








Prediction Model of Suspect Number Based on Deep Learning

Chuyue Zhang , Manchun Cai  , Xiaofan Zhao  , Luzhe Cao ,
and Dawei Wang 

People's Public Security University of China, Beijing, China
caiman Chun@ppsuc.edu.cn

Abstract. With the development of public security informatization, crime prediction has become an important tool for public security organs to carry out accurate attacks and effective governance. In this paper, we propose an algorithm to predict the number of suspects through the feature modeling of historical data. We use Deep Neural Networks (DNN) and machine learning algorithms to extract features of different dimensions of case data. We also use Convolutional Neural Networks (CNN) to extract the text features of case description. These two types of features are combined and fed into fully connected layer and softmax layer. Compared with the DNN model which only uses numeric data, the DNN-CNN model combined with text data has improved the precision rate by 20%. The addition of text data significantly improves the precision and recall rate of prediction. To the best of our knowledge, it is the first time to combine numerical and textual data of case information in crime prediction.

Keywords: Crime prediction · Deep learning · Feature selection

1 Introduction

At present, big data and artificial intelligence have become national strategies. Informationization revolution is fast and steady. The informationization of public security is also in full swing. Crime prediction provides assistance for crime prevention, public security prevention and control, case detection and police decision-making and has become a hot research topic nowadays. More than 25% of the criminal suspects prosecuted by prosecutors were burglars, ranking first [1]. The occurrence rate and quantity of theft crime is high. It causes the state, the collective and the citizen's individual property to suffer heavy loss.

Nowadays, the detection of theft crimes relies on traditional detection ideas and police experience. However, the police experience has limitations, subjectivity and one-sidedness. In this paper, we use the deep learning algorithm to predict the number of thieves through historical data. When a new crime occurs, it can provide the police with a prediction of the number of people involved in the crime.

This work is supported by National Key R&D Program Project (Grant No. 2018YFC0809802), the Fundamental Research Funds for the Central Universities (Grant No. 2018JKF609), Specialized Research Fund of Higher Education of China (Grant No. 2019ssky012).

2 Related Works

Crime prediction has gone through a process from qualitative to quantitative, from simple to complex. The qualitative research of crime prediction includes Delphi method, correlation factor analysis and so on. According to the different object scope, the quantitative research of crime prediction can be divided into two aspects: macro prediction of crime and micro prediction of crime. Macro prediction of crime mainly reveals the dynamic regularity of crime phenomenon in terms of quality and quantity. It is primarily a broad and comprehensive forecast of the number of long-term crimes. Predicting the number of crimes based on historical crime data and Geographic Information System data. Methods include grey model, markov chain, association rule mining [2], Support Vector Machine, hybrid model of Long Short-Term Memory and Spatial-Temporal Auto Regressive and Moving Average [3] etc. It is applied in the service command of prevention and control patrol. Micro-prediction is mainly the prediction of individual behaviors and attributes under certain space-time conditions. It is applied to the key personnel of the crime risk analysis, auxiliary case detection, suspects features identification. For example, based on criminal population conviction histories of recent offenders, Tollenaar and van der Heijden [4] use statistical method to predicate general recidivism, violent recidivism and sexual recidivism. Based on the case information and victim information, Li, Sun and Ji [5] use Support Vector Machine algorithm to predict the suspect's gender, age, race, etc. Based on date and location, crime type, criminal ID and the acquaintances, Vural and Gök [6] use Naive Bayesian Model to predict criminal of particular crime incident. Based on the features of criminals in criminal case, Sun, Cao and Xiao [7] uses random forest model to predict possible suspects. But some of the input data in these studies is only known after solving the case, such as suspect age, criminal history, acquaintances, etc. In most cases, the identity of the suspect is unknown after the theft. In this paper, we only use case information to predict the number of suspects.

This paper focuses on the micro-prediction of individual behavior. Extract case features such as the time of the case, loss amount, method, places and so on, and covert them to discrete values. For numerical features, machine learning methods are used to rank and delete the features with lower contribution. After feature selection, Deep Neural Networks (DNN) algorithm is used for feature processing. Some information may be lost if only discrete numerical values are used to represent case information. Therefore, text information of case description is added. We use the natural language processing method to extract the case information in the text and combine them to predict the number of suspects. When a new case occurs, the model can provide investigators with a prediction of the number of suspects.

3 Data Preprocessing

Statistics show that about 90% of the total number of thefts are “pickpocketing”, “theft of property in the car”, “household theft” and “theft of non-motor vehicles”. In this paper, we selected more than 20,000 cases of “pickpocketing”, “theft of property in vehicles”, “housebreaking” and “theft of non-motor vehicles” detected in X city as experimental data. The case category is numerically coded in this paper. That is 1, 2, 3, 4 code for “pickpocketing”, “theft of property in the car”, “household theft”, “theft of non-motor vehicles” respectively.

3.1 Time Data Processing

According to different time scales, the time information extracted in this paper includes year, quarter, month, ten days, day, week and time period. The data processing rules of time period are shown in Table 1:

Table 1. The data processing rules of time period.

Label	Time (24 h)	Note
1	0:00–8:00	Including 8:00
2	8:00–12:00	Including 12:00
3	12:00–14:00	Including 14:00
4	14:00–18:00	Including 18:00
5	18:00–22:00	Including 22:00
6	22:00–24:00	Including 24:00

3.2 Location Data Processing

In the criminal case information system, the location of the crime is recorded as the longitude and latitude of the place and the location name of crime. By the location name of the crime, the region of the crime and the place of the crime can be obtained. The region where the case occurs is the administrative division of the place. In the data set of this paper, X city is divided into 16 administrative districts. Therefore, this paper encodes the districts 1–16. There are more than 90 original categories of places involved, such as subway stations, shopping malls, Internet cafes, hotels and so on. According to the police's individual experience, this paper divides the places involved into four categories, namely residential, traffic area, office area and entertainment area. The data processing rules of places are shown in Table 2:

Table 2. The data processing rules of places

Label	Categories	Instructions
1	Residential	Residence, dormitory
2	Traffic area	Subway station, bus, roadside, expressway, etc.
3	Office area	Schools, hospitals, parking lots, offices, etc.
4	Entertainment area	Shopping malls, vegetable markets, street shops, catering places, Internet cafes and so on

3.3 Method Data Processing

The method of theft refers to the method used by thieves when they steal, such as technical method unlocking, climbing over walls, smashing glass, etc. Considering the different crime experience of different types of suspects, the methods may be different. This paper divides the theft methods into 12 categories. The data processing rules of methods are shown in Table 3.

Table 3. The data processing rules of methods

Label	Categories	Instructions
1	Others	Stolen goods from trucks, Unarmed climbing, Decoy
2	Pickpocketing	Pickpocketing, By the way to steal
3	Technical unlock	Insert card unlock, Poke the lock, Tin foil unlock, Technical unlock
4	From the window to enter	Break Windows, Break glass
5	From the roof to enter	From the roof, From the vent
6	From the wall to enter	Climb over the wall, Break walls
7	From the door to enter (Door damage)	Expand seam, Destroy anti-theft net (column), Destroy the door body
8	From the door to enter (No damage to the door)	Deceive into the door, Follow others into the door
9	Violence unlocked	Break the lock core, Pliers cut the lock, Break the lock
10	Violence hit the car	Smashing car windows, Pry the trunk of the car
11	Theft through the window	Get the key through the window with a pole, Open the window and steal
12	Stolen vehicle	Towing, dragged, Traction

3.4 Loss Amount Data Processing

The amount of damage is the value of the stolen goods. It has the best effect by comparing min-max normalization, zero-mean normalization and feature normalization of loss amount. Therefore, this paper uses the feature encoding method to process the loss amount data. We take 1/4, median and 3/4 of the loss amount as the dividing point and divide the loss amount into four categories. The data processing rules for loss amount are shown in Table 4.

Table 4. The data processing rules of loss amount.

Label	Money (RMB)	Note
1	0–900	Including 900
2	900–1880	Including 1880
3	1880–3600	Including 3600
4	3600–1000000	More than 3600

3.5 Weather Data Processing

In order to consider all kinds of factors when the case happened, the data of day weather, night weather, day temperature, night temperature, day wind and night wind were added in this paper. For day temperature, night temperature, day wind, and night wind, we extract their values by regularization. For the text description information of daytime weather and night weather, this paper uses keyword matching method, and the data processing rules for weather are shown in Table 5.

Table 5. The data processing rules of weather.

Weather	Sunny	Rain or snow	Others
Daytime weather	1	2	3
Night weather	1	2	3

4 Model Building

4.1 Feature Selection

In numerical data, this paper describes the case from 20 dimensions. If the information in the data is irrelevant or noisy, the prediction results will be affected. In this paper, the random forest classification (RFC) and Linear Regression recursive feature elimination (LR_RFE) are used to calculate the feature contribution and filter the variables so as to optimize the prediction model. The processing of numeric data is shown in Fig. 1.

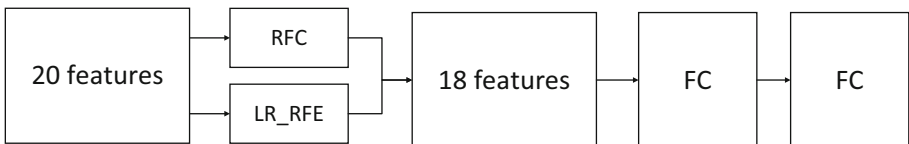


Fig. 1. Model flow for numeric data

4.1.1 Feature Ranking of Random Forest

Random forest is an integrated learning algorithm based on decision tree. It has good effect on classification and regression. In this paper, the grid search method is used to find the optimal solution of random forest parameters and return the contribution ranking of each feature. The experimental results show that the optimal parameters of grid search are criterion = ‘gini’, max_depth = 50, min_samples_leaf = 2, min_weight_fraction_leaf = 0.0 and n_estimators = 1000. Under the current parameter setting, the least contributing feature is night temperature.

4.1.2 Feature Ranking of Recursive Feature Elimination

Recursive feature elimination (RFE) is used in this paper to train Linear Regression in several rounds. After each round of training, the feature with low weight coefficient is removed, and the next round of training is carried out based on the new feature set until the remaining feature number reaches the required feature number. In this algorithm, the least contributing feature is night temperature.

By deleting the feature with the lowest contribution in the above two algorithms, this paper describes the case through 18 dimensions. They are Longitude, Case classification, Latitude, Quarter, Places, Loss amount, Years, Months, Methods, Days, Daytime wind, Ten days, Regions, Time period, Night weather, Night wind, Daytime weather and Week.

4.2 Text Data Model Construction

This paper uses 18 characteristics to describe case information. DNN algorithm is used to further explore the relationship. However, the data after numerical dispersion is easy to cause the loss of features. For example, the name, type and victim information of the stolen item cannot be represented. Therefore, the text description information of the case is added in this paper.

Brief case is a simple description of the case, ranging from 5 to 150 words. After transforming text information into word vectors, the feature is extracted by using convolution neural network algorithm [8]. CNN has two main operations: convolution and pooling. Convolution focuses on local features. Each hidden layer node only connects to an input point that is small enough, instead of connecting to every input point. At the same time, the weights of some neurons in the same layer are shared, which can greatly reduce the weight parameters that need to be trained.

The convolution layer uses multiple $n \times h$ convolution kernel (n is the dimension of the embedding vector, h is the dimension of the filter’s window size). Using different size of convolution kernel allows the network to extract different width features automatically. By connecting each convolution kernel, the output of the convolution layer can be obtained. The calculation is shown in formula (1).

$$m_i = f(w * x_{i:i+h-1} + b) \quad (1)$$

m_i represents the i -th feature of the convolution operation, f represents the activation function. w represents the weight of a filter. It performs a convolution operation on the input feature x whose window size is h , and get a new feature. The output features of

the convolution layer are obtained by connecting all the features obtained above. The calculation is shown in formula (2).

$$M = [m_1, m_2, \dots, m_{l-h+1}] \tag{2}$$

l represents the input length. The max-pooling used in this model is to take the maximum value for each vector of the convolutional layer output. The calculation is shown in formula (3).

$$z = \max\{M\} \tag{3}$$

z represents the output of max-pooling. By connecting all the maximum pooling results, the output of the pooling layer Z is obtained. k represents the number of convolution kernel.

$$Z = [z_1, z_2, \dots, z_k] \tag{4}$$

The model in this paper uses convolution kernel size of 2, 3 and 4 to acquire features of different widths, and then putting into max-pooling layer. After that, the full connection layer is added. The text data model structure is shown in Fig. 2.

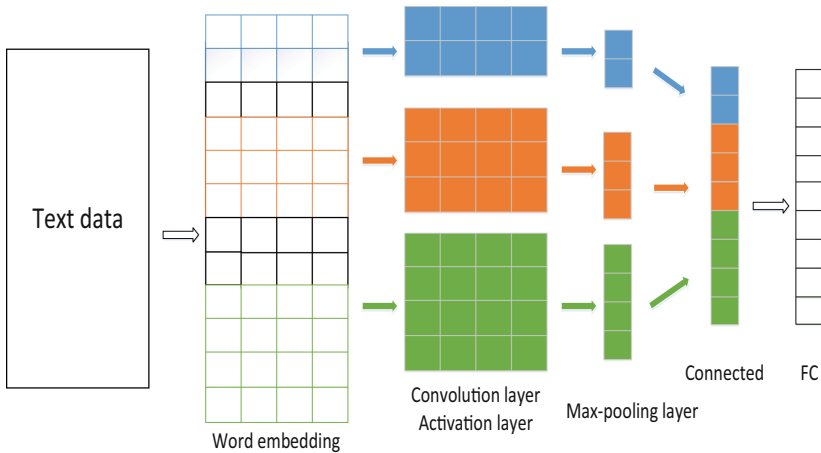


Fig. 2. Text data model structure

4.3 Model Construction

In this paper, numerical features processed by deep neural network and text features processed by convolutional neural network are combined and input into the full connection layer. Finally, classification tasks are completed through softmax layer. The model structure is shown in Fig. 3.

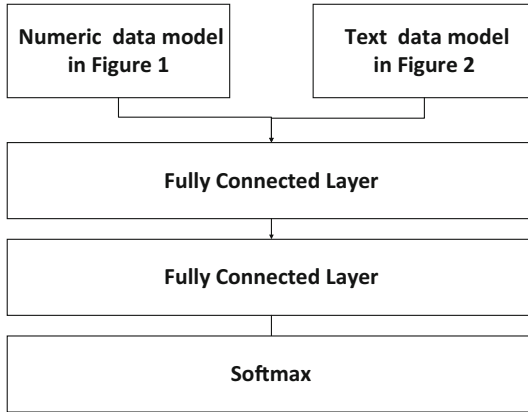


Fig. 3. Model structure

5 Experimental Results and Analysis

5.1 The Experimental Data

The data sets used in this paper are all from the criminal case information system of public security Intranet. The number of suspects in this data set is extremely unbalanced. In order to balance the data, this paper unifies the label of more than or equal to two people committing the crime as 2, and the label of single person committing the crime as 1.

The input data consists of numeric data which is shown in Table 6 and text data. The text data is description of the case, examples are shown in Fig. 4.

****年*月*日*时,报警人***,身份证号****,联系电话****,现住址****,称其在***地丢失****,被盗物品预估价值***,有(无)发票。↵

Fig. 4. Text data examples

In this paper, training sets, validation sets, and test sets are allocated according to the ratio of 7:2:1. The optimal model is determined by training set and cross validation set. The precision rate, recall rate, and F value of the test set are calculated as model evaluation.

5.2 DNN Parameter Setting

This paper uses the grid search method to adjust and optimize the parameters in DNN network. Parameter value optimal parameter values in 3-layer DNN network are shown in Table 6.

Table 6. Parameter value and optimal parameter value

Parameter	Parameter value	Optimal parameter value
batch_size	16, 32, 64, 128	64
epochs	10, 20, 30, 40, 50	10
init_mode	'he_normal', 'he_uniform', 'glorot_normal', 'glorot_uniform', 'lecun_normal'	'glorot_normal'
learning_rate	0.001, 0.01, 0.0001	0.001
momentum	0.0, 0.2, 0.4	0.2
activation	'relu', 'tanh', 'softmax', 'linear', 'hard_sigmoid', 'softplus', 'selu'	tanh
dropout	0.0, 0.1, 0.2, 0.3, 0.4, 0.5	0

5.3 Experimental Results and Analysis

Two comparative experiments are designed in this paper. One is an algorithm without text data, and the other is an algorithm with text data but only one convolution size in CNN. The experimental results are shown in Table 7.

The experimental results are analyzed as follows:

- (1) How to convert case information into values is the most important part of this paper. The experimental results show that it is feasible to convert the case information into the corresponding value according to the method presented in this paper. By modeling the features of the case data, good results can be achieved.
- (2) Through feature modeling, case information can be described with numerical data, but the information description is incomplete. The accurate in predicting the number of people who committed the crime is 71%. There is still room for improvement.
- (3) After adding text information, the precision rate and recall rate of the algorithm has been greatly improved. Especially in terms of precision rate, it has increased by 20%. In the experimental results, the precision rate and recall rate are relatively balanced, which indicates that the model is good at predicting the number of suspects.
- (4) Compared to the size of a single convolution, the model that has different sizes of convolution cores is better to extract feature information.

Table 7. Experimental results

Algorithm	Precision rate	Recall rate	F-values
DNN	71	73	72
DNN-CNN (filter size = 3)	90	90	90
DNN-CNN (filter size = [2, 3, 4])	91	91	91

6 Conclusion

In this paper, the number of suspects is predicted by modeling the features of real case information. As far as we know, this is the first time to combine the numerical and textual features of the case information. In terms of experimental results, this model has a better effect, with 91% precision and recall rate. When new cases occur, it can provide the police with more reliable prediction results. This model has a strong practical significance.

In terms of text data processing, how to use natural language processing algorithm to minimize text information in a deeper level is the next step of research. In addition, the application of this model in predicting other characteristics of suspects can be further explored. For example, suspect age, suspect domicile place.

References

1. Jin, G., Zhu, S., Lin, X.: Analysis and prediction of criminal situation in China (2017–2018). *J. Chin. People's Public Secur. Univ. (Soc. Sci. Ed.)* **2**(5), 99–110 (2016)
2. Asmai, S.A., Roslin, N.I.A., Abdullah, R.W., et al.: Predictive crime mapping model using association rule mining for crime analysis. *Sci. Int.* **26**, 1703–1706 (2014)
3. Liu, M., Lu, T.: A hybrid model of crime prediction. *J. Phys: Conf. Ser.* **1168**(3), 032031 (2019). <https://doi.org/10.1088/1742-6596/1168/3/032031>
4. Tollenaar, N., van der Heijden, P.G.M.: Which method predicts recidivism best: a comparison of statistical, machine learning and data mining predictive models. *J. Roy. Stat. Soc.* **176**(2), 565–584 (2013)
5. Li, R., Sun, C., Ji, J.: Suspect characteristics prediction based on support vector machine. *Comput. Eng.* **43**(11), 198–203 (2017)
6. Vural, M.S., Gök, M.: Criminal prediction using Naive Bayes theory. *Neural Comput. Appl.* **28**(9), 2581–2592 (2017)
7. Sun, F., Cao, Z., Xiao, X.: Application of an improved random forest based classifier in crime prediction domain. *J. Intell.* **33**(10), 148–152 (2014)
8. Kim, Y.: Convolutional neural networks for sentence classification. Eprint Arxiv (2014)