Precise Issuance of Meituan Merchants' Coupons with Machine Learning

Precise Issuance of Meituan Merchants' Coupons with Machine Learning

Xue Zhang
School of Business, Shenyang University of Technology
China
358625531@qq.com

Jie Qiu
Ningbo China Institute for Supply Chain Innovation
China
jie.qiu@nisci.edu.cn

Bo Li*
Ningbo China Institute for Supply Chain Innovation
China
libo@alum.mit.edu

Abstract

With the popularity of mobile Internet, the "Online-to-Offline" (O2O) business model has become popular. Issuing coupons to attract new customer registrations and keep old customers active is an important marketing tool for O2O companies. But the random distribution of coupons can be annoying to those non-target customers. For merchants, the transition of issuing coupons to merchants will not only increase the promotion cost but also have a negative effect on their brand reputation. The purpose of this study is to analyze transaction data and build a model to predict the redemption of coupons, so as to achieve the precise issue of coupons by merchants. We use machine learning to analyze the consumption data and extract features from five categories: coupons, merchants, consumers, consumers-merchants, and other categories. A total of 44 features are extracted and the XGBoost (eXtreme Gradient Boosting) model is adopted. It has been verified that the prediction results of the application of the XGBoost model can nearly increase 50% net profits of the merchants.

CCS Concepts

• **Applied computing** → Operations research; Forecasting.

Keywords

XGBoost, coupon, forecasting, online-to-offline

ACM Reference Format:
Xue Zhang, Jie Qiu, and Bo Li. 2024. Precise Issuance of Meituan Merchants' Coupons with Machine Learning. In *The International Conference on Machine Learning, Pattern Recognition and Automation Engineering (MLPRAE 2024), August 07–09, 2024, Singapore, Singapore*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3696687.3696700

1 Introduction

O2O (Online to Offline) is a new e-commerce model that refers to online marketing and online purchases that drive offline (non-online) operations and offline consumption. That is, the merchant provides service information online, and consumers search for the merchant information they need online to realize offline consumption according to their needs [1]. Online-to-offline (O2O) commerce is a business model that attracts potential customers from online channels like websites and mobile apps to make purchases in brick-and-mortar stores. Among O2O companies, Meituan, Ele.me and other head companies stand out. This paper focuses on coupon redemption prediction via analyzing these data. Many merchants issue coupons to reactive old consumer and attractive new ones, which is an important marketing method in O2O (online-to-offline) area. However, the random release of coupon will disturb most consumers. To merchants, the overissue of coupon will damage the brand reputation and will make it hard to calculate the marketing cost. This research aims to build a model to forecast the coupon usage through analyzing the consumer record via machine learning tool.

This paper is composed of six sections. Section 1 introduces the background. Section 2 reviews related literature. Section 3 presents the dataset and methods. Section 4 outlines the prediction results. Section 5 discuss the results and concludes the paper.

*Corresponding author.

MLPRAE 2024, August 07–09, 2024, Singapore, Singapore

ACM ISBN 979-8-4007-0987-6/24/08
https://doi.org/10.1145/3696687.3696700

2 Literature Review

The literature review of this paper sorts out the existing research results, research methods and problems that need to be further studied from coupon function and redemption factor, contemporary forecasting methods, predictive models for O2O coupon redemption.

2.1 Coupon function and redemption factors

This section provides a theoretical basis for the feature extraction in the following modeling process through the analysis and research of the existing research results from the marketing function of coupons, the influencing factors of traditional coupon redemption and the influencing factors of electronic coupon redemption.

2.1.1 Marketing function of coupon. Meituan and some big merchants establish information platform and data management platform, and actively carry out precision marketing. They issue coupons with marketing purposes. First, couponing has become an increasingly significant promotional tool for merchants. Ref [2] find that the electronic coupon could promote both online and offline sales. Second, Ref [3] confirmed that coupons have a promotional effect of increasing sales in short-term sales. Third, Coupons have an advertising effect.

2.1.2 Influencing factors of redemption behavior of traditional coupon users. Coupons can attract price-sensitive customers (customers with a preference for coupons) and increase customer flow at retail stores, while retailers can earn more profits from the consumer behavior of customers using coupons. Many households switch brands on the basis of price and promotion. Ref [4] mentions that the merchants should adopt aggressive pricing and location strategies to solve the problem of unpredictable offline demand variations. The psychological characteristics of "coupon-prone" consumers are also an important research direction. For consumers who are prone to coupons, psychological incentives play an important role. Consumers with high value awareness are more motivated by economic benefits in the purchase context (gain utility), while consumers with low value awareness are more likely to be affected by psychological incentives (transactional utility). Empirical research shows that consumer loyalty is also an important factor affecting consumer redemption behavior.

2.1.3 Factors influencing the redemption behavior of electronic coupon users. When merchants use coupons for marketing, they usually limit whether consumers can use the coupons depending on their purchases. The most common ones are setting coupon thresholds, limiting product types, and restricting people. If merchants use threshold-less coupons or lower the threshold, coupons may be more attractive to consumers, thereby increasing coupon redemption rates.

Nowadays the electronic coupon customization for each customer is more and more important, and it impact directly on coupon redemption [5]. Consumers' intention to use mobile coupons requires compatibility and can be amplified by social networks.

2.2 Contemporary forecasting methods

After reviewing some research on forecasting, we find that there are lots of methodologies to predict. And the general prediction method includes time series analysis, panel data models, machine learning, and stochastic models [6]. Machine learning is a branch of artificial intelligence [7]. Machine learning has been widely used in data mining, computer vision, biometrics [8], search engines [9], medical diagnostics [10], etc.

2.3 Predictive models for O2O coupon redemption

For O2O giant Meituan, on the one hand, it is focusing on the digital upgrade of merchants to further improve the business capabilities of merchants. On the other hand, it begins to expand the types of services and build a local living service ecosystem. All these need to mine and analyze the existing user data. The prediction of coupon write-off is an indispensable part. Its accuracy can not only improve the profitability of merchants and platforms, but also further increase the stickiness of users. Based on terabytes of massive customer data, there has been an increasing trend of several orders of magnitude, and it is appropriate to use machine learning to make predictions.

2.4 Summary

In this section, we reviewed stochastic models, time series models, and machine learning models. Stochastic models and time series models have been widely used and optimized in financial, engineering, and social science problems. Machine learning methods are considered the best way to solve prediction problems, especially the data set is very huge. Coupon redemption forecast is a binary classification problem. For a binary classification problem which has a big data set, to apply a machine learning model to classify, the data set must have a target variable to "feed" the training set. Let it find the mapping rules of feature attributes to target variables. If consumers' demand for coupons can be accurately predicted, the reduction of merchant marketing costs and the increase in consumer stickiness to O2O platforms can have a significant impact. What models or algorithms should we use to forecast the coupon redemption? What if the chosen model doesn't work? How to evaluate the chosen models? In this paper, we are going to discuss and solve these problems.

3 Dataset and Research Methodology

This section will define, clean, and split the O2O coupon redemption data obtained from Tianchi.

3.1 The O2O coupon redemption dataset

3.1.1 Data collection. We selected the data from the competition of Offline to Online (O2O) Prediction of Coupon Redemption on website Tianchi (https://tianchi.aliyun.com/dataset/59). The data are outlined in Table 1. This data shows online and offline user consumption data from 2016/01/01 to 2016/06/30 (This paper's date format is year/month/day). The research aims to predict the probability of customers redeeming a coupon within 15 days of receiving it. To protect the privacy of users and merchants, data is desensitized and under biased sampling. The results are evaluated based on the average AUC value. That is, the AUC value is calculated for every coupon_id. The average of each AUC value is the evaluation score. The dataset contains 44 features in total.

3.1.2 Data cleaning (preprocessing). In order to use machine learning to answer our prediction questions, we need to clean the data to match the input format of the algorithm [7]. For this reason, we should the clean data set before analyzing. The data provided by Tianchi has been cleaned. Take the training set "ccf_offline_stage1_train" as an example. The entire data set has 1,754,884 rows of data, involving "User_id", "Merchant_id", "Coupon_id", "Discount_rate", "Distance", "Date_received" and "Date" 7 fields. When the data of "Coupon_id", "Discount_rate", "Date_received" is "NULL" indicates that the customer did not receive the coupon, when the data of "Date" is "NULL" indicates that the customer did not apply for the coupon, and when the data of "Distance" is "NULL" indicates that the customer did not provide GPS information, but based on the other fields being not null, indicating that the data still has sample significance. Another two fields "User_id" and "Merchant_id" do not have "NULL" data, which means that there is no invalid data in the offline data set. In addition, through the repeatability check, it was found that there are 36,307 rows duplicate data sets in the offline data set, accounting for about 2% of the total data set, which will not have a significant impact on the prediction results. Similarly, the same data clean was performed on the "ccf_offline_stage1_test_revised" and "ccf_online_stage1_train" datasets, no invalid data was found, and the data duplication was less than 2%. Therefore, it can be inferred that the data provided

Table 1: Data fields.

Offline Data		Online Data	
Field	Description	Field	Description
User_id	User ID	User_id	User ID
Merchant_id	Merchant ID	Merchant_id	Merchant ID
Coupon_id	Coupon ID, when coupon_id = null, this means that a coupon has not been redeemed. In such case, Discount_rate and Date_received don't matter.	Action Coupon_id	0 - click, 1 - buy, 2 – get coupon when coupon_id = null, this means that a coupon has not been redeemed. In such case, Discount_rate and Date_received don't matter.
Discount_rate	Discount rate, range in [0,1]	Discount_rate	Discount rate, range in [0,1]. "fixed" means Limited Time Offer.
Distance	500x, the distance from the nearest shop around the user for locations in which a user is most active.	Date_received	Date the coupon is received
		Discount_rate	Purchase date, When Date=null & Coupon_id!= null, users receive coupon but don't redeem it; When Date!=null & Coupon_id= null, purchase happened but no coupon had been received; When Date!=null & Coupon_id!= null, "Date" in which the coupon was used.

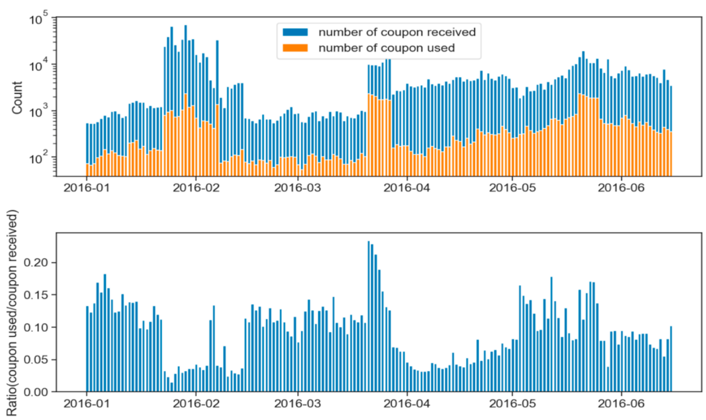

Figure 1: Coupon receiving and used condition per day.

by Tianchi has been cleaned up. Figure 1 shows the number of coupons received and used per day.

3.1.3 Data set partition. It was mentioned above that the prediction of O2O coupon redemption is a binary classification problem. To apply a machine learning model for classification, a target variable must be in the data set, so that the training set can train the machine learning model to find the mapping rules of feature attributes to the target variable. Therefore, first we need to construct the target variable in the training set. Our goal is "Predict the redemption of coupon within 15 days after customer receiving coupons in July 2016", so we need to construct a "label" attribute for the training set, and "label" identifies a record whether the coupon was redeemed within 15 days. Appendix B shows detail of data overview and set partition.

It is worth pointing out again that after the construction, the training set and the test set are identical in terms of field types. The only difference is that the training set is historical data and therefore contains the target variable (that is "label"). The set is future data for 2016/07/01-2016/07/31 and therefore does not contain the target variable. What a machine learning model does is to learn the mapping relationship between the feature set on the training set and the target variable, generate specific discrimination rules, and achieve the ability to predict the value of the target variable based on the feature set on the test set.

Summary of above discussion, this paper use sliding window method to split the raw data. The smaller the feature interval, the more trained data sets are obtained. By drawing multiple training sets, on the one hand, training samples can be added on the other hand, cross-validation experiments can be performed to facilitate the adjustment. The interval of forecast excludes the period 2016/06/16-2016/06/30, because our prediction is "Predict the redemption of coupon within 15 days after customer receiving coupons", the raw data doesn't have full record of coupon redemption for the period 2016/06/16-2016/06/30.

3.2 Feature extraction

A feature is a digital representation of the original data. There are many ways to convert raw data into digital metrics, which is why the features look like a lot. Naturally, raw data for extracting features must be available. Some features may not have a clear correlation with the model, and some models have a high degree of fit to the features, and vice versa. Good features are highly correlated with the target and can be easily obtained through the model (Zheng, 2018). Feature engineering is the process of refining accurate features based on known raw data, prediction targets, and models. And the feature engineering is most important process in the forecasting via machine learning.

The number of features is also important. If there are not enough features to provide information, the model will not be able to obtain ideal prediction results. However, if there are too many functions or their correlation is not high, the model will become difficult to train (this will consume too many resources and the training results are not ideal), which will reduce the prediction accuracy of the model. Some features are deformed during training, which can affect the performance of the model.

In this paper, in order to predict coupon redemption in July, consider using many features to build a good model. Based on intuition and knowledge of O2O demand modeling, we identified five factors that may affect coupon redemption. User, merchant, coupon, user-merchant and other features (leakage) are five categories of features which are likely to be correlative to the prediction model.

3.3 XGBoost model

XGBoost is a novel choice for this application due to its **efficiency**, ability to handle **sparse data**, **regularization** to avoid overfitting, **flexibility** in defining custom objectives, **tree pruning**, and built-in **cross-validation**. It often outperforms other models like SVMs, Random Forests, or Neural Networks, especially on structured data, and is easier to tune. However, the best model always depends on your specific project needs.

During training we found that the higher of the boost number will cause higher train-AUC, but this will cause over fitting issue which will cause prediction AUC decrease. After our repeated debugging, we found that 3500 is best value. Parameter values are: Booster=gbtree, Objective=rank:pairwise, Eval_metric=auc, Gamma=0.1, Min_child_weight=1.1, Max_depth=5, Lambda=10, Subsample=0.7, Colsample_bytree =0.7, Colsample_bylevel=0.7, Eta=0.01, Tree_method=exact, Seed=0, Nthread=12.

4 Results

In the prior chapter, we mentioned that we set the number of boosts as 3500. During training, we found that the higher of the boost number will cause higher train-AUC, but this will cause over fitting issue which will cause prediction AUC decrease. After our repeated debugging, we found that 3500 is best value. The AUC for XGBoost is 0.7961, which is far greater than 0.5, indicating that the prediction results of the model have considerable accuracy [11].

Each feature plays a role in XGBoost model, the Figure 2 shows the top 10 important features, and the higher score means the more important the feature is. We find that five of the top 10 features belong to Merchant, four belong to Coupon, one belongs to User, and nine of them are directly affected by the merchant (Merchant 5 + Coupon 4). This is in line with normal business logic, because although the main users of coupons are consumers, the factors of merchants play a leading role. Results are illustrated in Figure 2. Among the Top 10 features, one has a score of over 8,000, three has a score of over 6,000, three has a score of over 5,000, and three has a score of over 4,000. The median is 5,604.5 points and the average is 5,812.8 points.

The above top10 features can be classified into three categories, namely the number of coupons issued, the discount strength of coupons, and the restrictions on the use of coupons, which is in line with business logic. For the new Meituan merchants, they can formulate corresponding coupons based on the ranking of the characteristics of the coupons that are influenced by this article, so as to realize the accurate distribution of coupons. In addition, for those large and powerful businesses, such as the Meituan platform, it is possible to accurately segment the consumer groups and design customized coupons for each type of consumer, so that each customer could receive the desired coupon.

5 Conclusion

This paper uses 6 months consumption data of a domestic O2O platform as a research sample and uses machine learning modeling methods to predict the probability of the user's redemption within 15 days after receiving the coupon in the next half month. Through the feature engineering construction of the five aspects of merchants, users, coupons, user-merchants, and other aspects, the AUC value of the prediction result obtained by the XGBoost model is 0.7961. The methodology can be applied in other O2O settings as well due to its satisfactory performance and computational efficiency.

Through research and analysis of the prediction results of the user's coupon redemption probability, it is found that if the coupons are accurately distributed according to the probability prediction results, the net profit margin on sales of merchants will be nearly 50% higher than random distribution. It is found that the customer flow, coupon issuance volume, discount rate, coupon usage ratio, distance between merchants and users, coupon verification interval, coupon usage restrictions, the proportion of coupons used in merchant sales and the time of receipt of coupons are strongly correlated with the coupon redemption.

With the research conclusions of this paper, we give out the following recommendations:

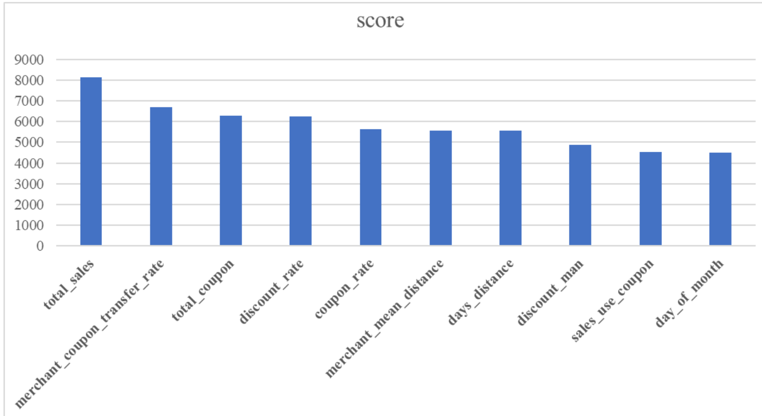

Figure 2: The features scores rank for XGBoost.

(1) When determining the type and plan of the coupon, the merchant's marketing purpose should be considered. If it is just to increase sales or attract new users, that is, merchants want to obtain a higher coupon redemption rate without considering profits, they should issue coupons with larger denominations and no thresholds. But if merchants want to increase the profit of via coupon issuances, they need to consider the balance between the discount rate of the coupon, the amount of issuance, the threshold of use, and the business cost and production cost. When these relations reach the optimal solution, merchant can achieve the maximum profit.

(2) When the merchant or platform analyzes and predicts the user's consumption data, it can accurately issue coupons to the target customer, the merchant can roughly predict the customer's demand in the next period, and the merchant can use this to achieve precise preparation to reduce inventory costs, operating costs, and labor costs, thereby further improving profitability.

(3) For those newly opened merchants, because they have not accumulated customer consumption data, they can formulate coupon to meet their potential target customers based on this paper analysis of the ten factors that affect the coupon redemption rate and their own characteristics. And according to the consumption data accumulated in the later period, they could make predictions through the prediction model, to continuously improve the accuracy rate of coupon issuance and optimize their inventory and capacity [12-14].

Acknowledgments

Grant sponsor: National Natural Science Foundation of China [Grant no: 72271133].

References

[1] Wang, C., Leng, M. and Liang, L. 2018. Choosing an online retail channel for a manufacturer: Direct sales or consignment? International Journal of Production Economics. 195, (Jan. 2018), 338–358. DOI:https://doi.org/10.1016/j.ijpe.2017.11.001.
[2] Phang, C.W., Tan, C.-H., Sutanto, J., Magagna, F. and Lu, X. 2014. Leveraging O2O commerce for product promotion: An Empirical investigation in Mainland China. IEEE Transactions on Engineering Management. 61, 4 (Nov. 2014), 623–632. DOI:https://doi.org/10.1109/tem.2014.2354056.
[3] Belch, G.E. and Belch, M.A. 1997. Advertising and Promotion: An Integrated Marketing Communications perspective.
[4] He, Z., Cheng, T.C.E., Dong, J. and Wang, S. 2016. Evolutionary location and pricing strategies for service merchants in competitive O2O markets. European Journal of Operational Research. 254, 2 (Oct. 2016), 595–609. DOI:https://doi.org/10.1016/j.ejor.2016.03.030.
[5] Venkatesan, R. and Farris, P.W. 2012. Measuring and Managing Returns from Retailer-Customized Coupon Campaigns. Journal of Marketing. 76, 1 (Jan. 2012), 76–94. DOI:https://doi.org/10.1509/jm.10.0162.
[6] Li, B. and Arreola-Risa, A. 2022. Managing a bone marrow transplant centre to maximise patients' health benefits. International Journal of Production Research. 61, 6 (Mar. 2022), 1771–1795. DOI:https://doi.org/10.1080/00207543.2022.2047239.
[7] Yuan, D. 2015. Applications of machine learning: consumer credit risk analysis.
[8] Mitra, S. *et al.* 2008. Introduction to machine learning and bioinformatics.
[9] Noble, S.U. 2018. Algorithms of oppression: How Search Engines Reinforce Racism. NYU Press.
[10] Ligeza, A. 1995. Artificial Intelligence: a Modern approach. Neurocomputing. 9, 2 (Oct. 1995), 215–218. DOI:https://doi.org/10.1016/0925-2312(95)90020-9.
[11] Fawcett, T. 2006. An introduction to ROC analysis. Pattern Recognition Letters. 27, 8 (Jun. 2006), 861–874. DOI:https://doi.org/10.1016/j.patrec.2005.10.010.
[12] Li, B., Tan, Z., Arreola-Risa, A. and Huang, Y. 2023. On the improvement of uncertain cloud service capacity. International Journal of Production Economics. 258, (Apr. 2023), 108779. DOI:https://doi.org/10.1016/j.ijpe.2023.108779.
[13] Lin, S., Li, B., Arreola-Risa, A. and Huang, Y. 2022. Optimizing a single-product production-inventory system under constant absolute risk aversion. Top. 31, 3 (Dec. 2022), 510–537. DOI:https://doi.org/10.1007/s11750-022-00650-4.
[14] Ji, Q., Zhou, S., Li, B. and Chen, Z. 2020. Component ordering strategies in assembly systems with uncertain capacity and random yield. Applied Mathematical Modelling. 88, (Dec. 2020), 715–730. DOI:https://doi.org/10.1016/j.apm.2020.06.065.