

# A Novel Digital Coupon Use Prediction Model Based on XGBoost

Lu SONG<sup>1,a</sup>, Wen-Sheng YANG<sup>1,b,\*</sup>

<sup>1</sup> School of Economics and Management, Nanjing University of Science and Technology,  
Nanjing 210094, China

<sup>a</sup>116107000819@njust.edu.cn, <sup>b</sup>wensheng\_yang@163.com

\*Corresponding author: Wen-Sheng YANG

**Keywords:** Digital coupon, Extreme Gradient Boosting, Feature engineering, Prediction.

**Abstract.** It was found that traditional predicting models of coupons could only process small-scale data, thus were unable to precisely predict whether consumers would use new digital coupons in the future. By introducing the feature engineering and the machine learning algorithm of Extreme Gradient Boosting (XGBoost), a novel Digital Coupon Use Prediction Model (DCUPM) was established to deal with vast amount of historical data and to provide high-precision prediction of the future digital coupons use of consumers. In this paper, an example of this DCUPM was then presented and analyzed based on the data provided by Alibaba Cloud. The result shows the Area under the Curve (AUC) is calculated to be 0.896, which demonstrates the reliability of the proposed DCUPM. And features with higher scores can be used to achieve targeted delivery of digital coupons for businesses.

## Introduction

In recent years, the use of digital coupons has exploded. A survey conducted by e-Marketer, a leading global market research institute, showed that in 2015, a total of 121.3 million people redeemed e-coupons in the U.S. grocery retail industry, an increase of 7.6% over the same period last year [1]. Since digital coupons are recordable, a large amount of use data will generate in real time when consumers purchase. And the data contains valuable information about consumer behavior which can be used to identify consumers who are more likely to be attracted to discounts. If companies directly distribute digital coupons to this group of consumers, they can not only effectively reduce the delivery cost, but also avoid the damage of brand reputation, resulting from excessive delivery. Therefore, it is important to predict whether consumers will use digital coupons in the future. In reality, Didi, a mobile app for taxi, and Alipay are typical cases.

However, it is not easy to handle large-scale data and extract features of consumers, companies, digital coupons and so forth. Large amount of data and complexity of behavior analysis make traditional methods in Econometrics ineffective. Fortunately, the rising and development of data science in recent years provides a solution to this problem. Machine learning is a new way to analyze data and performs well in prediction [2]. In this paper, the XGBoost algorithm in machine learning is used to establish a novel Digital Coupon Use Prediction Model.

Since the primary challenges in digital coupon use prediction is the lack of information between current moment and predictive moment, we solve this problem by selecting part of historical data as test data set based on the time sequence, which is not used at the beginning. Then, the prediction results will be compared with the test set data to verify the model.

As training set, the remaining historical data is used for feature extraction. The raw data can be transformed into consumer feature, company feature, digital coupon feature, and their interactive feature as the input to train the XGBoost model. The trained model is utilized to predict whether consumers appearing in the test set will use digital coupons in the next month. Meanwhile, feature scores will be given by model.

## Literature Review

Previous studies have shown that companies always exceed competitors by understanding consumers demand, and consumers demand is also reflected in the use of coupons. However, very few studies were conducted from the perspective of coupons before, if any, they seldom analyzed behavioral features of consumers through data. The published studies on coupons can roughly be divided into two types: (1) one holds that demographic and psychological characteristics affect the use of coupons, and (2) the other holds that attributes of coupons affect consumers' use of coupons. Among which, demographic characteristics include gender, age, education level, employment, income and other factors [3,4]. And consumers psychological characteristics include satisfaction, motivation, perception, intention and other factors [5,6]. For example, Kitchen use factor analysis to prove that consumers' income, education level will affect coupon usage [7]. Dickinger and Kleijnen used structural equations to verify that consumers' perceptions and attitudes have a significant impact on coupon use [8]. Lee and Yeu used structural equations to analyze the impact of coupon discount and expiration dates on the use of coupons [9].

With the increasing size of digital coupons data, traditional methods in Econometrics doesn't work. Therefore, new models and algorithms are urgently needed for highly-efficient mining and analysis of coupon data as well as feature extraction. The XGBoost algorithm is one of the new achievements in machine learning that runs more than ten times faster on a single computer than other solutions [10]. In practice, XGBoost is widely used in various data-driven classification and prediction such as bankruptcy prediction [11], telemarketing success rate prediction [12] and etc. Therefore, XGBoost is suitable for analyzing digital coupons data.

## Digital Coupon Use Prediction Model

The Digital Coupon Use Prediction Model (DUPM) was designed to predict whether consumers would use given digital coupons in the next month. To obtain the short-term prediction for each consumers appearing in test set, the following steps are required: first step is to extract features according to the participants in the use of digital coupon. Second step is to import features into the XGBoost for model training, in which exact greedy algorithm and approximate Algorithm are applied to find optimal value. Last step is to output the feature score.

## Feature Engineering

Feature engineering is the process of transforming raw data into features which can better represent the prediction model, thereby increasing the accuracy of the model for invisible data [13]. After transformation, each feature is evaluated for validity. If the feature is highly correlated with the predicted result, the feature is regarded as more important. A feature can be assessed by the accuracy with which the prediction result is improved.

This paper focuses on the issue of digital coupons and conducts feature engineering through digital coupons use data from January 1st to July 31st, in 2016. Since digital coupon data is difficult to obtain as corporate privacy, we can only use public data provided by Alibaba Cloud.

First, the raw data is divided into several training sets and test sets. Then the prediction interval is used for label extraction, and the feature interval is used for feature extraction. According to the participants in the use of digital coupons, the features are divided into four types as the input: features of digital coupons, features of consumers, features of companies, interactive features of consumers and companies.

After training the model, the feature scores are obtained, and then the original features can be appropriately added, removed and combined according to the feature scores until the prediction accuracy reaches expected value.

## Model of XGBoost

XGBoost is an efficient implementation of the Gradient Boosting algorithm. The base learner in XGBoost can be both linear booster and tree booster. When training data, XGBoost can construct

boosting trees in parallel with all the CPU cores, and use distributed computing to train very large models. In general, the XGBoost algorithm is more than 10 times faster than similar algorithms [10].

Previous studies mainly used interpretive statistical techniques to build models, while the need to apply predictive analytics to marketing research has been increasing. In this article, XGBoost is used to study the variables that affect the use of digital coupons. The key advantages of the XGBoost are high prediction accuracy, fast computation. XGBoost provides a Python interface that means a technical guarantee of predictive implementation.

### Regularized Learning Objective

In order to obtain the best parameters of the training dataset, we define an objective function  $Obj$  in XGBoost that contains training loss  $l$  and regularization  $\Omega$  to measure the performance of the model. The training loss represents the degree to which the model fits the data and the regularization controls the complexity of the model. XGBoost is known as ‘regularized boosting’ technology [10].

$$\text{Eq.1 } Obj^{(t)} = L^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (1)$$

Where,  $\Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2$ ,  $y$  is the dependent variable,  $T$  is the number of leaf nodes,  $w$  is the weight of leaf nodes and  $\lambda$  indicates regularization parameters. After the Taylor expansion of the  $Obj$ , the first derivative  $g_i = \partial_{\hat{y}^{t-1}} l(y_i, \hat{y}^{t-1})$  and the second derivative  $h_i = \partial_{\hat{y}^{t-1}}^2 l(y_i, \hat{y}^{t-1})$  of the objective function are obtained respectively, and the optimal weight of the leaf node is given as (2), and the optimal value of the objective function is transformed into formula (3) [10].

$$\text{Eq.2 } w_j^* = - \frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \quad (2)$$

$$\text{Eq.3 } \hat{L}^{(t)}(q) = - \frac{1}{2} \sum_{j=1}^T \frac{(\sum_{i \in I_j} g_i)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma T \quad (3)$$

### Gradient Tree Boosting

Although XGBoost has both tree and linear boosters, the tree booster performs far better than a linear booster [14]. Therefore, the tree booster is used in this paper. The solution of XGBoost is to find the optimal tree structure. In a tree structure, there are many potential splitting points, and whether the nodes split depends on whether the objective function is better. XGBoost uses a structure score to evaluate the structure of a tree. The lower the score, the better the tree structure. The algorithm constantly enumerates different tree structures evaluated by structure scores until it finds an optimal one.

Let  $I_L$  and  $I_R$  be the left and right nodes of the split sample set. Supposed, if the node is split, the gain is:

$$\text{Eq.4 } L_{split} = Gain = \frac{1}{2} \left[ \frac{(\sum_{i \in I_L} g_i)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{(\sum_{i \in I_R} g_i)^2}{\sum_{i \in I_R} h_i + \lambda} - \frac{(\sum_{i \in I} g_i)^2}{\sum_{i \in I} h_i + \lambda} \right] - \gamma \quad (4)$$

In formula (4), the left item, the middle item and the right item refer to the score of the left subtree, the score of the right subtree and the score in the case of no splitting respectively. And  $\gamma$  is the cost of complexity after adding the new leaf node.

In this study, the maximum depth of the tree is set as 5, and the XGBoost is applied to train model repeatedly. The algorithm gets the weights of the new features directly after each boost calculation. XGBoost allows cross-validation in each round of boosting iterations.

## Empirical Analysis

In this paper, the algorithm of XGBoost is implemented with Python 3.6 with the package of XGBoost. Feature Engineering is used to transform the data into four types of feature: feature of digital coupon, feature of consumer, feature of company, and interactive feature. These extracted features are then imported into XGBoost to verify the correctness of the prediction result.

## Dataset

The data in this paper is from Alibaba Cloud which records shopping information of consumers from 2016/01/01~2016/07/31 shown as table 1. It should be noticed that Alibaba does not provide some specific information of goods for corporate privacy protection. The public data from Alibaba Cloud contains a total of 1754884 records, involving 539438 consumers, 8415 merchandise and 1053282 kinds of coupons.

Using sliding window method to get multiple training sets. Three training sets have been divided by date: 2016/01/01~2016/04/13, 2016/02/01~2016/05/14, 2016/03/15~2016/06/30. Accordingly, test sets were one month later: 2016/04/14~2016/05/14, 2016/05/15~2016/06/15, 2016/07/01~2016/07/31. In particular, the digital coupons in the test sets did not appear in the training sets.

Table 1. Description of digital coupon data.

Name	Description
User_id	The consumer's ID
Merchant_id	The company's ID
Coupon_id	The digital coupon's ID
Discount_rate	Discount rate of digital coupon,
Date_received	Date of receiving coupon
Date	Date of consuming
Action	0: click, 1: purchase, 2: collect coupon
Null	No coupon consumption

## Extracted Features

According to the participants in the coupon usage process, the features are divided into four types: feature of digital coupon, feature of company, interactive feature of consumer and company. The data in the feature interval is transformed into four types of features and imported into the XGBoost model for repeated training. After completing the training, the data of test set is imported into the model for verification. XGBoost provides feature scores. Those features with low scores are suggested to be removed, combined or changed.

According to experience, active consumers are more likely to use coupons, and coupons issued by active companies are also more likely to be used. Therefore, in the process of feature extraction, more attention is paid to the frequency of consumption, the number of digital coupons received or used as indicators of consumer activity. The four types of features covered in this article have been adjusted many times. The final results of feature engineering in this study are shown in the table 2, and detailed feature scores are shown in the next section.

Table 2. Four types of features.

Features	Name	Description
Features of digital coupons	F1	Digital coupon (discount: 0, threshold: 1)
	F2	Discount rate of digital coupon
	F3	Digital coupon's available time
	F4	The day in the month when digital coupons are used
	F5	The number of the same digital coupon being received
	F6	The historical number of the same digital coupon being used
Features of consumers	F7	The number of digital coupons consumers receive
	F8	The number of coupons consumers receive and use
	F9	The using rate of digital coupons after receiving
	F10	Mean discount rate of used coupons
	F11	The number of companies with coupons being used/The total number of companies
	F12	The number of used coupons/the total number of digital coupons
Features of companies	F13	The number of received companies' digital coupons
	F14	The number of sales
	F15	The number of used companies' digital coupons
	F16	The using rate of companies digital coupons after received
	F17	Mean discount rate of used companies' digital coupons
	F18	The mean consumer-company distance with used digital coupons
Interactive features of companies and digital coupons	F19	The number of times consumers receive companies' digital coupons
	F20	The number of times consumers receive companies' digital coupons
	F21	The number of times consumers receive companies' digital coupons
	F22	The rate of digital coupons being received and used
	F23	The number of unused digital coupons of each consumer/the number of unused coupons of the total consumers
	F24	The number of used digital coupons of each consumer/the number of unused coupons of the total consumers

## Performance Indices

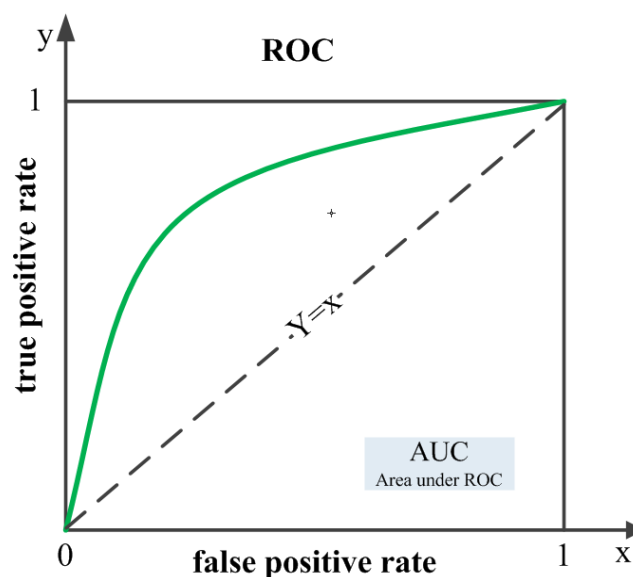


Fig. 1. An illustration of ROC curve.

The core purpose of this paper is to give an approach to predict whether consumers will use a new digital coupons in the next month based on their own historical behavior. For convenience, during the experiment, 1 means that consumers use digital coupons and 0 means not use. Therefore, the digital coupon use prediction has actually been converted into a dichotomous issue.

The diagnostic ability of a binary classifier system is commonly illustrated by Area under the Curve (AUC) of Receiver Operating Characteristic (ROC) [15]. The abscissa of the ROC curve is false positive rate (FPR), and the ordinate is true positive rate (TPR) as shown in Fig 1. And AUC value is calculated by the area under ROC curve. Generally, AUC is applied to evaluate the accuracy of classification because the ROC curve does not clearly indicate which classifier works better. While as a numerical value, AUC can obviously represent the effect of classifier. Since the ROC curve is generally above the line  $y=x$  (FPR=TPR, a random classifier with AUC=0.5), the AUC value ranges between 0.5 and 1. The closer the AUC is to 1, the higher the classification accuracy is.

## Result and Discussion

The features extracted above will be assessed by the prediction accuracy of model which is expressed as AUC. In order to verify the proposed DCUPM model, two additional algorithms of Gradient Boosting Decision Tree (GBDT) and Random Forest (RF) are also applied for comparison.

As shown in Fig 2, the AUC of XGBoost is 0.896 after 4000 times of iteration, which indicates that the presented prediction model in this paper is of high precision and reliability. However, GBDT and RF are not performing as well as XGBoost with AUC of 0.861 and 0.857 respectively. This satisfactory result of prediction benefits from both the advantageous machine learning algorithm of XGBoost and the rational selection of features.

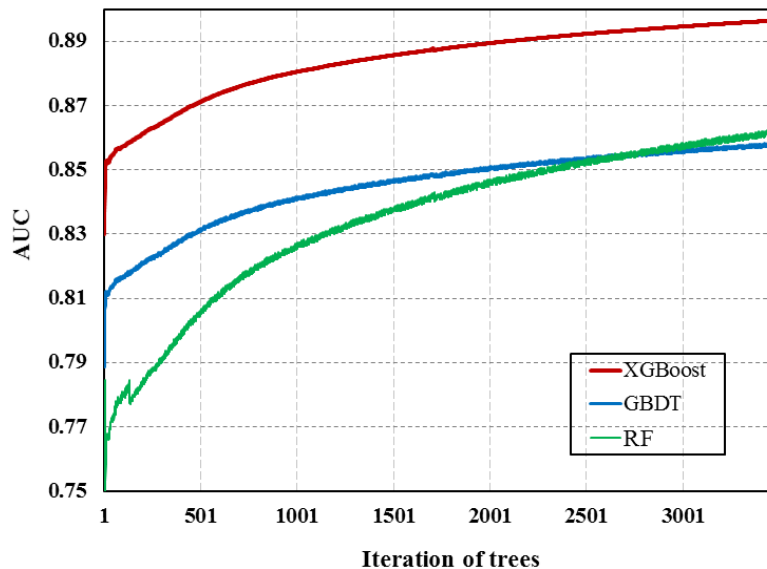


Fig. 2. Performance indices of three algorithms.

After confirming the reliability of the presented Digital Coupon Use Prediction Model, current feature scores were output by XGBoost. The significance of predictors is generally measured by feature scores. Due to space limitations, Fig 3 shows the 10 most important features.

According to the ranking of scores, it is clear that the top 10 predictors are more related to the features of digital coupons, followed by the features of companies. Among them, the variable F16 in the feature of companies is the most important predictor which means the using rate of companies digital coupons after received with the highest score 46. That is, digital coupons issued by active businesses are more likely to be used by consumers. Similarly, F14 measures the activity of companies from the total sales.



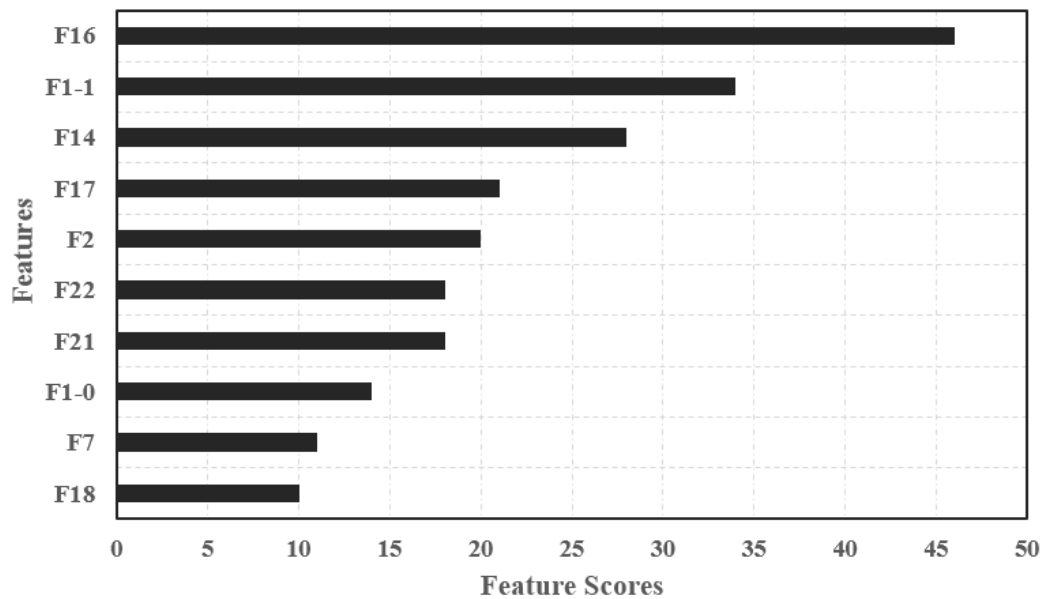


Fig. 3. The scores of top 10 features.

The digital coupons involved in the experiment are conditional with the threshold of lowest consumption. According to statistics, the average minimum consumption denoted by F1-1 is 88.12 yuan with a maximum of 300 yuan and a minimum of 5 yuan, as shown in table 3. The average discount for digital coupons denoted by F1-0 is 11 yuan, of which the maximum discount is 50 yuan and the minimum is 1 yuan. F17 means that consumers are more likely to use digital coupons with higher discount rates. And the mean discount rate of digital coupons used by consumers is about 13% off. Surprisingly, the average number of digital coupons received by a consumer is 5,837.32.

Finally, if the distance between the offline companies and consumers exceeds 1.43km, it will be difficult to attract consumers. These results are very useful for companies when they are trying to attract consumers with digital coupons.

Table 3. Statistical results of top 10 features.

Rank	1	2	3	4	5	6	7	8	9	10
Features	F16	F1-1	F14	F17	F2	F22	F21	F1-0	F7	F18
Mean	0.15	88.12	64.02	0.038	0.87	0.039	0.06	11.11	5837.32	1.43
Max	1	300	22636	1	0.99	1	70	50	22281	10
Min	0	5	1	0	0.33	0	2	1	1	0

## Conclusion

The popularization of smart terminals has promoted the rapid change of digital coupons. However, traditional methods in Econometrics are difficult to deal with large-scale digital coupons data. This paper uses XGBoost algorithm in machine learning with feature engineering to conduct Digital Coupon Use Prediction Model which has been confirmed reliable during the experiment, due to its high accuracy. The features with high scores should be firstly considered when companies decide to issue digital coupons. The features with the highest scores should be firstly considered when companies decide to issue digital coupons. This study provides decision-making basis for companies to achieve accurate marketing.

## Acknowledgement

This work is supported by Jiangsu Graduate Student Scientific Research and Innovation Project under the Grant KZ07022.

## References

- [1] eMarketer, //US Mobile Coupons Stat Pack: Consumer Usage in Grocery Retail, February 2, 2017. [online]. Available: <https://www.emarketer.com/Report/US-Mobile-Coupons-StatPack-Consumer-Usage-Grocery-Retail/2001975>.
- [2] Bollen J, Mao H, Zeng X. Twitter mood predicts the stock market. *Journal of Computational Science*, 2011, 2(1):1-8.
- [3] Kang H, Hahn M, Fortin D R, et al. Effects of perceived behavioral control on the consumer usage intention of e-coupons. *Psychology & Marketing*, 2006, 23(10):841–864.
- [4] Bawa K, Shoemaker R W. The Coupon-Prone Consumer: Some Findings Based on Purchase Behavior across Product Classes. *Journal of Marketing*, 1987, 51(4):99-110.
- [5] Im H, Ha Y. Is this mobile coupon worth my private information?. *Journal of Research in Interactive Marketing*, 2015, 9(2):92-109.
- [6] Tang Q, Zhao X, Liu S. The effect of intrinsic and extrinsic motivations on mobile coupon sharing in social network sites. *Internet Research Electronic Networking Applications & Policy*, 2016, 26(1):101-119.
- [7] Pei Y L, Cheha N, Kitchen P J, et al. Coupon redemption behaviour: a Malaysian cross-segment investigation. *Marketing Intelligence & Planning*, 2014, 32(1):66-88.
- [8] Dickinger A, Kleijnen M. Coupons going wireless: Determinants of consumer intentions to redeem mobile coupons. *Journal of Interactive Marketing*, 2010, 22(3):23-39.
- [9] Lee H S, Min S Y. Factors Influencing the Intention to Redeem Coffee Shop Coupons in Korea. *International Journal of Business & Management*, 2010, 5(7):61-69.
- [10] Tianqi Chen, Carlos Guestrin. XGBoost: A Scalable Tree Boosting System, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016: 785-749
- [11] Tomczak S K, Tomczak J M. Ensemble boosted trees with synthetic features generation in application to bankruptcy prediction. Pergamon Press, Inc. 2016.
- [12] Moro S, Cortez P, Rita P. A data-driven approach to predict the success of bank telemarketing. *Decision Support Systems*, 2014, 62(1246):22-31.
- [13] Kobdani, Hamidreza, et al. Relational feature engineering of natural language processing. *ACM International Conference on Information and Knowledge Management ACM*, 2010:1705-1708.
- [14] Chen T, He T. Higgs boson discovery with boosted trees on *International Conference on High-Energy Physics and Machine Learning*. JMLR.org, 2014:69-80.