[1,*]Qi Zhang
[2]Xuechen Li
[3]Tuo Shi
[4]Zijun Liu

Research on Risk Assessment Model for Brushing Type Telecom Network Fraud Victim Based on Bayesian Network

Journal of Electrical Systems

Abstract: - Assessing the risk level and exploring the risk characteristics of victims of brushing type telecom network fraud is of practical significance for crime risk prevention and control. Starting from the demographic characteristics of victims and case characteristics, this paper establishes a Bayesian network model with a tuple composed of loss amount and contact duration as the evaluation index of victim risk level, aiming to provide ideas for police to implement precise anti-fraud propaganda. The research shows that the tuple victim risk assessment model has a high prediction accuracy and can take into account both the loss amount and contact duration, which is feasible as a victim risk assessment model; there is no significant single influencing factor characteristic that affects the victim risk level; among the key victim groups of women who are commerce service personnel and use social media platforms, police should focus on highly educated groups, and among people with the same educational background, police should focus on young people under 28 years old.

Keywords: Shunt Brushing type telecom network fraud; Bayesian network; Tuple; Risk assessment.

I. INTRODUCTION

With the rapid development of Internet, big data and artificial intelligence, the contemporary society has entered a digital society in which the physical space and the cyberspace are interwoven and highly integrated. Under the background of the current era, China's crime situation is gradually changing from "city-attracting crime" to "network-attracting crime" [1]. The network has become the tool, the space and the scene of the crime, with numerous illegal activities such as the telecommunication network fraud, the network gambling, the network violence emerging, and the criminal methods constantly renewing. The Law of the People's Republic of China on Anti-Telecommunication Network Fraud, which was implemented on December 1, 2022, requires that front-end prevention and comprehensive administration shall be carried out for the crime of telecommunication network fraud and network black and gray industry, that precise prevention and control shall be carried out, publicity and education on anti-telecommunication network fraud shall be carried out in a targeted manner to the society, and early warning and dissuasion shall be conducted to potential victims. Among the five types of frequently occurred telecommunication network fraud cases published by the Ministry of Public Security in May 2022, the incidence rate of fraud in the form of click farming and rebate is the highest, accounting for about 30% of the total number of cases. In the implementation process of telecommunication network fraud crimes such click farming, criminals take the click farming and rebate as a verbal trap, fully grasp the victim's desire for benefits and voluntary participation, and utilize the drainage information such as "part-time work", "easy earning", "free delivery" and other drainage information for script-type precise fraud. The drainage platforms include various platforms such as social media platform, short video software, search engine, job-seeking software and the like, and the criminal methods are highly concealed and confusing, which brings great challenges to the precision prevention and control of public security organs. Based on the characteristics of the victim and the process of the crime related to click farming, this paper constructs the victimization risk assessment model, which aims at evaluating the risk of victimization for different populations of people with different characteristics, finding out the key influence characteristics of the people with higher risk level and providing ideas for the public security organs to carry out the accurate anti-fraud propaganda.

In recent years, the academic research on the crime of telecommunication network fraud has focused on the research on the cooperative governance concept [2], countermeasures technical measures [3], and execution linkage mechanism [4]proposed by the Anti-Telecommunication Network Fraud Law; on the other hand, research is focused on the ideas and methods of detecting the telecommunication network fraud [5-7]. However, there are relatively few studies on victims of telecom network fraud. Yin Ming selected 363 records of victims, analyzed

[1] Beiijng Police College, Beijing, China
[2] Beiijng Police College, Beijing, China
[3] Beiijng Police College, Beijing, China
[4] Beiijng Police College, Beijing, China
*Corresponding author: Qi Zhang

and studied individual characteristics [8]; Zhang Zhi et al. explored patterns by selecting factors such as gender, age, occupation based on the characteristics of the victims [9]; Luo Wenhua et al. analyzed the characteristics of victims by using Bayesian network [10], but the sample data had fewer features; Zhang Jiece et al. extracted the characteristics of interaction process between victims and criminals based on the reported materials of click farming fraud victims, and used differential analysis methods to test the relationship between the demographic characteristics and interaction process characteristics of victims, as well as the reaction speed and amount of property damage after being reported. The aim was to provide countermeasures and suggestions for clarifying the filing standards, rectifying crime channels, and carrying out precise publicity [11]; Gu Haiyan conducted an in-depth analysis of the characteristics of online click farming fraud and suggested that we should strengthen network supervision, improve the comprehensive management system, strengthen the public opinion publicity [12]. At present, there are few research achievements on the characteristics of the victims of telecom network fraud, and because of the limitation of sample data quality, the selected case characteristics are limited, and the practical application value of the analysis results is not high. There are even fewer empirical studies on victims of telecommunications network fraud related to click farming. This paper focuses on the telecommunication network fraud crime related to click farming, relies on the follow-up data of the first-line unit on such cases, and the data characteristics are relatively comprehensive, selects a specific index to evaluate the victim risk level of the telecommunication network fraud crime related to click farming, utilizes the Bayesian network to evaluate the victim risk level of the telecommunication network fraud crime related to click farming, aims to establish a victimization risk assessment model of telecommunication network fraud crime related to click farming based on victim characteristics, and deeply digs the victim risk level and characteristic rules of different characteristic populations, provides ideas for the public security organs to formulate accurate prevention and anti-fraud propaganda strategies, so as to realize more effective prevention and treatment of such crimes.

II. DATA SOURCE AND FEATURE SELECTION OF VICTIMIZATION RISK PERCEPTION OF TELECOMMUNICATION NETWORK FRAUD CRIME RELATED TO CLICK FARMING

The research object of this paper is the follow-up data of the public security organs of a certain large-scale city in in northern China aiming at the telecommunication network fraud case related to click farming, with a total of 443 pieces of data and high accuracy. According to the characteristics of the follow-up data, the characteristics describing the victim risk assessment can be divided into the characteristics of the victim and the characteristics of the case, wherein the characteristics of the victim include gender, age, education level, occupation category and jurisdiction, and the characteristics of the case include the drainage platform (the platform that first contacts with fraud information), the contact duration (the time from the first contact with the fraud information to the start of "hooking" the first task, which can reflect the risk prevention consciousness of the victim to a certain extent) and the loss amount. 340 pieces of complete data are obtained by removing null items and filtering the 443 pieces of telecommunication network fraud related to click farming. The data were divided into training set and test set according to the ratio of 306: 34(10: 1) to verify the rationality of the constructed Bayesian network model.

A. Characteristics of the Victim

Regarding the victim's own characteristics, this paper selects gender, age, education level, occupation category and jurisdiction as characteristic items, and classifies each characteristic item respectively. First of all, the sex of the victims is analyzed by descriptive statistics. The majority of the victims are female, accounting for 69% of the total number of victims, and 31% are male. Regarding the age division of victims, considering that 28 years old is the average marriage age of the current population in China, and 45 years old is the dividing line between young and middle-aged people defined by the World Health Organization, this paper divides the age into adolescents (28 years old and below), young and middle-aged people (29-44 years old), and middle-aged and elderly people (45 years old and above). Through descriptive statistical analysis, the proportions of each age group are 51%, 43%, and 6%, respectively; In this paper, the educational level is divided into secondary and lower education levels and higher education levels, accounting for 45% and 55% respectively; For the occupational broad category, based on the of similar occupational categories and the characteristics of the follow-up data, the occupational categories are divided into unemployed, the fourth major category of occupations (business service industry personnel), Other categories (including school students, professional technical personnel, inconvenience classification personnel, etc.), The proportions of the three major occupational types are 31%, 46% and 23%,; According to the geographical characteristics of the large-scale city, the jurisdiction is divided into four zones

from inside out: urban center zone, ring center zone, suburban zone, and outer suburban zone, and the proportions of the four geographical feature groups are 14%, 45%, 34%, 7% respectively.

B. Case Characteristics

According to the characteristics of sample data, the drainage platform, contact duration and loss amount of the case are selected as the case characteristics, and each data item is classified respectively. According to the different types of drainage platforms, the public security organs divide the telecommunication network fraud crime of click farming into seven categories: traditional communication tools (telephone and short message), short video software, job-seeking software, social contact software, search engine advertisement, live broadcast platform, pornography, online games, etc. In view of the small number of the last four categories, this paper combines the last four categories into one category, and the combined categories are traditional communication tools, short video software, job-seeking software, social contact software and other types, accounting for 11%, 15%, 9%, 60% and 5% respectively. Social software is the main drainage platform for telecommunication network fraud related to click farming. For ease of analysis, the contact duration is divided into short (0-1 hours), medium (2-13 hours), long (more than 13 hours), with each group accounting for 34.9%, 33%, 32.2% respectively; The loss amount is classified as low (below 15,000 yuan), medium (RMB 15,000-46,000), high (above 46,000 yuan), with each group accounting for 33%, 33%, and 33.9%, respectively..

III. The Risk Model of Victims in Telecommunication Network Frauds Related to Click Farming Based on Bayesian Network

A. Overview of Bayesian Networks

Bayesian networks express the causal relationship and influence degree among various information elements in the complex uncertainty system network according to the graph, and realize the modeling reasoning of the complex system based on strict mathematical foundation and probability analysis reasoning ability [13]. In a Bayesian network, each node represents an event or a random variable. Assuming that there are N events (nodes) in the Bayesian network $\{E_1, E_2, \cdots E_N\}$, all the states of the event E_i constitute the state space of E_i. Assuming that the event E_i has M_i states, then the event E_i can be represented as $S_{E_i} = \{S_{E_i}^1, S_{E_i}^2, \cdots S_{E_i}^{M_i}\}$, then the joint state space of N events can be expressed as a Cartesian product: $S_{E_1 E_2 \cdots E_N} = S_{E_1} \otimes S_{E_1} \otimes \cdots \otimes S_{E_N}$. where $\otimes$ represents the Cartesian product. The state space of the event E_i corresponds to a Conditional Probability Table (CPT), where the directed edges between nodes represent the dependency between them, and the value of CPT represents the strength of the relationship. The expression of factors in complex uncertain system is the advantage of Bayesian network, which is helpful to analyze the characteristic law of victimization risk of fraud in telecommunication network. In this paper, the Bayesian network is used as the modeling tool, the victim characteristics and case characteristics of telecommunication fraud cases of click farming are taken as node events in the Bayesian network, and the interaction relationship among the characteristics is taken as the directed edge of the node events in the network. Netica is widely used Bayesian network software developed based on Java, with a simple and reliable design, high performance [14], and a high visualization level, which can visually observe the CPT value of network node events.

B. Construction of Victimization Risk Assessment Model

The node events in the Bayesian network constructed in this paper give full consideration to the intuition of data and the sociality of telecommunication networks fraud cases related to click farming[15], and obtain the quantitative index of each node event (see Table 1-9). The prior probability of each event node in the Bayesian network is obtained by learning the training data of telecommunication network fraud cases related to click farming.

Table 1: Quantitative indicators of gender

Gender	Node Name (Gender)	Value
Male	Male	1
Female	Female	2

Table 2: Quantitative index of age

Age	Node Name (Age)	Interval	Value
Adolescents	Teenager	0~28 years old	1
Young and middle-aged	Young Adult	29~44 years old	2
Middle-aged and elderly	Middle Old	Over 45 years old	3

Table 3: Quantitative indicators of education level

Degree of education	Node Name (Education)	Value
Secondary education and below	Low Middle	1
Higher Education	High	2

Table 4: Quantitative Index of Occupation Category

Occupation Category	Node Name (Career)	Value
Unemployed	Unemployed	1
Business Service Personnel	Business	2
Other occupations	Others	3

Table 5: Quantitative Index of Jurisdiction

Jurisdiction	Node Name (Region)	Value
Urban center	Centre	1
Ring center zone	Near	2
Suburban zone	Modest	3
Outer suburban zone	Far	4

Table 6: Quantitative Index of Drainage Platform

Drainage Platform	Node Name (Platform)	Value
Traditional Communication Tools	Tradition	1
Short Video	Short video	2
Job-seeking category	Job hunt	3
Social Contact Media	Socialize	4
Other	Others	5

Table 7: Quantitative Index of Loss Amount

Amount of loss	Node Name (Damage)	Interval	Value
High	High	Above 46000 yuan	1
Medium	Middle	15000`46000 yuan	2
Low	Low	Below 15000 yuan	3

Table 8: Quantitative Index of Contact Duration

Duration of contact	Node Name (Duration)	Interval	Value
Short	Low	0-1 hour	1
Medium	Middle	1-13 hours	2
Long	High	Over 13 hours	3

Table 9: Tuple Quantization Index of Loss Amount and Exposure Duration

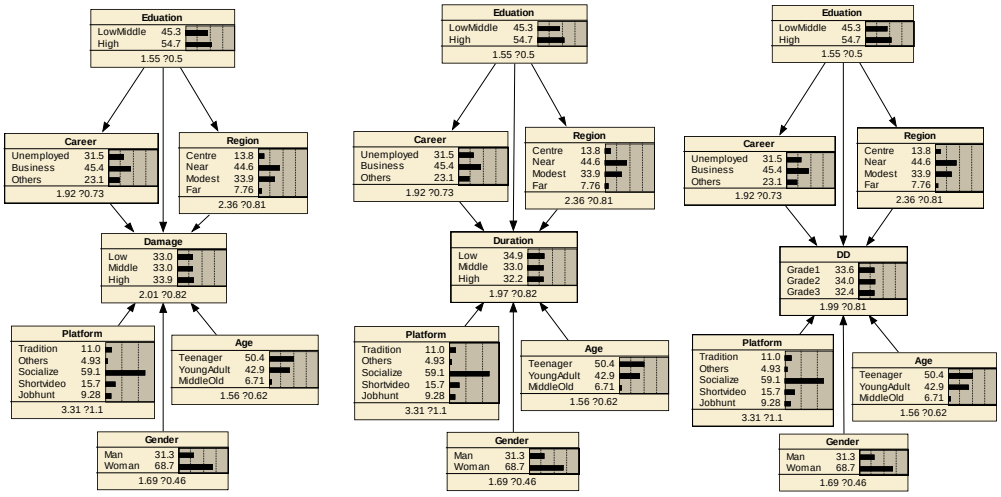
Risk level:(amount of loss, contact duration)	Node Name (DD)	Value
High:(high, short)(high, medium)(medium, short)	Grade1	1
Medium:(high, long)(low, short)(medium, medium)	Grade2	2
Low:(Medium, Long)(Low, Medium)(Low, Long)	Grade3	3

In order to further construct a reasonable Bayesian network, through conducting in-depth investigation on the front line, absorbing experts' opinions and consulting relevant academic documents, the characteristics of Bayesian network are obtained: the victim's own characteristics (including education level, gender, age, occupation and jurisdiction) directly affect the loss amount of telecommunication network fraud cases, the contact duration from the first contact with criminals to the first task ("hooking"), the occupation and the jurisdiction affected by the education level in the victim characteristics; The "drainage platform" of the case characteristics affects the amount of loss and the duration of contact. According to the above relationship, the loss amount,

contact duration and the binary group consisting of the two are respectively taken as the risk evaluation index of the victim: the larger the loss amount is, the higher the victim risk level is, the lower the loss amount is, and the lower the victim risk level is; the shorter the contact time is, the higher the victim risk level is, and the longer the contact time is, the lower the loss risk level is; the binary group is divided into three grades of high, medium and low on the basis of the classification of the loss amount and the contact duration (see Table 9), so as to comprehensively measure the risk level of the victim from the two indexes. In this paper, the “contact duration”, “loss amount” and binary groups are classified according to Tables 7, 8 and 9 respectively as the result nodes of the Bayesian network, and the characteristics of the victim (gender, age, education level, occupational category and jurisdiction) and drainage platform are taken as the influence nodes, and the Bayesian network model of the victim risk assessment is constructed respectively. After the Bayesian network model is determined, the sample data is imported to enable each node to automatically acquire the CPT value.

IV. Analysis of Victimization Risk Assessment Model

In order to find out the impact of different influence nodes on the risk level of the victim as much as possible and eliminate the interference of irrelevant factors, this paper adopts the research method of controlling other variables to observe the influence of target variables. The processed 306 pieces of data are imported into the Bayesian network model for training, and three victimization risk assessment models with "loss amount", "contact duration" and binary group as the result nodes are obtained (see Figure 1). The method comprises the following steps of: firstly, evaluating the prediction capability of three risk assessment models by using test data; secondly, carrying out sensitivity analysis on the three models, finding a characteristic node which has the most influence on the victimization risk and analyzing the influence of the characteristic node; Then, analyze the victimization risk levels of specific populations with different combination characteristics, respectively use 'loss amount', 'contact duration’ and binary group as indexes of the victimization risk levels, evaluate and sequence the risk levels of the specific population; Finally, carry out portrait on victims with different levels of risk levels based on risk sequencing analysis, analyze the key characteristics that affect the level of risk in a specific population.



(a) “amount of loss” model (b) “duration of contact” model (c) binary group model

Figure 1 Victimization risk assessment model with “loss amount” (a), “contact duration” (b) and binary group (c) as result nodes respectively

A. Evaluation of Prediction Ability of the Victimization Risk Assessment Model

Substitute the remaining 34 test data in the sample data into a loss amount victimization risk assessment model, contact duration victimization risk assessment model and binary group victimization risk assessment model one by one, compare the loss amount grade truth value, the contact duration grade truth value and the binary group risk grade truth value of the test data with the predicted values of the three models, and calculate the accuracy of the model predictions. The accuracy of the three models' predictions is 85.29%, 79.41%, and 82.35%, respectively. The accuracy of the binary victimization risk assessment model is lower than the loss amount victimization risk assessment model, but higher than the contact duration victimization risk assessment model. The model takes into account both the amount of loss and the duration of contact, and can be used as a comprehensive model of the victim risk level.

B. Sensitivity Analysis of Victimization Risk Assessment Model

Sensitivity analysis can determine the parent nodes in a Bayesian network that have a significant impact on the child nodes. In this paper, the importance of parent node to child node is shown as the importance of pair factor characteristics to the occurrence of victim risk. According to Shannon's information theory, mutual information (MI) is an index to measure the importance of a parent node to a child node in a network [16]. Netica regards MI as an important indicator for sensitivity analysis [17], in addition, measures such as percentage of variance and variance of Beliefs [18] are used for sensitivity analysis. In the binary group victimization risk assessment model, for sensitivity analysis with the binary group “DD” as the query node (see Table 10), when the prior probability of occurrence of each feature is given and the posterior probability of each feature is combined, the specific characteristic probability and the MI value of the binary group risk evaluation index can be calculated, and the importance of each influencing feature can be evaluated accordingly. According to Table 11, the most influential node of the binary group “DD” is occupation (Career, MI=0.00227), followed by age (Age, MI=0.00085), and the quotient subtraction percentage and variance of the two are also the largest, indicating that the most influential node on the binary group “DD” is occupation, followed by age. Similarly, the loss amount "Damage" and the contact duration "Duration" in the loss amount risk assessment model and the contact duration risk assessment model are used as query nodes for sensitivity analysis. The top two nodes of influence on the loss amount are respectively age (Age, MI=0.00254) and occupation (Career, MI=0.00172), and the top two nodes of influence on the contact duration are respectively age (Age, MI=0.00122) and occupation (Career, MI=0.00099). This indicates that the two most influential nodes are age and occupation, and age and occupation are the sensitive factors affecting the three risk evaluation indexes when the loss amount, the duration time and the binary group are taken as the risk evaluation indexes respectively.

According to Table 10, in the binary risk assessment model, the influence of occupation is much higher than that of other nodes. The posterior probability of the binary group is analyzed by the prior probability of changing the occupation. The posterior probability of each grade of the binary group changes within 5%. Similarly, observing the posterior probability of changing professions reveals that the change in the posterior probability of both levels is also within 5%, and the change in the posterior probability of the two is not significant. Therefore, compared with other factors, although the occupational factor has the greatest influence on the binary, the influence of occupational probability on the binary probability is not significant. Therefore, in the binary group risk assessment model, the individual influence of the six feature nodes on the binary group is small. Similarly, in the victimization risk assessment model of the loss amount and the duration time risk assessment model, the individual influence of the six characteristic nodes on the result node is small.

Table 10: Sensitivity analysis of events using "DD" as the query node in a binary group risk assessment model

Node	Mutual Info	Percent	Variance of Beliefs
DD	1.58466	100	0.4443052
Career	0.00227	0.143	0.0003502
Gender	0.00075	0.0472	0.0001129
Education	0.00067	0.0421	0.0001004
Region	0.00051	0.0322	0.0000768
Age	0.00085	0.0533	0.0001312
Platform	0.00082	0.0515	0.0001284

C. Risk Analysis of Specific Population Victimization

Through the evaluation of the prediction ability of the three models, the difference of the prediction ability of the three models is small, and all of them can be used as the models for evaluating the risk level of the specific victims, and the binary group risk assessment model is more comprehensive. Through sensitivity analysis, the sensitivity characteristics of the three models are similar, and the influence of the six characteristic nodes on the three result nodes is small. Next, the posterior probability of the risk level of the specific population is analyzed by adjusting the prior probability of different combination factor nodes, and the key influencing factors of the risk level of the specific population are analyzed. In order to facilitate the discussion, the characteristics of the prior probability of each influencing factor shall be fully considered and the discussion shall be focused on: for any one factor node, when the difference between the prior probabilities of each state is obvious, only the state with obviously high prior probability shall be analyzed, and when there is more than one state with higher prior probability, the states shall be discussed in different cases. First, analyze the prior probability characteristics of each factor node (see Figure 1): in terms of gender, the ratio of male to female is 1: 2, and there is a significant difference in the sample size between the two, so we fix the status that the victim is female; From the perspective

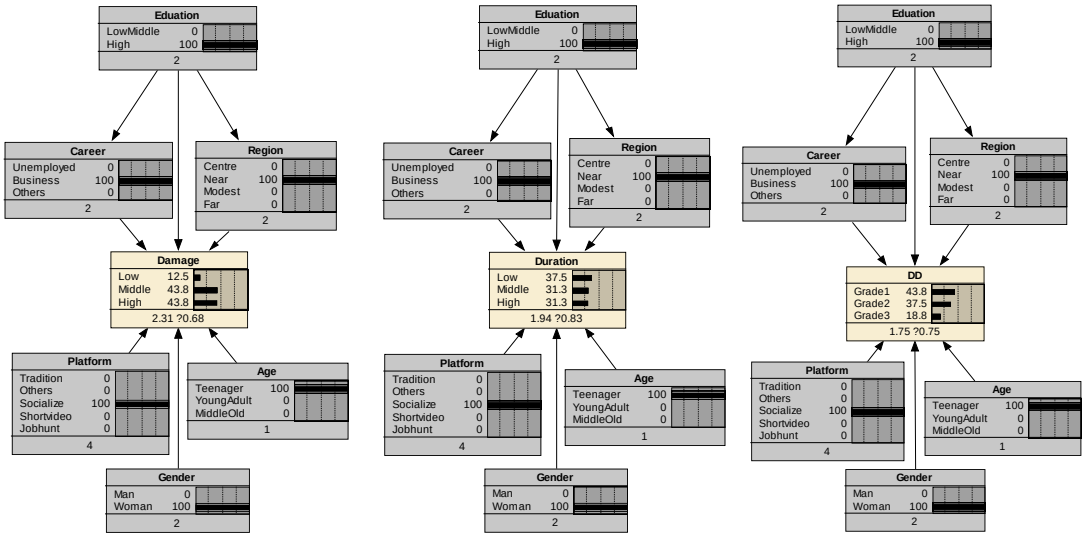
of age, more than 90% belong to the group under 45 years old, while the adolescents under 45 years old and the middle-aged and young people are relatively close to 1: 1, so the two statuses need to be discussed separately; From the perspective of education level, the prior probability between the education level of medium and below and the higher education level is close to 1: 1, so the two statuses need to be discussed separately; From the perspective of occupational characteristics, the prior probability of business service personnel is absolutely dominant, so the fixed occupation is the status of business service personnel; From the perspective of jurisdiction , the prior probability of the central area (Near) and the suburban area (Modest) is relatively high, and the prior probability of the two is close to 1: 1. Therefore, the two states of the urban center zone and the suburban zone are discussed separately. From the perspective of drainage platform, the prior probability of the social media platform is overwhelming, so the fixed drainage platform is in the state of the social media platform. Three Bayesian network models are used to analyze and evaluate the victimization risk level of the eight categories of special populations by arranging and combining the eight categories of key populations formed by various states (see Table 11).

Table 11: Eight specific populations formed by the combination of factor node states

Serial No.	Gender	Occupation Category	Drainage Platform	Age	Education Level	Zone
1	Female	Business Service	Social Media	Adolescents	Higher Education	urban center
2	Female	Business Service	Social Media	Adolescents	Higher Education	Suburban
3	Female	Business Service	Social Media	Adolescents	Secondary and below	urban center
4	Female	Business Service	Social Media	Adolescents	Secondary and below	Suburban
5	Female	Business Service	Social Media	Young and middle-aged	Higher Education	urban center
6	Female	Business Service	Social Media	Young and middle-aged	Higher Education	Suburban
7	Female	Business Service	Social Media	Young and middle-aged	Secondary and below	urban center
8	Female	Business Service	Social Media	Young and middle-aged	Secondary and below	Suburban

1) Risk Analysis of Victimization for Population 1

The victimization risk of the victim population 1(see Figure 2) is analyzed by using the victimization risk assessment model for the amount of loss, the victimization risk assessment model for the duration of contact and the victimization assessment model for the binary group. The posterior probability of the group losing more than 46,000 yuan reaches 43.8%, the posterior probability of the loss more than 15,000 yuan is 87.6%, while the posterior probability of the contact duration less than 1 hour is 37.5%, and the posterior probability of less than 13 hours is 68.8%. The group is likely to face a large amount of loss and quickly “hooking”. Using the binary risk assessment model to analyze the risk of the population, the posterior probability of the high risk level is 43.8%, belonging to the high risk population. It can be seen that the binary risk assessment model can comprehensively consider the loss amount and the contact duration, and can comprehensively evaluate the risk level of the victim.

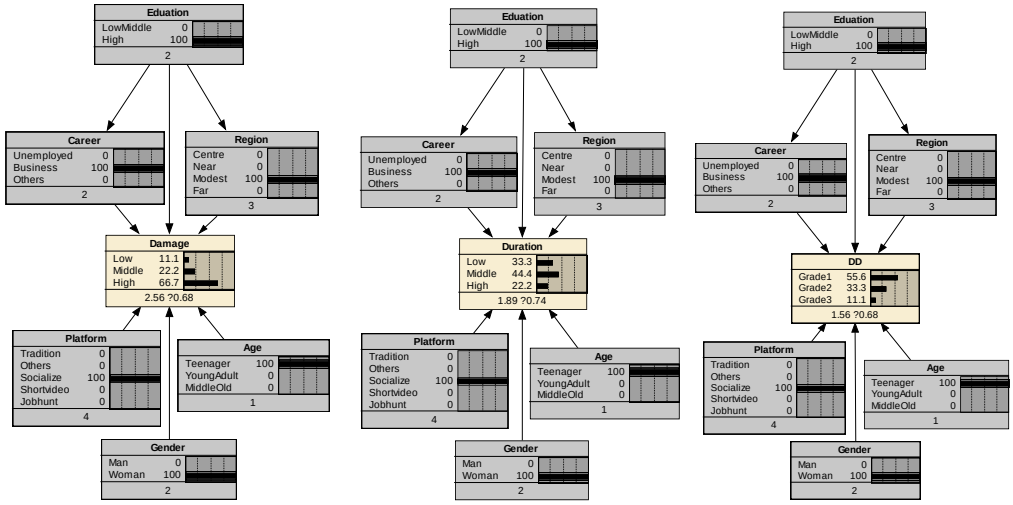


(a) "amount of loss" model (b) "duration of contact" model (c) binary group model

Figure 2 Analysis and assessment of victimization risk of population 1 based on three models

2) Risk Analysis of Victimization for Population 2

Similarly, the victimization risk of the victim population 2 (see Figure 3) is analyzed by using the risk assessment model for the amount of loss, the risk assessment model for the duration of contact and the risk assessment model for the loss of binary groups. The posterior probability of the amount of loss greater than 46,000 yuan is as high as 66.7%, and the posterior probability of the loss amount greater than 15,000 yuan is 88.9%, while the posterior probability of the contact duration less than 1 hour is 33.3%, and the posterior probability of less than 13 hours is 77.7%. The group is more likely to face huge losses, but the probability of quick "hooking" is stable compared with population 1. Using the binary risk assessment model to analyze the victimization risk of this population, the posterior probability of a high risk level is 55.6%. This is because this population faces a high risk of huge losses, which increases its risk level.



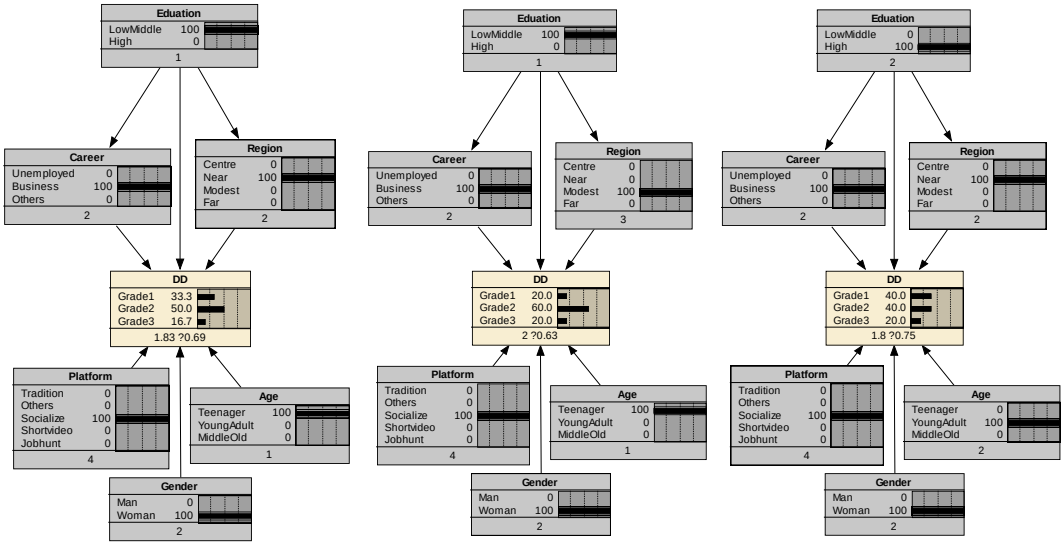
(a) "amount of loss" model (b) "duration of contact" model (c) binary group model

Figure 3 Risk Analysis and assessment of victimization in population 2

3) Risk Analysis of Victimization for population 3 to population 8

The victimization risk characteristics of population 3 to population 8 were analyzed in sequence and the victimization risk of the victims were analyzed by the loss amount, the contact duration and the binary group victimization risk assessment model respectively. The results of the three models were different, mainly manifested in the differences in the assessment results of the eight specific population's victimization risks based on the loss amount victimization risk assessment model and the contact duration victimization risk assessment model, while the binary victimization risk assessment model can balance the risk levels of loss amount and contract duration, and can reconcile the differences between the two. For example, the "high" risk level of the

binary is a combination of loss amount and contact duration of (high, high), (high, medium) and (medium, high), and a small posterior probability above the intermediate level of either of the two indicators reduces the posterior probability of the high risk level of the binary, while a higher posterior probability of any one of the two indicators is "high" increases the posterior probability of the high risk level of the binary. Therefore, it is more comprehensive to assess the risk of victimization of a specific population using the binary risk assessment model. Figure 4 shows that the risk of victimization of population 3 to population 8 is analyzed using the binary risk assessment model, and the posterior probability of each risk level of each population is obtained. Combined with the results shown in Figure 3 and Figure 4, it is found that the victimization risk level of eight specific populations can be determined with definite risk level.



(a) Population 3 (b) Population 4 (c) Population 5

(d) Population 6 (e) Population 7 (f) Population 8

Figure 4 Risk Analysis and assessment of victimization from population 3 to population 8

4) Risk assessment and analysis of eight special populations

The posterior probability of each risk level in the binary risk assessment model is used to rank the risk of eight populations. According to the higher the posterior probability of the high risk level, the higher the victimization risk of the group, the order of the victimization risk of the eight populations is obtained: population 2>population 1>population 6>population 5>population 7>population 3>population 4>population 8. Key contributing factors were further identified based on the results of the risk ranking. It can be seen that, among the three characteristics of age, education experience and jurisdiction, population 2, 1, 6 and 5 belong to those with higher education level, while those of population 7, 3, 4 and 8 belong to those with medium or lower education level, i. e. those with higher education level have higher victimization risk level, while those with medium or lower education level have lower victimization risk level; in population 2, 1, 6, 5 and 7, the victimization risk of youth group is generally higher than that of young and middle-aged victims (population 2, 1 belongs to adolescents, and population 6, 5,

7 belongs to young and middle-aged); in population 3, 4, 8, The risk of victimization among young people is also higher than that among young and middle-aged groups (population 3, 4 belongs to teenagers, the population 8 belongs to the young and middle-aged). In other word, the victimization risk of youth group is higher than that of young and middle-aged group. However, there is no obvious difference in the level of jurisdiction victimization risk. The above results can be interpreted as that the victimization risk of the group with higher education level is higher than that of the group with medium education level or below among the female victims who are professional for business service and use of social media platform. The victimization risk of the adolescent group is higher than that of the young and middle-aged group in the group with the same education level. Further verification was conducted through sensitivity analysis. The sensitivity analysis of the binary indicators was conducted on binary indicators with fixed gender characteristics as female, occupational characteristics as business service personnel and drainage platform characteristics as social media shows that the influence of the remaining three indicators on the binary images is ranked from high to low as education level, age and jurisdiction (with mutual information values of 0.01891, 0.01711 and 0.01354 respectively), which is consistent with the above analysis results for the characteristics of the eight types of victims.

V. Conclusions

In this paper, three Bayesian network models of victimization risk assessment for telecommunication network fraud crime related to click farming are constructed through expert opinions and literature data. The sample data are divided into training data and test data to train and test the models, and the results show that the prediction ability of the three models is not significantly different. Further, the sensitivity analysis of the three models shows that the sensitivity of the three models is similar, and the influence of the six characteristic nodes on the result nodes is relatively small. Considering the consistency between the two groups of loss amount and contact duration as the victimization risk evaluation index in sensitivity analysis, and the victimization risk assessment of specific populations can give consideration to both the loss amount and the contact duration, it is considered to focus on using the binary group victimization risk assessment model to evaluate the victimization risk level of the specific populations, and analyze the characteristics of the victimized populations of different levels, so as to provide ideas for the public security organs to realize the accurate anti-fraud propaganda.

(1) The binary group victimization risk assessment model can be used as the victimization risk assessment model for telecommunication network fraud related to click farming. The three models are constructed by expert opinions and literature materials, which are reasonable to some extent. In terms of test effect, the prediction effect accuracy of the three models is relatively high. From the sensitivity analysis, the sensitivity of the three models is similar. The binary group victimization risk assessment model can give consideration to the loss amount and the contact duration, and can evaluate the risk level of the victim more comprehensively, so it can be used as the victimization risk assessment model of telecommunication network fraud crime related to click farming. Based on the binary group victimization risk assessment model, public security organs can evaluate the victimization risk level of specific populations, and provide reference for formulating anti-fraud propaganda strategies.

(2) The individual influence of the victim's own characteristics and the drainage platform on the risk level of the victim is relatively small. According to the sensitivity analysis of the three models, the influence of the victim's own characteristics and the drainage platform on the loss amount, the contact duration and the binary group formed by the two is relatively small, which indicates that the telecommunication network fraud crime related to click farming does not have a single characteristic factor with great influence, which poses a great challenge for the public security organs to carry out accurate prevention and anti-fraud propaganda. Therefore, it is more feasible to evaluate and rank the risk level of the specific population by considering the specific population with a high proportion of influencing factor characteristics, and focus on the prevention on the basis of the risk ranking.

(3) Among the female victims who use social media platform and are engaged in business service, those with higher education levels have a higher risk of victimization than those with moderate or lower education levels. Among the population with the same education level, the risk of victimization is generally higher among adolescents under the age of 28 than among middle-aged and young people between the ages of 28 and 45. However, the area where the victim lives has no significant influence on the risk of victimization. When public security organs carry out prevention, publicity, and governance of high-risk female victims who work as business service personnel and use social media platform, they should pay more attention to the youth groups among the same educational population, and strengthen the key prevention and anti-fraud propaganda of the groups with these characteristics.

ACKNOWLEDGMENT

Organized Research Project Sub Topics of Beijing Police College [2023KYZZ01-2].

REFERENCES

[1] Shan yong. The Transformation of Digital Society to Front-End Preventive Crime Governance-A Case Study of A Draft Law on Anti-Telecom and Online Fraud. Journal of Shanghai Normal University (Philosophy & Social Sciences Edition), 2022, 51(03): 58-66.

[2] Sun Xiulan. Multiple Co-governance: Governance Research on the Black and Grey Production Involving Telecom Network Fraud. Public Governance Research, 2023, 35(01): 84-89.

[3] Lu Tianliang, Tu Junao, Du Yanhui, etc. Design of telecommunication network fraud case analysis based on big data technology. Experimental Technology and Management, 2020, 37(10): 50-55.

[4] Wang Yi, Di Xiaohua. Study on Connotation and Realization of Law-governing of "Counter Technical Measures": Citing the "Anti Online Fraud Law" as the Typical Example. Nanjing Journal of Social Sciences, 2022(10): 73-83.

[5] Wang Xiaowei. Exploration of Information Flow Investigation Methods for Telecom Network Fraud Crimes. People's Forum · Academic Frontier, 2022(15): 93-95.

[6] Wang Xiaowei, Zhao Zhao. Research on the Composition and Investigation Methods about the Personnel Flowing of Fraud Crime in Telecom Network. Journal of People's Public Security University of China (Social Sciences Edition), 2022, 38(04): 53-64.

[7] Zhi Jiayi. Research on Evidence Issues in Cases of Telecom Network Fraud. Journal of Law Application, 2022(09): 168-176.

[8] Yin Ming. An Empirical Study on the Victims of Telecommunication Fraud Cases - Quantitative statistical analysis based on victim records.Journal of Criminal Investigation Police University of China, 2017(03): 57-62.

[9] Zhang Zhi, Chen Feng. On the Problems and its Countermeasures in Telecom Network Fraud Prevention Publicity-Based on the Empirical Analysis of Victims' Characteristics. Journal of Guangxi Police College, 2021, 34(02): 41-49.

[10] Luo Wenhua and Zhang Yaowen. On Characteristics of Victims in Telecommunication Network Frauds: Comparative Analysis Based on Data from Different Regions. Netinfo Security, 2021, 21(12): 25-30.

[11] Zhang Jiece, Liu Yunxiao, Qi Chenhang. On the characteristics of Profitably Swiping Cards Fraud and Preventative and Controlling Countermeasures -- Based on an Analysis of 211 Cases in J District. Journal of Jiangxi Police Institute, 2022(04): 31-38.

[12] Gu Haiyan. Analysis and Countermeasures of Click Farming Fraud Behaviors. Journal of People's Public Security University of China (Social Sciences Edition), 2020, 36(01): 22-28.

[13] Chen Xuelong, Jiang Kun. Modeling Method of Concurrent Emergency Chain Based on Bayesian Network. Chinese Journal of Management Science, 2021, 29(10): 165-177.

[14] Chen Jing, Jiang Zhengkai, Fu Jingqi. Construction of Self-Learning Bayesian Network based on Netica. Journal of Electronic Measurement and Instrument, 2016, 30(11): 1687-1693.

[15] Luo Wenhua and Ma Xiaohan. An Empirical Study of Multi-stage Network Public Opinion Early Warning Based on Bayesian Network. Information Science, 2021, 39(07): 68-74.

[16] Liam Paninski. Estimation of entropy and mutual information. Neural Computation 15, Massachusetts Institute of Technology, 2003,15(6):1191-1253.

[17] Wang Peng, Xu Jianliang. Research on Risk Assessment of Information System Based on Bayesian Network. Journal of Ocean University of China (Natural Science Edition), 2022, 52(05): 131-138.

[18] Peng H, Long F, Ding C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Transactions on Pattern Analysis and Ma-chine Intelligence, 2005,27(8):1226-1238.