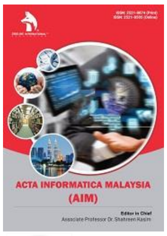
ZIBELINE INTERNATIONAL PUBLISHING

ISSN: 2521-0874 (Print)
ISSN: 2590-4043 (Online)
CODEN: AEMCDV

Acta Informatica Malaysia (AIM)

DOI: http://doi.org/10.26480/aim.01.2024.01.04



CrossMark

RESEARCH ARTICLE

RESEARCH ON THE CONSTRUCTION OF A KNOWLEDGE GRAPH OF INTERACTION RISK BETWEEN HOME INVASION THEFT OFFENDERS AND VICTIMS BASED ON INFORMATION EXTRACTION TECHNOLOGY

Tuo Shi, Danyang Li, Qi Zhang*

Beijing Police College, Beijing, China
**** Corresponding Author Email: zhangqi@bjpc.edu.cn***

This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ARTICLE DETAILS

Article History:

Received 28 September 2023
Revised 02 November 2023
Accepted 05 December 2023
Available online 08 December 2023

ABSTRACT

This thesis explores the force exerted by victims on criminals and crime phenomena based on the theory of "interactive relationship", providing a new perspective for the study of crime risk and crime formation. Through researching and applying information extraction technology, a knowledge graph of the interaction risk characteristics between home invasion theft offenders and victims is constructed, which serves as a reference for comprehensively, accurately, and dynamically understanding the formation of crime risk. This enables those in charge of crime control to anticipate the interactive patterns of such crime risks in advance, so as to intervene and control the crime risk before the crime fact emerges.

KEYWORDS

Knowledge Graph; Interaction Risk; Information Extraction

1. INTRODUCTION

For a long time, whether in academic research or in public security police practice, our attention has been more focused on studying crime phenomena centered on criminals, considering the criminal as the core of the cause-and-effect relationship of the entire crime and the main constituent element of the crime. Under this view of crime, scholars concentrate more on criminals dissecting the motivation and mechanism of crime through the study of the cognition, behavior, and psychology of the criminals, in the hope of achieving the goal of preventing and controlling crime (Li, 2015). Similarly, public security organs that carry out crime control also place more importance on cracking down on criminals (Piroozfar,2019). However, the continuous high incidence of new non-contact network crimes such as telecommunications network fraud and online gambling make people increasingly realize that this single-dimensional crime research perspective focusing on criminals is incomplete. Clearly, the generation of crime risk is not only the responsibility of the criminals; victims also "influence and shape" the crime risk (Schneider,1992). Therefore, focusing on the risk behavior of the criminal while ignoring the risk behavior of the victim is biased and unlikely to achieve effective prevention and control of crime.

The aforementioned conclusions reveal that the concepts of crime and victimhood are interdependent. Crime is not an isolated individual behavior that has nothing to do with others; it is the result of constant interaction between the offender and the victim.Thus, we believe that in most crimes, it is precisely because there is some kind of potentially risky interactive relationship between the offender and the victim. When appropriate spatiotemporal conditions trigger and activate this risk interaction, a crime phenomenon will occur (Zhou et al., 2022). Therefore, starting with the dynamic interaction relationship between the offender and the victim, and analyzing the crime risk from a dynamic, systematic perspective, can more effectively serve as a crime warning.

Based on the above views, this thesis takes the home invasion theft cases in B city as an example, cuts into the perspective of the interactive relationship between the victims and the criminals, and comprehensively applies named entity recognition and relationship extraction technology models to construct a knowledge graph oriented to the interactive relationship between the two parties (Bayley and Garofalo, 1989).This forms interactive knowledge rules between the criminal and the victim, providing a knowledge reference for effectively controlling such crimes.

2. DATA SET CONSTRUCTION

In this study, we obtained detailed case information of 18,534 home invasion theft cases in B city in 2020 through the internal resource platform of public security, and collected static attribute data of criminals and victims, dynamic behavior attribute data, and detailed description process data of crime occurrence. These data include both structured data and text data of brief case descriptions. The table contains 8409 case data of solved cases, and the sample case data are shown in Table 1.

In addition to the case data mentioned above, this empirical study also includes individual data of criminals. Due to the sensitive nature of this data, it is not shown here. It is difficult to obtain individual characteristic data of victims in this scenario, mainly because the public security organs protect the privacy of citizens, and detailed multi-dimensional features of victims of house theft have not been collected. Therefore, limited by the lack of data resources, in this empirical study, we extract a few features related to the victims based on case data.

In addition to the case data mentioned above, this empirical study also includes individual data of criminals. Due to the sensitive nature of this data, it is not shown here. It is difficult to obtain individual characteristic data of victims in this scenario, mainly because the public security organs

Quick Response Code	Access this article online	
	Website: www.actainformaticamalaysia.com	**DOI:** 10.26480/aim.01.2024.01.04

Cite the Article: Tuo Shi, Danyang Li, Qi Zhang (2024). Research on The Construction of A Knowledge Graph of Interaction Risk Between Home Invasion Theft Offenders and Victims Based on Information Extraction Technology. *Acta Informatica Malaysia, 8(1): 01-04.*

protect the privacy of citizens, and detailed multi-dimensional features of victims of house theft have not been collected. Therefore, limited by the lack of data resources, in this empirical study, we extract a few features related to the victims based on case data.

Table 1: Sample Instance of Case Data

Case Name	Unit of Record	Crime Location	Time of Crime	Brief Case Description
Theft Case of Yin XX	TZFJJKS Police Station	Building XX, Unit X, No. 1802, XX Community	2021/1/1 22:56:00	At approximately 20:40 on December 28, 2020, our station received a call from the victim Yin XX in TZQ community reporting a burglary at his home. Upon investigation, the reporter placed 2,000 yuan in cash in the shoe cabinet on the lower right side about seven or eight days ago and found the 2,000 yuan missing on December 28. A diamond ring and a crystal ring in the drawer of the shoe cabinet on the upper right side of the entrance are also gone, so he called the police.

3. Data Preprocessing

To achieve the fusion of multi-source heterogeneous data, and to deeply mine the risk characteristics, susceptibility characteristics, and interaction characteristics of criminals and victims, this section comprehensively applies named entity recognition and relationship extraction techniques to extract information from unstructured case text data. A pipeline model is used, that is, first training an entity extraction (NER) model, and then training a relationship extraction (RE) model, with the two models not affecting each other (Sherman, 1992).

3.1 Data Annotation

For the text analysis of the home invasion theft case data, it is necessary to prioritize partial sample sequence annotation of the data. The annotation method used in this empirical study is the BIO method. When performing named entity recognition in this study, the main approach is to annotate each character in the Chinese text. The types of labels for annotation are shown in the following table 2:

Table 2: Explanation of Text Sequence Annotation for Home Invasion Theft Case Data

Type	Description
B-PER	Beginning of a person entity
I-PER	Middle or end of a person entity
B-LOC	Beginning of a spatial location entity
I-LOC	Middle or end of a spatial location entity
B-TIM	Beginning of a time entity
I-TIM	Middle or end of a time entity
B-TRI	Beginning of a theft method entity
I-TRI	Middle or end of a theft method entity
B-DAM	Beginning of a lost property entity
I-DAM	Middle or end of a lost property entity
O	Non-entity

3.2 BERT + BiLSTM + CRF Entity Extraction

The model used for named entity recognition in home invasion theft case text is composed of 4 modules, as shown in Figure 1. The BERT layer generates dynamic word vectors through pre-training of the input text and uses the obtained word vector information as the input of the BiLSTM layer for bidirectional training to further extract text features. The attention mechanism mainly extracts the feature information that plays a key role in entity recognition from the results output by the BiLSTM layer, assigns weights to the feature vectors output by the upper layer, highlights the features that play a key role in entity recognition, and ignores irrelevant features (Van et al., 2021).

Through weight inspection, it directly assesses which embeddings are preferred for specific downstream tasks. Finally, the CRF layer effectively constrains the dependency relationship between predicted labels, models the label sequence, and obtains the globally optimal sequence. Through this information extraction model, factors such as personnel, location, time, and items in the case can be extracted to lay the foundation for subsequent establishment of relationship extraction.

3.3 R-BERT Based Relationship Extraction Model

In home invasion theft cases, the victim and the criminal usually do not have direct interaction. Most interactive risk behavior features come from the combined relationship rules of risk behaviors of both parties. Therefore, we use the relationship extraction model here to identify general risk interactive behavior knowledge rules. For this purpose, we chose the R-BERT model to extract relationships in the text data of home invasion theft cases. The overall architecture of the model is shown in Figure 1. The application of this model is mainly used to extract similar interaction relationship information in cases such as "criminal -- victim's residence", "criminal -- victim's property", etc.

3.4 Risk Interaction Relationship Classification

The interaction between the perpetrator and the victim in a home invasion theft crime belongs to a utilization relationship pattern. There is no direct interactive behavior between the criminal and the victim, but an indirect interactive relationship is formed through the corresponding carrier. Therefore, this study, according to the experience of field experts, divides the interactive relationship from the perspective of the media that produces the interaction into the following categories.

(1) According to different criminal methods, home invasion theft methods can be divided into technical lock picking entry, sneaking in through the door, and climbing and entering through the window.

(2) Depending on the different residences of the victims being robbed, the type of robbery can be divided into bungalows, old buildings, and luxury buildings.

(3) According to the organization behavior of the criminals, it can be divided into gang crimes and solo crimes.

(4) According to the behavior of the criminals' timing of the crime, it can be divided into night and day.

(5) Depending on the items chosen for theft by the criminals, it can be divided into valuable jewelry, cash, and electronic products.

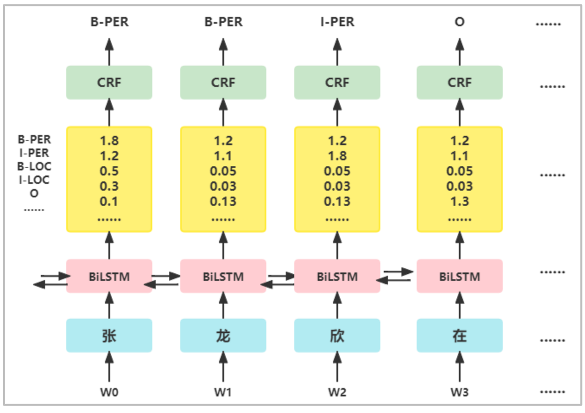


Figure 1: BERT + BiLSTM + CRF Entity Extraction Model

4. Construction of Criminal and Victim Knowledge Graph Based on R-BERT

4.1 R-BERT Model Architecture

R-BERT, as the most efficient and practical method in the scenario of relationship extraction, not only depends on the information of the two target entities, but also relies on the semantic and grammatical information of the sentence itself to carry out the task of relationship extraction. It uses the powerful coding ability of BERT to discover the rules of relationships between entities under the premise of extracting the above two kinds of features simultaneously. The advantages of R-BERT are significant in the task of extracting interactive relationships in home invasion crimes. The overall architecture of the R-BERT relationship extraction model is shown in Figure 2:

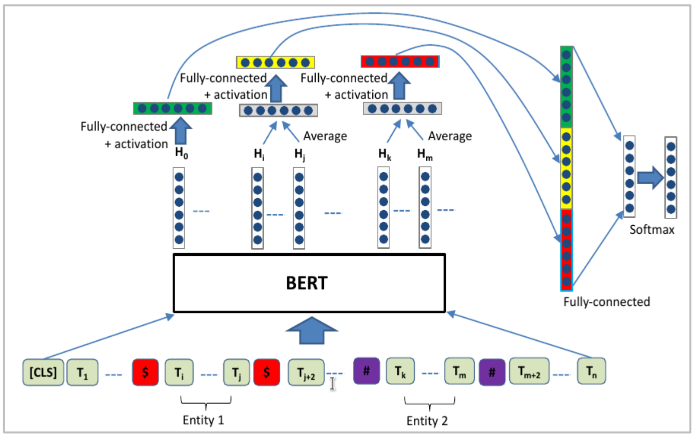


Figure 2: R-BERT Relationship Extraction Model Architecture

(1) In order to enable the BERT model to locate the positions of two home invasion entities, it is first necessary to add "[CLS]" at the beginning of each sentence, add a special character "$" before and after the first entity, and add a special character "#" before and after the second entity. The example is as follows:

Example: [CLS] Peng XX left an $iPhone 13$ in employee dormitory 402 of building #1# in the B area of Fuxing Jiayuan, Chengguan Street, Fangshan District, and found it stolen when he went to the toilet at about 3 o'clock.

(2) Then, this marked sequence is sent to the BERT pre-training model to extract the vector representation of the theft crime case text. The output of the BERT model part is divided into three parts: the hidden state vector of [CLS] and the hidden state vectors of the two entities. The first part retains the semantic content of the sentence, and the latter two parts retain the information of the two entities.

(3) Finally, use concat to join them into a softmax for classification. The classification result of the relationship between the entities is obtained.

In this study, the parameters of our R-BERT model are as shown in Table 3:

Table 3: Parameters of the R-Bert Model for Extracting Relationships of Entities in Home Invasion

Model Parameters	Value
Batch Size	32
Max Sentence Length	128
Adam Learning Rate	2e-5
Epochs	10
Dropout	0.1

4.2 Construction of Knowledge Graph for Perpetrators and Victims in Home Invasion Crimes

Through information extraction from unstructured text, as well as association rule mining algorithms for structured data, in this chapter, we construct a knowledge graph of home invasion that covers both individual risk features and interactive risk features of criminals and victims. The construction of the graph mainly uses triples to organize entities and relationships. The triples mainly include two description forms, which are "entity—relationship—entity" and "entity—attribute—attribute value"(Andresen and Malleson, 2013). The final constructed knowledge graph of risk for perpetrators and victims in home invasion crimes is shown in Figure 3. Based on this graph, further research and judgment can be carried out, integrating the experience of practical experts, and ultimately forming knowledge rules for the analysis of persons involved in home invasion crimes, providing a reference of domain knowledge for scientific control of criminals and precise warning of victims.

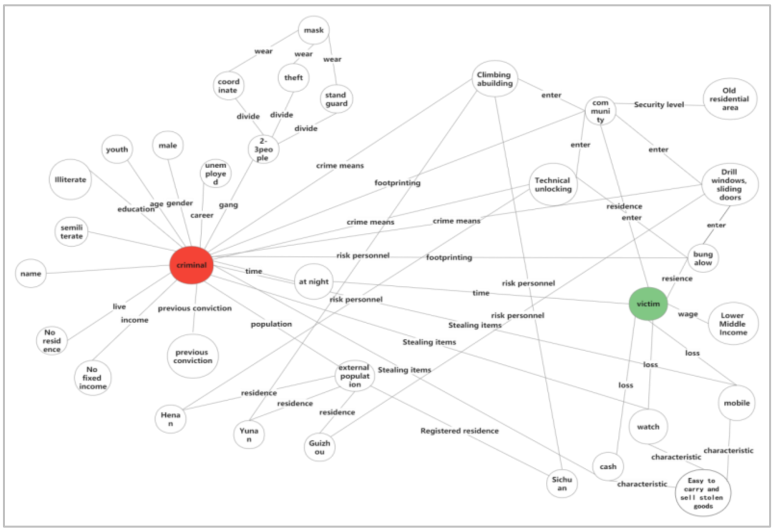


Figure 3: Knowledge Graph of Risk Perception for Perpetrators and Victims in Home Invasion Crimes

4.3 Analysis of Interactive Risk Patterns between Criminals and Victims

By integrating the above knowledge graph, expert summaries, and practical experience, we can infer the risk features related to the victims and criminals in the generation of home invasion crime risk. Through analysis and judgment, this knowledge graph can provide the following early warning knowledge about crime risk generation:

(1) From the perspective of criminal means, home invasion thefts mainly use methods such as nighttime climbing and window penetration (hand climbing residential buildings), night-time "sliding" lock picking (cutting plastic bottles into "L" shape slides to open the door through the door seam), daytime foil lock picking (inserting foil into a key mold to open the lock), and night-time sneak-in (choosing households that do not lock their doors). The residences of the victims are mainly bungalows and old residential areas with overall poor defense levels, such as lack of surveillance, absence of burglar-proof windows, or low-level human defense. Overall, these types of residential areas have poor human defense, physical defense, and technical defense levels

(2) From the perspective of crime timing, most home invasions occur at night or during the day when no one is at home, meaning that during these periods, households are either uncontrolled by the victims or under very weak control. Further observation reveals that general cases of climbing burglary tend to occur more in the summer. This is because the hot weather in summer leads victims to open windows during sleep at night, thus providing great convenience for criminals to enter and commit theft. That is, the victims unknowingly facilitate the criminals in committing the crime, increasing the risk of crime. This is a representative category of utilitarian interactive risk behavior in the interactive relationship.

(3) Looking at the criminals, those who commit home invasion theft are usually part of a gang, including 2-3 young men, with clearly defined roles. Some are specifically responsible for standing guard, some are engaged in home invasion theft, and some are designated for assistance. The high-risk group among these criminals has been involved in home invasion theft for many years, and most of them have a criminal record of more than five times, constantly "contending" with the investigative department and increasingly becoming aware of anti-investigation measures. For instance, crime suspects always wear gloves, hats, and masks when committing crimes to prevent leaving fingerprints and DNA, increasing the difficulty of on-site investigation. Crime suspects frequently change their mobile phone numbers, or do not carry a mobile phone during the crime, or turn off the phone when they get to city B, adding to the difficulty of the investigation.

(4) Looking at the victims, they are generally from middle to low income groups. Their living conditions are poorly defended, providing opportunities for criminals. Furthermore, further knowledge sorting reveals that victims with higher losses from home invasions generally like to store cash, jewelry, mobile phones, and other valuables at home.

(5) Looking at the lost items, most victims lose items such as mobile phones, jewelry, expensive watches, cash, tobacco, alcohol, and electronic products. These stolen items are portable and easy to fence, reducing the risk of criminals being discovered, and also making the case more difficult to solve.

(6) Looking at the characteristics of interactive risk behaviors, the interaction between the two parties in a home invasion theft fits the typical category of exploitative interactions. For instance, criminals often climb into homes for theft mostly because the victims open the windows without being aware of the risk, providing convenient conditions for the crime. In addition, victims' behaviors of storing valuable jewelry, cash, and watches at home, are implicitly interactive risk behaviors with criminals. Further analysis reveals that many victims living on high floors do not install burglar-proof windows. They hold a fluke mentality, thinking that there is no risk of theft on high floors, hence there are subjective careless risk interactions. These daily risk behaviors, unnoticed by the victims, ironically become exploitable relationships for the criminals. Therefore, crime risks ultimately transform into criminal facts.

5. Conclusion

In fact, knowledge induction and rule extraction based on the above graph go far beyond the aforementioned six aspects. Through in-depth analysis and judgment, one can obtain more knowledge about the interaction risk behaviors between criminals and victims, and extract more mechanisms of interaction that generate burglary risk and richer characteristics of individual risk. Understanding the interactive risk patterns between criminals and potential victims has strong guiding significance for practical work in public security. We can fully explore this risk interaction relationship and continuously optimize and improve prevention strategies in order to reduce crime.

Acknowledgement

The work was supported in part by a grant from 21CSH005 project.

References

Andresen, M.A., Malleson, N., 2013. Crime seasonality and its variations across space. Applied Geography, 43, Pp. 25-35.

Bayley, D., and Garofalo, J., 1989. The Management of Violence by Police Patrol Officers. Criminology, 27, Pp. 1-25.

Li, Q.Y., 2015. On the Practical Application of the Interactive Relationship between Crime and Victimization. *Law and Society*, (32), Pp. 278-279.

Piroozfar, P., 2019. Crime Prevention in Urban Spaces through Environment Design: A Critical UK Perspective. Cities J, 95, Pp. 1–11.

Schneider. 1992. Victims in an International Context (Z.R., Xu, et al, Trans.). Beijing: China People's Public Security University Press.

Sherman, L.W., 1992. Attacking Crime: Policing and Crime Control. In: Tonry, M. & Morris, N. (eds.), Modern Policing. Chicago, IL: University of Chicago Press.

Van Sleeuwen, S.E.M., Ruiter, S., Steenbeek, W., 2021. Right place, right time? Making crime pattern theory time-specific. Crime Science, 10 (1), Pp. 1-10.

Zhou, J.J., Hu, C.K., Chi, J.L., 2022. Behavior-Aware Account De-Anonymization on Ethereum Interaction Graph. [2022-07-08].