TELKOMNIKA, Vol.15, No.1, March 2017, pp. 351~356
ISSN: 1693-6930, accredited **A** by DIKTI, Decree No: 58/DIKTI/Kep/2013
DOI: 10.12928/TELKOMNIKA.v15i1.4269

A Two Stage Classification Model for Call Center Purchase Prediction

Kai Shuang, Kai-Ze Ding, Xi-Hao Liu, Xiao-Le Wen
State Key Laboratory of Networking & Switching Technology, Beijing Universiy of Posts & Telecommunications, Beijing, China
Corresponding author, e-mail: shuangk@bupt.edu.cn, kzding@bupt.edu.cn, lxh56b@bupt.edu.cn, xlwen@bupt.edu.cn

Abstract

In call center [1] product recommendation field, call center as an organization between users and telecom operator, doesn't have permission to access users' specific information and the detailed products information. Accordingly, rule-based selection method is common used to predict user purchase behavior by the call center. Unfortunately, rule-based approach not only ignores the user's previous behavior information entirely, and it is difficult to make use of the existing interaction records between users and products. Consequently, it will not get desired results if we just use the basic selection method to predict user purchase behavior directly, because the problem is that the features straightly extracted from the interaction data records are limited. In order to solve the problem above, this paper proposes a two-stage algorithm that based on K-Means Clustering Algorithm [2] and SVM [3, 4] Classification Algorithm. Firstly, we get the potential category information of products by K-Means Clustering Algorithm, and then use SVM Classification Model to predict users purchasing behavior. This two-stage prediction model not only solves the feature shortage problem, but also gives full consideration to the potential features between users and product categories, which can help us to gain significant performance in call center product recommendation field.

Keywords*: call center, K-Means, purchase prediction, SVM*

1. Introduction

With the popularity of smart phone in the world, the call center can recommend products to users via phone calls, which contains huge business opportunities. But as an organization between users and telecom operator, call center does not have permission to access specific information about the users, and it will not keep detailed product information in a log, which causes some difficulties to improve products recommendation accuracy. Call center selects users mainly by simple rules so far. However, the success ratio in this way is low, and valuable information cannot be fully excavated. Therefore, inappropriate recommendation call not only causes highly cost, but also annoys the users. So we have to solve the problem that how to improve the accuracy of the recommendation.

Predicting user purchase behavior [5, 6] usually can be transformed into a binary classification problem [7, 8]. Although the usage of classification model can increase the prediction success rate, there are still some obvious limitations in the call center dataset for model-based methods. There are two important problems that we will address:

User and product information deficiency: features that can be extracted from the original data is limited. What call center has is only behavior records between users and products that recorded by telephonist.

Category based features deficiency: because there is no information about the category of products in the data records, and category based features cannot be extracted directly from the original data records.

From a more general point of view, this kind of problem [9-11] is not only encountered by product recommendation field in call center, and in many cases, when we want to make a recommendation to existing users some products, the information we have about users, as well as the products may not reliable or not available. We can only predict users purchase behavior based on existing user behavior records (whether or not purchased certain product). However,

Received November 25, 2016; Revised January 7, 2017; Accepted January 26, 2017

the existing data can be used to extract feature is very limited. The limited number of features that can be extracted lacks product category information will lead to unsatisfying prediction result. How to increase the number of useful features effectively and get the categories features will become the key point of this kind of problem when we hope to improve the accuracy of prediction model. In order to solve the problem above, this paper proposes a two-stage prediction model that based on Extended K-Means Clustering Algorithm and SVM Classification Algorithm. This two-stage classification prediction model firstly gets the classification information by Extended K-Means Clustering Algorithm, then extract classification features and use SVM classification algorithm to predict users purchasing behavior. Through our experiments, we make a comparison among results of our prediction model, random selection, basic LR classification model, and basic SVM classification model, to verify the result of our approach is better than other methods.

2. Related Works

Typically, to predict whether a user will purchase a certain product or not can be transformed into a classification problem, more accurately, a binary classification problem [12]. Researchers have proposed a series of classification methods to predict the purchasing behavior by supervised learning method. Logistic Regression (LR) and Support Vector Machine (SVM) classification model are commonly used methods to predict user purchase behavior.

Logistic Regression [13] is a kind of classification model which is widely used and has a good result. It is been proved effective in the field of product recommendation in recent years. Allenby used LR to predict household purchase behavior [14]. Logistic Regression can fit to the data very well in linear dataset, and has the capacity to find the parameters through the gradient descent algorithm [15]. The model is simple, and the training speed is fast. But under the circumstance of nonlinear, logistic Regression cannot fit data well, and in call center product recommendation which features are relatively insufficient, there is no obvious linear separable condition to guarantee that the data is linearly separable, logistic regression is difficult to meet these two conditions. So we need a method which is able to fit nonlinear data.

Support Vector Machine is proposed by Vapnik in 1998, and has been widely used in the 90's. SVM model can use the kernel method mapping the features from low dimensional space to high dimensional space [16]. This characteristic can transform a nonlinear classification problem into a higher dimensional linear classification which suits the insufficient features in call center product recommendation problem. Xu [17] used SVM model build a recommendation system for TV programs and acquire a desirable result.

Although the support vector function can ease the problem that feature dimension is too low to classify, but it is not enough to solve this problem only in model aspect, we still need more useful features if we want to get better results. SVM model will not gain better result when features are insufficient [18]. So we should not only use features that can be extracted directly, and do not take into consideration the potential information behind original data, such as the product category features.

In order to introduce category based features, we should split all products into different categories. Classification algorithms and clustering algorithms are commonly used to accomplish this kind of work. While supervising classification algorithms require labeled dataset to train a classification model. So we choose unsupervised clustering model to get product categories.

In next section, we will introduce the two-stage SVM classification prediction model-based on K-means feature extending to overcome above problems.

3. A Two-Stage Prediction Model

The key problem of purchase prediction in call center is how to improve the accuracy of model prediction result in the case of insufficient features. In order to improve the prediction model effect further on the basis of the existing features, we need to extract the potential features as far as possible in the existing data sets [19]. At the same time, choosing a appropriate classification model which can alleviate insufficient feature problem will be helpful to improve the predictive effect.

The most serious problem of data which stored in call center is that we cannot extract enough useful features directly, because the call center does not like telecom operators has the authority to save the detailed information of users, and even product attributes can also not be obtained. Therefore, we should extract features from users purchasing records as much as possible, and use cluster model to get the category features of product to achieve the purpose of feature expansion.

3.1. Feature Extraction

The most important premise of model training is to extract the effective features in the existing data, and appropriate feature selection can improve the performance of the model prediction. The feature vector is composed of a set of features, and each feature vector can be represented as:

$$< \overrightarrow{user_i, item_i} > = (v_{i1}, v_{i2}, v_{i3}, v_{i4}, \ldots)^T \tag{1}$$

The problem studied in this paper is to predict the user purchase behavior, so the feature should be extracted based on the behavior records between users and products. Each feature vector is uniquely identified by < user, item > (user* represent all users, item* represent all products), and (user, item) represents a group of features between one user and one item (user* represent all users, item* represent all products).

Through analysis, the features that we can extract for every group is four features: success rate(S rate), failure rate (F rate), call reject rate(R rate), call duration. We will firstly extract these features in three groups in this paper. Features can be extracted are as follows:

Tabel 1. Feature Set 1

(user, item*)	S rate	F rate	R rate	duration
(user, item)	S rate	F rate	R rate	duration
(user*, item)	S rate	F rate	R rate	duration

From above table, we can see that these features do not contain product category features, which ignore the potential information in the call center dataset. If we just use these features to train the prediction model, we cannot get ideal result. So the key step is to obtain the category features of product. We will introduce a method use K-means model to solve this problem.

3.2. Feature Extraction Based on K-Means Clustering

This section focus on category based features extraction problem.

How to get different product categories? It can be seen as a multiclass classification problem [20]. There are many different supervising multi-classification models can be used. But supervising model need a labeled dataset to train the prediction model. Since we only have the interaction records between users and products, an un-supervising approach should be more appropriate.

Clustering algorithm [21-23] as an un-supervising model can help us extract category feature information and then add new useful features to improve the prediction result by dividing the existing data sets into several clusters.

K-Means clustering algorithm is a kind of non-supervised real-time clustering algorithm proposed by MacQueen. The principle of K-Means is to divide the data into a predetermined clusters number K based on minimizing error function. The algorithm is easy and appropriate to deal with a large number of data. K-Means algorithm is one of clustering algorithms which based on data division, it has good clustering effect and the advantage of fast convergence.

The K-Means algorithm can divide data into K clusters by unsupervised method. In the field of call center product recommendation, we can get a large number of user behavior data from the call center. K-Means method converges fast, and can maintain good results in large amount of data. In this section, we use the K-Means model to cluster the products, and then after getting the corresponding product clusters, we will extract the relevant category features as the input of the final prediction model.

The features we used in this paper for clustering is:
1. The purchase rate of the product.
2. Average call duration of the product.
3. Average call rejection rate of the product.

After clustering, we can add product category features to the original feature set in the previous section. (item+ means products in the same category):

Tabel 2. Feature Set 2

(user, item*)	S rate	F rate	R rate	duration
(user, item)	S rate	F rate	R rate	duration
(user*, item)	S rate	F rate	R rate	duration
(user, item+)	S rate	F rate	R rate	duration

By adding the category features, the original feature set is enriched, and the potential product category information is added, so that the predict effect of this classification model can be promoted to a certain extent.

3.3. SVM Model Prediction

By using the K-means clustering algorithm to get product category features, we can extend the feature set which classification prediction model need. As we know, products purchase prediction problem can be transformed into a binary classification problem. In this way, we can predict purchase behavior through a variety of classification models effectively.

Logistic Regression and Support Vector Machine are commonly used binary classification models. As a rational model in insufficient feature circumstance, the SVM model usually performs surprisingly well by feature mapping. In the next section, we compare different methods to choose a better prediction model. Based on the results, we have decided to use the SVM model to predict the purchasing behavior of the call center problem.

Based on K-Means clustering feature extension, we add the product category features into original feature set, then SVM classification model can be trained using training set. And we can use the trained model to predict new purchase behavior.

4. Experiments

The data sets which we use are recommendation sales records from the call center to the mobile phone users. As this paper mainly for the regular users for product sales prediction, so we selected 50000 users from who had as least 5 or more different product interactive records to predict. And the product number is 200.

The experiment uses rule-based prediction as the basic baseline. The other experiments all rely on classification model, so we need to build the training and test dataset. Assuming that each user has N product interaction records, we can predict the Nth result through the previous n-1 times behavior records. Therefore, we use the previous n-2 sales records as the training data set, and the n-1 time sales result as the label of training data set; while the previous n-1 sales records as the test data set, and the last sale result as the label of test data set. We train the model through the training data set, and test the effect on the test data set. The recommendation result of the model is measured by the accuracy and recall rate as well as the f1-score.

The experimental results are divided into the following five categories: 1) Rule-based prediction; 2) LR model-based prediction; 3) SVM model-based prediction; 4) LR model prediction with K-Means feature extension (K-LR); 5) SVM model prediction with K-Means feature extension (K-SVM).

Through experiments, we obtained the prediction results of each method in the existing data set, and the results are shown in Table 3. It should be noticed that classification features which obtained from K-Means clustering algorithm extending are not included in the feature set that used in LR or SVM model directly. Moreover, the K-LR and K-SVM perform best when cluster number is set 10.

Tabel 3. Results of Each Method

Method	Precision	Recall	F1-score
Rule	24.9%	30.2%	27.3%
LR	28.8%	43.3%	34.6%
SVM	30.4%	42.6%	35.4%
K-LR	30.2%	44.2%	36.3%
K-SVM	32.2%	43.8%	37.1%

From above table, we first note that the model-based prediction results is better than rule-based selection in all aspects. As a basic baseline, the rule-based approach cannot fully consider previous purchase behavior between users and products. And since we cannot access user and product information, rule-based selection usually performs poorly. On the other hand, previous purchase behavior information can be add in the features which we used in model-based methods, and the recall of model-based methods is much better than random selection.

Next, the result shows that SVM model performs better than LR, the K-SVM result is better than K-LR as well. It indicates that SVM can deal with the insufficient feature problem better. SVM transforms a nonlinear classification problem to a linear classification problem by its feature mapping property, SVM usually performs well when feature dimension is too low. But we notice that the improvement is not significant, more useful features should be added to train a better classification model.

Moreover, The K-LR and K-SVM shows that the results of model-based methods can be improved by adding product category features through K-means clustering algorithm. And SVM model is still better than LR model after adding category features from K-means clustering extension. So our proposed two-stage prediction model (K-SVM) is a rational choice when the feature dimension is too low.

At the same time, we also observed the changing trend of prediction results when the cluster number changes. Experimental results are shown in Figure 1.

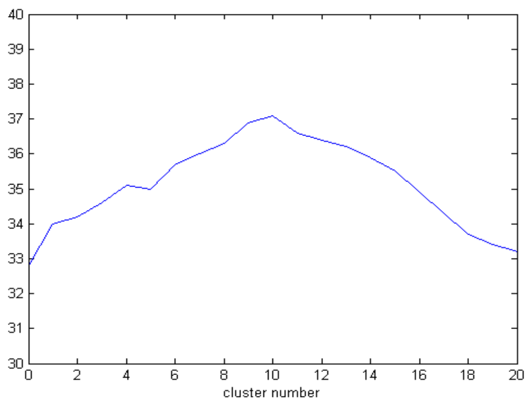

Figure 1. F1-Score

During the experiment, we found that the classification result is best when the number of clusters is 10 for all 200 products. How to choose a reasonable cluster number should be studied in our future work.

5. Conclusion

This paper according to the product recommendation scenes in call center which have limited useful feature between users and products, proposes a two-stage prediction model-based on K-Means Clustering feature extension and SVM classification model prediction bring a desirable result.

Firstly, the proposed model clusters products by means of K-Means clustering, and then extracts the features of users to product categories. Furthermore, this model will add category features to original classification feature set for classification prediction. The experiments section shows that this approach has certain improvement in effect compared to rule-based selection and basic classification model prediction. Therefore, when we face the insufficient feature problem or lacking of product and user information in product recommendation, we can try to obtain more useful classification features by way of K-Means feature extending and use SVM to train a prediction model. This method can be used to predict the purchasing behavior of users under certain circumstances, and improves the performance of prediction.

References

[1] Mauricio A Valle, Samuel Varas, Gonzalo A Ruz. Job performance prediction in a call center using a naive bayes classifier. *Expert Systems with Applications*. 2012; 39(11): 9939-9945.

[2] Witten IH, Frank E, Hall MA. Data Mining: Practical Machine Learning Tools and Techniques. 3rd ed. United States: Morgan Kaufmann Publishers. 2010: 147-187.

[3] Chang CC, Lin CJ. LIBSVM: A library for support vector machines. *ACM Trans. on Intelligent Systems and Technology (TIST)*. 2011; 2(3): 27.

[4] Cristianini N, Shawe-Taylor J. An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. United Kingdom: Cambridge University Press. 2000.

[5] Dirk Van den Poel, Wouter Buckinx. Predicting online-purchasing behavior. *European Journal of Operational Research*. 2005; 166(2): 557-575.

[6] F Ricci, L Rokach, B Shapira, PB Kantor. Recommender Systems Handbook. Springer. 2011.

[7] Eunju Kim, Wooju Kim, Yillbyung Lee. Combination of multiple classifiers for the customer's purchase behavior prediction. *Decision Support Systems*. 2003; 34(2): 167-175.

[8] Moon S, Russell GJ. Predicting product purchase from inferred customer similarity: An autologistic model approach. *Management Science*. 2008; 54(1): 71-82.

[9] Bobadilla J, Serradilla F. *The effect of sparsity on collaborative filtering metrics.* Proceedings of the Twentieth Australasian Conference on Australasian Database. Darlinghurst. 2009; 92: 9-18.

[10] Grčar M, Mladenič D, Fortuna B, Grobelnik M. Data Sparsity Issues in the Collaborative Filtering Framework. Springer, Heidelberg. 2005; 4198: 58-76.

[11] Xiaohua Sun. Research of Sparsity and Cold Start Problem in Collaborative Filtering. Hangzhou: Zhejiang University. 2005.

[12] Duen-Ren Liu, Ya-Yueh Shih. Hybrid Approaches to Product Recommendation Based on Customer Lifetime Value and Purchase Preferences. *The Journal of System and Software*. 2005; 77: 181-191.

[13] Wang GF, Ren Y, Duan LZ, et al. *An Optimized Collaborative Filtering Approach with Item Hierarchy Interestingness*. International Conference on Business Computing and Global Information (BCGIN). 2011: 633-636.

[14] Greg M Allenby, Peter J Lenk. Modeling household purchase behavior with logistic normal regression. *Journal of the American Statistical Association*. 1994; 89(428): 1218-1231.

[15] Jan Snyman. Practical mathematical optimization: an introduction to basic optimization theory and classical and new gradient-based algorithms. Springer Science & Business Media. 2005: 97.

[16] Sarwar BM, Karypis G, Konstan JA, Riedl J. *Application of dimensionality reduction in recommender system-A case study*. ACM WebKDD 2000 Workshop. 2000: 82-90.

[17] Truong KQ, Ishikawa F, Honiden S. Improving Accuracy of Recommender System by Item Clustering. *IEICE Transactions on Information and Systems*. 2007; 90(D-I).

[18] D Agarwal, BC Chen. Regression-based latent factor models. *KDD*. 2009; 19-28.

[19] Jin An Xu, Kenji Araki. *A svmbased personal recommendation system for TV programs*. Proceedings of the Multi-Media Modelling Conference. 2006; 4.

[20] Christopher M Bishop. Pattern recognition. *Machine Learning*. 2006.

[21] Karypis G. *Evaluation of item-based top-n recommendation algorithms*. Proceedings of the 10th International Conference on Information and Knowledge Management. New York, USA: ACM Press. 2001: 247-254.

[22] Guha S, Rastogi R, Shim K. ROCK: A Robust Clustering Algorithm for Categorical Attributes. *ICDE.* 1999; 15.

[23] B Mehta, W Nejdl. Unsupervised strategies for shilling detection and robust collaborative filtering. *User Modeling and User Adapted Interaction*. 2009; 19(1): 65-97.