Computer Science and Information Systems 22(4):1599–1615 https://doi.org/10.2298/CSIS250601063S

HRSP: A High-Risk Social Personnel Risk Assessment Model Based on Graph Attention Label Propagation Algorithm

Xin Su[1], Heng Zhang[2], Xuchong Liu[1], Chunming Bai[3], Wei Liang[4], and Ning Jiang[5]

[1] Hunan Police Academy
[2] Zhuzhou Public Security Bureau Economic Development Zone Branch
[3] Xiangtan University
[4] Hunan University of Science and Technology & School of Computer Science and Engineering
[5] Jingshan People's Hospital, no.448 xinshi road Jingshan city 431899, Hubei, China
hbjsxrmyy@163.com

Abstract. The current society is complex and changeable, and the post-pandemic era profoundly affects people's work and life. Identifying the potential risks of high-risk individuals in society and carrying out early warning and control work effectively is the focus of current public security work and is also the key to maintaining social stability and people's peace. This work first analyzes and constructs a knowledge graph of high-risk individuals based on their backgrounds, trajectories, and related information. Subsequently, we propose a high-risk personnel risk assessment model based on a graph attention-label propagation algorithm. The model employs a multi-label feature selection method, a basic classifier based on a graph attention network for the label propagation algorithm, and an adversarial data augmentation algorithm to enhance the gradient-based adversary during training. In the experiment, we train the model using a public-security-field personnel dataset, and the accuracy of the proposed method reaches 90.2%, Ablation experiments demonstrate the effectiveness and stability of the proposed method. Constructing a knowledge graph specifically for high-risk individuals based on backgrounds, trajectories, and related data,Proposing a risk assessment model using a graph attention-label propagation algorithm, incorporating multi-label feature selection and adversarial data augmentation, which enhances training effectiveness.

Keywords: Knowledgegraph, Graph Attention Label Propagation, Data Augmentation.

1. Introduction

Vicious incidents caused by individuals with extreme or mental issues have been a severe threat to social security and stability in recent years. The risk assessment of High-Risk Social Personnel (HRSP) still depends on the subjective experience of public security police, making it difficult to determine the risk changes promptly and accurately determine the risk. The limitations make the polices lack real-time and precise risk management for high-risk personnel. The urgent issue that the public security organization needs to solve is how to effectively assess the risks of high-risk individuals and identify potential-risk individuals[1,11].

High-risk personnel risk assessment technology is an essential area of smart policing. Efficient and accurate risk assessment technology can provide more intelligent and efficient support for public security departments to carry out risk prevention and control work and contribute to the response speed and accuracy of policing. The risk assessment of high-risk personnel based on deep learning is of great significance at the theoretical level[20,4,7]. First, it provides a new direction for developing risk assessment techniques for high-risk individuals. Most of the traditional risk assessment methods rely on the subjective experience judgment of police in the actual situation and carry out qualitative analysis of the risk characteristics of high-risk personnel while using deep learning technology to the risk assessment task of high-risk personnel can use algorithm model to conduct a more accurate and objective analysis of the risk characteristics and relational data of high-risk personnel. Second, the deep learning algorithms can identify potential high-risk personnel and risk rules that police cannot find from massive data on high-risk personnel, help police obtain a deeper understanding of the formation and evolution of the risk of high-risk personnel, and provide a new perspective for risk management and control.

In recent research, [17] use ordinal value quantification to calculate the risk factors for terrorist risk assessment. [19] use the data mining method to mine hidden risk factors from big data and assess the recurrence risk of correctional personnel. However, in those studies, two significant challenges need to be addressed. First, in the era of big data, the characteristics of high dimension, large magnitude, and multiple redundant features of human data are directly applied to machine learning, leading to efficiency decline and dimensional disasters. Second, a person has risk factors and numerous complex relationships with others, which also affect their risk coefficients. Therefore, the direct application of machine learning in risk assessment methods is ineffective in identifying potentially at-risk personnel.

In response to the first issue, we propose a Relief-GAs multi-label [10] feature selection method for feature selection to achieve feature dimensionality reduction. To address the second issue, we apply the knowledge graph to the risk assessment model, taking people as entity nodes and relationships between people as node edges [13]. Building a knowledge graph of high-risk individuals expands the scope of personnel data and mines more potential information by leveraging the relationships between nodes, reflecting the coupling risk factors among individuals. To improve the model's accuracy, we propose an enhanced graph attention network model [14]. The model serves as the base classifier for the label propagation algorithm, to predict and classify the risk level of high-risk personnel nodes with multiple relationships for risk assessment. Graph attention network models suffer from overfitting when trained on large-scale datasets, and real-world graph datasets involve many test nodes. We add the FLAG algorithm to iteratively add node features during training, allowing the model to maintain stability in response to input data's small fluctuations, enabling it to generalize to out-of-distribution samples and improve its performance during the test. Finally, comparative testing and ablation experiments on multiple datasets demonstrate the efficiency and effectiveness of our model [12].

In summary, the main contributions of this work are as follows:

1. To propose a new feature selection method, the Relief-GAs multi-label feature selection method first uses Relief to remove irrelevant features and then uses a genetic algorithm to find the optimal feature subset;

2. To construct a knowledge map of high-risk individuals, effectively exploring potential risks among high-risk individuals;
3. To improve the graph attention network and replace the basic predictor of the label propagation algorithm with the improved model, resulting in higher prediction accuracy.

The remainder of this work is organized as follows. The related research is explained in the second and third sections, including performing feature selection and improving graph attention networks for risk level prediction. Section 4 evaluates the effectiveness of the proposed method, and Section 5 is a summary.

2. Related works

In this section, we summarize the risk assessment methods of high-risk individuals and those in other fields.

2.1. High-risk Personnel Risk Assessment

Among the existing research methods, domestic and foreign scholars mainly analyze personnel risks through qualitative methods. There is few research on evaluating personnel risks by quantitative methods. [8] use naive Bayesian networks combined with four risk characteristics (static risk, violation score, sudden risk, psychological risk) manually annotated based on police experience to predict unknown risks of supervised personnel, with an accuracy rate of 84% and a recall rate of 86%. [3] construct a judgment matrix based on drug users' typical characteristics (physiological characteristics, social characteristics, drug exposure characteristics, and data characteristics) using the Analytic Hierarchy Process to predict the risk of social drug users. [18] propose a risk assessment method for key personnel by using random forest screening dataset features, selecting the optimal feature combination as the evaluation indicator, establishing a risk assessment system, using the AHP method to determine indicator weights, and combining the evaluation indicator scoring table.

2.2. Research In Other Fields

Several studies in other fields are related to risk assessment; for example, in food safety, food safety risks are analyzed and evaluated. Currently, research focuses on employing machine learning methods for risk assessment analysis. For example, [16] use BP neural network algorithm to combine features (food category, production province, sampling location) to classify and predict food risks with an accuracy rate of over 95%. Lou et al. [15] propose a risk prediction model based on differential automatic regression moving average and support vector machine (ARIMA SVM). Geng et al. [5] proposed an improved hierarchical clustering radial basis function (AHC-RBF) neural network for food risk prediction and early warning.

3. Methodology

3.1. Construction Of A Knowledge Map For High-risk Individuals

Based on the existing data on individuals and their relationships, we can construct a high-risk personnel knowledge graph, as shown in Figure 1.

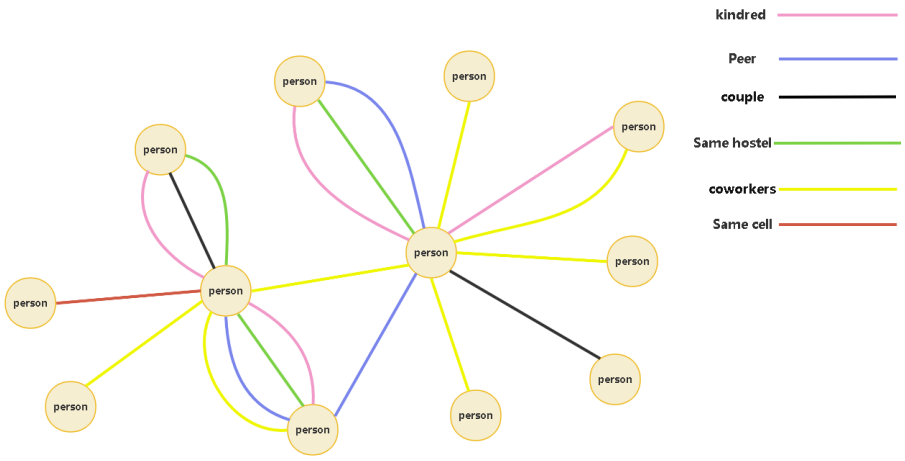

Fig. 1. Schematic diagram of high-risk individual knowledge graph construction

The entity of the knowledge graph is a person, and the attributes of the entity node are the person's risk features (criminal record, recent violation, mental illness, disreputable person, historical disputes, registered personnel, and basic personal information). The various relationships between people serve as node relationships. The entity set of the knowledge graph for high-risk individuals is $X = person(v1)$, with only one type of entity labeled as circles in the graph. The relationship type set $E = \{relative(r1)$, husband and wife$(r2)$, colleague $(r3)$, colleague $(r4)$, same hotel $(r5)$... same ward $(r6)\}$. In the figure, different colors represent different types of relationships. It can be seen that there are multiple edges between two entity nodes. By constructing a knowledge graph of high-risk individuals, we can explore their potential connections.

3.2. Overall Process Of Attention Label Propagation Method

This work proposes a graph attention label propagation method based on the graph attention mechanism consisting of five parts and the algorithm depicted in Figure 2. This method first uses the Relief-GAs multi-label feature selection algorithm to reduce the dimensionality of feature data (section 3.3). The dimensionality-reduced feature data is then used as the basis for a simple prediction classifier using an improved graph attention network. The resulting prediction results are refined using a residual propagation algorithm to correct errors in the prediction, and then the final prediction results are smoothed

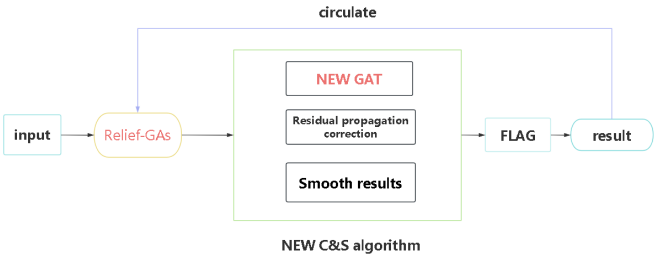

Fig. 2. Describes the five components of the label propagation method based on the graph attention mechanism

using a graph structure (section 3.4). During model training, the FLAG data augmentation technique improves adversarial interference based on gradients and achieves data augmentation (section 3.5).

3.3. Relief GAs Multi-label Feature Selection Method

In the era of big data, personal data has multiple characteristics, such as high dimensionality, large data volume, and numerous redundant features [9]. To avoid dimensionality disasters and improve data value density and model evaluation efficiency, feature selection methods can be used to reduce data dimensionality. Existing feature selection methods can be classified into filter-based, wrapper-based, and embedded-based. Relief is an efficient filter-based feature weight algorithm, but it assigns high weight to all features with high correlation with the class and cannot effectively remove redundant features. We propose a Relief-GAs multi-label feature selection method to address the shortcomings. The algorithm removes irrelevant features using

$$W(A) = W(A) - \sum_{j=1}^{K} \frac{diff(A, R, H_j)}{mK} + \sum_{cclass(R)} \frac{\frac{p(C)}{1-P(class(R))} \sum_{j=1}^{K} diff(A, R, m_j(C))}{mK}. \tag{1}$$

Then uses a genetic algorithm to find the optimal feature set.Among them, $p(C)$ is the proportion of the category, $p(class(R))$ is the proportion of the category of a randomly selected sample, $diff(A, R_1, R_2)$ represents the sample R_1,R_2. The distance on feature A, where m is the number of samples, k is the number of nearest neighbor samples, and $M_j(C)$ represents the $j-th$ nearest neighbor sample in class C.

The algorithm process is as follows:

1. Randomly select a sample R from the training set using the ReliefF algorithm, extract K nearest neighbor samples $H_j(j = 1, 2, \ldots k)$from similar sample sets of R, and then find K nearest neighbor samples $M_j(C)$ from different sample sets, Finally, update the feature weights according to $W(A)$to obtain the average weight of each

feature in a single label dataset and select features with a mean greater than 0 as relevant features to filter the features.
2. Randomly generate feature sets and train the model.
3. Evaluate the fitness of each feature set and remove feature subsets with poor adaptability.
4. Cross-construct a new feature set from the remaining feature sets.
5. Repeat iterations to achieve optimal results.

3.4. Risk Profile

The core of the model is divided into three parts: basic predictor, residual propagation correction, and smoothing prediction results. First, the basic predictor is used to predict the risk of personnel, and the error between the predicted and the truth is corrected by residual propagation. Finally, based on the assumption of the label propagation algorithm, adjacent nodes with similar labels are used to smooth the final prediction results. Each part is introduced as follows.

Basic predictor. The basic predictor is divided into four layers, as shown in Figure 3. Firstly, input the feature data and relational data. Since relational data cannot be directly applied to machine learning, we use the relational quantization layer to quantify relational data as the edges' weights in the graph. The obtained multiple quantitative relationship data and feature data pass through the entity layer, where the feature data is aggregated based on each relationship data, and the numerous aggregated feature matrix is obtained. After the relationship layer, the multiple aggregation matrices are fused to obtain the final feature matrix, which is then used to obtain the classification result in the classification layer. The detailed workflow for each layer is provided below. Since character graph data differs from other graph data during node categories prediction, there are usually multiple relationships between two-character nodes, and node categories are more dependent on the node's relationships. Therefore, this work modifies the graph attention network to serve as the fundamental predictor for label propagation algorithms.

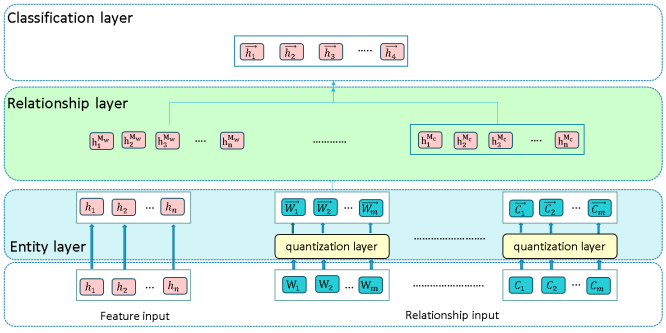

Fig. 3. Detailed structure of the basic classifier used for the label propagation algorithm diagram

Relationship quantification layer,Due to the knowledge graph of high-risk individuals with single entity types and multiple relationship types, different relationship types have different degrees of impact on risk coefficients, and relational data cannot be directly applied in machine learning. Therefore, it is necessary to quantify the relationships between individuals as

$$s_{ij}^{M_K} = a^{M_K} \cdot \frac{h_i \cdot h_j}{\sqrt{\sum_{f=1}^{n} h_{if}^2} \cdot \sqrt{\sum_{f=1}^{n} h_{jf}^2}} \,. \tag{2}$$

In the equation: $s_{ij}^{M_K}$ is the between entity i and entity j quantized value of the relationship under relationship matrix M_K, a^{M_K} is Mapping function under relationship matrix M_K , h_{if}^2 is the value of the $f-th$ dimension of the feature vector of entity i.

Physical attention network layer, The size of a person's risk is related to their risk characteristics and interpersonal relationships. For example, if a person has an intimate relationship with high-risk individuals, their risk level will be increased; that is, by designing an entity attention network layer to achieve the goal. Firstly, learn the attention coefficient between entity a and all connected neighboring entities b in the relationship matrix M_K , where the relationship quantification value is greater than the preset threshold, and calculate it according to equation

$$e_{ij}^{M_K} = a\left(\left[W^{M_K}\ h_i\right| \left|W^{M_K}\ h_j\right]\right). \tag{3}$$

where h_i and h_j are the original feature vectors of entities i and j , and M_K is the weight matrix of the entity attention network layer in the relationship matrix M_K . $||$ is a symbol that represents a connection operation, a is the mapping function , $e_{ij}^{M_K}$ is the attention coefficient in the relationship matrix M_K . Subsequently, normalize the attention coefficient and calculate it in accordance with equation :

$$\partial_{ij}^{M_K} = \frac{\exp\left(LeakyReLU\left(e_{ij}^{M_K}\right)\right)}{\sum_{s \in N_i^{M_K}} exp\left(LeakyReLU\left(e_{is}^{M_K}\right)\right)}. \tag{4}$$

where LeakyReLU is a nonlinear activation function $s \in N_i^{M_K}$, M_K where all relationship quantization values of entity i are greater than the threshold, ∂_{ij} is the normalized attention coefficient.By Using equation $\partial_{ij}^{M_K}$, we can obtain the attention coefficient. By linearly combining this coefficient with adjacent points whose relationship quantization value exceeds the threshold, we aggregate the features after the entity attention network layer under the relationship matrix M_K and get a new feature vector $h_i^{M_K}$. To enhance model training stability, we set the attention mechanism head to $k = 8$, and averaged the feature vectors as equation :

$$h_i^{M_K} = \sigma\left(\frac{1}{K}\sum_{K=1}^{K}\sum_{j \in N_i^{M_K}} \partial_{ij}^{M_K} \cdot W^K \cdot h_j\right). \tag{5}$$

In the above equation, the σ is a nonlinear activation function; the J is the adjacency point j where all relationship quantization values under the relationship matrix M_K are greater than the threshold, $\partial_{ij}^{M_K}$ is the attention coefficient between entities obtained from the

$k-th$ attention mechanism head under the relationship matrix M_K. W^K is the feature transformation matrix corresponding to the $K-th$ attention mechanism head.

Relationship attention network layer,In the knowledge graph of high-risk individuals, each relationship between individuals corresponds to a feature matrix. It is essential to merge new feature vectors from each relationship feature matrix to obtain more valuable entity feature vectors. We design a relationship-level attention network layer. The input for the entity attention network layer serves as the input for the relationship attention network layer, which is calculated using equation:

$$\gamma^{M_K} = \frac{exp\left(\frac{1}{N}\sum_{i\in h} a \cdot sigmoid\left(W \cdot h_i^{M_K} + b\right)\right)}{\sum_{K=1}^{K} exp\left(\frac{1}{N}\sum_{i\in h} a \cdot sigmoid\left(W \cdot h_i^{M_K} + b\right)\right)}. \tag{6}$$

where γ^{M_K} is the attention coefficient, A is the attention mechanism vector, W is a parameterized matrix, b is an offset vector, and sigmoid is a nonlinear activation function. Perform linear combination like the physical network layer, set the attention mechanism header to $p = 8$, and perform averaging, calculated according to equation:

$$\vec{h_i} = \partial\left(\frac{1}{p}\sum_{p=1}^{p}\sum_{k=1}^{k}\gamma^{M_K} \cdot h_i^{M_K}\right). \tag{7}$$

Entity classification laver,The last layer of the model is the entity classification network layer. The core is to aggregate the features of entity I from $\vec{h_i} \in R^F$ to R^C according to equation:

$$\vec{h_i'} = sigmoid(W_c \cdot \vec{h_i}). \tag{8}$$

The set of feature vectors aggregated through these four layers of networks is $\vec{h_i} = \{f_1, f_2, \ldots, f_c\}$, which corresponds to the final classification feature values of the entity. The feature dimension C is used as the risk category to be classified. According to the risk level, the risk is divided into 4 categories (high risk, medium risk, low risk, no risk), and $c = 4$. By normalizing with the softmax function, the probability of the risk level of entity i can be obtained, according to equation :

$$P_i^{f_x} = \frac{\exp\left(f_x\right)}{\sum_{c=1}^{c}\exp\left(f_c\right)} \quad , x \in [1, C]. \tag{9}$$

where $P_i^{f_x}$ is the probability value that entity i belongs to f_x , f_x is an eigenvalue in the entity I eigenvector. After obtaining the basic prediction results, proceed to the next level for error correction.

Error correction in basic prediction by residual propagation. The basis for label propagation is the assumption that adjacent nodes have the same label, which means that the label information of nodes is positively correlated along the graph edges. Therefore, the prediction error of nodes is also positively correlated along the edges, which can improve the accuracy of basic prediction results by combining label information association errors. Firstly, define an error matrix E, where the error of the training set is the residual between its predicted results and the actual label, and the remaining errors are zero:

$$E_T = Z_T - Y_T \; EV = 0 EU = 0. \tag{10}$$

where E_T represents the error of the training set, only when the predicted results are identical to the real labels, the residual of the corresponding training node is zero, and the validation set error E_V and test set error E_U are zero.

Using label propagation technology to smooth errors on the graph and the optimization target is:

$$E = argmintrace\left(W^T\left(1-S\right)W\right) + \mu|\left|W-E\right|\right|_F^2. \tag{11}$$

where W is the final error matrix to be obtained, and S is the normalized adjacency matrix $S = D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$, The first term of the formula promotes the smoothness of the error on the graph, which is equivalent to $\sum_{j=1}^{c} W_j^T(I-S)W_j$, where W_j represents the jth column of matrix W, The second term of the formula controls the degree of deviation between the final solution and the initial error. The solution can be obtained by iterating the equation $E^{t+1} = (1-\alpha)E + \alpha SE^t$ to rapidly converge to E, where $\alpha = \frac{1}{1+\mu}$, $E^{(0)} = E$. This iteration is the process of error propagation, where the smoothed error is added to the basic prediction and the corrected basic prediction $Z^r = Z + E$ is obtained. This is a post-processing technique that does not participate in the training process of the basic predictor.

Smooth the final prediction using graph structure. The corrected basic prediction result $Z^{(r)}$ is obtained after the Correct process. To obtain the final prediction and fully utilize the structural information of the graph, further smoothing processing is needed on the corrected prediction $Z^{(r)}$. Its motivation comes from the assumption in label propagation algorithms that adjacent nodes are likely to have similar labels. Therefore, another label propagation process is used to promote the smoothness of label distribution on the graph. Start from the best prediction of labels $\widehat{Y} \in R^{n\times c}$, $\widehat{Y}_T = Y_T$, $\widehat{Y}_{V,U} = Z^{(r)}_{V,U}$. Replace label information of the trained node information $\widehat{Y}_T$ with real label Y_T, and replace Label information $\widehat{Y}_{V,U}$ in validation and testing sets with the corresponding label information $Z^{(r)}_{V,U}$ in basic prediction $Z^{(r)}$. Iterate the equation $Y^{(t+1)} = (1-\alpha)\widehat{Y}_T + \alpha s\widehat{Y}^{(t)}$ until convergence to obtain the final prediction result $\widehat{Y}$ and perform row normalization to obtain the final label distribution and the node label prediction. Like the Correct process, the smooth process is also a post-processing process and does not participate in the training process of the basic predictor. Then obtain the final prediction result.

3.5. FLAG Data Augmentation

As a semi-supervised learning task, graph node classification often faces a low proportion of labeled nodes [2]. Whenever there is a large difference in the distribution of labeled and unlabeled nodes, the model is prone to overfitting, resulting in a significant difference between the prediction results of unlabeled nodes and the truth, and insufficient generalization ability of the model. Therefore, a data augmentation algorithm called FLAG based on gradient-based adversarial perturbation is used in the model in this work. During training, the node features are iteratively enhanced to make the model invariant to input data's small fluctuations, allowing the model to generalize to out-of-distribution samples and improve its performance, as depicted in Algorithm 1 the flowchart of the FLAG algorithm. The basic idea of the algorithm is to increase the number of projection gradient

descents during each model training iteration and reduce the number of model training iterations. During each model training process, a perturbation matrix of the same size as the feature matrix is defined and sent to the model together with the feature matrix for training. During each gradient descent step in the training process, the perturbation matrix is updated based on its gradient. The gradients are added together, and finally, the parameters are updated by backpropagation. The algorithm adds perturbations to labeled and unlabeled nodes, and adding this algorithm to the high-risk personnel risk assessment model can improve the model's accuracy. In addition, FLAG can alleviate the model's over-smoothing problem and enable the design of deeper graph neural networks.

Algorithm 1 FLAG: Free Large-scale Adversarial Augmentation on Graph

Require: Graph $G = (V, E)$; input feature matrix X; learning rate τ; ascent step M; ascent step size ω; training epochs N; forward function on graph $f_\theta(\cdot)$ denoted in $H^{(k)} = f_\theta(X; G)$; $L(\cdot)$ as objective function. We omit the READOUT$(\cdot)$ function in $h_G = \text{READOUT}(\{h_v^{(k)} \mid v \in V\})$ for the inductive scenario here.
Initialize θ
for epoch $= 1$ **to** N **do**
  $\delta_0 \leftarrow U(-\omega, \omega)$
  $g_0 \leftarrow 0$
  for $z = 1$ **to** M **do**
    $g_z \leftarrow g_{z-1} + \frac{1}{M} \cdot \nabla_\theta L(f_\theta(X + \delta_{z-1}; G), y)$
    $g_\delta \leftarrow \nabla_\delta L(f_\theta(X + \delta_{z-1}; G), y)$
    $\delta_z \leftarrow \delta_{z-1} + \omega \cdot \frac{g_\delta}{\|g_\delta\|_F}$
  end for
  $\theta \leftarrow \theta - \tau \cdot g_M$
end for

4. Evaluation

4.1. Experimental Environment and Dataset

The experimental computer server is composed of Intel Core i7-9700F CPU, NVIDIA GeForce RTX 3070Ti GPU, 8GB memory, 500GB SSD, running Windows 10 operating system and PyTorch, open-source deep learning framework.

The dataset used for training and evaluation is the public-security-field personnel dataset, which contains three types of personnel data (mental illness, criminal record holders, and drug users). The public-security-field personnel dataset selects three types of personnel information data and personnel relationship data that have been desensitized. The structure of personnel information data and relationship data significantly affects the model's accuracy, and appropriate data preprocessing can make the prediction results more accurate. For personnel information data, based on the Relief-Gas multi-label feature selection method in this paper, the optimal feature combination was obtained, which takes "whether there is a case history, recent violations, mental illness, dishonest individuals, historical disputes, and registered individuals" as a strong correlation factor affecting

human risk. For personnel relationship data selection, "peers, same supervision room, and same hotel" are taken as strong correlation relationships. At the same time, people are divided into four categories based on risk levels, personnel information data, and personnel relationship data. Table 1 provides detailed information corresponding to each dataset.

Table 1. Provides detailed information about the public-security-field personnel dataset, including entities, relationships, features, and categories

Data	entities	Relationship	sides	features	category
Drug	1000	6	7689	12	5
Ex-offenders	2000	6	14302	12	5
Psychopath	2000	6	10394	12	5

4.2. Experimental Content

This work compares the model with mainstream methods based on the neural network model, including selecting Graph Attention Network (GAT), C&S, and GraphSAGE to compare accuracy and recall in the public-security-field personnel dataset. To demonstrate the effectiveness of the model algorithm in different situations, several ablation experiments are conducted to analyze the other modules' impact on performance.

4.3. Results Of Classification Accuracy And Recall

The experimental results of the proposed method are compared with mainstream methods, as depicted in Figure 4. Figure 5 shows that using the proposed model has significant accuracy improvement compared to GAT, C&S, and GraphSAGE. Also, as depicted in Figure 6, it can be observed that the PR curve of this model completely covers those of other models, indicating that the recall and accuracy of the proposed model are superior to others. This is because the proposed model adopts the Relief-Gas multi-label feature selection method, an improved graph attention network as the basic predictor, and incorporates the FLAG algorithm to add gradient-based adversarial perturbations to the input node features to enhance the data. Thus, the model can be generalized to samples outside the distribution and improve the performance. Finally, it is more robust and discriminative. The personnel data validation set consists of 500 nodes, each with a risk level label. The model is used to predict each node's risk, then compare them with the risk level label to calculate the model accuracy. The detailed classification results are shown in Table 2.

Considering that in real life, there are two unfavorable conditions with human data: one is edge missing (missing relationships between people) and the other is node information missing [6] (missing human node features), and this work conducts relevant experiments for these two situations.

Edge missing. In the experiment, preserve 20%, 40%, 60%, 80%, and 100% of the edges in the graph and compare them with other models. In the experiment, 70% of the dataset is taken as the training set and 30% as the testing set.

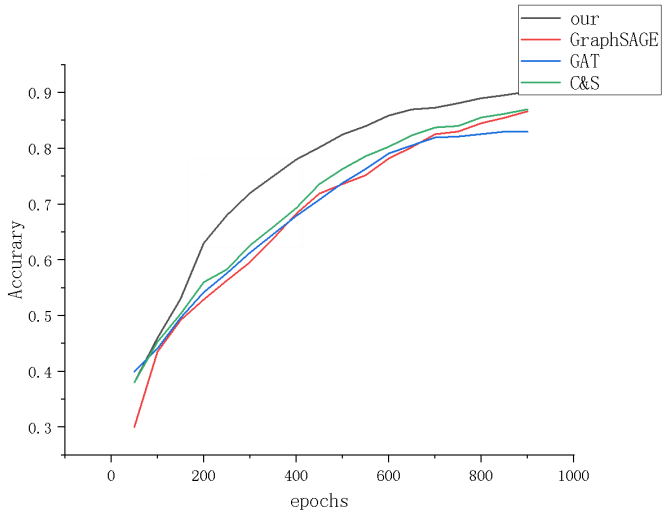

Fig. 4. Training accuracy of different models over epochs. Shows the training accuracy curves for ”our”, GraphSAGE, GAT, and C&S models

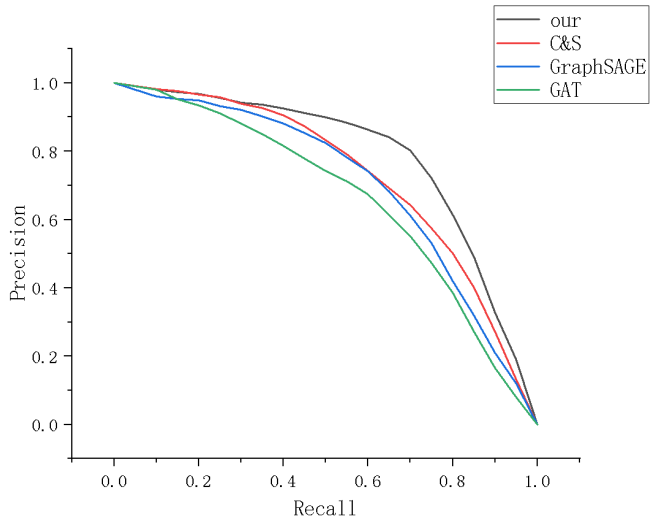

Fig. 5. Compares the PR curves of different models on the public-security-field personnel dataset

The experimental results of different edge missing rates are shown in Figure 6. The graph shows that the more complete the edge information in the graph structure, the higher the accuracy. It verifies that the structural information in the graph can assist in completing node classification tasks. The more complete the structural information, the better the assistance effect. At the same time, it can be found that under different edge missing rates, compared to GAT, C&S models, the classification accuracy of proposed model is higher. This is benefitted from our model's ability to finely mine the interaction information between nodes and the specific interaction information contributing to better classification.

Table 2. Shows the number of correct and incorrect classifications for different risk levels by the model

CATEGORY	CORRECT	INCORRECT
HIGH	31	2
MID	56	7
LOW	46	5
NO	318	35

Missing node information. In the experiment, 40%, 60%, and 80% of non-isolated nodes in the knowledge graph are empty, indicating that they only have structural information. Comparing this model with others, it can be seen from Figure 7 that the classification performance decreases as the amount of data for missing nodes increases. It indicates that node attribute information has an important impact on node classification performance. In the case of varying degrees of information loss, our model performs better compared with GAT and C&S models. This is benefitted from the fact that our model can more finely infer the interaction information between two nodes and thus infer the information of missing nodes.

From the above experiments, it can be concluded that the proposed model utilizes the information and structural information of nodes in the knowledge graph to guide node classification and effectively mines the implicit information between nodes, bringing better classification performance.

4.4. Ablation Studies

From the results of the ablation experiment in Table **??**, the NEW GAT performs feature aggregation on multiple relational nodes, resulting in higher accuracy compared to results of those using traditional GAT as the basic predictor for C&S. At the same time, the FLAG data augmentation algorithm can help the application of graph algorithms and improve model performance. The best feature combination can be found by incorporating the Relief-Gas multi-label feature selection method, and the model's accuracy can be improved. Combining these four methods can significantly improve the accuracy and recall of the model.

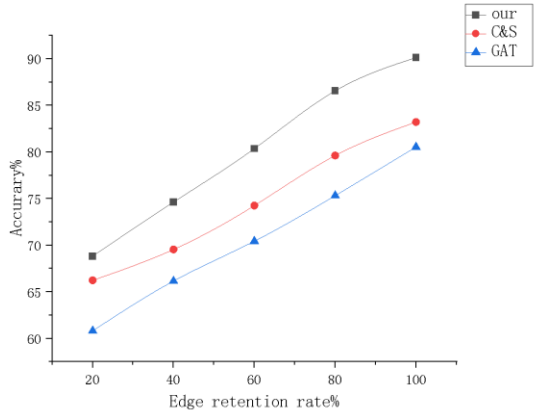

Fig. 6. Shows the impact of different edge retention rates on model classification accuracy

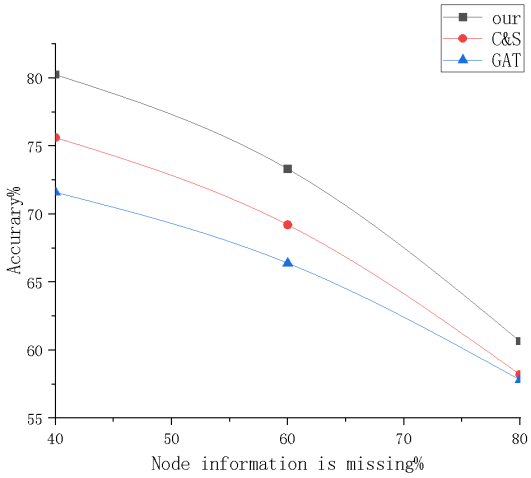

Fig. 7. Shows the impact of different node information missing rates on model classification accuracy

5. Conclusions And Future Work

In the post-pandemic era, how to tap into the potential dangers of high-risk individuals in society and better carry out early warning and control work is the focus of current public security work. With the development of technology and the increase in data volume, traditional risk assessment methods for high-risk personnel can no longer meet the requirements of police officers. This article aims to design a more efficient risk assessment model for high-risk personnel.

A High-Risk Social Personnel Risk Assessment Model Based on Graph Attention Label Propagation Algorithm. By analyzing the main risk factors and relationships that affect the risk coefficient of high-risk individuals, a knowledge graph of high-risk individuals is constructed, and a risk assessment model for high-risk individuals based on graph attention label propagation algorithm is proposed. At the same time, the Relief-Gas label selection algorithm is used to identify the optimal feature set. we train the model using a public-security-field personnel dataset, the accuracy of proposed method reaches 90.2 the recall of proposed method reaches 90.6%. Through comparative experiments with other mainstream graph neural networks, the experimental results show that the proposed model performs better in the graph node classification task.

The experimental results show that the improvements proposed in this paper for graph attention network and label propagation algorithm can improve the performance of the model in the graph data classification task, and the combination of all the improvements in the risk assessment of high-risk personnel can accurately predict the risk level of high-risk groups, improve the work efficiency of public security personnel, and play an important guiding role in social security work.

This model's innovations in the field of high-risk personnel risk assessment lie in its ability to leverage graph attention mechanisms to analyze complex relationships within the constructed knowledge graph, enhancing the accuracy of risk prediction. The practical value of this model is evident in its potential to assist public security agencies in identifying potential threats more efficiently, thereby enhancing public safety. Compared to existing technologies, the superiority of this model is demonstrated through its higher accuracy and recall rates, as well as its robustness in handling real-world data with varying degrees of incompleteness.

In the future, we will continue to optimize the algorithm proposed in this study to improve its performance and generalization ability. Specifically, the following aspects will be considered.

1. Improving the size and quality of the dataset will continue to expand its size and incorporate more scenarios and situations, enhancing the adaptability and generalization ability of the algorithm.
2. The ultimate goal of this paper is to improve the computational efficiency and speed of the algorithm, and successfully apply it to the intelligent policing platform to help public security officers better maintain social order. Therefore, we will continue to optimize the computational efficiency of the algorithm.
3. Explore more network structures and algorithms, continue to focus on relevant algorithms in related fields, and explore more algorithms to improve the effectiveness and performance of the model.

Acknowledgments. This research was supported by the Key Research and Development Program of Hunan Province of China under Grant No. 2022SK2109, Hunan Provincial Department of Education Outstanding Youth Fund under Grant No. 21B0850, Zhejiang Provincial Natural Science Foundation of China under Grant No. LTGG23F020004, Natural Science Foundation of Fujian Province under Grant No. 2023J011460.

References

1. Cai, J., Liang, W., Li, X., Li, K., Gui, Z., Khan, M.K.: Gtxchain: A secure iot smart blockchain architecture based on graph neural network. IEEE Internet of Things Journal 10(24), 21502–21514 (2023)
2. Cai, J., Liang, W., Li, X., Li, K., Gui, Z., Khan, M.K.: Gtxchain: A secure iot smart blockchain architecture based on graph neural network. IEEE Internet of Things Journal (2023)
3. Cai, L.: Research on risk assessment of social drug users. People's Public Security University of China (2019)
4. Diao, C., Zhang, D., Liang, W., Jiang, M., Li, K.: A novel attention-based dynamic multi-graph spatial-temporal graph neural network model for traffic prediction. IEEE Transactions on Emerging Topics in Computational Intelligence 9(2), 1910–1923 (2025)
5. Geng, Z., Liu, F., Shang, D., Han, Y., Shang, Y., Chu, C.: Early warning and control of food safety risk using an improved ahc-rbf neural network integrating ahp-ew. Journal of Food Engineering (2021)
6. Hao, Z., Ke, Y., Li, S., Cai, R., Wen, W., Wang, L.: Node classification method in social network based on graph encoder network. Journal of Computer Applications 40(1), 188–195 (2020)
7. Hu, N., Zhang, D., Xie, K., Liang, W., Li, K., Zomaya, A.: Multi-graph fusion based graph convolutional networks for traffic prediction. Computer Communications 210, 194–204 (2023), https://www.sciencedirect.com/science/article/pii/S0140366423002785
8. Li, S., Li, P., Li, S., Zhao, S.: Research on the application of risk assessment and early warning model for supervised personnel. Police Technology pp. 38–41 (2021)
9. Li, Y., Liang, W., Xie, K., Zhang, D., Xie, S., Li, K.C.: Lightnestle: Quick and accurate neural sequential tensor completion via meta learning. In: IEEE INFOCOM 2023 - IEEE Conference on Computer Communications. pp. 1–10. IEEE (2023)
10. Liang, W., Li, Y., Xie, K., Zhang, D., Li, K.C., Souri, A., Li, K.: Spatial-temporal aware inductive graph neural network for c-its data recovery. IEEE Transactions on Intelligent Transportation Systems 24(8), 8431–8442 (2022)
11. Liang, W., Li, Y., Xie, K., Zhang, D., Li, K.C., Souri, A., Li, K.: Spatial-temporal aware inductive graph neural network for c-its data recovery. IEEE Transactions on Intelligent Transportation Systems 24(8), 8431–8442 (2023)
12. Liang, W., Li, Y., Xu, J., Qin, Z., Zhang, D., Li, K.C.: Qos prediction and adversarial attack protection for distributed services under dlaas. IEEE Transactions on Computers (2021)
13. Liang, W., Xie, S., Cai, J., Xu, J., Xu, Y., Hu, Y.: Deep neural network security collaborative filtering scheme for service recommendation in intelligent cyber-physical systems. IEEE Internet of Things Journal 9(22), 22123–22132 (2021)
14. Liu, Y., Liang, W., Xie, K., Xie, S., Li, K., Meng, W.: Lightpay: A lightweight and secure off-chain multi-path payment scheme based on adapter signatures. IEEE Transactions on Services Computing (2023)
15. Lou, H., Cao, Q., Li, H.: The prediction of food safety risk of china's export to eu based on the arima-svm combination model. Food Industry 41(1), 334–339 (2020)
16. Wang, X., Zuo, M., Xiao, K., Liu, T.: Data mining on food safety sampling inspection data based on bp neural network. Journal of Food Science and Technology 34(6), 85–90 (2016)

17. Wang, Y., Dang, J., Hu, X., Zhang, Y.: Research on the application of matrix method and sequential value method in risk assessment of terrorist suspects. Journal of Zhejiang Police College 2021(01), 83–91 (2021)
18. Wu, R., He, Y., Jia, Y.: Key personnel risk assessment method based on improved ahp. China Safety Science Journal 31(10) (2021)
19. Wu, Z., Fang, Y.: Risk assessment of community correction personnel recidivism based on data mining technology. Guizhou Social Sciences 2016(7), 82–87 (2016)
20. Zhang, S., Ren, F., Liang, W., Li, K., Ling, N.: Gpvo-fl: Grouped privacy-preserving and verification-outsourced federated learning in cloud-edge collaborative environment. IEEE Transactions on Network and Service Management pp. 1–1 (2025)

Xin Su was born in Hunan, China in 1983. He received the Ph.D from Hunan University, Changsha, China, in 2015. He is currently an professor in the Department of Information Technology at Hunan Police Academy, Changsha, China. His research interest is Large Language Model, Machine learning, and Data Analyze.

Heng Zhang was born in Hunan, China in 1980. He received the bacheler degree from People's Public Security University of China, Beijing, China, in 2002. He is currently an political commissar in Zhuzhou Public Security Bureau Economic Development Zone Branch, Zhuzhou, China. His research interest is Intelligence policing.

Xuchong Liu was born in Hunan, China in 1974. He received the Ph.D from Central South University, Changsha, China, in 2010. He is currently an professor in the Department of Information Technology at Hunan Police Academy, Changsha, China. His research interest is Intelligence policing, Big data analyze.

Chuming Bai was born in Shanxi in 2000. He is currently a third-year graduate student at the School of Computer Science at Xiangtan University. His research interests include federated learning, large-scale language models, and deep learning.

Wei Liang was born in Hunan, China in 1978. He received the Ph.D from Hunan University, Changsha, China, in 2013. He is currently an professor in the College of Computer Science and Engineering at Hunan University of Science and Technology, Xiangtan, China. His research interest is Blockchain, Quantum Secure Computing.

Ning Jiang was born in Hubei, China in 1987. He received the ma.eng from wuhan University, wuhan China, in 2012. He is currently as the Director of the Information Department at Jingshan People's Hospital in Hubei Province. His research interests include software engineering, big data, and hospital information management.

Received: June 01, 2025; Accepted: August 22, 2025.