Comparison of Machine Learning Algorithms for Predicting Crime Hotspots

ABSTRACT

Crime prediction is of great significance to the formulation of policing strategies and the implementation of crime prevention and control. Machine learning is the current mainstream prediction method. However, few studies have systematically compared different machine learning methods for crime prediction. This paper takes the historical data of public property crime from 2015 to 2018 from a section of a large coastal city in the southeast of China as research data to assess the predictive power between several machine learning algorithms. Results based on the historical crime data alone suggest that the LSTM model outperformed KNN, random forest, support vector machine, naive Bayes, and convolutional neural networks. Therefore, future crime prediction should take advantage of both historical crime data and covariates associated with criminological theories. Not all machine learning algorithms are equally effective in crime prediction.

EXISTING SYSTEM

The focus of crime hotspot prediction is to forecast future concentration of criminal events in a geographical space. Theoretical criminology provides the necessary theoretical basis. Specifically, several related criminological theories not only provide guidance for us to understand the important influence of location factors in the formation and aggregation of criminal events, but also provide a basic mechanism for the police to use information of crime hot spots for crime prevention or control. It mainly includes routine activity theory, rational choice theory, and crime patterns theory.

Datasets are available in Kaggle.com

PROPOSED SYSTEM

In this paper, random forest algorithm, KNN algorithm, SVM algorithm are used for crime prediction. First, historical crime data alone are used as input to calibrate the models. Comparison would identify the most effective model. Second, built environment data such as road network density and poi are added to the predictive model as covariates, to see if prediction accuracy can be further improved.

CONCLUSION

Our models have improved prediction accuracies, compared with other models. In empirical research on the prediction of crime hotspots, Rummens et al. used historical crime data at a grid unit scale of 200 m×200 m, In the biweekly forecast, the highest case hit rate for the two-robbery type is 31.97%, and the highest grid hit rate is 32.95%; Liu et al. Used the random forest model to predict the hot spots in multiple experiments in two weeks under the research scale of 150 m

×150 m. The average case hit rate of the model was 52.3%, and the average grid hit rate was 46.6%. and the average grid hit rate was 57.6%, which was improved compared with the previous research results,

REFERENCES

[1] U. Thongsatapornwatana, "A survey of data mining techniques for analyzing crime patterns", Proc. 2nd Asian Conf. Defence Technol. (ACDT), pp. 123-128, Jan. 2016.

[2]J. M. Caplan, L. W. Kennedy and J. Miller, "Risk terrain modeling: Brokering criminological theory and GIS methods for crime forecasting", Justice Quart., vol. 28, no. 2, pp. 360-381, Apr. 2011.

[3] M. Cahill and G. Mulligan, "Using geographically weighted regression to explore local crime patterns", Social Sci. Comput. Rev., vol. 25, no. 2, pp. 174-193, May 2007.

[4] A. Almehmadi, Z. Joudaki and R. Jalali, "Language usage on Twitter predicts crime rates", Proc. 10th Int. Conf. Secur. Inf. Netw. (SIN), pp. 307-310, 2017.

[5] H. Berestycki and J.-P. Nadal, "Self-organised critical hot spots of criminal activity", Eur. J. Appl. Math., vol. 21, no. 4, pp. 371-399, Oct. 2010.

[6] K. C. Baumgartner, S. Ferrari and C. G. Salfati, "Bayesian network modeling of offender behavior for criminal profiling", Proc. 44th IEEE Conf. Decis. Control Eur. Control Conf. (CDC-ECC), pp. 2702-2709, Dec. 2005.

[7] W. Gorr and R. Harries, "Introduction to crime forecasting", Int. J. Forecasting, vol. 19, no. 4, pp. 551-555, Oct. 2003.