TrORF: Building Trading Areas around Organizations Based on Machine Learning Techniques

Yuwei Xia[1], Min Fu[2, 3], Xi Zheng[3], Binqing Wu[4], Zhihao Ke[5]
[1]School of Economic Information Engineering, Southwestern University of Finance and Economics, Chengdu, China
[2]Alibaba Group, Hangzhou, China
[3]Department of Computing, Macquarie University, Sydney, Australia
[4]Department of Computer Science, Leeds Joint School, Southwest Jiaotong University, Chengdu, China
[5]State Key Laboratory of Traction Power, Southwest Jiaotong University, Chengdu, China
xiayuwei0226@yahoo.com, hanhao.fm@alibaba-inc.com, james.zheng@mq.edu.au, wubinqing@my.swjtu.edu.cn, 1048001458@my.swjtu.edu.cn

***Abstract*—Nowadays, making investments in trading areas around organizations is becoming increasingly pervasive since more and more organizations (e.g. universities) have been relocated in suburban-districts far from downtown. It generally requires a fairly long time for investors to make decisions and establish mature trading areas around organizations. Existing researches that focus on optimal location selection ignore the latent relationship between organizations and their surrounding business types. Therefore, this paper proposes a machine-learning based methodology of determining business types and trading areas, named TrORF, to help investors select business types and build trading areas around organizations. We implement our approach and evaluate it by using 13 real-world universities as our case studies. We compare the business type recommendation accuracy of the 13 real-world universities between our method and the other four existing methods. The experiment results indicate that our approach has strong ability to build the suitable business types around each university's trading area since its average recommendation accuracy is 98.23%, and it improves on other approaches by at least 13%.**

Keywords—Machine Learning, Data Mining, Trading Area Building, Business Types Determination

I. INTRODUCTION

Trading areas around organizations have become more significant for people due to the fact that land resources are limited and every inch of space should be utilized carefully to obtain high payback. A trading area has various business types, and if investors could find out the suitable business types for an organization, a well-developed trading area could be built in a relatively short time. However, this is not an easy task since it is usually done by conducting a survey [8] or by expert experience and intuition.

Most of the researches on business decisions making for e-commerce businesses focus on optimal location selection[9]. They mainly aim to find business types based on the surrounding location's characteristics [2], or make location recommendations based on mass check-in data on the internet [4], or select the most important factors that affect the location determination [5], or combine location selection with other factors including the prices of alternative products [11]. The main problem is that they all ignore the underlying latent relationship between organizations (e.g., universities) and surrounding business types.

In this paper, we propose a novel and innovative methodology, named TrORF, to help investors build trading areas around organizations. TrORF is based on historical data analysis and machine learning techniques. We first collect data related to organizations and business types around each organization under the assumption that there exists latent correlation between organizations and business types around them. After data preprocessing, we rely on machine learning techniques to train the historical data to attain a suitable business types prediction model, which is used to predict whether a business type is suitable for an organization. In our proposed method, we choose Random Forest (RF) as the machine learning model to predict trading areas around organizations because RF model has great capability in performing complete cross analysis on all the data features and determining the most correct relationship between features and labels.

We implement and evaluate our method by using 13 real-world universities as our case studies. We first compare the accuracy and f1-score of our model with other four methods including Gradient Boosting Decision Tree (GBDT), Support Vector Machine (SVM), Deep Neural Network (DNN) and Naive Bayes (NB) [12][15][16] [7]. The results show that our method performs best where the accuracy of our method is 99.75%, GBDT 88.14%, and SVM 71.59%, DNN 62.61%, and NB 64.83%, and the f1-score of our method is 99.59%, GBDT 88.97%, and SVM 71.31%, DNN 61.83%, and NB 65.45%. Then we use our model as well as the other four methods to make recommendation of business types for 13 universities, and compare the business type recommendation accuracy between our method and the other four methods. The results show that our model is capable of predicting the suitable trading areas for universities. The average accuracy comes up to 98.23%, and the mean accuracy of GBDT, SVM, DNN and NB is 85.77%, 67.58%, 60.42% and 64.93%, respectively.

The main research contributions of this paper are: 1) we propose an intelligent method for building trading areas around organizations; 2) we demonstrate the feasibility of our approach through experimental evaluation.

0-7695-6728-2/19/$31.00 ©2019 IEEE
DOI 10.1109/ICICAS48597.2019.00032

The remainder of this paper is organized as follows: Section II introduces the background of our research; Section III illustrates our proposed method of determining trading areas for organizations; Section IV provides the experimental evaluation of our proposed method; Section V discusses some related work; Section VI provides the conclusion and our future work.

II. Background

A. Economy Background

The economic development level is a key element in deciding whether people would spend their extra money on risky activities like starting a business or just save their money in the bank. Only people of vision would make some investments like starting a business. The high-growing economy in these years stimulate an increasing amount of people to become more and more willing to take chances to earn more through doing business. Furthermore, some effective government proposals also help in encouraging people to start a business.

B. Current Situation of Organizations

Due to rapid economic development of China, land resources have become more and more limited and every square inch of land should be properly used. Moreover, people nowadays are getting unsatisfied with the facilities around their living areas, which brings commercial opportunities to investors. A large number of organizations (e.g. universities) have been relocated in suburban areas far from city center. It could mean a lot of difficulties if the surroundings of those organizations could not meet the needs of residents. If the trading areas around the organizations are properly built, it can even accelerate the economic development of this district.

III. Our Proposed Methodology

Our proposed methodology is designed to have the ability to help predict suitable business types and build trading areas around organizations. We define two ***research questions***: 1) how to decide whether a business type is suitable for an organization? 2) what is the latent relationship between an organization and business types around it? Our proposed methodology, TrORF, addresses these two research questions. As far as we know, we are the first to propose such a method. The problem can be defined in the following way:

Suppose the total number of business types is M, and the total number of organizations is N, the set of business types is B, the set of organizations is O, the predicted suitable set of business types of each organization j is P_j, and whether a business type i is suitable for a given organization j is $L_{ij}(L_{ij} \in \{0,1\})$, then we need to determine P_j based on the following formula:

$$P_j = \{B_i | B_i \in B \cap L_{ij} = 1\} \quad (1)$$

where:

$$L_{ij} = f(B_i, O_j), 1 \le i \le \mathrm{M}, 1 \le j \le \mathrm{N} \quad (2)$$

where $f(.)$ is the function which stands for the machine learning model used for anticipating whether a business type is suitable for a give organization.

A. Overview

The process of our methodology mainly includes four procedures:

procedure 1: Data collection. This procedure contains two steps: 1) we first utilize various methods to collect data of all kinds of aspects; 2) we then determine the characteristics we will use to represent each business type and features of each organization.

procedure 2: Data preparation. This procedure contains four steps: 1) we conduct data preprocessing which includes careful data scrutiny, data transformation and data categorization; 2) then we use the approach of data filling to fill in the blanks of some data. Moreover, we add ***Labels*** to the data; 3) we apply the technique of data expansion to expand the total amount of data to make it more adequate for training; 4) finally, transductive learning is performed to obtain the complete dataset.

procedure 3: Data mining. This procedure contains three steps: 1) we select the best performance model from several candidates; 2) we adjust the optimal values of the key parameters for the model selected; 3) we train the model and use it to predict whether a business type is suitable for an organization.

procedure 4: Evaluation. This procedure is composed of two steps: 1) we cluster the business types predicted by our model and compare with the actual business types; 2) we then compare the prediction accuracy of real-word organizations between our method and the other four methods.

B. Data Collection

We use the case of 13 real-world universities to represent organizations and help explain all steps in detail. Business Type (BT), Open Time (OT), Close Time (CT), Daily Turnover (DT) and Chain Property (CP, which means whether a store is chained or not) are determined as the characteristics of business types.

For organizations, features that are most relevant to them are selected. In the case of universities, we use 10 features in total to represent a university: Establishment Year (EY), Student Number (SN), Teacher Number (TN), Major Number (MN), Department Number (DN), University Type (UT), Master Program Number (MPN), Doctoral Program Number (DPN). Additionally, in order to take excellence of each university into consideration, in this case of universities in China, we consider whether a university is one of ‘Project 211’ universities (P211) and whether it is one of ‘Project 985’ universities (P985).

Based on the actual situation, we suggest stores within the radius of one kilometer from the center of an organization should be considered [13]. During our research, we noticed that almost around every university, there were specific types of stores on the verge of bankruptcy or even closed down, while in contrast, some types of stores enjoyed high popularity.

C. Data Preparation

After collecting data, we now have two datasets: Organization dataset (O) and Business Type dataset (B). However, since the amount of real-world data is relatively small and is not sufficient for model training, we conduct data augmentation to obtain more data. The whole procedure of data preprocessing is shown in Fig. 1 and it contains four steps:

step 1: Data preprocessing. We first scrutinize the dataset to make sure it is totally valid. Then we transform the strings of datasets into numerical forms. Besides, we apply approaches including one-hot encoding [14] to convert categorical variables into machine learning oriented vectors.

step 2: Data filling. We need to complete the business type dataset by supplementing data of unsuitable business types to each organization using the average value or model number of each feature. Then we add ***Label***, which is the prediction target of the machine learning model, for each business type and each organization. Value 1 means this business type really exists around this organization, and value 0 means this business type does not exist around this organization but may exist around some of the others. The overall algorithm is shown in Algorithm 1. We then join the organization data and the business type data which is related to each respective organization to obtain the original dataset.

Algorithm 1: Data Filling Algorithm

Inputs: the business type dataset ***B(BT, OT, CT, DT, CP)***, the length of ***B*** is ***L***, the number of distinct business types ***M*** ,the business types $\boldsymbol{B_i}(\boldsymbol{1 \le i \le M})$, the average level of each business types $\boldsymbol{B_{a_i}}(\boldsymbol{1 \le i \le M})$
Outputs: the final business type set

```
function DataFilling(B, Bi) {
   for (1 → j ; j <= L ; j++) {
      replace B[j] with B[j](BT, OT, CT, DT, CP,1); }
   for (1 → j ; j <= M ; j++ ) {
      if ( Bj ∉ B(BT) ) {
         insert Baj(Bj, OT, CT, DT, CP, 0) into B; }}
   return B; }
```

step 3: Data expansion. we use the technique of data expansion to attain more data of organizations, and the overall algorithm is outlined in Algorithm 2. We get a new dataset by joining the new organization dataset and the business type dataset.

Algorithm 2: Data Expansion Algorithm

Inputs: the real organization dataset ***S***, the expansion multiple ***R***, the number of organizations ***N***, the set of features ***O***
Outputs: the new organization dataset

```
function DataExpansion(S,R,O) {
   for(1 → i ; i <= N;  i++) {
      for(1 → j ; j<= R; j++) {
         change the value of 1 or 2 features of O;}
            return F;}
```

step 4: Transductive learning. we conduct transductive learning to obtain the ***Labels*** of the newly-gained data. RF (Random Forest) is selected to predict the ***Labels*** of the new dataset. After that, we combine the original dataset and the new dataset to acquire the complete dataset and it contains approximately 200,000 pieces of balanced data. The data is divided into the training set, the validation dataset and the prediction dataset with a proportion of 7:2:1.

D. Training Model Selection and Parameter Tuning

In our preliminary study, we notice that RF is able to behave perfectly when conducting binary classification because it has great capability in performing complete cross analysis on all the data features and determining the most correct correlation between features and labels, therefore we choose RF as our training model to make a prediction.

RF is an ensemble supervised machine learning technique which generates a bagging of de-correlated decision trees and allowing them to vote for the most popular class [17][18]. In order to attain higher accuracy, we use the approach of GridSearchCV to adjust the key parameters [10]. When training RF with default parameters, the results show that the accuracy is 95.65%, the precision is 96.37%, and the recall is 95.24% and the f1-score is 95.58%. Optimal parameters of the RF model are as follows: number of trees is 70, maximum depth of the tree is 13, minimum number of samples is 70, and minimum number of features is 9. The importance of features are displayed in Fig. 1. Then we use the approach of 10-Fold Cross Validation on training dataset to measure the validity of our method and the mean accuracy is 99.65%, which means our method is valid.

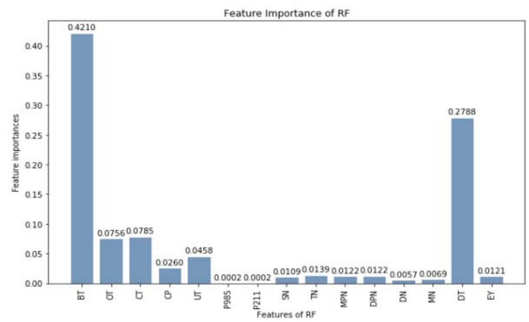

Fig. 1. Feature importance of Random Forest.

E. Data Prediction

After tuning the parameters, we apply our model to the testing dataset and prediction dataset. We use accuracy, precision, recall and f1-score to evaluate the performance of our model. The results are shown in TABLE I.

TABLE I. VALIDATION RESULTS AND PREDICTION RESULTS OF RANDOM FOREST.

	Accuracy	Precision	Recall	F1-score
Validation	99.66%	99.99%	99.43%	99.64%
Prediction	99.75%	99.95%	99.62%	99.59%

F. Business Types Prediction and Trading Areas Building

After we have obtained the label of each business type around each organization, we can determine what business types should be contained in a trading area around each organization. We cluster all the business types whose prediction label is 1 and compare this cluster with the cluster of business types of each organization. The algorithm is described in Algorithm 3.

Algorithm 3: Trading Area Building Algorithm

Inputs: the number of organizations ***N***, the number of distinct business types ***M***, the average level of each business type $\boldsymbol{B_{a_i}}$, the prediction label of each business type for each organization $\boldsymbol{L_{ij}}$.
Outputs: the clusters of business types in predicted trading area of each organization

```
function TradingAreaBuilding(Bai,Lij) {
   for (1 → n ; n <= N ; n++) {
      for( 1 → m ; m <= M; m++) {
         if (Lmn == 1) {
            Cn.add(Bam); }}}
   return C; }
```

IV. EXPERIMENTAL EVALUATION

We implement our methodology by building trading areas around real universities as our case studies. We collect data of 13 real-world universities from reliable websites like

Meituan (https://www.meituan.com/) and Dianping (https://www.dianping.com/) providing information of off-line stores, like the opening time, closing time, customer reviews and so on, and all data we use are real and effective. Thus, our experimental evaluation is based on a real-word scenario. In the following subsections, we describe the experimental procedure, analyze the experimental results and make some discussion about the results.

A. Experimental Procedure

Since the target of this research is to build a suitable trading area around a given organization, it is not enough only to predict whether a single business type is suitable for an organization. We need to go further by clustering all the business types whose prediction label is 1 and comparing the cluster with the trading area in the real world. The accuracy of our prediction measures the ability of TrORF to build the proper trading area around an organization like a university. Consequently, the experimental procedure is as below:

step1: We apply our method, RF, and other four methods, which are GBDT, SVM, DNN and NB, to predict whether a business type is appropriate for a given organization. Still, 70% of the data is divided as training dataset, 20% validation dataset and 10% prediction dataset. Then we compare the accuracy, precision, recall and f1-score of our method and the other four methods.

step2: After predicting labels for all business types of organizations, we cluster those business types whose prediction label is 1 for each organization.

step3: Ultimately, we compare the experimental results between our method and the other four methods. If our prediction enjoys higher accuracy, our model could be regarded as successful.

B. Experimental Results

Following the steps above, we first compare the accuracy, precision, recall and f1-score of our method with other four existing methods. The results demonstrate that our method has higher accuracy, precision, recall and f1-score than the other four, which means our method performs best when predicting whether a business types is suitable for a university. The results of validation dataset are shown in TABLE II, and the results of prediction dataset are shown in TABLE III.

TABLE II. VALIDATION RESULTS OF TRORF, GBDT, SVM, DNN AND NB.

Method	TrORF	GBDT	SVM	DNN	NB
Accuracy	99.66%	88.88%	71.12%	62.35%	65.47%
Precision	99.99%	89.90%	71.94%	62.80%	66.05%
Recall	99.43%	88.67%	70.66%	61.79%	65.21%
F1-score	99.64%	89.77%	71.63%	62.55%	65.89%

TABLE III. PREDICTION RRESULTS OF TRORF, GBDT, SVM, DNN AND NB.

Method	TrORF	GBDT	SVM	DNN	NB
Accuracy	99.75%	88.14%	71.59%	62.61%	64.83%
Precision	99.95%	89.95%	72.31%	62.72%	65.72%
Recall	99.62%	88.01%	70.79%	61.04%	65.15%
F1-score	99.59%	88.97%	71.31%	61.83%	65.45%

Next, we display the comparison between business types of trading areas predicted by our method and those in reality in TABLE V. We use alphabets A to M to represent 13 real-world universities and numbers 1 to 61 to replace the business types. Some of the business types and their corresponding numbers are shown in the legends table of TABLE V. Then we compare the average business type recommendation accuracy of our method with the other four existing methods and display the results in TABLE VI.

In order to clarify the comparison between our prediction and the reality, the differences of each university's two clusters are shown in boldface and blue in TABLE V. For universities A, C, E, G, H, K, L and M, only a few business types are wrongly added, like one or two, and for A, B and F, only one business type is deleted incorrectly. Meanwhile, it is remarkably noticed that for universities D, I and J, all the business types are correctly predicted, which demonstrates that the trading area built by TrORF is exactly the same as the one in reality. We use accuracy to evaluate the prediction ability of our model and the results are shown in the fourth column of TABLE IV. The average accuracy of our model is 98.23%, which is a very satisfying result. What' s more, from TABLE V, we could find that TrORF has higher business type recommendation accuracy than the other four existing methods. Hence, it is clear that our model does provide high accuracy when building trade areas around universities.

TABLE IV. BUSINESS TYPE RECOMMENDATION ACCURACY RESULTS OF OUR MODEL FOR 13 UNIVERSITIES.

Legends				
B_Type	Beef	Bobofish	Boiledfood	Bakery
Number	1	2	3	4
B_Type	Breakfast	Casserole	Chickenpot	Chinese
Number	5	6	7	8
B_Type	Clothing	……		
Number	9	……		

Name of Universities	Business types predicted by our method	Business types in reality	Accuracy
A	{1,2,4,5,6,7,8,9,12,13,15,17,18,19,20,21,25,28,32,33,35,36,38,39,40,47,50,51,52,53,54,57,59,60,61}	{1,2,4,5,7,8,9,12,13,15,17,18,19,20,21,24,25,28,32,33,35,36,38,39,40,47,50,51,52,53,54,57,59,60,61}	96.72%
B	{1,2,5,6,7,8,17,18,20,25,29,30,35,36,38,39,40,42,49,50,53,55,60,61}	{1,2,5,6,7,8,17,18,20,25,29,30,35,36,38,39,40,42,47,49,50,53,55,60,61}	98.36%
C	{2,4,6,7,8,9,12,13,14,17,18,19,20,22,24,29,34,35,36,39,47,48,49,53,55,58,61}	{2,4,6,7,8,9,12,13,14,17,18,19,20,22,24,29,34,35,36,39,47,48,49,53,58,61}	98.36%
D	{1,2,6,7,8,17,18,20,26,29,35,36,39,40,42,43,47,50,52,53,61}	{1,2,6,7,8,17,18,20,26,29,35,36,39,40,42,43,47,50,52,53,61}	100%
E	{1,2,7,8,17,18,25,29,35,37,39,40,43,47,53,61}	{1,2,7,8,17,18,25,35,37,39,40,43,47,53,61}	98.36%
F	{4,6,7,9,15,17,18,20,23,25,31,35,36,37,39,45,46,47,49,51,53,61}	{4,6,7,9,15,17,18,20,23,25,31,32,35,36,37,39,45,46,47,49,51,53,61}	98.36%
G	{2,4,6,7,8,11,17,18,25,31,35,36,37,39,40,47,49,53,61}	{2,4,6,7,8,11,17,18,25,31,35,36,37,39,47,49,53,61}	98.36%
H	{2,3,7,8,17,18,20,27,35,36,38,39,47,49,52,61}	{2,3,7,8,17,18,20,35,36,39,47,49,52,61}	96.72%
I	{7,8,16,17,20,21,25,31,33,35,36,38,39,42,47,50,51,52,61}	{7,8,16,17,20,21,25,31,33,35,36,38,39,42,47,50,51,52,61}	100%
J	{4,7,8,9,12,13,17,18,20,25,33,35,36,39,40,42,47,49,51,53,61}	{4,7,8,9,12,13,17,18,20,25,33,35,36,39,40,42,47,49,51,53,61}	100%
K	{7,8,9,14,17,26,32,35,38,39,47,51,61}	{7,8,9,14,17,32,35,38,39,47,51}	96.72%
L	{6,7,8,9,12,17,20,21,24,25,29,31,33,36,38,39,43,44,47,49,50,51,53,54,61}	{6,7,8,9,12,17,20,21,24,25,29,31,33,36,38,39,44,47,49,50,51,53,54,61}	98.36%
M	{4,8,9,10,12,17,18,19,20,21,24,25,27,33,36,38,39,41,49,51,54,56,61}	{4,8,9,10,12,17,18,20,21,24,25,27,33,36,38,39,41,49,51,54,56}	96.72%

TABLE V. COMPARISON BETWEEN THE AVERAGE ACCURACY OF OUR METHOD AND THE OTHER FOUR METHODS.

Model	RF	GBDT	SVM	DNN	NB
Average Accuracy	98.23%	85.77%	67.58%	60.42%	64.93%

C. Discussion

The experimental results really show that our model has great ability of building trading areas around universities. Although sometimes it might mistakenly add or delete some business types, the number of wrongly labeled business types is always small like one or two. Despite of that, 100% prediction accuracy for some universities truly makes us believe that our method can help in determining business types and building trading areas.

V. RELATED WORK

A. Business Location Selection using R²Net

Xu et al. in [1] proposed a location selection mechanism for business utilizing Regression-and-Ranking Combined Neural Network(R²Net) technique in 2018. This proposed approach resolves the difficulty of exploiting the highly available satellite data as well as urban data for business location selection. The regression part of the model aims to predict popularity of locations, and the ranking part regularizes the regression part to caculate and predict popularity scores by the preserved ranking order [1]. However, it ignored human activities, which is of importance in determining the optimal location.

B. Discovering Optimum Location for a Business using OptiLocator

Bembenik et al. in [3] proposed a method named OptiLocator in 2017. The proposed method combines the area characteristics indicators and the spatial data exploration to discover the optimal location for conducting a business of a given type. Unlike other researches which analyze alike factors like foot traffic, neighborhood structure, space rent and so on, spatio-temporal data is used as a new direction in this research to find groups of people who often appear in a given location in similar time.

C. Predicting Optimal Retail Store Placement using SVR

Karamshuk et al. in [6] proposed a Support Vector Regression(SVR) based approach for predicting the optimal retail store placement in 2014. An SVR model based on location-based social networks was proposed, and features based on two general signals (geographic and user mobility) were utilized to find out the common best performing features of three different commercial chains. Furthermore, multiple features were combined in the supervised learning algorithm using Support Vector Regression and the optimal performance was achieved for the predictions [6].

VI. CONCLUSION & FUTURE WORK

Trading areas around organizations have become more and more popular because of the needs and great consumption power of people working, studying or living nearby. However, it would take a long time to wait for a trading area to become mature enough to meet the needs of people. Meanwhile, it is not easy for investors to choose the suitable business types for a given organization. In this research, we propose a machine learning based methodology, called TrORF, to address the problem of determining suitable business types around organizations. We implement the prototype of our approach and evaluate it by using 13 universities as our case studies. We compare the performance of our method with the other four methods, which are GBDT, SVM, NB and DNN. The experiment results indicate that our method has strong capability to build suitable trading areas around organizations like universities since its average recommendation accuracy is 98.23%, improving on other approaches by at least 13%.

Our future work includes: 1) recommending Top-K most popular business types to investors around an organization; 2) obtaining more data of all-kinds of organizations to improve the generality of our methodology.

ACKNOWLEDGMENT

This research work is a collaboration among Southwestern University of Finance & Economics, Alibaba Group and Macquarie University.

REFERENCES

[1] Y. Xu, Y. Shen, Y. Zhu and J. Yu, “Fusing Satellite Data and Urban Data for Business Location Selection: A Neural Approach”, DASFAA 2018, pp. 589-605, 2018.

[2] A. M. B. M. Rohani and FF. Chua, “Location Analytics for Optimal Business Retail Site Selection”, ICCSA 2018, pp. 392-405, 2018.

[3] R. Bembenik, J. Szwaj and G. Protaziuk, “OptiLocator: Discovering Optimum Location for a Business Using Spatial Co-location Mining and Spatio-Temporal Data”, ISMIS 2017, pp. 347-357, 2017.

[4] B. Eravci, N. Bulut, C. Etemoglu and H. Ferhatosmanoğlu, “Location Recommendations for New Businesses Using Check-in Data”, ICDWM 2016, pp. 1110-1117, 2016.

[5] E. KoÃ§ and H. A. Burhan, “An Application of Analytic Hierarchy Process (AHP) in a Real World Problem of Store Location Selection”, Advances in Management and Applied Economics, pp. 41-50, 2015.

[6] D. Karamshuk, et al., “Geo-spotting: mining online location-based services for optimal retail store placement”, ACM SIGKDD 2013, pp. 793-801, 2013.

[7] S. Wu, et al., “Website Defacements Detection Based on Support Vector Machine Classification Method”, ICCDE 2018, May 2018.

[8] M. Zhong, et al., “Sample Location Selection for Efficient Distance-Aware Influence Maximization in Geo-Social Networks”, DASFAA 2018, vol 10827, pp. 355-371, 2018.

[9] M. Fu et al, “DAliM: Machine Learning Based Intelligent Lucky Money Determination for Large-Scale E-Commerce Businesses”, ICSOC 2018, pp. 740-755, 2018.

[10] H. Ishwaran and M. Lu, “Standard errors and confidence intervals for variable importance in random forest regression, classification, survival”, Workshop on Recent Advances and Challenges in Statistical Methods, vol. 38, pp. 558-582, 2018.

[11] W. Shan, et al., “Optimization of competitive facility location for chain stores”, Annals of Operations Research, vol. 273, pp. 187-205, 2019.

[12] G. Nagarajan and L. D. D. Babu, “An Emperical Comparison of Six Supervised Machine Learning Techniques on Spark Platform for Health Big Data”, Smart Innovation, Systems and Technologies, pp. 299-301, 2018.

[13] L. H. David, “Defining and Estimating a Trading Area”, Journal of Marketing, vol. 28, pp. 4-38, 1964.

[14] Q. Zhang, K. Lee, H. Bao, Y. You and D. Guo, “Category coding with neural network application”, Information Theory, 2018.

[15] A. Das, et al., “Deep Learning Based Liver Cancer Detection Using Watershed Transform and Gaussian Mixture Model Techniques”, Cognitive Systems Research, vol. 54, pp. 165-175, May 2019.

[16] V. B. Semwal, et al., “An Optimized Feature Selection Technique Based On Incremental Feature Analysis for Bio-metric Gait Data Classification”, Multimed Tools Appl, vol. 76, pp. 24457-24475, 2016.

[17] M. Pal, “Random Forest Classifier for Remote Sensing Classification,” International Journal of Remote Sensing”, vol. 26, pp. 217-222, 2005.

[18] V. Y. Kulkarni and P. K. Sinha, “Random Forest Classifiers: A Survey and Future Research Directions”, International Journal of Advanced Computing, vol. 36, pp. 1144-1153, 2013.