ISSN: 2688-9323

The Importance of Brand Communication in Improving Product Power based on the Catboost Model

Bowen Zhao

China Automotive Technology&Research Center Co., Ltd. Tianjin, China

zhaobowen@catarc.ac.cn

Abstract

Strengthen the brand building of self-owned brands and build a strong country of automobile enterprises. The article first introduces what the catboost model is and explains the advantages of the model; secondly, it collects sample data on the Internet and uses the sample data to train the CatBoost model; thirdly, it uses the best training parameters for calculation; finally, it conducts a comparative analysis. Mainly from the four aspects of product, service, marketing, and brand communication. Comparative analysis. Contribution of sub-level indicators for analysis. Based on the research conclusions of this article, it can have a specific guiding role for enterprises to build strong brands.

Keywords

CatBoos; Training; Brand Communication; Contribution.

1. Background Introduction

Carry forward the national automobile spirit. First of all, we must improve the brand power of federal automobiles. Enhancing brand power relates to products, culture, services, and brand communication. This article only studies the importance of brand communication to brand building.

"Brand communication" is a series of communication activities about brand information carried out by brand owners with internal and external target audiences through various communication strategies such as advertising communication, marketing campaign communication, public relations communication, interpersonal communication, and various communication tools. It aims to build a brand and maintain a positive relationship between the brand and consumers, and other stakeholders. It aims to promote the target audience's understanding, recognition, trust, and experience of the brand to increase brand equity optimally. As a kind of corporate communication behavior, brand communication has the typical characteristics of all communication activities, and it is a dynamic process. Brand communication determines that the brand communication of an enterprise is an open, systematic, and long-term work. Enterprises should do an excellent job of brand strategic planning. Enterprises make brand strategic planning, maintain continuity, and continuously communicate brand information to target audiences; simultaneously, enterprises make timely adjustments to the brand according to market changes to keep the brand alive [1].

Brand communication is a part of a company's marketing activities, and its ultimate purpose is to promote sales. However, brand communication is different from other communication behaviors. It is responsible for implementing brand marketing, establishing and maintaining the brand. Therefore, companies often use brand communication from brand marketing to achieve the following three different levels of purpose.

ISSN: 2688-9323

1.1. Build Brand Awareness

Consumers often face many products of the same type among the thousands of brands. In this case, the most important thing is to establish and spread the brand's unique cognitive system. The primary cognitive system mainly includes two aspects: visual identification system, such as brand name, good logo, image logo, standard color, et cetera.; concept identification system, such as advertising slogan, corporate song, brand concept, brand story, et cetera. These are the essential communication elements that run through all brand communication activities of the company. By delivering the brand information of the visual identification system and the concept identification system, we help to enhance brand awareness, make the target audience familiar with the brand, and differentiate our brand from many other brands [1].

1.2. Build a Deep Brand Identity

Building brand awareness is not enough. We also need to show the target audience what makes the brand different: the brand identity system, which includes unique brand personality, image, and association. The enterprise has formulated a strict brand communication plan based on a long-term brand strategy. Through various communication activities, the enterprise deepens the target audience's understanding of the brand. The audience can believe that their brand can provide unique benefits that other brands do not have. The audience has a good understanding of the brand. Cognition is transformed into brand value recognition.

1.3. Build Relationships that Strengthen Brand and Audience

After consumers perceive the brand, brand marketers begin to build and strengthen the relationship between the brand and consumers. At this level, brand communication focuses on the brand's core values, such as the brand's functional, social, emotional, cultural, psychological benefits, et cetera. Through brand communication, brand marketers let the target audience know and understand the brand and make them feel that it is customized for them and their good friend. They not only see the brand themselves but also introduce the brand to them. More people. Therefore, we need to build a stronger relationship between brands and consumers and turn it into brand loyalty. Two-way communication between brands and consumers can achieve this goal, and consumers can also increase the brand's experience.

CatBoost is a machine learning library open-sourced by Russian search giant Yandex in 2017. It is a kind of Boosting family of algorithms. CatBoost, XGBoost, and LightGBM are the three mainstream artifacts of GBDT. They are all improved algorithm modes under the framework of the GBDT algorithm. XGBoost is widely used in the industry. LightGBM effectively improves the computational efficiency of GBDT, and Yandex's CatBoost is an algorithm that performs better than XGBoost and LightGBM in terms of algorithm accuracy.

Catboost has achieved actual results in academia. For example, Zhang Fan [2] used scenario-based thinking and the CatBoost model to analyze factors influencing satisfaction with new energy vehicles. It has a significant influence, and consumers' attributes are also an essential factor affecting product satisfaction. The research conclusions of the article provide a reference and basis for improving enterprise product satisfaction and precision marketing; Wang Qiang et al. proposed a TSA method based on multi-layer CatBoost, which further improved the accuracy and reliability of transient stability assessment. TSA method. Firstly, with the steady-state operating variables before the power system fault as the input features, a maximum correlation minimum redundancy integration scheme is adopted to screen several critical components from the input features. Then, using these key features to train multiple CatBoost models separately, build multiple CatBoost-driven TSA models, and combine various trained CatBoost models to construct a comprehensive TSA model. In the transient stability analysis, the analysis of numerous CatBoost models is synthesized, and the final evaluation result is determined by majority voting. The test results show that the TSA comprehensive model has

ISSN: 2688-9323

not only highly high prediction accuracy but also has good generalization ability and robustness. Dang Cunlu [4] and others proposed a short-term load forecasting method based on the combination of long-short-term memory neural network and CatBoost, aiming at the time series and nonlinear characteristics of power load data. The long-term and short-term memory networks cannot directly deal with categorical features, establish the LSTM load forecasting model and Cat Boost load forecasting model for the processed power load data; use the inverse variance method to determine the weighting coefficient to obtain the predicted value of the combined model of LSTM and CatBoost; finally, use the actual load data to evaluate the effectiveness of the algorithm. The verification and prediction results show that the method using the combined model of LSTM and Cat Boost has a significant improvement in the load prediction accuracy.

Therefore, the catboost model proposed in this article studies the importance of brand communication in improving product power. The purpose is to analyze the role and importance of brand communication in brand building.

2. Method Research

The research method in this article is shown in Figure 1, which mainly introduces the Catboost model, collecting data, data preprocessing and bringing it into the catboost model.

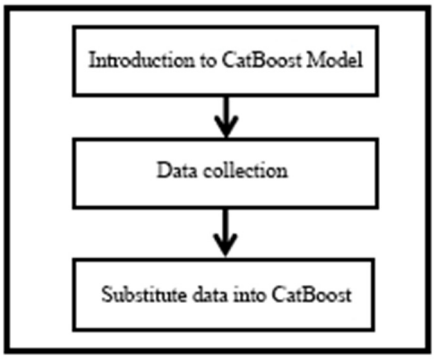

Figure 1. Research method in the article

CatBoost is a GBDT framework with fewer parameters, support for categorical variables, and high accuracy based on symmetric decision trees (oblivious trees) as the base learner. As can be seen from its name, CatBoost is composed of Categorical and Boosting, which can efficiently and reasonably handle categorical features. In addition, CatBoost also solves the problem of gradient bias (Gradient Bias) and prediction shift (Prediction shift), so it reduces the occurrence of overfitting, thereby improving the accuracy and generalization ability of the algorithm.

Many model training samples are obtained after the outlier and missing value processing in the above process. The purpose of training the model is to calculate the satisfaction of the survey questionnaire more conveniently, accurately, and efficiently and to calculate the questionnaire in real-time.

2.1. Introduction to CatBoost Model

The CatBoost algorithm is a profoundly improved version based on the GBDT framework. Its main feature is that it uses a unique way to process categorical features. First, do some mathematical statistics on the category features, calculate the frequency of a particular category feature, and generate a new floating-point feature. In addition, to prevent the model from overfitting, the L2 regularization hyperparameter can be added at the end of the model. The CatBoost model can use the relationship between features to combine into a new set of

ISSN: 2688-9323

category features, which significantly enriches the feature dimension of the model and improves the accuracy of the model. The base model of the CatBoost model uses an asymmetric tree, which can also prevent the model from overfitting.

The most important principle of the CatBoost model is the processing of eigenvalues, which mainly includes two aspects: the first aspect is the processing of category features. The main processing methods are as follows [2]:

(1) Randomly permutes the set of input eigenvalues to generate multiple random permutations;

(2) We calculate the average sample value for samples of the same class;

(3) Convert all categorical eigenvalues to numerical values using the following formula:

$$ave\_target = \frac{CountInClass + \alpha \bullet P}{TotalCount + \alpha} \tag{1}$$

The second aspect is the recombination of features. The number of combinations grows exponentially with the number of classes features in the dataset. All situations cannot be considered in this algorithm. When the current tree considers a new split, the CatBoost model will adopt a greedy method. The calculation method of the cheap split method is shown in the following formula (2), and the specific parameters are shown in Table 1. Show [2].

$$Gain = \frac{1}{2}\left[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{G_L + G_R + \lambda}\right] - \gamma \tag{2}$$

Table 1. Specific parameters of a greedy segmentation method

Parameter	Implication	Remark
$Gain$	Optimal Gini coefficient when splitting	The split with the smallest Gini coefficient is the best.
$\frac{G_L^2}{H_L + \lambda}$	Fraction to split the left subtree	
$\frac{G_R^2}{H_R + \lambda}$	A fraction that divides the correct sub number	
$\frac{(G_L + G_R)^2}{G_L + G_R + \lambda}$	Score when not split	
γ	Increase model complexity from new leaves	Each split increases the complexity of the model.

The specific operations are as follows:

(1) The first split (root node split) does not consider any combination in the tree

(2) For the second classification, in the current tree of the dataset with all the category features, the CatBoost model contains all the combined and categorical features, and the combined value is converted into a numeric feature;

(3) The CatBoost model is to generate a combination of numerical and categorical features: all splits selected in the tree as categories with two values and used similarly in the mix.

ISSN: 2688-9323

After recombining the eigenvalues, the Gini coefficient must be calculated for each segmentation scheme. The partition with the smallest Gini coefficient is the best partition of the node, and its model is shown in equation (3):

$$y_i = \varphi(x_i) = \sum_{k=1}^{k} f_k(x_i), fk \in F \tag{3}$$

Among them, k is the number of basic tree models, and F is the primary tree model. On this basis, assuming that the learning errors of the bare model trees are independent of each other, add a new function F(t) to the model. The learning goal of the CatBoost model is to find and minimize the objective function, as shown in the following formula (4):

$$\hat{y}_i^{(t)} = y_i^{(t-1)} + f_t(x_i) \tag{4}$$

The sample with the sequence number i is learned in the t-round model, and it needs to combine the results of the previous t-1 round with learning the current round number model, so the objective function of the sample with the sequence number i during the t-round is the formula (5):

$$\begin{aligned} L^{(t)} &= \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t)}\right) + \Omega(f_t) \\ L^{(t)} &= \sum_{i=1}^{n} l\left(y_i, y^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) \end{aligned} \tag{5}$$

2.2. Data Collection

In this article, big data online technology collects data about a product, service, product, brand communication, and other four dimensions. It uses keyword group technology to obtain text data of four sizes and then statistical techniques. The number of mentions for each measurement is taken as an eigenvalue.

2.3. Substitute Data into the Catboost Model

2.3.1. Determine the Parameters of the CatBoost Model

Determining the best parameters of the CatBoost model is the premise to ensure the accuracy and efficiency of the model. In this article, the parameters are optimized based on the control variables. The specific parameters are described in Table 2.

Table 2. Specific parameters of the CatBoost model

Parameter	Implication
Learning rate	Learning rate
depth	Tree depth
iterations	Number of iterations
eval_metric	Evaluation Criteria
random_seed	Random seed number
bagging_temperature	Bayesian bagging control strength
od_type	Overfitting Detection Type
od_wait	The number of iterations after minimizing the loss function

ISSN: 2688-9323

2.3.2. Model Calculation

After (1), determine the parameters of the model. Then, the importance of each feature, such as product, service, marketing, and brand communication, can be analyzed through the CatBoost model. The importance of each element to the model can be expressed as a percentage.

3. Conclusion

Shape the brand value of car companies, especially the brand power of self-owned brands. Products, services, marketing, brand communication, and other aspects have become essential factors affecting brand satisfaction; The indicators are (in proportion to new energy): safety, battery life, economy, charging, comfort, drivability, and power; in terms of service, the sub-indices that affect service satisfaction are pre-sales service, after-sales service Maintenace and other aspects.

References

[1] https://www.sohu.com/a/249113181_100060617.

[2] Fan Zhang, Jing Yang, BenShuang Liang,Hongjing He. Analysis of Influencing Factors of New Energy Vehicle Satisfaction Based OnScenario Thinking and Catboost Model, IOP Conference Series: Earth and Environmental Science, 2021. Science Technology and Engineering.

[3] Dang Cunlu , Yang Hailan , Wu Wencheng .Short-term Load Forecasting Based on Combination Model of LSTM and CatBoost[J]. Journal of Electrical Engineering,2021,16(03):62-69.

[4] https://zhuanlan.zhihu.com/p/102540344.