Consumer Behavior Prediction Based on Machine Learning Scenarios

Zhengyi Hu(✉)

SHU Business School, Sheffield Hallam University, Sheffield, UK
c1052106@my.shu.ac.uk

Abstract. Consumer behavior prediction is an important part of the business analysis and strategy development process. With the development of AI technology, machine learning has been also applied to consumer behavior prediction. This paper introduces some common and novel consumer behavior prediction methods, including decision tree algorithm, recurrent neural networks (RNNs) algorithm, and the application of fMRI in machine learning. The advantages and limitations of these methods are clarified by analyzing the practical cases based on these methods. This paper first introduces the background of consumer behavior prediction and review some existing prediction methods. Subsequently an introduction and review of the three selected prediction methods and the cases they participated in are demonstrated accordingly. Afterwards, the advantages of the three methods and the drawbacks of them are evaluated by analyzing the three cases. Then, some measures are proposed to improve the performance of the models to pave the path for future development of consumer behavior prediction in terms of machine learning supports. These analytical results offer guideline for the future development of consumer behavior prediction.

Keywords: Consumer Behavior Prediction · Machine Learning · Decision Tree Algorithm · Recurrent Neural Networks · FMRI

1 Introduction

"Consumer behavior" describes the journey a consumer goes on as they research, select, and buy a product or service. Enterprises need to know why customers choose to buy or reject their products, which is helpful to better design the own products, so as to achieve the goal of improving profits and the development of the company. This is important for gaining profits, so consumer behavior is required to be predicted. Culture, education, life style and many other things affect consumer behavior. The prediction of consumer behavior by the selection of key factors is the core of this project, namely, the screening process and the analysis process.

For each enterprise, the factors most closely connected to their consumers were different, hence the main discussion of this paper is the method of consumer behavior prediction. The most traditional method is to interview the consumers of enterprises and the customer service departments, so as to get the customer needs. Nevertheless,

D. Qiu et al. (Eds.): ICBBEM 2022, AHIS 5, pp. 410–419, 2023.
https://doi.org/10.2991/978-94-6463-030-5_43

sometimes consumers are unable to describe their demanding exactly, and maybe they can choose the one they prefer if companies offer choices. With the development of AI, machine learning has become the main analysis tool, but the alternative methods are still difficult to count, and some may have only very small differences. In this paper, three more typical analysis methods are selected for discussion.

The oldest "method for predicting consumer behavior" was developed by Howard in 1963, which was the first consumer decision-mode [4]. This model was developed further in 1969 by Howard and Sheth to become the 'Theory of Buyer Behaviour' (or Howard and Sheth Model) [6]. This model has systematically proposed methods to predict consumer behavior for the first time, providing analytical methods in response to various behaviors of consumer sociology, psychology, and operational means such as advertising.

The decision tree algorithm comes from the notion of a single tree. The age of this concept is almost comparable to the age of consumer behavior analysis, and there were "how to simplify the decision tree" published back in 1987 [14]. Decision tree algorithms are the product of referring this concept in machine learning. Its core is screening for factors that may affect the final results and ranking them by importance, which is the easiest analytical algorithm to understand and learn.

Recurrent neural networks (RNN or RNNs) are a kind of artificial neural networks where connections between nodes form a directed or undirected graph along a temporal sequence. Recurrent neural networks are theoretically Turing complete and can run arbitrary programs to process arbitrary sequences of inputs, which means it is a completed artificial intelligence product. Therefore, RNNs has become one of the most commonly used analytical algorithms. Due to its high popularity, many scholars have improved it in different directions, including second order RNNs, independently RNNs (IndRNNs), long short-term memory (LSTM) networks.

In addition to the above business analysis/machine learning models, knowledge of neurology has also been developed to make consumer behavior prediction. By directly measuring the consumer brain response to different products, the method most intuitively gives accurate conclusions, which was the application of functional magnetic resonance imaging (fMRI) in machine learning. However, if successful consumer behavior predictions are to be implemented, researchers need to determine the meaning of those activities in the brain and whether "scanning the instrument" itself can affect consumer minds and ultimately lead to biased results. The idea was proposed within two decades and has not been fully developed which makes it still a "novel" approach.

The rest part of the paper is organized as follows. The Sect. 2 will basically introduce the state-of-art machine learning scenarios. Subsequently, the application of three selected models will be demonstrated one by one. Afterwards, the features and characteristics of the above models will be discussed and a future outlook will be presented. Eventually, a brief summary will be given in Sect. 5.

2 Machine Learning Approaches

2.1 Decision Tree Algorithm

Decision tree algorithm is an algorithm that appeared in 1963 [11], which is a structure model of tree classification and can be applied in business analysis, machine learning and many areas else. The decision tree algorithm establishes data analysis rules by splitting the variable value and forming the conceptual path through the tree structure. The mathematical descriptions of the decision trees are given as follows. The output function is

$$f(x) = \sum_{m=1}^{M} C_m I(x \in R_m) \tag{1}$$

where

$$C_m = ave\{y|x \in R_m\} \tag{2}$$

The rationale of the decision tree algorithm: when using the decision tree algorithm, first needs to create a model, generating a series of bifurcation trees, each fork represents a node, adding a node when the input variable is found to have a significant correlation with the predictive variable.

For example, consumers of different ages have different possibilities to buy a certain product, and the decision tree algorithm model needs to increase the nodes according to the age. Overall, the decision tree algorithm is to classify different information and can be predicted based on existing experience when variables similar to the historical data appear. Whether the decision tree algorithm is effective depends on the correct judgment of the split properties. For example, for crutches, splitting by age is obviously more appropriate than splitting by consumer city.

2.2 Recurrent Neural Networks

There is a fully machine learning-based consumer behavior prediction model that was born in 1993 [16], which evolved from David Rumelhart's work in 1986 [18]. It has solved a very large "Very Deep Learning" task as its first task, which has required more than 1,000 layers at the runtime. Besides, it was impossible for traditional consumer behavior analysis methods that were not using machine learning.

The structure of RNNs is relatively simple, and RNNs models can be constructed through the input layer-hidden layer-output layer. The biggest difference between the RNNs and the traditional holy boundary network is that the previous output is each brought to the next hidden layer for machine learning together.

In contrast to vector-based methods, RNNs took sequences $X = (x_1, \ldots, x_t)$ of varying length T directly as the input. RNNs were established as the junction sequence of the computing cells. The cells in step t accepted the input x_t and kept a hidden state $h_t \in \mathrm{R}^{\mathrm{d}}$. This hidden state would be calculated from the cell state of the input x_t and the cell state at the former time-step h_{t-1} as:

$$h_t = \sigma(W_x x_t + W_h h_{t-1} + b) \tag{3}$$

where W_x and W_h were learned weight matrices, b was a learned bias vector and σ was the sigmoid function. A hidden state h_t captures information from the input sequence ($x_1,\ldots, x_t$) up to the current time-step t. The calculation formula of the RNNs algorithm has used these parameters to preserve those earlier data. This allowed the RNNs calculation formula to analyze the time series data, which could give it completely one more "dimension" than the other algorithms. The dimension d of the hidden state was a hyperparameter chosen based on the complexity of the temporal dynamics of the scene.

2.3 fMRI in ML

Advances in machine learning as applied to functional magnetic resonance imaging (fMRI) data offer the possibility of pretesting and classifying marketing communications using unbiased pattern recognition algorithms [2].

The core of this method is measuring the human brain to determine the attitude of subjects to a certain condition through the signal transmitted by the brain. In predicting consumer behavior, measuring subjects' brain activity directly is much more directly measured than inviting them to fill out the questionnaire, and analysis combined with machine learning yields more accurate prediction results.

The fMRI uses strong magnetic fields to create images of the brain to reflect the brain response of the scanner for different content, thus achieving an accurate analysis of consumer behavior and further completing the prediction. The magnetic field in the fMRI scanner does not change over time, so it can be called "static", with a Tesla as the unit of measurement. A typical fMRI scanner produces a static magnetic field intensity of 1.5 to 3.0 T, while the experimental scanner produces a stronger magnetic field intensity of 7.0 T or more [15]. Human beings are known to contain about 70% of water and are higher in the brain. Strong magnetic field can stimulate water to achieve the purpose of "line into an image". Outside the fMRI scanner, the hydrogen atoms in the scanned body rotate in a random direction. Once the scanned person enters a static magnetic field formed by the test instrument, the hydrogen atoms are drawn in a uniform direction.

The Blood oxygen level-dependent (BOLD) signal, which was also known as the hemodynamic response function (HRF), was one of the key effects observed in fMRI [5]. Generally speaking, the instrument emitted a series of unseen pulses to analyze the brain activity in a way similar to echolocation. The main contents analyzed were oxygenated and deoxyated blood.

The results of pure fMRI were just images of the human brain, which could not be analyzed and get the results that the researchers need, but there would be a great deference if they introduced machine learning into fMRI. Through the development of neurology, it is now much easier to determine which parts of the human brain and which activities represent the different emotions of the participants. In this method, researchers can judge the emotional responses of the participants when seeing different products, whether they are happy or disgusted; whether it is a product already introduced in the market, or a fictional product that still has only one concept.

3 Application

3.1 Decision Tree Algorithm

The study conducted by Li et al. in 2019 has used a decision tree algorithm, including data mining techniques and three data analysis methods such as decision tree analysis [9]. They used data-mining techniques to analyze their products' consumer behavior predictions. In this case, the decision tree algorithm was one of the mainly used algorithms that have successfully analyzed the common characteristics of those target groups that were likely to perform consumption behavior. The researchers also made a simple analysis and comparison between the three algorithms used in the study.

These researchers compared and analyzed the effectiveness of the three consumer behavior prediction algorithms they had chosen and the accuracy of the prediction results, thus enabling businesses to understand what services were targeted and attractive to their consumers, to further meet their personal needs and improve customer satisfaction. The steps of their work have revealed that consumer behavior analysis was mainly an analysis of past experience. In classifying and predicting consumer groups, practical results show that the decision tree model was the best practical and was significantly better than the other two other algorithms chosen by the researcher (cluster analysis and naive Bayesian model). The prediction results had provided practical information for the production and sales of their enterprises, and had provided research ideas to meet the personalized needs of the users. It still has a broad market application prospect that though the decision tree algorithm has a long history (Fig. 1).

3.2 RNNs

Lang & Rettenmeier performed the data analysis using RNNs algorithms in 2017 and predicted consumer behavior, finding some advantages of RNNs algorithms [8]. The

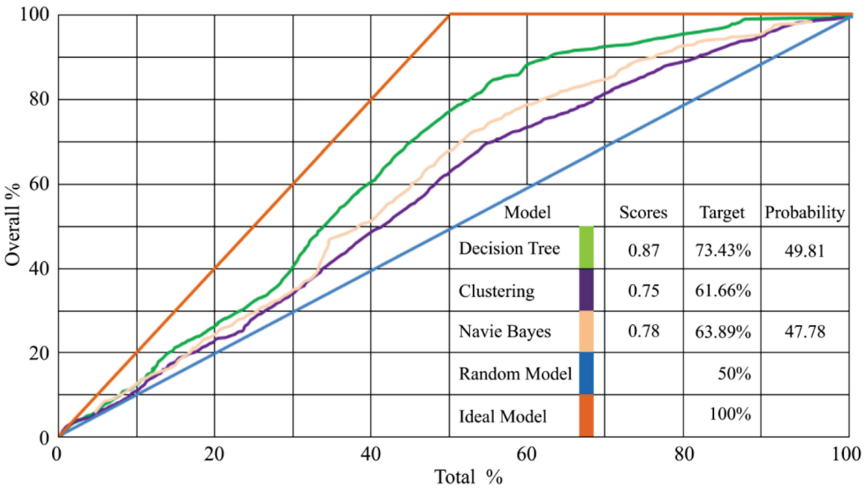

Fig. 1. Lifting chart for different models

researchers believed that RNNs algorithm was the inevitable consequence of using machine learning for consumer behavior analysis. The modeling method of the RNNs algorithm was of an unsurpassed advantage. Based on the results, it can be found that in contrast of the RNNs algorithm they performed with the actual production system, the RNNs algorithm could analyze data more comprehensive and can be processed in chronological order.

Direct application of RNNs algorithm to consumer behavior sequences can yield the same or higher prediction accuracy as the vector-based methods such as logistic regression. Unlike the latter, the application of the RNNs algorithm does not require substantial feature engineering. Furthermore, it is found that RNNs algorithm can help us directly relate individual behavior to prediction in an intuitive way. This allows researchers to understand the impact of consumer behavior on the predictive probability in the historical development of consumers. They demonstrate the advantages of RNNs algorithm on empirical data from a large online fashion platform in Europe.

They proposed an approach to apply RNNs to predict future consumer behavior in e-commerce. Consumer behavior was inherently sequential which made RNNs a perfect fit. They were employing RNNs in production which offered significant advantages over existing methods: reduced feature engineering; improved empirical performance; and better prediction explanations.

3.3 fMRI

Deppe et al. applied fMRI to machine learning to study how individual economic decisions are affected by their implicit memory contributions [3]. Twenty-two participants were asked to make binary decisions between almost indistinguishable different brands of consumer goods. Functional magnetic resonance imaging (FMRI) is used to detect changes in brain activity in the presence or absence of a specific target brand for comparative decisions (Fig. 2).

Only when the target brand was the participant's favorite one, did the authors find reduced activation in the dorsolateral prefrontal, posterior parietal, and occipital cortices and the left premotor area (Brodmann areas [BA] 9, 46, 7/19, and 6). Simultaneously, activity was increased in the inferior precuneus and posterior cingulate (BA 7), right superior frontal gyrus (BA 10), right supramarginal gyrus (BA 40), and, most pronounced, in the ventromedial prefrontal cortex (BA 10).

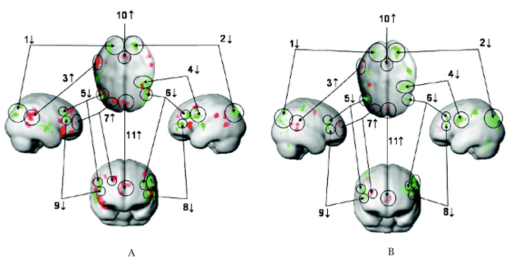

Fig. 2. The different activities of the (A) male's brain and the (B) female's brain for the emotion of "favorite brand" were found in the research.

They revealed that participants' favorite brands had nonlinear winner-take-all effects for products that were distinguished primarily by brand information, with on the one hand reduced activation in brain regions associated with working memory and reasoning, and increased activation in areas related to processing emotion and self-reflection during decision-making.

4 Limitation and Future Perspective

4.1 Advantages and Limitations

As for decision tree algorithm, by reviewing the above research, Customer segmentation is mainly based on company history data. However, as mentioned previously, the factors influencing consumer behavior are highly diverse, so the research conducted by Li et al. was not comprehensive. Besides, it was also the problem of many existed cases of decision tree algorithm, and diversification should also be the focus of the future development of decision tree algorithm. As a matter of fact, decision tree algorithm possesses lots of advantages. Primarily, it is suitable for classifying non-linear problems, with unfixed model structure and higher flexibility; for the influence order of attribute, the nodes closer to the starting point; for various types of predictive variables, and when the data is large, they generally will not greatly affect the overall structure of the decision tree if there are individual outliers present in the variable; the rule expression of the algorithm was relatively complete, clear concept, easy to explain. On the other hand, there are also drawbacks for such an algorithm. Decision tree algorithm was unstable. When splitting the decision tree, the location of each node needs to be accurate, not leading to completely different results; more difficult to process continuous data; to process them, prior to convert them into discrete data before establishing the decision tree; insufficient ability to process missing value data; unable to obtain rules for multi-feature combination (Fig. 3).

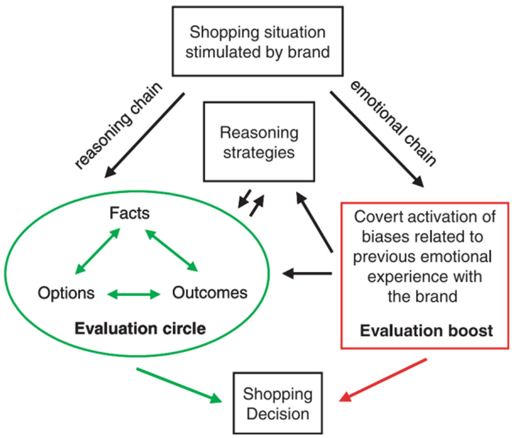

Fig. 3. Diagram of the hypothesized interactions involved in buying decisions

Regarding for RNNs algorithm, it can handle input variables of any length; the model structure is stable and will not change with the input variables; historical information will be used in the calculation and sort the data in time; the importance of information is related to time, enabling the algorithm can progress with the data. However, RNN algorithms are slow; archaic information may not be obtained; any future input for the current time cannot be considered; a gradient disappearance problem occurs while LSTM has been developed to partially address the issue.

For fMRI, existing neurological techniques can provide clearer images for machine learning and further improve the accuracy of the analysis results, which is also designed to gain a more accurate understanding of the participants' psychological activities. The fMRI in a machine learning environment can locate changes in consumer choice and brain activity during consumption to achieve accurate prediction of consumer behavior. This principle makes the fMRI in a machine learning environment reliable and valid measure for cognitive and affective responses able to detect changes in chemical composition or changes in the flow of fluids in the brain [17]; fMRI is a non-invasive method. Nevertheless, its should be noted that fMRI do have a lot of drawbacks. The limitations of fMRI are also obvious, and the most notable point is its high cost, not only to cost each participant to test, but also to cost the purchase and maintenance costs of analytical equipment. Operating costs are around 80.000–200.000 € per year [13], and the analysis cost around 100–50 € per subject [1]. This drawback leads to another limitation: its sample size is much smaller than other algorithms, (e.g., RNNs algorithms), and its sample size is even lower than the number of layers that the first task of the RNNs algorithm has. Moreover, fMRI also has the most common disadvantage of MRI, which is the instability of the participants. Participants have the potential to have additional emotions about the instrument, such as tension and fear, which will lead to fluctuations in the test results, in addition, because the MRI machine works, it is obviously unable to make the body containing metal parts into the participants, which to some extent reduces the range of fMRI in consumer behavior prediction. The way the fMRI machine operates also leads to its situational functionality and products that require actual use to get correct feedback cannot be accurately expressed by fMRI results. The next drawback is that the data analysis of fMRI is extremely complex, which is not a simple process of analyzing inputs, which is an analysis process of the human brain, which is interconnected with the "expensive" disadvantage mentioned earlier. The last drawback is that this analytical approach may have ethical hazards, such as whether participants would feel that fMRI violated their privacy.

4.2 Future Perspective

As for decision tree algorithm, ensemble learning will significantly boost the performance. Using a combination of multiple weak learning methods and making multiple classification models, i.e., multiple decision trees (analogy to multiple expert voting), to obtain better classification results than the original method. To improve the classification effect of the decision tree, K fold cross check, bagging method, lifting method, random forest can be applied. To improve feature accuracy, the core of the decision tree algorithm is the choice of key features. The existing methods are to learn from the feature selection methods once used. If this problem is to be improved, it can be achieved by the detailed definition of product attributes.

With regard to RNNs algorithm, Long Short-Term Memory Networks (LSTM) is an upgraded version of it, which has higher control-ability and better results [12]. The biggest improvement made by LSTM on the basis of RNNs algorithm is to solve the problem of gradient disappearance, using the concept of three gates: forgetting gate, input gate, and output gate. In other words, this model can filter out the irrelevant information and forget it, but as long as the front and forth information is closely related.

Additionally, Bidirectional RNNs (BRNNs) is also a further development of RNNs. BRNNs network structure enables that both the previous and subsequent information can be obtained at a point in the sequence [7]. However, this leads to another drawback that BRNNs can only predict after all the data is entered.

For fMRI, increased spatial resolution is pretty crucial. Increasing spatial resolution is an obvious approach to obtain more detailed fMRI data, which might additionally help to shed some light on the fine-grained functional organization of small areas in a living brain [10]. For computational modeling of multi-sensor processing, a series of fMRI experimental data from our group provides a possible role for brain regions involved in alphanumeric integration and the flow of information between these regions.

5 Conclusion

In summary, this paper investigates consumer behavior prediction based on machine learning approach. Specifically, with the development of artificial intelligence and Internet technology, machine learning has become the primary means of consumer behavior prediction. This paper has introduced a variety of data analysis algorithms in different eras, then referenced and analyzed the practical applications of three typical algorithms. Decision tree algorithm is a very versatile model but has many unstable factors and needs to be manually adjusted repeatedly. The RNNs algorithm is a widely used and expanded algorithm model that can sort the data according to time, but the way the data is read needs to be strengthened. The application of fMRI in machine learning is just beginning, which is promising to become the most accurate consumer behavior analysis model with the development of neurology technology, but there are still fatal shortcomings that are difficult to obtain for analytic samples. A common problem with these algorithms is the inability to achieve perfect automation, and there are many links to be manually adjusted, which hinders their further development. These problems will be solved with the development of AI technology. Overall, the main advantages and defects of the three typical state-of-art analysis methods are demonstrated, which provides a reference for the path to choose appropriate machine learning algorithms for consumer behavior prediction.

References

1. Ariely, D. & Berns, G. “Neuromarketing: the hope and hype of neuroimaging in business”. (Nature Reviews Neuroscience, 11(4), 2010) pp. 284–292.
2. Calvert, & Brammer, M. J. “Predicting Consumer Behavior: Using Novel Mind-Reading Approaches”. (IEEE Pulse, 3(3), 2012). pp. 38–41.

3. Deppe, Schwindt, W., Kugel, H., Plaßmann, H., & Kenning, P. "Nonlinear responses within the medial prefrontal cortex reveal when specific implicit information influences economic decision making". (Journal of Neuroimaging, 15(2), 2005). pp. 171–182.
4. Du Plessis, PJ & Rousseau, GG & Blem, NH. "Consumer behaviour. A South African perspective". (Pretoria. Sigma, 1991).
5. G. Reimann, Schilke, O., Weber, B., Neuhaus, C., & Zaichkowsky, J. "Functional magnetic resonance imaging in consumer research: A review and application: fMRI in Consumer Research". (Psychology & Marketing, 28(6), 2011). pp. 608–637.
6. Howard, J. A. and Sheth, J. N. "The theory of buyer behavior," (John Wiley and Sons, Inc., 1969).
7. Jason Brownlee. "A Gentle Introduction to Long Short-Term Memory Networks by the Experts". (DeepAI, 2017).
8. Lang, T., & Rettenmeier, M. "Understanding consumer behavior with recurrent neural networks". (In Workshop on Machine Learning Methods for Recommender Systems, 2017).
9. Li, Pan, S., Huang, L., & Zhu, X. "A machine learning based method for customer behavior prediction." (Tehnički Vjesnik, 26(6), 2019) pp. 1670–1676.
10. Logothetis, Nikos K. "What we can do and what we cannot do with fMRI." (Nature 453, (7197), 2008) pp. 869–878.
11. Morgan, J. N., & Sonquist, J. A. "Problems in the analysis of survey data, and a proposal." (Journal of the American statistical association, 1963). 58(302), 415–434.
12. Palangi H and others "Deep Sentence Embedding Using Long Short-Term Memory Networks: Analysis and Application to Information Retrieval". (IEEE/ACM Transactions on Audio, Speech, and Language Processing, 24(4), 2016). pp. 694–707.
13. Plassmann, H., Ramsøy, T.Z., Milosavljevic, M. "Faculty and Research Working Paper: Branding the Brain - A Critical Review". (INSEAD The Business School of the World, 2011).
14. Quinlan, J. R. "Simplifying decision trees". (International Journal of Man-Machine Studies, 1987) 27 (3): 221–234.
15. Rohit Sharma, et al. "Tesla (SI unit)". (Radiopaedia, Web, 2018).
16. Schmidhuber, J. "A neural network that embeds its own meta-levels. In IEEE International Conference on Neural Networks" (IEEE, 1993). pp. 407–412.
17. Wang, Y.J. & Minor, M.S. "Validity, Reliability and Applicability of Psychophysiolgical Techniques in Marketing Research". (Psychology & Marketing, 25(2), 2008), 197–232.
18. Williams, Ronald J. & Hinton, Geoffrey E., Rumelhart, David E. "Learning representations by back-propagating errors". (Nature 323 (6088), 1986). pp. 533–536.