The Text Classification of Theft Crime Based on TF-IDF and XGBoost Model

Zhang Qi
Beijing Police College
Beijing, China

***Abstract*—Classifying theft crime data of a city from 2009 to 2019 based on text classification technology. Firstly, manually classifying and defining theft crimes based on legal view and criminal practice view, then selecting 2621 data at random from the whole data. Extracting features from pre-processed sample data by TF-IDF model, then training and testing text classification model by XGBoost algorithm, and comparing the test results of KNN algorithm, Naïve Bayes algorithm, SVM algorithm and GBDT algorithm. The results show that the XGBoost algorithm are better than KNN, Naïve Bayes, SVM and GBDT. Adjusting slightly various categories to improve the accuracy of classification, and the accuracy of each algorithm is improved by 2-5 percentage points and the accuracy of XGBoost is highest. So, the results show that, ①XGBoost algorithm is best to use as classifying the whole data. ②The influence of data quality on classification accuracy is obvious and can improve the accuracy of algorithms rapidly. The classified theft crime data of 2009-2019 through XGBoost algorithm can be used as based data for the prediction of various types of crimes.**

Keywords—Theft crime, Text classification, TF-IDF model, XGBoost algorithm

I. INTRODUCTION

Theft crime, as the main type of crimes, accounts for about 80% of all types of crimes. With a large number of cases and a low detection rate, theft crime has always been the focus of the public security organs to prevent and combat. Police big data digging is the basis of crime prediction, and time information and space information are two important data dimensions of police data. Therefore, it is a hotspot of police data research to mine spatio-temporal data and make crime prediction. The use of spatio-temporal data for crime prediction research beginning in the 1930s, these researches mainly included: spatial distribution and regional hot spots distribution of crimes, crime ecological surveying model [1], and environmental criminology theory research [2]. In recent years, the studies focus on crime time characteristics, influence factors analysis, trend prediction and so on, which are done by using these methods: statistical method [3], time series analysis method [4], method of machine learning (Decision tree and Naïve Bayes model [5], network analysis technology [6], random forest model [7], self-exciting point process model [8], long short term memory (LSTM) model [9]). The analysis based on statistical methods and time series mainly focuses on macro data, which cannot realize short-term prediction of crimes. Although machine learning method is based on micro data, its prediction accuracy is low (the prediction accuracy is between 35% and 63% in [5] and [7-9]). There are several reasons for the low accuracy of machine learning in predicting crimes. Firstly, the algorithm needs to be optimized. Secondly, the quality of data is low. Thirdly, the results are affected by random error and accidental factors. Among them, data quality is the key factor that affects the accuracy of prediction. The effect of improving data accuracy on improving the accuracy of crime prediction is much higher than that of optimizing algorithm. In real work, the public security data quality is low, and main problems are: data error, lack of data, data irregularities, data repeat, etc. For example, about 40%-50% of the theft crime data, which is object of study, are indefinite and inaccurate, which leads to the inaccurate prediction of a particular type of theft crime based on historical data, such as burglary. It takes a lot of time and manpower to correct historical data by manual work, while it can save a lot of manpower and time cost to classify massive data based on the relatively mature text classification technology. Therefore, the training of a good classification model can greatly improve the accuracy of data, lay a foundation for the analysis and research of various theft crimes from the perspective of criminal practice, and improve the accuracy of crime awareness and prediction.

At present, there are few achievements in text classification of crime data based on machine learning algorithms at home and abroad, but many achievements have been made in the classification of short texts in some fields, such as email, weibo, comments and short news based on various machine learning algorithms [10]. Short text classification refers to text with a content length of at most about 100 words, which has become an important research direction in the field of data mining. The description of "brief case" of theft crime is a typical short text, which can use machine learning short text classification technology in other fields to achieve accurate classification of case events. Text classification refers to that given document set $D=\{x_1,x_2,\ldots\ldots,x_n\}$ and category set $C=\{y_1,y_2,\ldots\ldots,y_N\}$, and a classification function $\theta(x)$ is obtained by using some classification algorithm to map the documents x_i in the document set D to one or more categories in the category C. In order to improve the classification accuracy, it is generally improved from two aspects: one method is to reduce the assumption of consistent importance of features, improving classification accuracy by amplifying the influence (weight) of some important features. This method is simple and can effectively improve classification accuracy, such as TF-IDF model, word2vec model [11]. Another method is to improve the algorithm, the existing essay this classification technology is mainly based on machine learning algorithm KNN (K-NearestNeighbor) algorithm, Naïve Bayes algorithm[12], Decision Tree Model[13] algorithm, SVM (Support Vector Machine)

algorithm[14], ANN (Artificial Neural Network) algorithm[15] and other algorithms, the researchers improve the classification accuracy through improving algorithm or combining algorithm. For example, tree augmented Naïve Bayes [16], improved RNN algorithm, Boosting [17] and Bagging [18] ensemble algorithm based on Decision Tree, among which the Boosting family includes GBDT (Gradient Boosting Decision Tree) algorithm, Adaboost algorithm, XGBoost (Extreme Gradient Boosting) algorithm [19]. Among them, XGBoost algorithm is the improvement of GBDT, which has the advantages of regularization, parallelization, and flexibility and so on, and the classification effect is remarkable. The most representative of Bagging are Random Forests [20] and various combination algorithms, which achieve good classification results.

II. Research Design

Through random sampling, 2622 pieces of preprocessed theft crime data from 2009 to 2019 are extracted as samples for preprocessing, and case categories of sample data are manually marked. The TF-IDF model is used to extract text features, and redefining each text data based on the extracted feature vector and inputting them into the trained model, respectively using XGBoost algorithm, KNN algorithm, the Naive Bayes algorithm, the SVM algorithm, GBDT algorithm to train and test data, and using the Precision (accuracy), Recall (recall rate), F1 - score (F1) to compare classification accuracy of each algorithm. Finally, selecting the optimal model to classify the whole data of 2009-2019. The design process is shown in Fig.1.

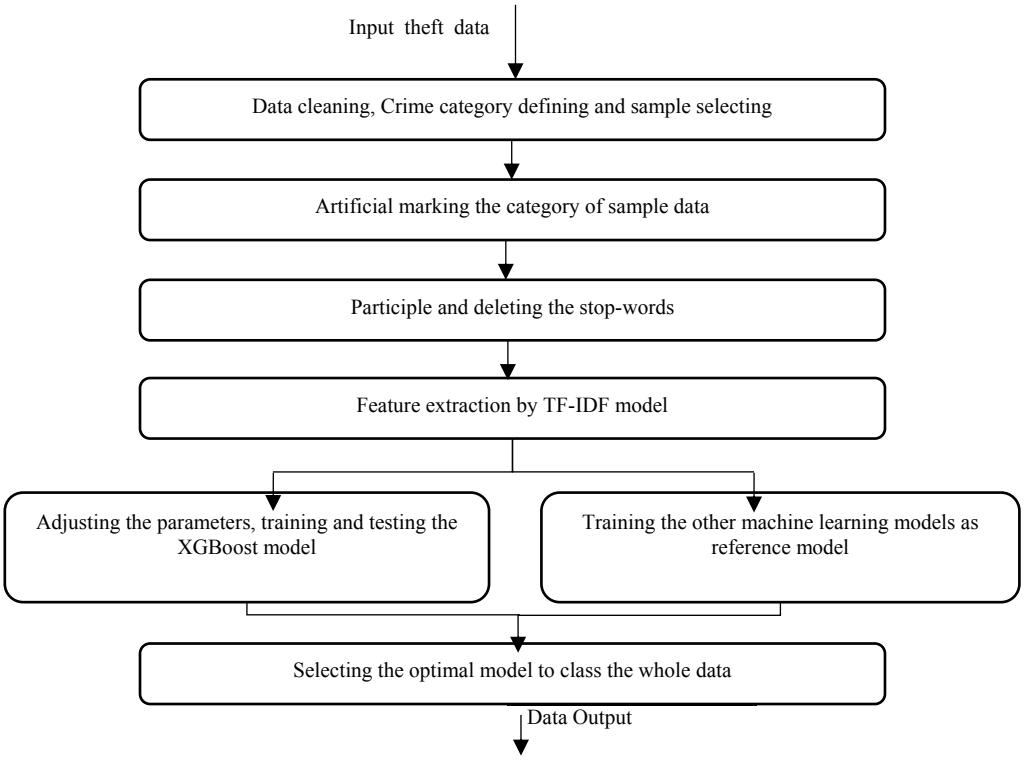

Fig. 1. Flow chart of text classification of theft crime category

A. Introduction of TF-IDF Model

TF-IDF (Term frequency-inverse document frequency) algorithm is an improvement of DF method. It is a kind of statistical method, used to evaluate the importance of a word in a file set. The importance of a word is proportional to the number of times it appears in the document and inversely proportional to the number of times it appears in the entire document set.

$TF_{i,j}$ represents the frequency of entry w_i in file x_j, as show in formula:

$$TF_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \quad (1)$$

In here, $n_{i,j}$ represents the number of occurrences of entry w_i in file x_j, and the denominator is the sum of all entries in file x_j.

IDF represents the entry w_i inverse document frequency index, which is divided by the total number of articles divided by the number of articles containing the keyword, and then the logarithm of the result is obtained, namely formula (2).

$$IDF_i = \log\left(\frac{|D|}{|\{j : w_i \in x_j\}|+1}\right) \quad (2)$$

In here, $|D|$ is the total number of files in the corpus, $|\{j : w_i \in x_j\}|$ is the number of files containing entry w_i, and 1 is added to avoid the occurrence of 0 in the denominator.

The TF-IDF value of the given word w_i is:

$$(TF-IDF)_{w_i} = TF_{i,j} \times IDF_i \quad (3)$$

B. Introduction of XGBoost Algorithm

The full name is Extreme Gradient Boosting, which is an improvement of the model of GBDT (Gradient Boosting Decision Tree). The basic principle is to combine multiple Decision trees with lower accuracy into a model with higher accuracy [21]. The XGBoost algorithm adopts the idea of gradient descent in the generation of each tree. Based on the tree generated in the previous step, it iterates to the direction of the minimum given objective function. Through the iteration of multiple decision trees, the loss error is continuously reduced, and the prediction model is finally obtained. The split nodes of each decision tree are constructed in accordance with the criteria of CART (regression) tree, and the least square loss and logarithmic function are commonly used. The specific algorithm principle is as follows:

For a given dataset $D=\{(x_i,y_i)\big|\,|D|=n,x_i\in R^m,y_i\in R\}$ with n samples and m eigenvalues, where y_i is the category truth value corresponding to sample x_i, the XGBoost algorithm takes (regression tree) as the basis classifier and predicts the output of the function of K CART trees summation, where the input value of $f_k(x)$ is the error value between the predicted value of $\sum_{t=1}^{k-1} f_t(x)$ and the truth value to reduce the loss function, and the prediction model of XGBoost is shown in formula (4).

$$\hat{y}_i=\theta(x_i)=\sum_{k=1}^{K} f_k(x_i) \tag{4}$$

XGBoost defines the complexity of each tree, and the complexity $\Omega(f)$ of each decision tree is shown in formula (5).

$$\Omega(f)=\gamma T+\frac{1}{2}\lambda\,\|\omega\|^2 \tag{5}$$

$\omega_j(j=1,2,\cdots,T)$ represents the weight value of each leaf node, T represents the number of leaf nodes, $\|\omega\|^2$ represents the module of leaf node vector, γ represents the difficulty of node segmentation, and λ represents L^2 regularization coefficient. The smaller the value of $\Omega(f)$, the lower the complexity, the stronger the generalization.

The formula of the target loss function is shown in formula (6).

$$L(\theta)=\sum_{i=1}^{n} l(y_i,\hat{y}_i)+\sum_{k}\Omega(f_k) \tag{6}$$

n represents the number of training samples, i represents the i sample, $\hat{y}_i$ represents the predicted value of the i sample x_i, and y_i represents the actual value corresponding to x_i, $\sum_{i=1}^{n} l(y_i,\hat{y}_i)$ represents training error. XGBoost can be used to set the second-order Taylor expansion for $l(y_i,\hat{y}_i)$ in the loss function. After the second Taylor expansion, the loss function is shown in formula (7).

$$\begin{aligned} L^{(t)}(\theta)&=\sum_{i=1}^{n} l(y_i,\hat{y}_i^{(t-1)}+f_t(x_i))+\Omega(f_t)+C_1\\ &\approx\sum_{i=1}^{n}(g_i f_t(x_i)+\frac{1}{2}h_i f_t^2(x_i))+\gamma T+\frac{1}{2}\lambda\sum_{j=1}^{T}\omega_j^2+C \end{aligned} \tag{7}$$

The formulas for g_i and h_i are shown in formula (8).

$$g_i=\partial_{\hat{y}^{(t-1)}} l(y_i,\hat{y}_i^{(t-1)}),\quad h_i=\partial^2_{\hat{y}^{(t-1)}} l(y_i,\hat{y}_i^{(t-1)}) \tag{8}$$

Definition $I_j=\{i\,|\,q(x_i)=j\}$ represents the sample set of leaf nodes with serial number j, and $q:\{x_1,x_2,\cdots,x_n\}\to\{1,2,3,\cdots T\}$ represents the mapping of sample space to category space serial number.

From formula (5) and the definition of ω_j, $f_t(x)=\omega_{q(x)}$, the formula (7) can be transformed the formula (9).

$$L^{(t)}(\theta)\approx\sum_{j=1}^{T}[G_j\omega_j+\frac{1}{2}(H_j+\lambda)\omega_j^2]+\gamma T+C \tag{9}$$

In here, $G_j=\sum_{i\in I_j} g_i, H_j=\sum_{i\in I_j} h_i$ The optimal solution ω_j^* is obtained by using maximum likelihood estimation, and there are: $G_j+(H_j+\lambda)\omega_j=0\Rightarrow\omega_j^*=-\frac{G_j}{H_j+\lambda}$, Thus, the optimal solution of the objective function is obtained as formula (10).

$$L^*(\theta)=-\frac{1}{2}\sum_{j=1}^{T}\frac{G_j^2}{H_j+\lambda}+\gamma T \tag{10}$$

III. Experimental Process

A. Data Preprocessing

Data preprocessing mainly includes the following steps. ① Cleaning all theft crime data from 2009 to 2019, remove duplicate data and delete invalid data.②Removing the types of crimes with small number of cases and little research significance, such as stealing logging forests and precious cultural relics, and 1123978 pieces of data are finally retained. ③Redefining criminal categories of the corpus from the perspective of judicial interpretation and criminal practice, theft crime can be divided into burglary, theft of properties of

companies and stores, theft of goods, theft of electrical equipment, theft of livestock, car theft, pickpockets, theft of non-motor vehicles, theft of car properties, a total of nine categories (all kinds of other definitions are shown in Table I below), the main purpose to do data preparation for crime prediction.

TABLE I. DEFINITION OF SEVERAL TYPES OF THEFT CASES FROM THE PERSPECTIVE OF CRIMINAL PRACTICE

The case category	Definition
Burglary	Refers to the household theft in the judicial sense, refers to the illegal entry into the living place where the family and its members are relatively isolated from the outside world for the purpose of illegal possession, including closed courtyards, rented houses for family life.
Theft of properties of companies and stores	Refers to the theft of the company, factories, stores, supermarkets, non-motor delivery vehicles and other local property behavior.
Theft of electrical equipment	Refers to the theft of cables, communication lines and other acts.
Theft of livestock	Refers to the theft of the owner's dogs, cattle, sheep and other livestock.
Theft of car and car properties	Refers to the theft of a private car or van, and theft of property inside a private car, minibus, good van.
Pickpockets	Refers to the theft of property carried by another person in a public place or on public transportation.
Theft of non-motor vehicles	Refers to the theft of bicycles, electric bicycles, motorcycles or the corresponding car parts.

⋆In practice, there are cases of goods theft, but because of there are no sufficient numbers and it's easy to mix them with Theft of properties of companies and stores and Pickpockets, it have impacted on the accuracies of five training models and reduce the accuracies (Fig.2). So it's suitable to merge the cases of stealing van goods into Theft of car and car property, the cases of stealing the goods in delivery non-vehicles and in public into Pickpockets, the cases of stealing the goods in private into Theft of properties of companies and stores. Doing these will not have influence to crime prediction in practice.

④Data are selected for manual marking, and 2621 pieces of sample data (including 7 categories) were finally formed, which were divided into training set and test set according to the proportion of 2089:532, as shown in Table II.

TABLE II. DISTRIBUTION OF TRAINING SETS AND TEST SETS FOR EACH TYPE OF CASE IN SAMPLE DATA

The case category	Training sets	Testing sets
Burglaries	302	77
Theft of properties of companies and stores	126	32
Theft of car and car property	554	141
Pickpockets	412	105
Theft of non-motor vehicle	573	146
Theft of electrical equipment	71	18
Theft of livestock	51	13

⑤The Jieba word segmentation tool is used to divide the sample data into words, stop-words and demystify the data (the names, id Numbers and time in the case were removed) to form the data set.

B. Evaluation Method

In order to test the performance of the algorithm, Precision, Recall and f1-score were used as evaluation indexes to evaluate the performance of the algorithm.

C. Parameter Setting of TF-IDF Model

The TF-IDF model uses the built-in default parameter settings for the CountVectorizes () and Tf-idf Transformer () methods in the python installed sklearn module.

D. Training and Testing Design

The brief case description of the sample data after preprocessing is designed by using the XGBoost model, and the classification results of the XGBoost, KNN, Naive Bayes, SVM and GBDT models are compared. The specific design is as follows:

Step1. The word vector sparse matrix space formed by the preprocessed sample data set is extracted by using the TF-IDF model to generate the TF-IDF feature vector space, and then the TF-IDF text vector of each sample is formed and input into the machine learning model for training.

Step 2. The TF-IDF sample data are divided into training set and test set at a ratio of 2089:532, and the training sample data are trained with five machine learning models including XGBoost algorithm, KNN algorithm, Naive Bayes algorithm, SVM algorithm and GBDT algorithm.

Step 3. After adjusting parameters, training the best model of each algorithm, analyzing the test results of test data, analyzing the main error points, and adjusting and improving them based on the definition of training data classification.

Step 4. Selecting the model with the best prediction effect as the classification model for the whole data from 2009 to 2019. After data cleaning and word vector transformation, inputting the above trained optimal model for classification to realize the classification of the whole data.

IV. RESULT ANALYSIS

A. Comparison Experiment

After adjusting parameters, the optimal models of the following five algorithms were finally trained, and there are three optimal results as follow show:

① The quality of artificial marking data has high influence to the accuracy of algorithms. As footnote 1shows, before cases of goods theft being divided and merged into the other three categories, the experimental results of XGBoost algorithm, KNN algorithm, Naive Bayes algorithm, SVM algorithm and GBDT algorithm, were compared. The results are shown in table 3. But after cases of goods theft being divided and merged into the other three categories, the accuracy of each algorithm has improved by 2%-4% (Table III, Fig. 2 and Table IV), and the accuracy of XGBoost has improved by about 5%. So the quality of sample data has obvious influence to the accuracy of each algorithm.

TABLE III. TESTING RESULTS OF FIVE MACHINE LEARNING MODELS ON CLASSIFICATION OF SAMPLE DATA

Machine Learning Model	Evaluation Standard		
	Precision	*Recall*	*F1-score*
SVM	78.2%	74.8%	76.4%
KNN	84%	84%	84%
Naïve Bayes	87.4%	88.2%	87.8%
GBDT	91.3%	90.5%	90.9%
XGBoost	92.3%	91.6%	91.9%

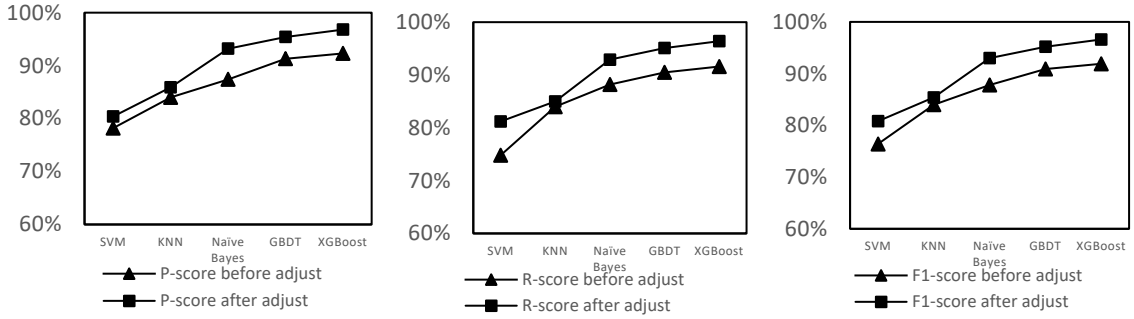

Fig.2. Comparison of classification evaluation indicators before and after adjustment

* The Y-axis in three graphs of Fig.2.represent respectively P-score, R-score and F1-score of algorithms.

②XGBoost training model is the best performance and has obvious advantages compared with the other four machine learning models. Based on Footnote 1, the sample data has been adjusted. Re-input the adjusted data into each algorithm for training and testing, and adjust the corresponding test set and training set, which are not listed here. The prediction results are shown in Table IV. Table IV shows that the XGBoost classification is still optimal. Comparing the two classification results, the accurate classes of classification were significantly improved (as shown in Fig.2), and the improvement of XGBoost, GBDT and Naive Bayes are higher, while the improvement of SVM and KNN are smaller. Table IV indicates XGBoost classification algorithm is the optimization algorithm. So XGBoost model is chosen as the best classification model for the overall data.

TABLE IV. TESTING RESULTS OF SAMPLE DATA CLASSIFICATION BY EACH ALGORITHM AFTER ADJUSTMENT

Machine Learning Model	Evaluation Standard		
	Precision	***Recall***	***F1-score***
SVM	80.4%	81.2%	80.8%
KNN	85.9%	85%	85.4%
Naïve Bayes	93.2%	92.9%	93%
GBDT	95.4%	95.1%	95.2%
XGBoost	96.8%	96.4%	96.6%

③The accuracy of each algorithm is different among seven theft crime categories, and the accuracies of burglary, pickpockets, theft of non-motor vehicles and theft of car and car properties is higher compared with the accuracies of theft of properties of companies and stores, theft of electrical equipment and theft of livestock (Table V). This is because that the former four have obvious category features, such as "door", "room", "home" which are obvious category features are involved in burglary, and these categories has more sample data which will improve the accuracy. On the contrary, it's easier to be mixed among the other categories theft of properties of companies and stores, theft of electrical equipment and theft of livestock, sometimes there would be similar features in them. At same time, the quantity of sample data also has influence to the accuracy.

TABLE V. THE COMPARISON OF TESTING RESULTS OF EACH CRIME CATEGORY

The Case Category	Evaluation Standard		
	Precision	***Recall***	***F1-score***
Burglary	97.4%	100%	98.68%
Theft of properties of companies and stores	81.25%	83.87%	82.54%
Pickpockets	97.14%	95.33%	96.23%
Theft of non-motor vehicles	100%	98.65%	99.32%
Theft of electrical equipment	88.89%	100%	94.12%
Theft of car and car properties	98.58%	97.89%	98.23%
Theft of livestock	92.31%	92.31%	92.31%

B. Analysis of Classification Results of the Whole Data

The above trained XGBoost model is used to classify the overall data from 2009 to 2019, and randomly sampling 1000 pieces of reclassified data and manually inspecting. It was found that Precision=96.9%, which was close to that of the test data, so it is determined that the combined algorithm of TF-TDF and XGBoost has a good classification effect on this data set. Through structural analysis of the data after classification, it was found that burglaries accounted for 22.65%, goods and property theft of company stores accounted for 7.34%, power equipment theft accounted for 1.61%, livestock theft accounted for 0.84%, car-related theft accounted for 12.95%, picking-up cases accounted for 29.11%, and theft of non-motor vehicles accounted for 25.51%. These classified data can be used as high quality data for the prediction of various types of crimes.

V. CONCLUSIONS AND PROSPECTS

The text classification of theft crime data based on the combination of TF-IDF and XGBoost algorithm achieves accurate and efficient classification effect of data. This is an effective attempt of machine learning algorithm for police data mining, and a basic work for police data governance and crime prediction. From the perspective of data classification effect, we can mainly draw three conclusions: ①The quality of sample data is an important factor affecting the accuracy

of data classification and crime prediction. The improvement of data quality improves the accuracy of data classification and crime prediction more than that of the optimization algorithm. ②Under the standard of high data quality, the combined algorithm based on TF-IDF and XGBoost is an effective text classification algorithm. The classification of all data from 2009 to 2019 based on the training model of TF-IDF and XGBoost combined algorithm can serve as the basis for studying the prediction of various types of crimes, which is also the next research direction. ③After Classification based on TF-IDF and XGBoost model, the whole data of 2009 – 2019 can be used to do temporal and spatial analysis of theft crime and predict theft crime. They can ensure the accuracy of crime prediction.

REFERENCES

[1] E. Griffiths, J. M. Chavez, "Communities, street guns and homicide trajectories in chicago, 1980-1995: merging methods for exmamining homicide trends across space and time", Criminology, vol.4, pp.941-978, 2004.

[2] Y. C. Li, S. Liu, F. M. Wang, "The prediction model for crime based on the environmental criminology", Journal of Intelligence, vol.2, pp.45-56, 2018.

[3] L. Z. Xiao, L. Liu, et al, "Impacts of community environment on residential burglary based on rational choice theory", Geographical Research, vol.12, pp.2479-2491, 2017.

[4] M. H. Qu, S. M. Hao, "Research on the prediction of the number of property crimes in China based on ARMA model", China Journal of Criminal Law, vol.2, pp.100-106, 2013.

[5] T. Almanie, R. Mirza, E. Lor, "Crime types and using spatial and temporal hotspots", Computer Science, vol.4, pp.1-19, 2015.

[6] Dash S K, Safro I, Srinivasamurthy R S, "Spation-Temproal prediction of crimies using network analytic approach", 2018 IEEE International Comference on Big Data (Big Dater), 2018.

[7] L. Liu, W. J. Liu, W. W. Liao, et al, "Comparison of random forest algorithm and space-time kernel density mapping for crime hotspot prediction", Progress in Geography, vol.6, pp.716-771, 2018.

[8] G. O. Monler, M. B Short, P. J. Brantingham, et al, "Self-Exciting point process modeling of crime", Journal of the American Statistical Association, 106(493), pp.100-108, 2011.

[9] H L Shen, H Zhang, Y F Zhang, et al, "Prediction of burglary crime based on LSTM", Statistics & Information Forum, vol.11, pp.107-115, 2019.

[10] X. G. Hu, C. Q. Yang, Y. H. Zhang, "Short text classification based on extension with its own features", Application Research of Computer, vol.4, pp.1008-1010, 2017.

[11] R. M. Xie, J. Chen, G.R. You, D.T. Xie, "A word2vec-based study of the classification of Chinese books", Journal of Yunnan Minzu University (Natural Sciences Edition, vol.4, pp.335-339, 2018.

[12] P. D. Hoff, "A first course in Bayesian Statistical Methods", Journal of the Royal Statistical Society, vol.3, pp.694-695, 2010.

[13] L. Huang, H. Chen, X. Wang, et al, "A fast algorithm for mining association rules", Journal of Computer Science and Technology, vol.6, pp.619-624, 2000.

[14] J. H. Lou, "Research on support vector machine algorithm", Dalian University of Technology, 2007.

[15] Y. D. Li, Z.B. Hao, H. Lei, "Survey of convolution neural network", Journal of Computer Applications, vol.9, pp.2508-2515, 2016.

[16] D.W. Li, X.J.Hu, C. J. Jin, et al, "Learning to detect traffic incidents from data based on tree augmented naïve Bayesian classifiers", Discrete Dynamics in Nature & Society, vol.1, pp.1-9, 2017.

[17] L. H. Dong, G. H. Geng, Y.Gao, "Survey of boosting", Computer Application and Software, vol.8, pp.27-29, 2006.

[18] M. S. Chen, X. Y. Lu, "Three random under-sampling based ensemble classifiers for Web spam detection", Journal of Computer Applications, vol.2, pp.535-539, 2017.

[19] X.F. Li, J. Ma, L. Chi, H.M. Zhu, "Identifying commodity names based on xgboost model", Data Analysis and Knowledge Discovery, vol.3, pp.34-41, 2019.

[20] M. Fernandez-Delfado, E. Cernadas, S. Barro, et al, "Do we need Hundreds of Classfiers to Solve Real World Classfication Problems", Journal of Machine Learning Research, vol.15, pp.3133-3181, 2014.

[21] T. Chen, C. Guestrin, "XGBoost: A scalable tree boosting system", Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 785-794, 2016.