Hindawi Mathematical Problems in Engineering Volume 2021, Article ID 6688750, 5 pages https://doi.org/10.1155/2021/6688750



Research Article

A Consumer Behavior Prediction Model Based on Multivariate Real-Time Sequence Analysis

Lin Guo , Ben Zhang , and Xin Zhao

¹School of Economics and Management, Changchun University of Science and Technology, Jilin 130022, China ²College of Engineering, Northeastern State University, Boston 02115, MA, USA

Correspondence should be addressed to Lin Guo; guolin@cust.edu.cn

Received 8 December 2020; Accepted 19 May 2021; Published 26 May 2021

Academic Editor: Tomas Balezentis

Copyright © 2021 Lin Guo et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the rapid development of online finance and social networks, a large amount of behavioral data is stored on the Internet, which can fully reflect the shopping tendencies and habits of real users. Using big data to analyze consumer behavior is more scientific and accurate than the traditional sampling survey method. Internet consumption behavior data are time series data. Therefore, this paper proposes a method of analyzing behavioral sequence data, which learns personal consumption interests and habits, and finally predicts payment behavior. The experiments compare the execution effect of different algorithms on multiple databases and verify the feasibility and effectiveness of the proposed algorithm SeqLearn.

1. Introduction

Internet e-commerce platforms contain a wealth of information about events, relationships, and attitudes. A series of technologies such as text mining, statistical theory, association analysis, and visualization are used to realize sentiment analysis, information extraction, and user influence analysis. Learning consumer behavior can help analyze the characteristics of consumers, the relationship between products, and so on. Therefore, it is a very valuable research topic to construct consumption structures according to the consumption behavior records of different consumers. The algorithm SeqLearn proposed in this paper extracts consumers' consumption interests and habits by analyzing the behavior data of e-commerce platforms and calculates the comprehensive probability to predict the next payment behavior. Forecast results can be used for product recommendations, advertising placement, and other applications.

Online shopping is a process in which users browse and search for related products to complete shopping or other related tasks in a virtual shopping environment. Payment decision is the integration and unity of consumer demand, motivation, activity, and afterthought. By analyzing the network data, the potential connections behind a series of

consumer behaviors can be mined out and the prediction function can be realized. The application of this method was quick and effective to identify and segment consumer groups and facilitated the mapping of the differences among these groups and the comparison of the consumption behavior expressed by consumers on different markets [1]. In the field of social network analysis, users are defined as nodes. Relationships between nodes can be abstractly represented by edges. Common techniques used to analyze relationships between nodes include causal inference [2-4] and Bayesian network [5, 6]. Different strategies are used for different purposes and needs, such as emotion-oriented analysis, information extraction, and user influence analysis. For example, Bollen analyzed the sentiment in tweets and found that the creation and dissemination of online information are closely related to stock market, future commodity prices, and major social events [7]; Asur of HP LABS successfully predicted the box office using Twitter data [8]. Based on the big data of social networks, González-Bailón [9] studied the dynamics of the protest movements and revealed its influences on political trends.

Group interaction, information dissemination, and other behaviors on the Internet can affect politics, economy, and society. Network data analysis methods are divided into

the following aspects: (1) analyzing the characteristics of nodes [10]; (2) analyzing relationships (such as advertising push relationship [11], intimate relationship [12], and other types of relationships [13, 14]) between nodes according to sociology [15–18]; (3) analyzing the law of information transmission; and (4) group behavior studies (such as topology analysis, normalization analysis, modular analysis, random data flow analysis, node clustering, and classification) [19–21].

The Internet has become a new media, which makes users have a great sense of participation. The convenient and efficient mode of communication is a driving force for Internet business to analyze and mine network data. A recommendation model based on statistical modelling is used to assist consumers facing choice overload by predicting their interests and consumption behaviors [22]. However, due to the complexity of network data, it is a huge challenge for any enterprise or operator to mine valuable information with commercial value efficiently and quickly. The development of the Internet has brought about tremendous changes in consumption concepts, consumption patterns, and consumer conditions.

In this paper, a large number of real network data are collected to support the analysis of consumer behavior. We study personal consumption interests and consumption habits and finally predict the next payment behaviors. Forecast results can be used for product recommendations, advertising placement, and other applications. In the experimental part, online consumer behavior data, such as Ali Data (an online shopping website in China), are used to compare the execution effects of different algorithms on different databases and verify the feasibility and effectiveness of SeqLearn.

The remainder of this paper is organized as follows. Section 2 proposes the predictive model of consumer behavior. Section 3 introduces the experimental results. Section 4 states the conclusions of research achievements.

2. Consumer Behavior Prediction Model

Online consumption data can reflect the actual consumption trends and consumption changes of users. This paper proposes a sequence analysis algorithm to study personal consumption preferences and consumption habits. Personal consumption preferences can be used to predict what consumers will buy over a period of time. Consumption habits reflect a time cycle, that is, how long it takes for consumers to pay attention to a certain commodity before they buy it.

The behaviors of Internet users can be divided into four categories: click, collect, add-to-cart, and payment. Among these four behaviors, payment data are considered outdated data, which only reflects the user's past shopping preferences and is used as validation data to verify the accuracy of the algorithm. Add-to-cart and collect behaviors are regarded as equal. They have the same effect on predicting shopping behavior. The amount of browsing behavior data is huge, and its correlation with user interests is weak. Experiments show that the accuracy of our prediction algorithm will not

be significantly improved by analyzing the browsing behavior data. Therefore, this paper does not consider browsing behavior data and obtains accurate analysis results with as little computational effort as possible. In summary, this paper only considers purchase behavior, add-to-cart behavior, and collect behavior. In order to verify the feasibility of the algorithm, we analyze a large amount of user data. It is found that these three behaviors are strongly correlated and often appear in a logical order. This proves that using only these three behavioral data to predict consumption trends is feasible.

In order to analyze the continuously updated behavioral data sequence and prevent the impact of stale data on analysis accuracy, the length of the dynamic sequence to be analyzed is selected as n. We transform the training sequence into a fixed-length sequence $s = (s_1, s_2, \ldots, s_n)$, where n represents the maximum length. If the sequence length is greater than n, then the nearest n actions are considered. If the sequence length is less than n, we will repeatedly add a "padding" item to the left until the length is n.

The most important task of sequential analysis is to learn each user's personal consumption preferences and consumption habits based on historical data. The prediction formula is shown as follows:

$$P\left(\widehat{i}, \widehat{a} \mid S_i^u, S_a^u\right) = P\left(\widehat{i} \mid S_i^u\right) \times P\left(\widehat{a} \mid S_a^u\right). \tag{1}$$

 $P(\hat{i}, \hat{a} | S_i^u, S_a^u)$ calculates the probability that user u performs action \hat{a} on item \hat{i} , where \hat{i} represents the predicted next item and \hat{a} represents the predicted user action. S_a^u represents the historical access record of user u. S_i^u represents the historical action sequence of user u.

The following sections introduce the formulas for learning personal consumption interests and habits.

2.1. Personal Consumption Preference Analysis. In a long time, the consumer's consumption trend is constantly changing, but in a short time, the consumption trend is stable. By analyzing user behavior over a period of time, we can find out consumers' consumption interests, which reflect the users' recent consumption trends. The following formula illustrates the process of quantifying user interests to predict user behaviors. $P(\hat{i} \mid S_i^u)$ in formula (1) reflects the consumption interest of a consumer over a period of time. The probability function formula is as follows:

$$P\left(\widehat{i} \mid S_i^u\right) = \frac{P\left(\widehat{i}, S_i^u\right)}{P\left(S_i^u\right)},\tag{2}$$

where S_i^u represents the sequence of access items for user i. If the length of the sequence is less than n, it is automatically filled. The number of occurrences of each item in the sequence is denoted by x. x corresponds to the Poisson distribution. x is related to the number of times users use online media and the amount of information generated [23], so $x \sim$ Poisson (λ), where $\lambda = ev \times c$, in which ev represents the number of times a user uses network media in time t and c represents the total amount of data generated in a certain

consumer interest area within time *t*. From this, the following probability density formula can be derived:

$$P(S_i^u) = \prod_{k=1}^n P\{\text{item}_k = x\} = \prod_{\substack{k=1 \ \text{item}_k = x}}^n e^{-ev \times c} \frac{ev \times c^x}{x!}.$$
 (3)

 $P(\hat{i}, S_i^u)$ in formula (2) represents the probability that \hat{i} and S_i^u appear at the same time. The specific formula is shown as follows:

$$P\left(\widehat{i}, S_i^u\right) = \frac{m+1}{n},\tag{4}$$

where n represents the sequence length and m represents the number of elements that appear in both S_i^u and the sequence being studied. Substituting formula (3) and formula (4) into formula (2), the following formula is obtained:

$$P\left(\hat{i} \mid S_i^u\right) = \frac{m+1}{n} \prod_{\substack{k=1 \\ \text{item}_k = x}}^n e^{-ev \times c} \frac{ev \times c^x}{x!}.$$
 (5)

2.2. Consumption Habit Analysis. The consumption habits of consumers are different, which can be roughly divided into impulse type and conservative type. Impulse consumers refer to consumers who purchase goods intuitively and impulsively under the influence of urgent purchase psychology. Conservative consumers have strict attitudes towards consumption. They are sensitive and wait-and-see attitudes towards products and are sensitive to prices. By comparing these two types of consumers, impulsive consumers will be influenced by the first suitable product and make quick purchase decisions instead of repeated selection and comparison. However, conservative consumers have long consumption cycles and need to compare many products before buying.

The consumption habits of any consumer are different. By studying historical data, the consumption habits of different consumers can be derived. The probability formula for predicting the next action of consumers is shown as follows:

$$P(\widehat{a} \mid S_a^u) = \frac{P(\widehat{a}, S_a^u)}{P(S^u)},\tag{6}$$

$$P(\widehat{a}, S_a^u) = \frac{t}{n}. (7)$$

Formula (7) calculates the frequency of purchase behavior, where n is the length of the sequence being studied and t is the number of purchases.

 $P(S_a^u)$ satisfies binomial distribution B (n, p), and let y be the number of times the purchase behavior A occurs in n user actions, so y is a random variable. y is equal to 0, 1, 2, ..., n. Set P(A) = p, $P(\overline{A}) = 1 - p$. So the distribution law of $P(S_a^u)$ is shown as follows:

$$P(S_a^u) = P\{S_a^u = k\} = C_n^k p^k (1 - p)^{n - k},$$
 (8)

where k is the number of times the purchase behavior occurs in sequence S_a^u . Then, formula (7) and formula (8) are substituted into formula (6), and the following formula can be obtained:

$$P(\widehat{a} \mid S_a^u) = \frac{t}{nC_n^k p^k (1-p)^{n-k}}, \quad k \in [0, n].$$
 (9)

Finally, by substituting formula (5) and formula (9) into formula (1), the following formula can be obtained:

$$P(\widehat{i}, \widehat{a} \mid S_{i}^{u}, S_{a}^{u}) = \frac{t(m+1)}{n^{2}C_{n}^{k}p^{k}(1-p)^{n-k}} \prod_{\substack{k=1 \ \text{iterm}_{k}=x}}^{n} e^{-ev\times c} \frac{ev\times c^{x}}{x!}.$$
(10)

3. Experiment

The parameters used in the experiment are shown in Table 1. Datasets used in the experiment are as follows:

- (1) Ali data (https://tianchi.aliyun.com/dataset/dataDetail?dataId = 46): this dataset is provided by Alibaba Group. It possesses a wealth of user data, such as user location information and access time. This dataset contains the following attributes: user ID, item IDs, types of behaviors (including click, collect, add-to-cart, and payment; the corresponding values are 1, 2, 3, and 4, respectively), locations, the category ID, and time.
- (2) EP data: it contains 15,890,209 pieces of data collected from http://www.dianping.com/ in August 2018. The dataset contains the following attributes: shop_id (unique), province, city, city_id, area, big_cate, big_cate_id, small_cate, small_cate_id, service_rating, all_remarks, very_good_remarks, good_remarks, common_remarks, bad_remarks, and very_bad_remarks.

The algorithms used in the experiment are as follows:

- (1) Statistical learning: statistical learning techniques are the tools we use to understand data, which are divided into supervised learning and unsupervised learning. Broadly speaking, statistical learning builds statistical models based on one or more inputs to predict or estimate outputs.
- (2) EC-Structure [23]: this algorithm grabs consumption data of the e-commerce platform, analyzes consumption structures, and predicts consumption behaviors according to consumption structures. The algorithm integrates multiple dimensions of network data to comprehensively study consumer behavior.

In order to show the relationship between user behavior and payment behavior, describes the behavior data of an Ali

TARIF	1.	Parameter	description.

This is a second to the second							
The number of behaviors that are correctly predicted							
The number of the correctly predicted behaviors that will not occur							
The number of incorrectly predicted behaviors that will not occur							
The number of incorrectly predicted consumption behaviors that will not occur							
A parameter to measure the accuracy of prediction results							
Precision							
Recall							
Synthesis of P and R							

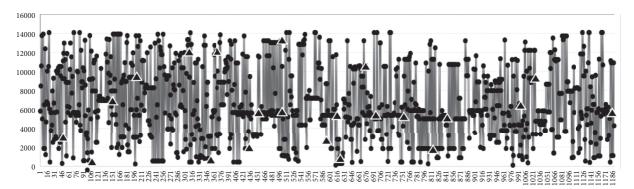


FIGURE 1: Predict consumer behavior based on Ali data.

Table 2: Confusion matrix (Top@10).

		Ac	tual
		Positive	Negative
Predicted	Positive	132910	20137
	Negative	28193	213461

Table 3: Comparison of algorithm execution effects.

Dataset T	T OM	Statistical algorithm			EC-Structure			SeqLearn					
	Top@N	P	R	FB	ACC	P	R	FB	ACC	P	R	FB	ACC
Ali data	Top@5	0.64	0.70	0.67	0.60	0.78	0.70	0.74	0.75	0.80	0.81	0.80	0.77
	Top@10	0.72	0.77	0.74	0.70	0.83	0.82	0.82	0.80	0.86	0.88	0.87	0.85
EP data	Top@5	0.62	0.60	0.61	0.61	0.81	0.70	0.75	0.80	0.82	0.86	0.84	0.81
	Top@10	0.73	0.77	0.75	0.72	0.83	0.85	0.84	0.81	0.88	0.90	0.89	0.87

user for one month. It visually shows the sequential relationship between different user behaviors.

In Figure 1, the abscissa indicates the order of appearance of the behavior data, and the ordinate indicates the consumption category number. The dots in Figure 1 represent the behaviors of collect, add-to-car, and payment. The triangles represent the predicted payment behaviors based on historical data. It can be seen from Figure 1 that there is a certain logical connection between users' online behavior and payment behavior, and it is verified that the method of integrating consumer interests and consumption habits can accurately predict the next payment behavior.

Table 2 illustrates the accuracy of the consumer behavior prediction algorithm SeqLearn (Top@N: among the results calculated by the algorithm, the first N are selected for precision analysis). The horizontal quantity represents the

actual value, and the vertical quantity represents the predicted value. In order to measure the accuracy of the algorithm, the following indicators are selected for measurement: precision P = tp/(tp+fp); recall R = tp/(tp+fn); FB parameters FB = $P \times R \times 2/(P+R)$, where tp is the number of behaviors that are correctly predicted, tn is the number of the correctly predicted behaviors that will not occur, fn is the number of incorrectly predicted behaviors that will not occur, and fp is the number of incorrectly predicted consumption behaviors that will not occur. In addition, parameter ACC = tp/(tp+tn+fn+fp) is used to measure the prediction accuracy of the algorithms.

Table 3 displays the performance comparison of statistical learning, EC-Structure, and SeqLearn using different datasets. According to Table 3, we can conclude that SeqLearn has the most stable execution effect and the best results according to FB and ACC.

4. Conclusion

If a questionnaire survey is used for research, problems such as low sample coverage, long data survey cycle, and lagging survey results will arise. In this paper, a large number of real network data is collected to support the analysis of consumer behavior. The results of this method are accurate. Internet consumption behavior data are time series data. Therefore, this paper adopts the method of time series analysis to study personal consumption interests and consumption habits and finally predicts next payment behaviors. The prediction results can be used in areas such as product recommendation and advertising push. The experiment shows the process of predicting consumer behavior and verifies the feasibility and efficiency of the SeqLearn algorithm on multiple datasets. In the future research, we hope to study the characteristics of long-term consumption and short-term consumption, so that we can accurately calculate the parameters of different types of goods in the recommendation method.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by Youth Program of The National Social Science Fund of China (Project name: Research on online behavior pattern of customers and multidimensional customer insight method under big data; Grant no.19CGL024).

References

- [1] K. Migda-Najman, K. Najman, and S. Badowska, "The GNG neural network in analyzing consumer behaviour patterns: empirical research on a purchasing behaviour processes realized by the elderly consumers," *Advances in Data Analysis and Classification*, vol. 14, pp. 1–36, 2020.
- [2] J. Pearl, "Causal inference in statistics: an overview," *Statistics Surveys*, vol. 3, pp. 96–146, 2009.
- [3] A. Belloni, V. Chernozhukov, I. Fernndez-Val, and C. Hansen, "Program evaluation and causal inference with high-dimensional data," *Econometrica*, vol. 85, no. 1, pp. 233–298, 2017.
- [4] G. Hong, "A review of explanation in causal inference: methods for mediation and interaction," *Journal of Educational and Behavioral Statistics*, vol. 42, pp. 34–53, 2017.
- [5] S. Mei, J. Zhu, and X. Zhu, "Robust regbayes: selectively incorporating first-order logic domain knowledge into Bayesian models," in *Proceedings of the International Con*ference on Machine Learning, Beijing, China, June 2014.
- [6] A. M. Oliveira and O. L. V. Costa, "An iterative approach for the discrete-time dynamic control of Markov jump linear systems with partial information," *International Journal of Robust and Nonlinear Control*, vol. 30, no. 2, 2020.

- [7] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *Journal of Computational Science*, vol. 2, no. 1, pp. 1–8, 2011.
- [8] S. Asur and B. A. Huberman, "Predicting the future with social media," in *Proceedings of the Web Intelligence and Intelligent Agent Technology(WI-IAT)*, IEEE, Toronto, Canada, September 2010.
- [9] S. González-Bailón, J. Borge-Holthoefer, A. Rivero et al., "The dynamics of protest recruitment through an online network," *Scientific Reports*, vol. 1, p. 197, 2011.
- [10] L. Backstrom and J. Kleinberg, "Romantic partnerships and the dispersion of social ties:a network analysis of relationship status on facebook," in *Proceedings of the 17th ACM Con*ference on Computer Supported Cooperative Work and Social Computing, pp. 831–841, Baltimore, MA, USA, February 2014.
- [11] C. Wang, J. Han, and Y. Jia, "Mining advisor-advisee relationships from research publication networks," in *Proceedings of the 16th ACMSIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 203–212, Washington, DC, USA, July 2010.
- [12] L. Aggeborn, N. Lajevardi, K.-O. Lindgren, P. Nyman, and S. Oskarsson, "Parents, peers, and politics: the long-term effects of vertical social ties," *Quarterly Journal of Political Science*, vol. 15, no. 2, pp. 221–253, 2020.
- [13] Y. Sun, R. Barber, and M. Gupta, "Co-author relationship prediction in heterogeneous bibliographic networks," in *Proceedings* of the International Conference on Advances in Social Networks Analysis and Mining, pp. 121–128, Kaohsiung, Taiwan, July 2011.
- [14] P. Rozenshtein, N. Tatti, and A. Gionis, "Inferring the strength of social ties: a community-driven approach," *Data Structures* and Algorithms, 2019, https://arxiv.org/abs/1902.01832.
- [15] K. Kamath, K. Kamath, and A. Sharma, "Detecting strong ties using network motifs," in *Proceedings of the International Conference on World Wide Web Companion*, pp. 983–992, Perth, Australia, April 2017.
- [16] K. Farrahi and K. Zia, "Trust reality-mining: evidencing the role of friendship for trust diffusion," *Human-centric Com*puting and Information Sciences, vol. 7, no. 1, 2017.
- [17] S. Otte, T. Schmitt, and K. Friston, "Inferring adaptive goal-directed behavior within recurrent neural networks," in Proceedings of the International Conference on Artificial Neural Networks, Alghero, Italy, September 2017.
- [18] R. Azizur, Q. Tie, and N. Zhaolong, "Social acquaintance based routing in vehicular social networks," *Future Generation Computer Systems*, vol. 93, pp. 751–760, 2019.
- [19] X. Li, B. Kao, Z. Ren, and D. Yin, "Spectral clustering in heterogeneous information networks," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 1, pp. 4221–4228, Honolulu, HI, USA, January 2019.
- [20] Y. Song, J. Lu, H. Lu, and G. Zhang, "Fuzzy clustering-based adaptive regression for drifting data streams," *IEEE Transactions on Fuzzy Systems*, vol. 28, no. 3, pp. 544–557, 2020.
- [21] L. Ma, J. Li, Q. Lin, M. Gong, C. A. Coello Coello, and Z. Ming, "Reliable link inference for network data with community structures," *IEEE Transactions on Cybernetics*, vol. 49, no. 9, pp. 3347–3361, 2019.
- [22] M. R. Haddad and H. Baazaoui, "An adaptive and interactive recommendation model for consumers' behaviours prediction," *International Journal of Data Mining, Modelling and Management*, vol. 10, no. 1, p. 89, 2018.
- [23] L. Guo and D. Z. Zhang, "EC-Structure: Establishing consumption structure through mining e-commerce data to discover consumption upgrade," *Complexity*, vol. 2019, Article ID 6543590, 8 pages, 2019.