

# Research on Intelligent Research Model of Clues in Cases of Production and Sale of Counterfeit Drugs

Linfeng GU<sup>1</sup>, Sisi LI, Yutian WANG, Zhihui LIN and Jingsong WU  
*School of information technology, Nanjing Forest Police College, Nanjing 210023, China*

**Abstract.** This model takes the vigorous promotion of the construction of "smart policing" and the construction of a modern police system as the background, creates a data-based, technology-based, vigorously improves the analysis and mining capabilities of public security organs on the hot issues of counterfeit drug crimes, and transforms the traditional post-processing and passive response methods of police work to the direction of pre-prediction and active intervention. At the same time, drawing on the design ideas of the system architecture, the model system is layered and decoupled, creating a multi-channel data collection, to case text preprocessing and standard definition, forming a "four-category personnel database", and then classifying according to the theme of public security business to achieve a criminal model of production and sales of counterfeit drugs The overall framework is reused vertically in the field and flexibly extended horizontally.

**Keywords.** Counterfeit drug crime, research and judgment model, active discovery, knowledge base, system construction

## 1. Research background

### 1.1. Research status and development trends at home and abroad

Counterfeit drug crimes are economic crimes related to people's livelihood, which is a new type of crime derived from the development of social and commodity economies, which has a double harm to the public's life and health and property safety, as well as the national economic management order and social stability, seriously corrupt the commercial atmosphere, and lead to the phenomenon of "bad money expelling good money" from low-quality products expelling high-quality products on the market.[1]

In recent years, the crime of counterfeit drugs nationwide is still in a serious situation, but the academic community has done less research on it, and even the new direction that has not yet been excavated in the geographical study of crime is in huge contrast with the urgent reality. Although the public security organs have strengthened daily patrols, strengthened linkage and co-governance, and increased publicity and guidance, they have shown obvious results, but they are still based on long-term work experience, for some sudden or new types of counterfeit drug hotspots, it is difficult for us to find a better entry point to prevent in advance, which may cause negligence and leakage, and may also be ineffective and waste of manpower. Therefore, we hope to explore ways to take

---

<sup>1</sup> Corresponding author: GU Linfeng, E-mail: 1533601364@qq.com.

precautions and prevent problems before they occur, and there are still many pain points in the fight against counterfeit drugs.

1.2. Realistic pain points

- **The manufacturers and sellers of counterfeit drugs are very hidden and have certain anti-investigation capabilities, and it is difficult to collect all the evidence of the crime.** Pharmaceutical production enterprises produce counterfeit drugs by not investing or less in the main materials, valuable raw and auxiliary materials, etc., and sell fake and inferior drugs through legal channels, and their behavior is quite hidden.
- **Circulation management problems.** Those who manufacture and sell counterfeit drugs mainly complete the "flow of goods" through post offices and logistics companies. There are four reasons for this: First, mailing and express delivery to sell counterfeit drugs quickly; Second, the review is not strict, and safety is guaranteed; The third is to adopt the cash on delivery method, which has a certain fraudulent nature and convenient payment collection;Fourth, it is not in direct contact with patients, and it is highly concealed.[2]
- Police data resources are relatively limited, and it is difficult to reflect the hot spots of social governance of counterfeit drugs; The research results of counterfeit drugs have focused on solving a type of problem and have not formed a system.[3] As shown in Figure 1.

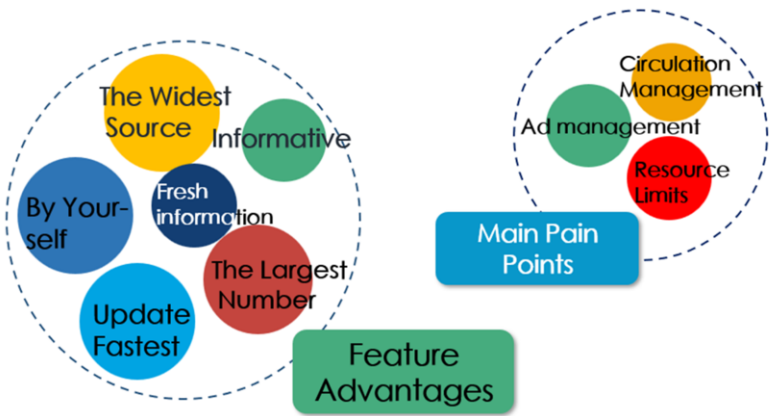


Figure 1. Feature Advantages and Main Pain Points

2. Technical realization

2.1. Multi-channel data access

In the selection of analysis samples, the idea of retrieving and collecting relevant data of counterfeit drug production cases involved in the whole country through the Chinese judgment documents and Openlaw legal document platform and local courts was determined, and then the division of labor was carried out according to different provinces, batch screening and downloading cases, and the data were integrated in

accordance with unified norms, and the results of case integration in different provinces were summarized separately for follow-up analysis.[4]

The research object of this project team is the criminal case of production and sale of counterfeit drugs recorded in the past five years in various provinces across the country recorded in the China Judgment Documents Network and Openlaw, and after data cleaning and preliminary review, 60,746 pieces of case data with complete data content and regular format are integrated and analyzed by case related variables. A total of 36 crimes and 60,746 cases were committed in four sections. The first part has 10 crimes against the production and sale of counterfeit and shoddy drugs, with a total of 18,182 case texts, the second part has 8 crimes of infringement of intellectual property rights, a total of 8,802 case texts, the third part has 2 crimes against public health, a total of 35 case texts, and the third part has 16 crimes against the protection of environmental resources, with a total of 33,727 case texts.

2.2. Definition of case text preprocessing and extraction standards

After using PivotTable function to preliminarily explore the correlation between various variables and case information, classify and sort the data according to provinces and case types, study the internal relationship between case frequency and time and space of each province, and complete the preliminary statistical analysis work. Through semantic analysis technology to realize the analysis of Internet social problems and police intelligence, realize the transformation of unstructured data to structured data.

This article extracts the key industries, key places, key parts, key personnel, main items, main species, and main crime methods involved in the case text in accordance with the standards. As shown in Table 1.

Example	According to the standard	Label the results
The suspect Tong is a pharmacist who colludes with Lin to sell counterfeit drugs	Key industry standards	Pharmaceutical manufacturing
Chen illegally obtained counterfeit drugs and shipped them to slimming pharmacies for sale	Standards for key places	Weight loss pharmacy
Under the pretext of beauty, Li sold counterfeit drugs to customers in the beauty shop he opened	Standard for key areas	Beauty shop
The suspect Wu mou is an individual and colludes with Lin to sell counterfeit drugs	Key personnel standards	individual
Ding illegally sold compound rheumatic capsules at outlets	The standard of the main items involved in the case	Compound rheumatic capsules
Sun illegally sold a cheap nine-valent vaccine	The criteria for the main species involved in the case	vaccine
Lin took the WeChat group as the starting point and sold counterfeit drugs through WeChat payment methods	Standards for the main modus operandi	Establish WeChat groups and sell online

Table 1. Label example table

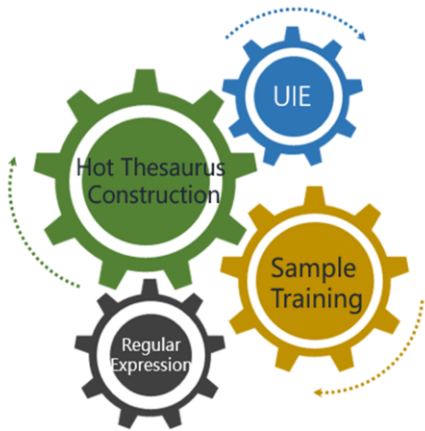


Figure 2. Push The Process

On the basis of structured data, it further relies on deep neural network to realize intelligent mining of keywords of network social problems and police information. Tagging text extraction as follows: people, places, organizations, time, emotion is negative, similarity information, the main body, keyword information, information, publishing time, publishers, original forwarding information, field category information, sensitive category information, level of sensitive information, garbage label information such as hundreds of tags. In terms of text features, Word2VEC based and Bert based text semantic feature extraction techniques are respectively adopted according to different scenarios. As shown in Figure 2.

By classifying and labeling the historical data, the training data sample library of fake medicine text was constructed. The text semantic feature modeling based on Word2VEC method is implemented, and Text\_CNN text classification technology is used to achieve the rapid extraction of local features, and the basic classification answer is output from the problem. The topic case, entity label, trigger word label and other important information are introduced for model fusion, and the accuracy of classification is further improved. Finally realize the classification of fake medicine text.

2.3. Integrate data sources and establish “Thesaurus of Four Types of Personnel Data”

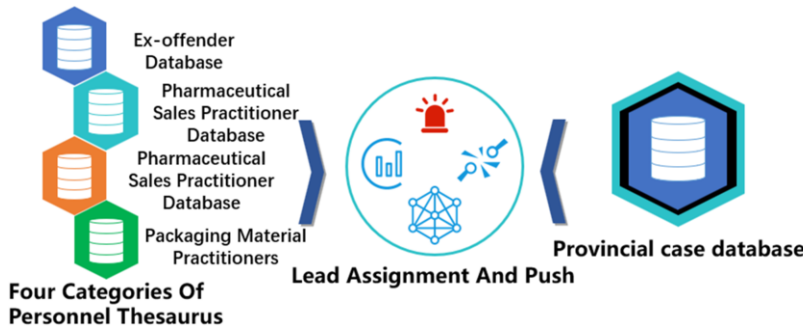


Figure 3. Thesaurus Building and Pushing

In the preparation process of manufacturing and selling fake drugs, raw materials such as inferior drugs, bottles, pill boxes, security codes and trademarks will be purchased from the suppliers of original wine and packaging materials. In the manufacturing process, to find production dens, hire packers to organize production; During the storage process, you need to find the storage dens. Temporary placement of unsold counterfeit medicines; In the sales process, it is necessary to contact the drug sales front line and support personnel for transportation and sales. Again, in this process, a large amount of relevant personnel information, funds, logistics and other key data will be generated, which can accurately depict and restore criminal activities, and formalize the evidence we need to combat such crimes. We use Word2VEC method to realize the semantic feature modeling of Text, and use text-CNN Text classification technology to achieve rapid extraction of local features, so as to output basic classification labels, and introduce multi-element information such as subject labels, entity labels and trigger word labels for model fusion, so as to further improve the accuracy of classification. Finally, the intelligent classification of police data is realized. We construct the required structured data thesaurus according to the following two points. As shown in Figure 3.

### *2.3.1 Aggregate and integrate internal and external data resources to establish "four types of personnel database" information*

- Database of ex-offenders: In the text of the case, the public security organs punished and punished ex-offenders for manufacturing and selling counterfeit drugs; Persons at large and evacuated for the crime of manufacturing and selling counterfeit drugs throughout the country; Administrative punishment personnel for the manufacture and sale of counterfeit drugs by the Market Bureau; Information on the production and sale of counterfeit drugs seized by the major departments of pharmaceutical companies.
- Drug sales practitioner database: self-employed drug personnel registered in the case text; E-commerce platforms open pharmaceutical stores.
- Logistics employee database: logistics express employees
- Packaging materials employees: Employees of relevant printing plants and packaging material factories involved in the case

### *2.3.2 In response to the characteristics of the crime, clarify the elements of research and judgment, and collect "third-rate information"*

- Capital flow information: account, online banking, alipay transaction information of the person involved in the case
- Communication flow information: spider web address book information, software address book information
- Logistics information: logistics delivery and delivery information

2.4. Clue assignment and push



**Figure 4.** The Assignment Of The Person Type

Through explicit and implicit two-dimensional reasoning, the model can reason social public opinion to the police topic, so as to achieve the automatic generation of social hotspots to police work goals, in the downstream of this model, is a steady stream of thematic business model, we for the time correlation, location correlation, character relevance, behavior correlation and other dimensions in accordance with the rules of the assignment, according to the total score value of the order, the branch is higher than 40 people are included in the scope of attention. According to the calculation results, according to the score level, the key objects are listed, the relationship chain is generated, and the associated personnel are included in the scope of early warning of the deployment control. Combined with its capital, logistics, communication upstream and downstream relationships to find out the sales port of counterfeit drugs and inferior drugs, test and buy identification, test the flow of funds, meet the standards for filing a case, and file a case for investigation. As shown in Figure 4.

3. Innovation highlights and practical results

3.1. Multi-channel construction of four types of personnel hot thesaurus, actively find and reduce human resource costs

By producing a large amount of data fusion of counterfeit crime subject keywords, combined with different associated personnel in each link of information data, to build a hot link library in the four types of personnel, through the model with the help of the public website relevant data real-time implementation of spiders crawling, real-time grasping library update supplement, realize the diversity of data source and thesaurus, For future cases to provide data support. This model carries out collision results of four types of personnel database data. After the results are generated, the data system of gang members, transaction record evidence and sales network is formed by using "third-rate information" to trace the source and relationship data, so as to further promote the investigation work. To find and collect clues of production and sales of fake drugs, and provide them to the corresponding functional departments in time, so as to provide safe and reliable technical support for law enforcement personnel. Daily police can pay attention to the real-time update of the keyword thesaurus, obtain the peak time and its

frequent areas, carry out timely actions, realize the process from passive attendance to active analysis, and greatly reduce the cost of human resources.

### *3.2. Intelligent analysis of public opinion and police intelligence to improve data processing efficiency*

Using semantic analysis technology, the scheme of the collected from the network to produce counterfeit drugs related case analysis classification, sorting out and produce counterfeit crime cases related elements, implementation to produce counterfeit case flows, communication records, logistics information and other elements, and through the above elements to achieve a more accurate judgment, targeted projections for the criminal process. The collected massive texts related to counterfeit drug cases were independently classified, and the heterogeneous sub-structured texts were extracted as unified vector features to facilitate the subsequent correlation analysis and model fusion of the cases. The different links of the whole model system were connected in series to realize the end-to-end closed-loop computing process. Through the model modeling, the model calculated the number of personnel risk labels: place labels 4; 68 labels of individuals involved in high-frequency cases; 124 criminal tags using wechat or QQ; 569 low-cost drug labels; 1,663 labels for low-priced supplements; Health care product label 9207; Form 47 case association model instructions. Through manual research, judgment and screening, it was transferred to the public security business department 42 items. After the second research, judgment and verification of the materials, the correct rate of related cases was 71.43%. [5]

### *3.3. Social public opinion is linked to policing and improve data processing efficiency*

This scheme draws on the design idea of the system architecture, decouples the model system in layers, realizes the vertical domain reuse of the overall model framework and the flexible expansion of the horizontal, and creates a "pyramid" model transmission system from the upstream problems of network public opinion mining to lead to social governance hotspots, to the processing factory that classifies the theme of public security business in the middle, and then to a series of corresponding practical solutions in the downstream. And through the thesaurus construction, the thesaurus will be the peak time of high-frequency words and their frequent occurrence of statistics, involving the production and sale of counterfeit drugs in the high incidence of crimes for rapid screening and screening, to achieve a complete closed loop from the discovery of problems to the solution of problems, for the police work intelligently proposed a complete set of solutions. To make up for the late start of police big data modeling applications in the field of public security, and the data related to counterfeit drug crimes is relatively complicated, which is not conducive to the efficient implementation of public security law enforcement work.

## **4. Summary and Prospect**

In this study, the construction of a thesaurus of fake drug manufacturing and sales cases was completed, and the intelligent investigation model of fake drug manufacturing and

sales cases was realized based on semantic analysis, knowledge extraction and other big data technologies. Specifically, the following work was carried out:

- The crime model, research status and development trend of counterfeit drugs at home and abroad were studied.
- Go into the police grassroots business, think and summarize the pain and difficulty in the actual combat of the police grassroots, and formulate corresponding solutions.
- Refer to the policies and instructions related to police intelligence to determine the research direction
- Completed the text analysis project of food and drug ring investigation cases, investigated related databases, and completed the construction of the domain corpus
- Construct and design the experimental architecture according to the pain and difficulty of public security experimental business, and test the model system

#### *4.1. Through the construction of a special thesaurus, improve the data utilization rate and the efficiency of the police's case handling*

This model timely labels the data after statistics, classifies the keywords according to the frequency of occurrence, and sets the hot words with more occurrences as high-frequency words, which are used to show the peak period of high-frequency words and their frequent occurrence areas. Through the use of the thesaurus, the public security police can quickly screen and identify areas with a high incidence of the crime of designing, producing and selling counterfeit drugs, promptly crack down on the crime of producing and selling counterfeit drugs, and improve the efficiency of case handling. At the same time, the thesaurus will continue to inject new case data, enrich and optimize the model, and continuously improve the data utilization.

#### *4.2. Through data intelligence, realize the transformation to science and technology to police force*

Compared with the lag in the fight against counterfeit drug crimes in the past, which led to the infringement of the people's right to life and health and property rights, this model intelligently calculates the data, builds a real-time updated data thesaurus, continuously updates the thesaurus, liberates the police from redundant case data, realizes the innovative transformation to science and technology to police forces, and greatly reduces human resource costs.

### **Acknowledgements**

This research was supported by the 2022 College Students Innovation and Entrepreneurship Training Program (Grant No. 202212213006Z).



## References

- [1] Liu Lihua. Research on the Supervision of Drug Network Sales[C].Hainan University, 2020:33-37
- [2] Gong Lu, Zhu Zhengwang, Investigation of Online Sales of Counterfeit and Inferior Drugs[J], Jiangsu Police Officers College, Yancheng Municipal Public Security Bureau, 2017
- [3] Yang Xuzheng, Investigation of Internet-related food safety crimes [D], People's Public Security University of China,2022
- [4] Xu Fengsheng, Research on the Prevention and Control Countermeasures of Counterfeit and Shoddy Drugs Sold online in China[C].Chinese Min Public Security University, 2017:35-40
- [5] Zhang Yuying, Research on the Attributes of China's Online Public Opinion Crisis Events [D]. Hunan University,2018