Brand Loyalty Measurement Model Based on Machine Learning Clustering Algorithm

Yueqiu Li*, Chunming You
Heihe University, Heilongjiang, Heihe, China

*Corresponding author: Yueqiu@163.com

Abstract. In the "new economy" represented by the Internet economy background, online consumption as the representative of the new forms of consumption is gradually changing people's consumption idea and way, online brand loyalty has the extremely important status in the field of online consumption, to stimulate consumption and achieve accurate enterprise marketing, risk control and decision support, improve efficiency and product design business model, business forms, and even change of business thinking, improve enterprise competitiveness is of great significance in the field of online. It is urgent and necessary to apply scientific and effective machine learning method to systematically analyse and study online brand loyalty. In the big data environment, facing the massive data information provided by online consumption, traditional technology methods have gradually failed to meet the competitive needs of enterprises to create and maintain brand loyalty. The traditional random sampling method is difficult to locate the consumer groups with high brand loyalty. At the same time, the traditional data processing technology can not deal with the online consumption behavior with the characteristics of massive, mixed and unstructured data. Traditional approaches have limitations when it comes to the sheer volume of online data and how to use it in real time to target the needs of a brand's consumer group. The purpose of this study is to build an online consumption of the era of big data model of artificial intelligence, machine learning model, by machine learning method, branded goods purchase behavior of consumers online clustering, building an online brand loyalty measurement model, achieve similar loyalty user clustering, at the same time realize the online measurement of brand loyalty. Among them, focus on machine learning path, machine learning algorithm, model construction methods, and model verification and optimization methods.

Keywords: Machine learning; clustering algorithm; Brand loyalty; Measurement model

1. Introduction
The modern commercial consumption economy characterized by online consumption big data analysis and processing needs the support of a series of advanced technologies [1]. The value of online consumption of big data does not lie in the massive nature of data information, but in the professional processing of valuable massive data. Among them, machine learning is an important technology for
online consumption of big data processing, is an interdisciplinary discipline involving probability theory, statistics, mathematics and computer algorithms and other fields. It can simulate and realize human learning behavior to acquire new knowledge and skills, and provide model methods for solving other similar problems by reorganizing the knowledge structure to continuously improve its own performance [2]. Due to the sheer volume of data online consumption big and big, fast processing speed, data structure, variety and huge value, at the same time, a large online consumption data also has high dimensions and data processing methods diversity, higher characteristic, data analysis, the complexity of algorithm makes the application of machine learning online consumption big data has irreplaceable advantages [3].

Online brand loyalty plays an extremely important role in the field of online consumer marketing, which plays an extremely important role in stimulating consumption and enhancing the competitiveness of enterprises in the field of online consumer marketing. Therefore, under the background of "new economy" represented by Internet economy, it is urgent and necessary to apply scientific and effective machine learning methods to systematically analyse and study online brand loyalty [4].

In the context of big data, faced with the massive data information provided by online consumption, traditional technical methods have gradually been unable to meet the competitive needs of enterprises to create and maintain brand loyalty [5]. The traditional random sampling method is difficult to locate the consumer groups with high brand loyalty. At the same time, the traditional data processing technology cannot deal with the online consumption behavior with the characteristics of massive, mixed and unstructured data [6]. Manually-driven approaches are difficult to cope with when there is a huge amount of online data and how to use it in real time to target the needs of a brand's consumer group. The purpose of this study is to build an artificial intelligence model - machine learning model in the era of online consumption big data. By using machine learning method, the brand purchasing behavior of online consumers is clustered, the users with similar loyalty are clustered, and the online brand loyalty is measured. Among them, focus on machine learning path, machine learning algorithm, model construction method, and model inspection and optimization method.

2. Algorithm and Code
The logic of theoretical research is that firstly, through analysing the characteristics of online brand loyalty big data, the necessity and advantages of applying machine learning method to online brand loyalty big data processing are summarized. Then through the analysis of online brand loyalty big data processing methods, summarize machine learning related technologies and methods; Finally, the machine learning path of online brand loyalty measurement is sorted out. The logic of model design is to follow the machine learning path of online brand loyalty measurement in theoretical research. Firstly, the online data of brand loyalty measurement is collected. Secondly, the online data of brand loyalty measurement is cleaned. Thirdly, the characteristics and models of brand loyalty measurement are constructed. Finally, the brand loyalty measurement model is tested. The logic of model optimization is based on model verification in model design to optimize the model. Firstly, optimize data processing; then feature engineering optimization is carried out. Finally, the algorithm is adjusted and optimized to obtain the optimal model. The optimal model is evaluated from two aspects: clustering of users with similar online brand loyalty and defining the degree of online brand loyalty. Specifically, clustering can be defined as: given a data set D, in which each element in the data set has n attributes, some clustering algorithm is applied to divide D into k subsets, and the similarity between the elements inside each subset is required to be as high as possible, while the element dissimilarity of different subsets is required to be as high as possible. As can be seen from the definition of clustering, the core of clustering implementation is the application of clustering algorithm. Among them, the most basic and commonly used clustering algorithm is partitioned clustering algorithm. The basic step is to give an initial subclass division scheme, and improve the subclass through repeated iteration, making the subclass after each improvement better than the previous one. The evaluation criteria are as
follows: the closer the record distance in the same subclass is, the better, while the farther the record
distance in different subclasses is, the better. Specifically, the clustering number of the scattered points
is determined first, then several points are selected as the initial particles, and then the data points are
iteratively reset according to the defined algorithm until the optimal clustering effect is finally
achieved. According to the algorithm function and process, the algorithm program is as follows:

```r
#feature behavior usrcount<-function(data_set,data_attribute_num)
{#data set:
#data attribute num:
usr count<-data_set[,data_attribute_num]
usr count temp<-length(sort(usr_count))
}
#feature behavior usrcount(protran_cckes data,ll)
usr sum_count<-data.frame(usrid=usr count temp$values,sum count=usr count temp$lengths)
#feature behavior usrcount(proe tran_cck sandisk_data,ll)
usr_sandisk_sum_count<-data.frame(usrid=usr count temp$values.sandisk,sum_count=usr count temp$lengths)
#feature behavior sumprice<-function(data_set.data_attribute_num)
#data_set:
#data attribute num:
usr}rice<-data setj,c(11.4)]
usr}rice sum+=aggregate(usr price[,2J.list(usr}rice[,ll],sum)
pro tran-cck data[is.na(pro_tran_cck_data)<-i0.34
pro tran_cck sandiskwe data[is.na(pro_tran-cck_sandisk_data)<-X0.41
#feature behavior sumprice(pro_trap-ecllida,4)
usr sum}rice<-usr-price_sum
names(usr_sum}rice)="usrid","sum}rice"
#feature behavior sumprice(proe tran-eclhdata,4)
usr sandisk}sum}rice<-usr}price_sum
names(usr_sandisk}sum}rice)="usrid","sandiskse sum_price"
}
```

As can be seen from the algorithm program, this algorithm includes three sub-algorithms. The core
is as follows: for the sub-algorithm f_feature_behavior_usrcount, it mainly focuses on the memory
card transaction records of the data set, conducts the record set count statistics according to the user
number, and the statistical result is the purchase volume data set with users as the unit. For the sub-
algorithm f_feature_behavior_sumprice, it mainly aims at the memory card transaction record of the
data set. It calculates the sum of the commodity amount according to the user number, and the statistical result is the data set of the total purchase amount of the memory card. For the sub-algorithm \( f_{\text{feature.behavior}} \): it mainly calculates the repurchase rate from the data set of the purchase amount of application memory card, and calculates the wallet share from the data set of the purchase amount of application memory card.

Clustering model is based on similarity and groups data sets, so that the similarity within the same class is as large as possible and the difference between different classes is as large as possible, which is the basic requirement for the effectiveness of clustering model. In order to test whether the results of clustering meet the above requirements, some methods to test the validity of clustering are proposed. In general, the evaluation of clustering effectiveness includes the measurement of clustering quality, the degree to which the clustering algorithm is suitable for a particular data set, and the optimal number of clustering for a certain partition. The commonly used evaluation methods of clustering effectiveness include external evaluation, internal evaluation and relative evaluation.

3. Conclusion
The data used to construct the brand loyalty measurement model has the characteristics of big data. The core of model building using big data is to make full use of the value of data, and machine learning is the key technology to give full play to the value of data, and as much data as possible will improve the accuracy of the model. In the study of model construction using machine learning, an experiment has effectively confirmed a theory: the larger the amount of data applied to machine learning model construction, the higher the effectiveness and efficiency of the model. For different algorithms, with the increase of the amount of data, the accuracy is constantly improved, and the better the effect of the model.

In the process of model construction, missing values and outliers in the data set will lead to low accuracy and poor effectiveness of the model. Therefore, dealing with missing values and outliers is very important for model building and optimization. Because the missing values have been detected and cleaned during the data cleaning phase. Therefore, in the optimization stage, the outliers in the data set are mainly processed. However, it should be noted that data processing is not completed once. In each stage of model construction, the state of data should be paid attention to at any time and corresponding processing should be provided.

According to the previous analysis, the optimization of the model can be achieved through data processing optimization. Data processing should include the processing of outliers in addition to the cleaning of wrong and missing data in the data cleaning stage. Combined with the research of the upcoming feature project, it is found that there is an abnormal record due to the wrong brand name of the memory card in the data set. In order to provide an accurate and effective original feature set for feature engineering optimization in the follow-up research and improve the effectiveness of the model, it is necessary to deal with the record anomalies caused by brand name errors during the process of building the original feature set.

Feature plays an important role in model construction and optimization. The selection and construction of high-quality features can make the model get a good structure in data set and make the model operation speed faster. Make the model results easier to understand; Make the model easier to maintain. Therefore, in order to make the model give full play to the above excellent results, it is necessary to optimize the model through the method of feature engineering. Its role lies in: First, it can improve the effect of the algorithm. The increase in the number of features will increase the difficulty of algorithm solving. In particular, with the increase of the number of a large number of unrelated features, the number of samples in the dataset will increase significantly. The algorithm running complexity caused by feature increase or redundancy will result in higher variance error and overfitting effect. Second, it can improve model performance. Model generation requires the ability to experiment as many features as possible in a shorter time and faster speed, and effective selection and adjustment of features can obtain better features and improve model performance. Thirdly, it can improve the interpretability of the model. The interpretability of the model is conducive to the
judgment of the stability of the model effect and the guidance and support of the application of the model results. It is an important way to improve the interpretability of the model to select the limited and reasonable features and eliminate the irrelevant features.

Acknowledgements
The Project of Year 2018 Scientific research business fee project of Heilongjiang Provincial Education Department New risk characteristics of border enterprises in Heilongjiang Province under the new normal Project number: 2018-KYYWF-1264.

References
[1] Labrinidis A, Jagadish H V. “Challenges and Opportunities with Big Data”. Proc Endowment, vol. 5, no. 12, pp. 2032-2033, 2012.
[2] Li J Z, Liu X M. “An important aspect of big data: Data usability”. Journal of Computer Research and Development, vol. 50, no. 6, pp. 1147-1162, 2013.
[3] Pholsena K, Pan L. “Traffic Status Evaluation Based on Possibilistic Fuzzy C-Means Clustering Algorithm”. 2018 IEEE Third International Conference on Data Science in Cyberspace (DSC). IEEE, 2018.
[4] Peng D, Xu Y, Wang Y, et al. “Soft-sensing in complex chemical process based on a sample clustering extreme learning machine model – ScienceDirect”. Ifac Papersonline, vol. 48, no. 8, pp. 801-806, 2015.
[5] Hu T, Zheng M, Li J, et al. “A scene-adaptive motion detection model based on machine learning and data clustering”. Multimedia Tools & Applications, vol. 74, no. 8, pp. 1-19, 2015.
[6] Tang M, Sun C, Liang L. “Clustering Analysis Method of Ethnic Cultural Resources Based on Deep Neural Network Model”. Machine Learning for Cyber Security, 2020.