

Crime Hotspots Analysis in South Korea: A User-Oriented Approach

Aziz Nasridinov¹ and Young-Ho Park^{2*}

¹School of Computer Engineering, Dongguk University at Gyeongju,
123 Dongdaero, Gyeongju, Gyeongbuk, 780-714, Korea
aziz@dongguk.ac.kr

²Department of Multimedia Science, Sookmyung Women's University,
Cheongpa-ro 47-gil 100, Yongsan-Ku, Seoul, 140-742, Korea
yhpark@sookmyung.ac.kr

*Corresponding Author

Abstract. Law enforcement agencies use various crime analysis tools to discover crimes. However, a big volume of crime data has made the process of analyzing crimes difficult. Previously proposed methods only consider improving the productivity of the detectives and other law enforcement agencies. In this paper, we argue that the crime analysis methods should be also useful for citizen, and propose a user-oriented crime analysis and notification method. The proposed method use k-means clustering algorithm to determine dangerous crime regions, and web-based Geographic Information System (GIS), so called Crime Map, for tracking crime hotspots in South Korea.

Keywords: crime analysis, crime hotspots, k-means clustering algorithm.

1 Introduction

Historically, South Korea is known as a land of calmness. However, the recent social and economic growth have resulted in significant increase of crimes. Meanwhile, the growing availability of information technologies has enabled law enforcement agencies in South Korea to collect the detailed data about various crimes and analyze them. However, a big volume of such crime data has made the process of analyzing crimes difficult.

Several researchers proposed methods to solve this problem. We can divide these methods in two types: machine learning techniques based methods and statistical analysis based methods. Machine learning techniques perform crime analysis by examining a big volume of crime data and determine patterns and trends. Chen et al. [1] proposed a data mining framework for detecting and predicting crimes using entity extraction and association rule mining. Nasridinov et al. [5] proposed a crime prediction method based on decision tree classification. Statistical analysis based methods perform crime analysis using various statistical and computational methods. Several researchers applied statistical analysis based methods to study the relationships between crime and places [2], visibility [3], environmental [4] and other

social factors [6]. However, these papers mainly deal with improving the productivity of the detectives and other law enforcement agencies.

In this paper, we argue that the crime analysis methods should be also useful for citizens in order to better understand the location where they live in. Thus, we first use a k-means clustering algorithm to determine dangerous crime regions. Next, we propose a web-based Geographic Information System (GIS) for tracking crime hotspots in South Korea. One of the main advantages of proposed method is that it serves as a useful tool for both police agencies to determine crime and citizens to be on the safe side of the places they live in.

The rest of the paper proceeds as follows. Section 2 discusses the related work. Section 3 describes the proposed method. Chapter 4 highlights conclusion.

2 Related Work

In this section, we discuss related work. We categorize these papers into two groups: machine learning techniques based methods and statistical analysis based methods.

Machine learning techniques use pattern recognition, statistical and mathematical techniques to determine patterns and trends in a big volume of crime data. Several researchers used these techniques to analyze the crimes. For example, Chen et al. [1] proposed a data mining framework for detecting and predicting crimes using entity extraction and association rule mining. The proposed framework is also able to determine the relationships between crime types and data mining techniques applied in crime data analysis. Nasridinov et al. [3] proposed a method for constructing a decision tree based classification model for crime prediction. The proposed method predicts crimes by analyzing the biological data of a person that are received by various sensor in his/her body.

Statistical analysis based methods analyze the relationship between crime and various social factors. Since our method focuses on crime analysis in South Korea, we discuss application of statistical analysis based methods in South Korea. Lee [2] proposed to consider the relationship between crime and places, such as apartment complexes, low-rise multi-family housing and neighborhood parks. Lee and Kim [3] proposed to study the relationship between crime and visibility, urban density and the scale of housing complexes. Lee and Cho [4] proposed to explore the environmental factors' impact on offenders' spatial decision making. Park [6] proposed to construct a correlation matrix between crime and various social factors, such as population density, income, education attainment, number of foreign residents, temperature, urbanization and alcohol consumption. According to this correlation matrix, the authors introduce a ranking criteria of crime types.

3 Crime Hotspots Analysis in South Korea

In this section, we describe the steps of the crime hotspots analysis. It has three steps. In the first step, we retrieve crime dataset from various sources, such as National Police Agency of South Korea and Supreme Prosecutor's Office. The dataset contains

detailed information of each region about security level, number of CCTV cameras, number of household, alcohol level, temperature, population, number of foreigners, and number of arrests. In the second step, we parse the extracted raw data according to the crime types and social attribute types, and store in a database. In the third step, we apply k-means algorithm using R statistical software.

We used k-means algorithm in order to cluster the Seoul regions into dangerous, average and safe. K-means algorithm clusters the regions into k groups, so that the total distance between the group's members and its corresponding centroid. The result of clustering using R Statistical Software is demonstrated in Fig.1. The clustering is performed between number of crimes, number of foreigners, security level and number of arrests. Red dots are attributes that indicate safe region. Black dots are attributes that indicate dangerous region. Green dots are attributes that indicate average region.

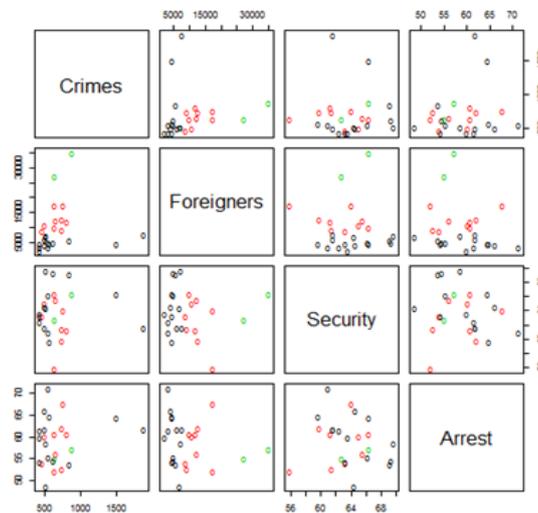


Fig.1. Clustering of Seoul regions in danger, average and safe zones.

Once the regions are clustered, we can map the crimes to the web-based GIS system for tracking crime hotspots as shown in Fig.2. It contains useful functions not only for law enforcements agencies, but also to ordinary citizens. For example, alert crime events function gives a warning about potential risk when the user is entering into region that is marked as danger. It also warns when the user is entering into a heavy traffic location and recommends different directions where the traffic is less heavy.



Fig.2. Web-based GIS for crime tracking.

4 Conclusion

In this paper, we proposed a user-oriented method for crime analysis. The proposed method is useful only for law enforcements agencies, but also to ordinary citizens. Particularly, ordinary citizens can observe the crime map and restrict himself from entering to the places that has potential danger.

Acknowledgement. This work was supported by the IT R&D program of MKE/KEIT. [10041854, Development of a smart home service platform with real-time danger prediction and prevention for safety residential environments].

References

1. Chen, H., Chung, W., Xu, J. J., Wang, G., Qin, Y., Chau, M., "Crime data mining: a general framework and some examples," *The Computer*, Vol. 37, No. 4, pp. 50-56, 2004.
2. Lee K. W., "Interpretation of Crime-Prone Locations through Spatial Analysis Theory: Burglary Case Studies in Apartment Complex," Master Thesis, Seoul National University, 2001.
3. Lee C., Kim E., "An Analysis on Structural Changes in Regional Crime Behaviors before and after Financial Crisis," *The Journal of the Korean Urban Administration Association*, Vol. 25, No.1, pp. 295-324, 2012.
4. Lee S. W., Cho J. K., "The Effects of Spatial and Environmental Factors on Crime Victimization," *Seoul City Research*, Vol. 7, No. 2, pp. 57-67, 2006.

5. Nasridinov, A., Ihm, S. Y., Park, Y. H., "A Decision Tree Based Classification Model for Crime Prediction," Proceedings of the 10th International Conference on Secure and Trust Computing, data management, and Applications, pp. 531-538, 2013.
6. Park, Y. H., "Relationship Analysis between Crime Types and Social Attributes in South Korea," Database Research, Vol. 29, No. 2, pp. 81-94, 2013.