\beta_1, \beta_2, \cdots, \beta_p$ and $\alpha_1, \alpha_2, \cdots, \alpha_q$are regression coefficients to be estimated from the data.

4 Data and Result

4.1 Data Analysis

This paper analyzes the sales data of cigarettes in a city in 2022. The number of cigarette brands sold in this city in 2022 was 193. The consumer conversion analysis in this paper refers to the situation where consumers switch brands after their previous purchase according to the natural time sequence, i.e., it is considered as a single brand conversion behavior. Based on this definition, the distribution of brand conversions among consumers in this city in 2022 can be statistically analyzed, as shown in Table 1.

Table 1. Distribution of Brand Types and Number of Brand Conversions among Consumers

Range of Purchased Brand Types Quantity	Number of Consumers (%)	Range of Brand Switch Quantity	Number of Consumers (%)
1-10	82.67	0	29.75
11-20	5.35	1-100	61.81
21-30	3.13	101-200	3.10
31-40	2.73	201-300	1.51
41-50	2.40	301-400	0.88
51-60	1.42	401-500	1.06
61-70	1.31	501-600	0.58
71-80	0.63	601-700	0.28
81-90	0.18	701-800	0.53
91-99	0.18	801-900	0.24
Total	100	901-973	0.26
		Total	100

As shown in Table 1, 82.67% of consumers purchased cigarette brands from 10 different types or less. The maximum number of brands a single consumer purchased was 99. The percentage of consumers who never switched brands in the past year was 29.75%, indicating that nearly one-third of consumers in the city are loyal to a certain brand. The proportion of consumers who switched brands fewer than 100 times was 61.81%, suggesting that more than half of consumers in the city would at least try out two brands.

4.2 Calculation Results

Based on the NMF model used in this paper, the probability of consumer brand switching is calculated. Taking the purchase records of a randomly selected consumer in 2022 for analysis, the calculation results are shown in Figure 1.

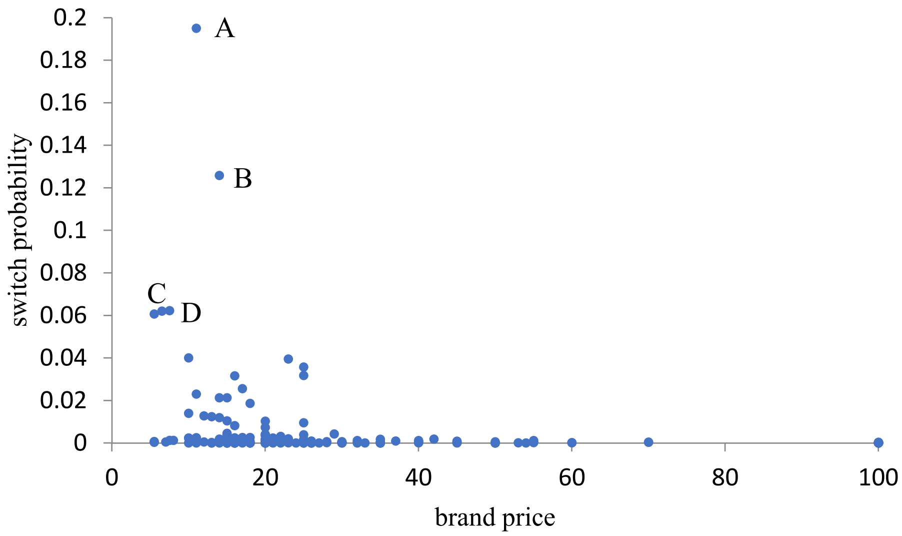

Figure 1. Probability Distribution of Single Consumer Brand Conversion

Most of the consumer's brand conversion probabilities are not high, and the price range is relatively concentrated. There are 175 brands with a conversion probability of less than 2%, and the average price is 33.77 yuan. The highest conversion probability is to brand A at a price of 11 yuan, reaching 19.5%; followed by switching to brand B at a price of 14 yuan; then there is a close difference in conversion probabilities between switching to brand D at 7.5 yuan and brand C at 6.5 yuan, which are 6.22% and 6.20%, respectively. For cigarette brands priced at 100 yuan, the average conversion probability for this consumer is 0.01%. It can be seen that this consumer does not have a tendency to switch to higher-priced cigarette brands.

4.3 Model Application

This paragraph discusses the application of a model developed based on the calculated consumer brand switching probability. The model can be used for consumer classification and innovative marketing strategies. It constructs a consumer behavior classification diagram based on two dimensions, i.e., brand switching probability and price trend change. Figure 2 classifies consumers into four categories: high switching probability and high price trend; low switching probability and high price trend; low switching probability and low price trend; and high switching probability and low price trend. Specifically, by collecting all the maximum values of consumer switching probabilities and their corresponding prices, we can obtain the overall consumer behavior in the market. This information can then be used to create consumer labels and develop targeted consumer attention marketing strategies based on these labels.

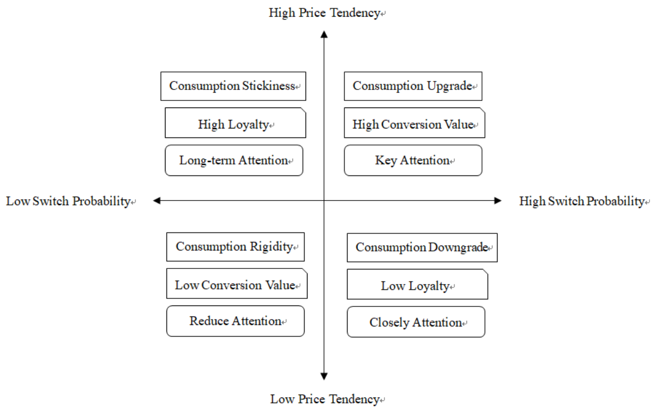

Figure 2. A consumer behavior classification diagram based on the perspective of brand switching by consumers

4.3.1 Consumer Trend

If a consumer has a high switch probability and tends to switch to higher-priced brands, it indicates that the consumer is in the stage of consumption upgrading. If a consumer has a high switch probability but tends to switch to lower-priced brands, it indicates that the consumer is in the stage of consumption downgrading. If a consumer does not frequently engage in brand switching behavior and is in the stage of switching to higher-priced brands, it indicates that the consumer has a strong attachment to higher-priced brands. If a consumer does not frequently engage in brand switching behavior but is in the stage of switching to lower-priced brands, it indicates that the consumer's purchasing behavior has solidified.

4.3.2 Transitional Value

For an individual consumer, if they are in the quadrant with high switch probability and high price trend, it means they have higher transitional value; if they are in the quadrant with low switch probability and low price trend, it means they have lower transitional value. For consumers in both quadrants with low switch probability and high price trend or high switch probability and low price trend, their transitional value needs to be analyzed specifically based on the actual situation.

4.3.3 Attention Degree

For consumers with characteristics of high switch probability and high price trend, companies should pay special attention to them because these consumers have a higher chance of becoming brand consumers for their own brands and can bring higher transitional value to their own brands. For consumers with characteristics of low switch probability and high price trend, companies need to pay long-term attention to them because these consumers belong to high net worth loyal consumers and need to maintain good customer relationships.

For consumers with characteristics of high switch probability and low price trend, companies should closely monitor them because these consumers have a relatively low level of loyalty and may switch to other brands at any time. Although there is a tendency to switch to lower-priced brands, if the number of such consumers is large enough, it may also affect brand image and company performance. For consumers with characteristics of low switch probability and low price trend, companies should reduce their attention as these consumers have relatively low transitional value and can reduce the investment of marketing resources by allocating limited resources reasonably.

5 CONCLUSION

In this paper, a consumer behavior classification model is constructed based on the perspective of brand switching from the angles of conversion probability and price trend changes, which is called consumer behavior switch tendency model (CBST model). Among the study, the NMF model is used to calculate the conversion probability of consumers' brand switching and the effect of the model is verified with cigarette brands as an example. Specifically, consumer behavior can be divided into four types: high switch probability, high price tendency; low switch probability, high price tendency; low switch probability, low price tendency; high switch probability, low price tendency.

For these four types of consumer behavior, targeted marketing strategies can be formulated from three aspects: consumption trend, transitional value, and attention degree. Consumers with high switch probability and high price tendency indicate that they are upgrading their consumption and have high transitional value, so they need to be given special attention. Consumers with low switch probability and high price tendency show strong consumption stickiness and high brand loyalty, so they need to be given long-term attention. Consumers with low switch probability and low price tendency indicate that their consumption has solidified and their transfer value is small, so they should be given less attention. Consumers with high switch probability and low price tendency indicate that their consumption is downgrading and their brand loyalty is low. If this consumer group has a large number of members, they need to be given close attention.

Acknowledgment : This work was funded by the China Tobacco Guangxi Industrial Co. Ltd.'s 2022 Technology Project Plan [number: GXZYCX2022H002].

Reference

[1] Ning Yang, Liu Yujie, Hou Yaohua. The characteristics, roots, and solutions of consumption alienation in the era of big data. Fujian Tribune(The Humanities & Social Sciences Monthly). 2023(03),140-148. https://ifgfy2b08d79e045e4fd4s65cxwxuk559x6p55ficg.res.gxlib.org.cn/kcms/detail/detail.aspx?FileName=FJLW202303012&DbName=CJFQ2023.

[2] Boone, T., Ganeshan, R., Jain, A., & Sanders, N. R. (2019). Forecasting sales in the supply chain: Consumer analytics in the big data era. International Journal of Forecasting, 35(1), 170-180. doi: https://doi.org/10.1016/j.ijforecast.2018.09.003.

[3] Alzate, M., Arce-Urriza, M., & Cebollada, J. (2022). Mining the text of online consumer reviews to analyze brand image and brand positioning. Journal of Retailing and Consumer Services, 67, 102989. doi: https://doi.org/10.1016/j.jretconser.2022.102989.

[4] Yan Lin, The Impact of Product Involvement on Consumer Brand Conversion Intention:Research on the Dual Role of Transformation Cost and Perceived Risk. Journal of Commercial Economics. 2022(23), 76-79. https://ifgfy2b08d79e045e4fd4s65cxwxuk559x6p55ficg.res.gxlib.org.cn/kcms2/article/abstract?v=l5E5JTlxS0s0iio2zvzRRHG-tjy4wc3ql6UvarOASSMlKcFLn-tXwMx9CmfiEl_HQK3_qIBFDnlJHIUtEbXkAZRF58SJX4rkTLOWGMp2lWp8AQE4aQqsnc65RWIr1Ndhotokl3MUI4k=&uniplatform=NZKPT&language=CHS.

[5] Lin Bingkun, Lv Qinghua. Research on the Impact of Consumer Confusion on Brand SwitchingThe Mediating Effect of Decision Postponement and the Moderating Effect of Switching Costs. China Business And Market. 2022, 36(05), 65-76. doi: https://doi.org/10.14089/j.cnki.cn11-3664/f.2022.05.006.

[6] Gönül, F. F., & Srinivasan, K. (1997). A dynamic model of repeat purchase and brand switching behavior in a consumer products category. Journal of Retailing and Consumer Services, 4(3), 185-191. doi: https://doi.org/10.1016/S0969-6989(96)00060-4.

[7] Liu Jianxin, Fan Xiucheng, Li Dongjin. An Impact of Brand Endorser's Scandal on Consumer Brand Switch-Based on a Model of Multiple Mediators. Luojia Management Review, 2021, (02),80-104. https://ifgfy2b08d79e045e4fd4s65cxwxuk559x6p55ficg.res.gxlib.org.cn/kcms2/article/abstract?v=l5E5JTlxS0sBonJ14wZrNdt8nvSXtU63BCIIFK6YMr5fEn1qypB74YYle6Z9gEnoP1QoKZOJQz8asGvUNYoK9i5WD_ZN0ayuOnloE9XVEICFDNOY64Swyn0c0V0lX86YjF11OL3QiJs=&uniplatform=NZKPT&language=CHS.

[8] Kobuszewski Volles, B., Van Kerckhove, A., & Geuens, M. (2023). Triggering brand switching in online stores: The effectiveness of recommendations for private labels versus national brands. Journal of Business Research, 164, 114020. doi: https://doi.org/10.1016/j.jbusres.2023.114020.

[9] Chen Kun. Research on Consumer Behavior Change and Business Model Innovation in the Era of Big Data. Journal of Shenyang Agricultural University (Social Sciences Edition), 24(06), 678-682. doi: https://doi.org/10.3969/j.issn.1008-9713.2022.06.008.

[10] Erevelles, S., Fukawa, N., & Swayne, L. (2016). Big Data consumer analytics and the transformation of marketing. Journal of Business Research, 69(2), 897-904. doi: https://doi.org/10.1016/j.jbusres.2015.07.001.

[11] Erkayman, B., Erdem, E., Aydin, T., & Mahmat, Z. (2023). New Artificial intelligence approaches for brand switching decisions. Alexandria Engineering Journal, 63, 625-643. doi: https://doi.org/10.1016/j.aej.2022.11.043.

[12] Rahim, M. A., Mushafiq, M., Khan, S., & Arain, Z. A. (2021). RFM-based repurchase behavior for customer classification and segmentation. Journal of Retailing and Consumer Services, 61, 102566. doi: https://doi.org/10.1016/j.jretconser.2021.102566.

[13] Anitha, P., & Patil, M. M. (2022). RFM model for customerpurchase behavior using K-Means algorithm. Journal of King Saud University - Computer and Information Sciences, 34(5), 1785-1792. doi: https://doi.org/10.1016/j.jksuci.2019.12.011.

[14] Monalisa, S., Nadya, P., & Novita, R. (2019). Analysis for Customer Lifetime Value Categorization with RFM Model. Procedia Computer Science, 161, 834-840. doi: https://doi.org/10.1016/j.procs.2019.11.190.

[15] Song, M., Zhao, X., E, H., & Ou, Z. (2017). Statistics-based CRM approach via time series segmenting RFM on large scale data. Knowledge-Based Systems, 132, 21-29. doi: https://doi.org/10.1016/j.knosys.2017.05.027.

[16] Caracciolo, F., Furno, M., D'Amico, M., Califano, G., & Di Vita, G. (2022). Variety seeking behavior in the wine domain: A consumers segmentation using big data. Food quality and preference, 97, 104481. doi: 10.1016/j.foodqual.2021.104481.
[17] Shiffman, S., & Goldenson, N. I. (2023). Changes in dependence over one year among US adults who smoke cigarettes and switched completely or partially to use of the JUUL-brand electronic nicotine delivery system. Drug and Alcohol Dependence Reports, 6, 100137. doi: https://doi.org/10.1016/j.dadr.2023.100137.
[18] Cowie, G. A., Swift, E., Partos, T., & Borland, R. (2015). Quitting activity and tobacco brand switching: findings from the ITC-4 Country Survey. Australian and New Zealand Journal of Public Health, 39(2), 109-113. doi: https://doi.org/10.1111/1753-6405.12323.