



Research and Analysis of Crime Prediction in China and Abroad Based on Knowledge Graph

Yixin Liu, Juan Wang^(✉), and Peng Zhang

China People's Police University, Langfang, Hebei, China
66267476@qq.com, zp_1981@aliyun.com

Abstract. The era of big data is changing the mode of police work. Crime prediction based on big data analysis has become a research hotspot in crime prevention and combat. In order to fully reveal and summarize the current development status and characteristics of the crime prediction field in China and abroad, and to provide reference for the next development and research direction of the crime prediction field. This article will use the documents in the CNKI and Web of Science databases as the research samples, using the knowledge graph visualization software Citespace, and focus on analyzing the research hotspots and research progress in the field of crime prediction from the perspectives of the number of articles, research institutions, and keywords. Through research, it is concluded that the research hotspots in the field of crime prediction in China and abroad focus on prediction methods, prediction objects, and factors that affect the prediction results.

Keywords: Crime prediction · CiteSpace · Visualization · Co-occurrence Map

1 Introduction

In the context of the information age, crime, as a serious and widespread social problem on a global scale, continues to have a negative impact on the normal order of social life and production; prediction is to pre-speculate the facts that will occur in the future. The phenomenon of crime has its own law of occurrence and development [1]. Crime prediction is based on comprehensive and accurate intelligence information data, through scientific research methods to speculate and judge the increase or decrease, type changes, and development trends of crimes that may occur within a certain social range.

The influential crime prediction research first appeared in 1828, when the Belgian statistician Ketterer took the lead in using probability theory to make predictions. The term “crime prediction” was first proposed by E.W.Bnngess from the United States 100 years later, but the scope of the prediction at that time was framed on the possibility of people who were released on parole and released after serving their sentence and committing a crime again. In 1960, the crime prevention conference held by the United Nations formally cited the term “crime prediction”, advocating a comprehensive study

of crime prediction. In recent years, my country's crime prediction scholars have done relevant research on crime prediction, summed up a large number of advanced international experience in crime prediction, and provided practical experience for China's crime prediction [2]. However, crime prediction in China started around the 1980s. Compared with foreign countries, crime prediction in my country started late and has not deep roots. Academic research also faces a lack of perspective of the times and is not sensitive to academic trends [3], research direction A series of problems such as insufficient refinement.

In view of this, this article uses knowledge graph visualization software to analyze the Chinese and foreign literature data in the field of crime prediction through institutional co-occurrence, keyword co-occurrence, author co-occurrence, cluster analysis and other methods to understand the distribution of institutions and keywords Distribution, core authors, etc. and sort out the research hotspots and development directions in the field of crime prediction, summarize the current research and provide references for future development.

2 Research Methods and Data Sources

2.1 Research Methods

The visual knowledge map is based on combining the theories and methods of graphics, information science and other disciplines with bibliometrics, co-occurrence analysis, etc., to deeply dig out the literature research context, research hot issues and potential frontier trends in a subject field. This article mainly uses Java-based scientific and technological text mining and information visualization software CiteSpace V (5.7.R1).

2.2 Data Sources

In order to ensure the quality of the literature, the literature collected in this article comes from China Knowledge Network (CNKI) and Web of Science (WOS). The data was retrieved and downloaded on January 24, 2021.

The scope of English document data collection is limited to the Web of Science Core Collection in WOS, but only documents in the category of Article are included. Start the CiteSpace software, select Import/Export under the data column, import the original data under WOS, use the option "Remove Duplicates (WOS)" to perform deduplication operations on the original data, and finally get 1737 valid sample data [4].

The search scope of Chinese data is all journals in CNKI, excluding English documents. Select "Export Document" in the "Export and Analysis" option under CNKI platform and export it to Refworks format. Run the CiteSpace software, select Import/Export under the data column, and import the exported raw data in CNKI, which can be converted into Web of Science format, and finally 405 valid sample data will be obtained.

3 Analysis of the Amount of Papers and Research Institutions

3.1 Paper Volume Analysis

Figure 1 shows the statistical visualization of the number of documents in the field of crime prediction at home and abroad from 1990 to 2020. Through analysis, it can be

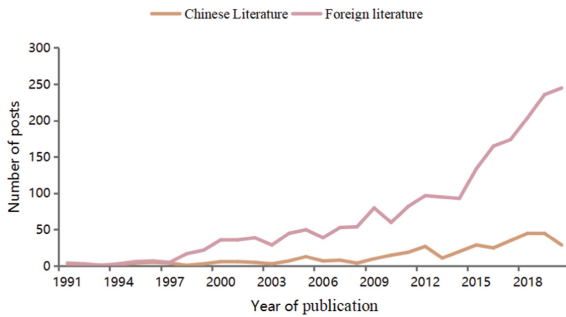


Fig. 1. Distribution map of research literature in the field of crime prediction in both Chinese and English

seen that there are certain differences in the amount of publications in the field of crime prediction between China and foreign countries. The number of English documents in the field of crime prediction is much higher than the number of Chinese documents in most years, especially in recent years. As the year progresses, the number of papers is on the rise and the growth trends of the two are roughly the same. The research on crime prediction can be divided into three stages: (1) The number of papers in this field published in 1997 and before is very small. The first Chinese document appeared in 1984, and the annual index volume was 2.07 in 1997; the first English document appeared in 1976, and the annual index volume was 1.375 in 1997 (except for 1984–1990, which was not included) (2) Since 1998–2012, the annual publication volume of Chinese and English documents has increased significantly, and the overall trend has shown a slow growth trend; (3) Since 2012, the volume of Chinese and English documents has increased rapidly. Increase, showing a rapid growth trend. It can be concluded from comprehensive analysis that the field of crime prediction has been continuously valued internationally as an emerging field. However, from the perspective of the amount of articles published, the research interest in English academic circles has been much higher than that in Chinese literature circles since 1997.

3.2 Analysis of Characteristics of Research Institutions

As can be seen from Fig. 2, the top five organizations with the frequency of cooperation in crime prediction are: Chinese People's Public Security University (11 articles), Zhejiang University of Technology Law School (9), Zhejiang Police College (4), Tongji University Law School (3), Central Judicial Police Academy (3). Except for the People's Public Security University of China, the frequency of publication of cooperative documents in this field by other institutions does not exceed 10. It can be seen that the number of connections between nodes in Fig. 3 is extremely small, indicating that China's institutions and institutions have not yet formed a pragmatic and efficient research cooperation network in the field of crime prediction, and there is no stable cooperation group.

In order to understand the distribution characteristics of research forces in the field of crime prediction in the world, the Citespace software was launched, and the data in WOS was used to analyze the cooperative co-occurrence map of institutions. From



Fig. 2. Co-occurrence map of cooperation among agencies in the field of China's crime prediction

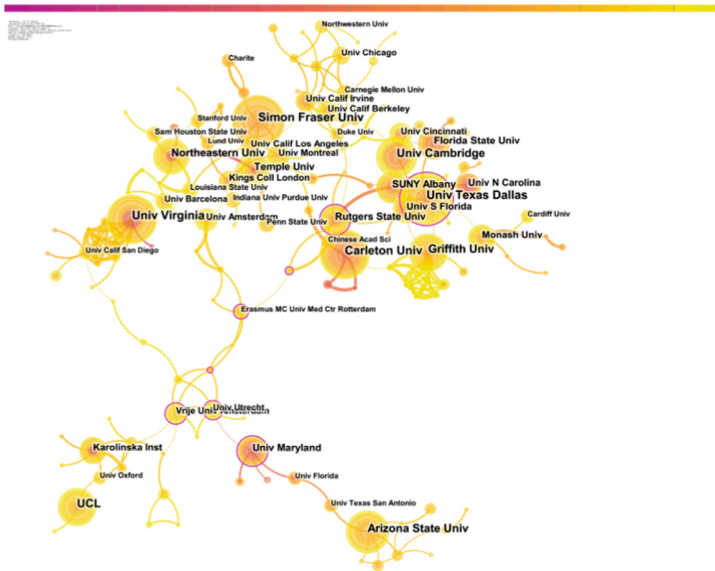


Fig. 3. Co-occurrence map of cooperation among institutions in the field of international crime prediction

the perspective of institutional distribution (Fig. 3), the top five institutions in terms of co-publishing volume are: Simon Fraser University (23 articles), Texas State University Dallas (23), Carleton University (22), University of Virginia (19), Arizona State University (19). Among them, only 4.5% of Chinese institutions have published more than 10 articles, and only 4.8% of cooperative publications; American institutions have the highest proportion of 55%. Most of the top cooperative institutions are colleges and

universities. While disseminating knowledge and culture, colleges and universities continue to innovate and think about new things, not only disseminating new technologies to all sectors of society, but also exporting high-level talents for social services. At the same time, it can be seen that there are still some research institutions outside the main cooperative institutions. Although there is a certain amount of publications, a stable cooperative group has not been formed [5].

4 Keyword Analysis

4.1 Keyword Co-occurrence Analysis

Keywords are often a high-level summary of a document, the extraction and concentration of research content, and the essence and core of the research theme [6]. We will adopt the method of keyword co-occurrence to visually and multi-dimensionally display the keyword co-occurrence maps studied in the field of crime prediction at home and abroad, so as to objectively observe and analyze them.

Start the CiteSpace software, select Keyword under the Node Types column, set the time slice to 1 year, set TOPN = 50, select Pathfinder for cropping, and the rest are the default values to get Fig. 4 and Fig. 5.

It can be seen from Fig. 4 that the top China’s keywords in terms of frequency are: crime prediction (Freq = 121), big data (Freq = 56), crime mapping (Freq = 30), crime hotspots (Freq = 16), crime prevention (Freq = 15), prediction (Freq = 15), data mining (Freq = 11), predictive policing (Freq = 10), among which the centrality of intermediary is higher: big data (0.59), crime prediction (0.51), Crime mapping (0.43), it can be seen that China’s research in this field focuses on crime categories and prediction methods.

According to Fig. 5, the top international keyword frequencies are: crime “crime” (Freq = 469), prediction “prediction” (Freq = 281), violence “violence” (Freq = 155), behavior “behavior” (Freq = 142), recidivism “recidivism” (Freq = 128), among which the middleness is higher: comprehensive analysis “meta analysis” (0.23), prediction

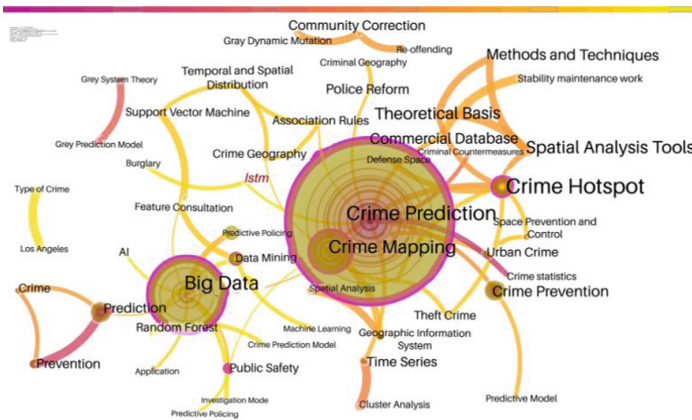


Fig. 4. Keyword co-occurrence map in the field of crime prediction in China

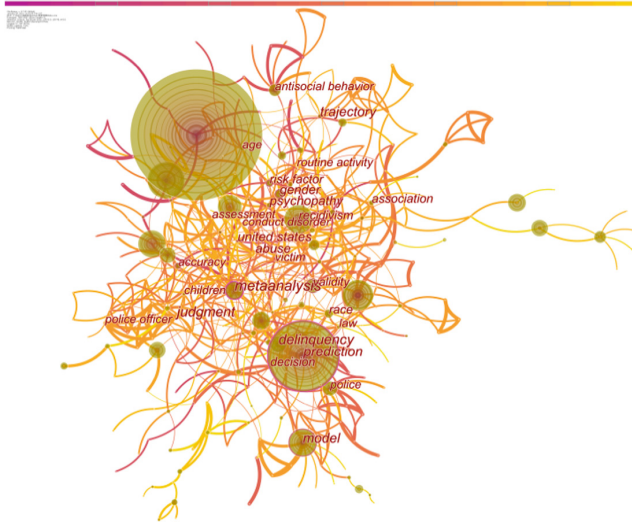


Fig. 5. Keyword co-occurrence map in the field of international crime prediction

“prediction” (0.21), human “men” (0.16). It can be seen that international research in this field focuses on prediction methods and evaluation objects. Words such as “crime”, “prediction” and “recidivist” in international keywords are also reflected in China’s high-frequency keywords.

At the beginning of the development of crime prediction, academic researchers predicted the possibility of people who were released on parole and sentenced to commit crimes again, so it is not surprising that these words are frequently used internationally. At the same time, it can be seen that the research hotspots in the field of crime prediction at home and abroad are consistent in the overall direction [7].

Prediction methods are a core focus of research in China and abroad, whether it is China’s “big data”, “data mining”, “machine learning” and other high-frequency words or international “risk assessment” and “comprehensive analysis”. The method used behind the crime prediction.

4.2 Keyword Emergence Analysis

Keyword emergence refers to the sudden increase in the frequency of its appearance in a certain period of time. The emergent words have two distinctive characteristics, one is suddenness, that is, the keyword appears suddenly in a certain period, and its frequency increases rapidly; the other is persistence. That is, the trend of keyword frequency growth will continue for a period of time [8]. By analyzing the emerging keywords, we can grasp how the research hotspots of a certain field change with the change of the times, and can also predict the future research development direction of this field.

Start the CiteSpace software, select the “Burstness” column of the control panel on the basis of keyword co-occurrence. The gamma values are 0.5 and 1, respectively. The

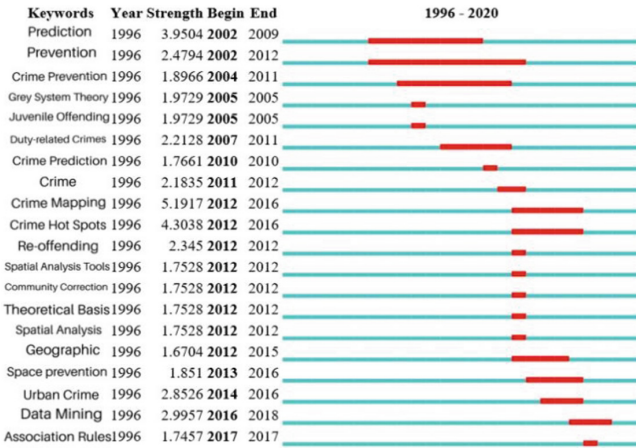


Fig. 6. Keyword emergence map in the field of China’s crime prediction

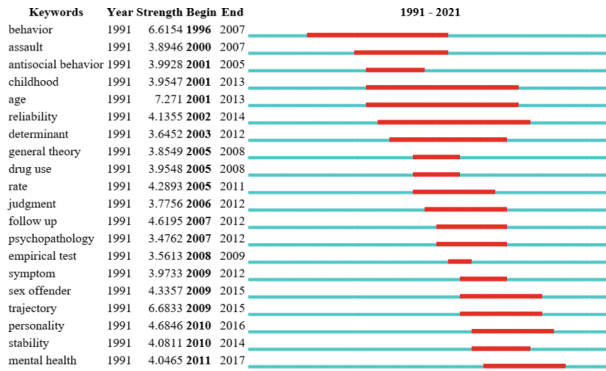


Fig. 7. Keyword emergence map in the field of international crime prediction

first 20 emergent words are selected, and the keyword emergence map is finally obtained (Fig. 6 and Fig. 7).

In the early exploration process in China, the keywords “juvenile delinquency” (Burst = 1.9729), “duty crime” (Burst = 2.2128), etc. appeared. This is because in the early days, my country paid more attention to macro forecasts in the field of crime forecasting, such as forecasting crimes in a certain industry, department, or region. With the continuous development of science and technology and the popularization of big data applications, the field of crime prediction in my country has changed from qualitative to quantitative, using mathematical prediction methods, such as gray system theory, spatial analysis, data mining, etc., to transition from macroscopic predictions to microscopic predictions [9]. This is why the China’s research hotspots in recent years are “data mining” (Burst = 2.9957) and “association rules” (Burst = 1.7457).

Judging from the strength of emergent words, the emergence strength of age “age” (Burst = 7.271), trajectory “trajectory” (Burst = 6.6833), and behavior “behavior” (Burst

= 6.6154) are all greater than 5. Among them, behavior prediction research emerged in 1996, age prediction research emerged in 2001, and trajectory prediction research emerged in 2009; from the perspective of the duration of emergence, the emergence of childhood “childhood”, age “age”, and recidivism “reliability” lasted for 13 years; Behavior “behavior” emerges for 12 years. This shows that the international literature has been paying attention to the influence of these factors on the results of crime prediction for a long time.

5 Conclusion

Since “crime prediction” was proposed in 1928, to the continuous exploration of this field in various countries in the early 21st century, to the rapid development of the field of crime prediction today, digital technology with big data as the core has been indispensable. This article uses 1737 pieces of data collected from Web of Science and 405 pieces of data collected by CNKI as samples. Using CiteSpace, it comprehensively reflects the current crimes against crimes in China and abroad in terms of the amount of publications, research institutions, and high-frequency keywords. The current status of research in the field of prediction also shows the hot topics and development trends in the field of crime prediction in China and abroad.

In this context, research in the field of crime prediction has gathered a large number of high-quality research forces. In this regard, the summary and recommendations are as follows:

First, there has not yet been an effective and stable cooperation network between institutions. At present, there is a lack of exchanges and cooperation among institutions in the field of research, and various results are limited to their respective discipline frameworks, there is overlap in research success, and there are few interdisciplinary innovative research.

Second, there is a lack of case analysis. At present, the main research topics in this field are concentrated on theoretical research, and the number of researches on examples is small, and the systematic and operability of the research needs to be improved.

Third, research hotspots are mainly concentrated in three areas: First, the methods used behind crime prediction, such as domestic: data mining, machine learning, and foreign risk assessment and comprehensive analysis, etc.; the second is the object of crime prediction: Repeat offenders, adolescents, etc.; the third is the changing factors that affect the prediction results: age, behavior, etc. At present, China’s policing model is in an important stage of shifting from business-driven to data-driven. Traditional business-driven policing is facing the marginal energy dilemma. And big data has the advantages of full volume and precision, combined with big data The crime prediction will be able to achieve multi-dimensional feature analysis and precise crime risk prediction for different targets and groups, thereby driving the development of police work from extensive to refined, and improving the effectiveness of police work.

Looking back at its development process, it can be foreseen that in the next few years, the upsurge of research on crime prediction at home and abroad will continue. In addition, driven by continuous technological innovation, research in the field of crime prediction will also move to a higher level.

Acknowledgment. Supported by National Statistical Science Research Key Project “Social Stability Risk Modeling and Early Warning Research Based on Public Opinion Big Data” (No. 2019LZ07).

Supported by Hebei Province Higher Education Teaching Reform Research and Practice Project “Research on the Training Mode and Teaching Practice of Police Big Data Application Professionals under the Background of New Engineering” (No. 2018GJJG450).

References

1. Huang Chao, Li Jihong. Methods of crime prediction[J]. Journal of Jiangsu Police Officer Academy, 2011, 26(01):107-110.
2. Yang Jincheng. The progress and trend of counter-terrorism intelligence research at home and abroad——visual measurement based on Citespace V [J]. Journal of Information, 2020, 39(01):45-55+145.
3. Zhao Jun. China's crime prediction and its research status, problems and development trends—Analysis of the content of “China Knowledge Network”[J]. Journal of Hunan University (Social Science Edition), 2011, 25(03):155-160 .
4. Yang Yang. Knowledge Modeling and Intelligence Analysis of Stakeholder Economic Crimes Based on Knowledge Graph[J]. Journal of Chinese People's Public Security University (Natural Science Edition), 2020, 26(02):87-95.
5. Zheng Ziwan, Jin Cheng, John E. Eck, Spencer Chainey, James G. Cameron, Michael Leitner, Ronald E. Wilson. Crime Mapping: Understanding Crime Hotspots (Part 2)[J]. Research on Juvenile Crime Prevention, 2012(05):67 -77.
6. Chen Peng, Ma Wei. The application of hierarchical clustering method in the analysis of spatial crime hotspots[J]. Journal of Chinese People's Public Security University (Natural Science Edition), 2013, 19(01):64-67.
7. Yan Yaojun, Zhang Ming. Construction of information management system for crime prediction spatio-temporal positioning[J]. Journal of Chinese People's Public Security University (Social Science Edition), 2013, 29(04):73-80.
8. Shi Shaochong, Chen Peng, Yuan Penghui, Hou Chao, Ming Hongxia. Application of Classification Learning Method in the Prediction and Recognition of the Regional Features of Criminals[J]. Journal of Chinese People's Public Security University (Natural Science Edition), 2019, 25(01): 59-64.
9. Zheng Wei, Xia YiXue. The Visual Correlation Analysis of my country's Internet Public Opinion and Internet Ideology Research [J]. Information Research, 2021(04):126-134.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

