



Research article

Intelligent crowd sensing pickpocketing group identification using remote sensing data for secure smart cities

Jing Zhang¹, Ting Fan^{1,*}, Ding Lang², Yuguang Xu¹, Hong-an Li¹ and Xuewen Li³

¹ College of Computer Science and Technology, Xi'an University of Science and Technology, Xi'an 710600, China

² College of Energy Engineering, Xi'an University of Science and Technology, Xi'an 710600, China

³ College of Safety Science and Engineering, Xi'an University of Science and Technology, Xi'an 710600, China

* **Correspondence:** Email: 21308223012@stu.xust.edu.cn.

Abstract: As a public infrastructure service, remote sensing data provided by smart cities will go deep into the safety field and realize the comprehensive improvement of urban management and services. However, it is challenging to detect criminal individuals with abnormal features from massive sensing data and identify groups composed of criminal individuals with similar behavioral characteristics. To address this issue, we study two research aspects: pickpocketing individual detection and pickpocketing group identification. First, we propose an IForest-FD pickpocketing individual detection algorithm. The IForest algorithm filters the abnormal individuals of each feature extracted from ticketing and geographic information data. Through the filtered results, the factorization machines (FM) and deep neural network (DNN) (FD) algorithm learns the combination relationship between low-order and high-order features to improve the accuracy of identifying pickpockets composed of factorization machines and deep neural networks. Second, we propose a community relationship strength (CRS)-Louvain pickpocketing group identification algorithm. Based on crowdsensing, we measure the similarity of temporal, spatial, social and identity features among pickpocketing individuals. We then use the weighted combination similarity as an edge weight to construct the pickpocketing association graph. Furthermore, the CRS-Louvain algorithm improves the modularity of the Louvain algorithm to overcome the limitation that small-scale communities cannot be identified. The experimental results indicate that the IForest-FD algorithm has better detection results in Precision, Recall and F1score than similar algorithms. In addition, the normalized mutual information results of the group division effect obtained by the CRS-Louvain pickpocketing group identification algorithm are better than those of other representative methods.

Keywords: crowdsensing; intelligent transportation; IForest-FD algorithm; CRS-Louvain algorithm

1. Introduction

Smart cities can make full use of remote sensing data to analyze and integrate various important pieces of information from the urban core system, so as to make intelligent responses to various services of public safety and transportation. However, some people seldom pay attention to their property when taking public transport in smart cities, which provides opportunities for pickpockets [1]. Although considerable manpower and material resources have been allocated, many pickpockets still do not get caught, which brings risks to the people's property safety. Therefore, it is very important to intelligently identify pickpockets and their groups in smart cities [2]. Smart cities use Internet of Things (IOT) to break down the data islands between devices and organizations, and combine technologies such as wireless sensor networks and cloud computing. It can provide various data for pickpocket detection [3, 4], which is of great significance to understand the identity characteristics and travel mode of passengers. Considering the huge differences between pickpockets and normal passengers in travel time, visited stations and stay time, it is possible to find potential pickpockets and groups through exploration of ticketing data provided by the IOT [5, 6]. However, it is difficult to accurately identify pickpocketing groups within massive data. Crowdsensing can solve this problem, through adopting the various real-time data collected by the IOT sensors, and then intelligently processing the data by machine learning methods running in the cloud to complete large-scale pickpocketing group identification [7, 8].

How can suspicious individuals be detected in a large amount of data to avoid false-positives? There are thousands of passengers traveling every day, and pickpockets are only a tiny part of them. To detect suspicious individuals in such large-scale data, we need to filter the passengers who have completely normal travel modes and are unlikely to be pickpockets [9]. Therefore, we use the collected ticketing data and geographic information data to extract the temporal, spatial and social features of passengers. Through the combined analysis of various features, the normal individuals are filtered out before pickpocketing individual detection, which greatly improves the performance of our algorithm.

How to accurately identify pickpocketing groups among suspicious individuals is a key research topic of this paper [10]. Members of groups often show strong correlations in temporal, spatial, social and identity features. The correlations exist not only in the space-time domain, but also fellow villagers and fellow criminals who are active in the same area and are more likely to form groups to commit crimes together. However, existing studies rarely pay attention to multi-factor correlation combinations for pickpocketing group identification [11]. In this paper, we use the weighted combination similarity of multi-factor features to measure the correlation between members. Furthermore, the pickpocketing groups are mainly small groups composed of two to three people, while the traditional algorithms tend to combine small groups into unified large groups.

The contributions of this paper are summarized as follows:

(1) We propose an IForest-FD pickpocketing individual detection algorithm. The IForest algorithm is used to filter the abnormal individuals using temporal, spatial and social features. Through the filtered results, the FD algorithm learns the combination relationship between low-order and high-order features to improve the accuracy of identifying pickpockets, which is composed of FM [12] and DNN [13].

(2) We propose a CRS-Louvain pickpocketing group identification algorithm. Based on the idea of crowdsensing, we measure the similarity of temporal, spatial, social and identity features among the pickpocketing individuals. We use the weighted combination similarity as an edge weight to construct the pickpocketing association graph. Furthermore, the CRS-Louvain algorithm improves the modularity of

the Louvain algorithm [14] to overcome the limitation that small-scale communities cannot be identified.

2. Related works

Based on massive sensing data, pickpocketing individuals and groups can be identified. It is mainly to be able to identify the behavioral characteristics differences between pickpocketing individuals and normal passengers and similar behavioral characteristics among members of the group. At present, relevant researches mainly include two aspects: pickpocketing individual detection and pickpocketing group identification.

2.1. Pickpocketing individual detection

The purpose of pickpocketing individual detection is to use the crime related data and intelligent analysis technology to find individuals who are significantly different from most individuals, so as to assist law enforcement agencies in detecting for criminals. Ramachandran et al. [15] proposed an intelligent automatic system to detect behavior of the human in public places including such as fighting, pickpocketing and threatening. the behavior of humans was monitored using a convolution neural network (CNN), and multi-classifier support vector machines (MSVMs) were used on the various features scored to be able to predict activities. Tsiktsiris et al. [16] used deep learning and unsupervised approaches to detect abnormal events in autonomous vehicles including pickpocketing, bag snatching, fighting etc. Selvi et al. [17] introduced an enhanced convolutional neural network (ECNN)-based suspicious activity detection system to detect shooting and stealing. Pascale et al. [18] tested the performance of multiple federated devices encompassing drones, closed-circuit television, smart phone cameras and smart glasses to detect real-case scenarios of potentially malicious activities such as mosh pits and pickpocketing. All the above are based on the detection of abnormal individual behaviors in surveillance video data. For the detection of abnormal individuals in non-surveillance video data, the traditional methods are mainly based on statistical analysis to detect the criminals, which is seriously influenced by expert experience. To find pickpockets intelligently, the method based on machine learning has been widely used. Chen et al. [19] used the logistic regression (LR) algorithm to analyze the people with criminal records and found the individuals with the tendency of being repeat offenders. Du et al. [20] used the support vector machines (SVM) algorithm to detect potential pickpockets in the public transit systems. Ogunleye et al. [21] applied the XGBoost algorithm to detect abnormal individuals by using various features, and the algorithm has high performance. Gu et al. [22] analyzed the collected data of passengers entering the subway station and found the differences between pickpockets and normal passengers in terms of ride time, ride frequency, getting on and off the station, etc. Chun et al. [23] applied DNN to predict the influence of gender, age, race and previous record on whether individuals will commit crimes in the next few years and the severity of their crimes, and obtained good prediction results. Lu et al. [24] proposed a scheme combining principal component analysis (PCA) and gradient lifting tree (GBDT), which found out the crimes by analyzing the time of theft and the selected objects of theft cases in ShenZhen city in the past seven years.

However, the pickpocketing individual detection algorithm based on machine learning is difficult to directly apply to practice because the experimental results tend to the negative ones with a larger proportion facing the imbalance of positive and negative samples. In order to overcome this problem, Xue et al. [25] pre-classified the data according to the features of common abnormal individuals “in-out

in the same station". Pradhan et al. [26] undersampled a large negative sample and oversampled a small positive sample in the experiment. Du et al. [20] proposed an unsupervised LOF algorithm for anomaly detection to reduce data imbalance, and then used decision tree (DT), SVM and other algorithms for classification. This paper reduces data imbalance by using the IForest algorithm [27], which filters the abnormal individuals of each feature extracted from ticketing data because of its linear time complexity and the idea of integrated learning. From the filtered results, the FD algorithm learns the combination relationship between low-order and high-order features to improve the accuracy of identifying pickpockets.

2.2. Pickpocketing group identification

The key-point of pickpocketing group identification is how to explore the similarity between individuals in the group. The relationship between passengers can be quantified as the similarity in temporal, spatial, social and identity features. Previous studies in the field of crime have clearly pointed out that pickpocketing groups are mostly linked by geography or similar criminal experience. Zhang et al. [28] measured the temporal similarity of individuals by drawing individual time distribution histograms and calculating the earth mover's distances (EMD) between histograms. Liu et al. [29] developed a novel measure that simultaneously considers multiple dimensions of travel behavior to quantify intrapersonal variability. Zhao et al. [30] calculated the cumulative cosine similarity distance between sites based on probability. The study of Gravel et al. [31] shows that there is also convergence effect among individuals in criminal groups, and they tend to recruit people who share the same living space and social environment.

To divide groups by the results of similarity measurement between individuals, an increasing number of scholars use machine learning. Wang et al. [32] proposed a density-based spatial clustering of applications with noise (DBSCAN) approach to recognize irregular travel groups based on cascade clustering, which divided the set of points with sufficiently high density into the same community. Troncoso et al. [33] considered the association between two criminal individuals, adopted the LiRAM algorithm based on the constrained shortest path length to classify criminal groups of social networks, and achieved certain results. Lim et al. [34] applied deep reinforcement learning (DRL) to process metadata such as wiretapping times, arrest warrants and judicial decisions, and constructed FDR-CNA algorithm. By taking the output result of DRL as the edge weight, the algorithm built a network to discover the relationship between criminals, and realized the effective identification of the internal membership relationship of criminal groups. Ma et al. [35] realized the community discovery of multi-layer networks by fusing nonnegative matrix factorization and topological structural information. Tayebi et al. [36] applied the Girvan-Newman algorithm based on a greedy strategy to detect offender groups as denser subgraphs of some co-offending network. Zhao et al. [37] applied the Louvain algorithm to find potential pickpocketing groups on buses, and achieved faster convergence speed and better division effect when there are more than five individuals. However, these are not ideal for identifying the small-scale groups. Considering that the small-scale groups of pickpockets account for a very high proportion, this paper proposes the CRS-Louvain algorithm to improve the modularity of the Louvain algorithm, which overcomes the limitation that small-scale groups cannot be identified.

3. Our approach

In this section, we introduce and describe the two algorithms proposed in this paper: IForest-FD pickpocketing individual detection algorithm and CRS-Louvain pickpocketing group identification algorithm. We mainly use the collected ticketing data, geographic information data and personnel identity data to analyze the ticket sales time, travel time, departure and arrival stations. Additionally, we extract the temporal, spatial, social features and passengers' identity features in the travel mode of passengers, and construct the model of pickpocketing individual detection and pickpocketing group identification.

This paper includes two research aspects: pickpocketing individual detection and pickpocketing group identification. First, we propose an IForest-FD pickpocketing individual detection algorithm. The IForest algorithm is used to filter the abnormal individuals by temporal, spatial and social features. Through the filtered results and identity feature, the FD algorithm learns the combination relationship between low-order and high-order features to improve the accuracy of identifying pickpockets. Second, we propose a CRS-Louvain pickpocketing group identification algorithm. Based on the idea of crowdsensing, we measure the similarity of temporal, spatial, social and identity features among the pickpocketing individuals, and then use the weighted combination similarity as an edge weight to construct the pickpocketing association graph. Furthermore, the CRS-Louvain algorithm improves the modularity of the Louvain algorithm to overcome the limitation that small-scale communities cannot be identified. The overall flow of the model is shown in Figure 1.

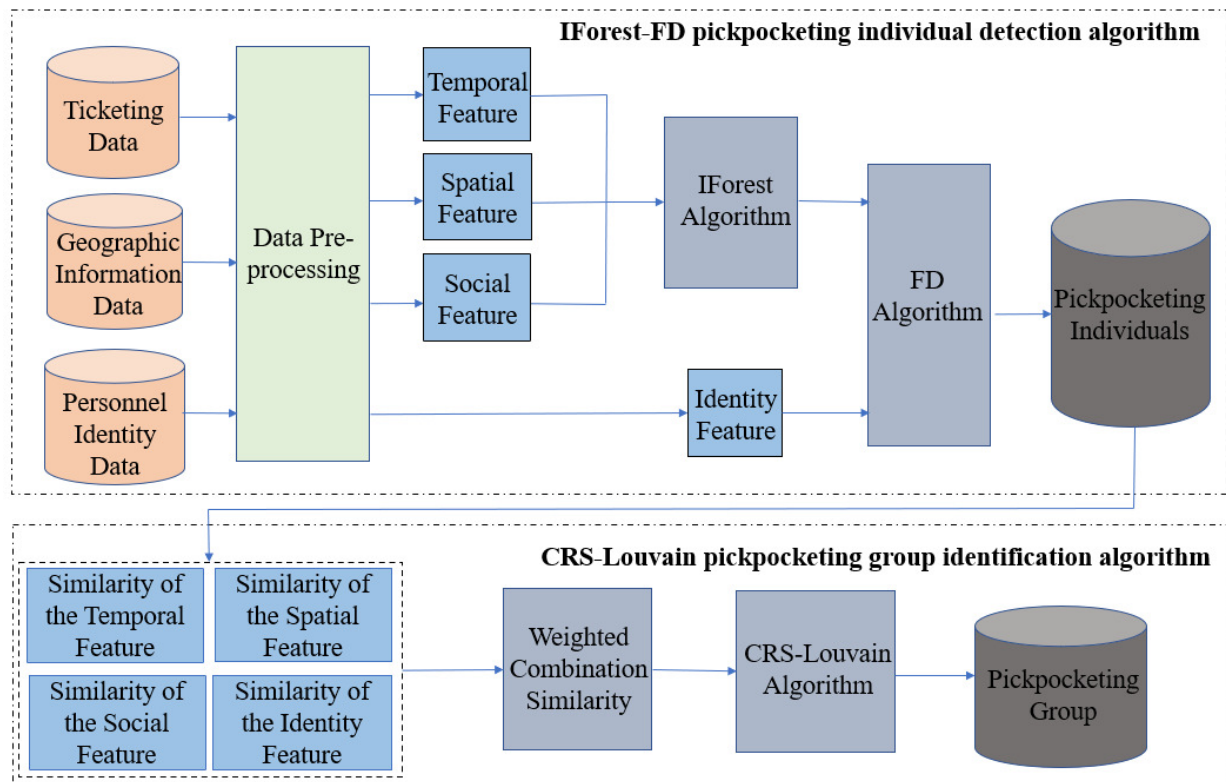


Figure 1. The process of pickpocketing individual detection and group identification algorithms.

3.1. IForest-FD pickpocketing individual detection algorithm

The IForest-FD pickpocketing individual detection algorithm consists of two parts. First, the IForest algorithm is used to filter the abnormal individuals by analyzing the temporal, spatial and social features. Second, the FD algorithm is used for individual pickpocketing detection using the filtered individuals. The FD algorithm learns the combination relationship between low-order and high-order features to improve the accuracy of identifying pickpockets.

3.1.1. IForest algorithm

The IForest algorithm [38] recursively divides the data space until only one tree in all subspaces reaches the upper limit height. Then, each tree is traversed from root to leaf, the average depth is calculated and the abnormal scores are estimated. The abnormal score of sample x is calculated by the following formula:

$$s(x) = 2^{-\frac{E[h(x)]}{c(\varphi)}} \quad (3.1)$$

where $h(x)$ is the height of sample x in the tree; $E[\cdot]$ represents the average of $h(x)$ of all trees, and $c(\varphi)$ is the average of path length when the number of samples φ is given. $c(\varphi)$ is computed as follows:

$$c(\varphi) = \begin{cases} 2H(\varphi - 1) - 2(\varphi - 1)/\varphi, & \varphi > 2 \\ 1, & \varphi = 2 \\ 0, & \varphi < 2 \end{cases} \quad (3.2)$$

where $H(k) = \ln(k) + \varepsilon$, ε is Euler's constant. From the definition of an abnormal score, it can be seen that:

$$s(x) = \begin{cases} 1, & E[h(x)] \rightarrow 0 \\ 0.5, & E[h(x)] \rightarrow c(\varphi) \\ 0, & E[h(x)] \rightarrow \varphi - 1 \end{cases} \quad (3.3)$$

If the value of $s(x)$ is close to 1, the possibility of abnormal data is higher; otherwise the value of $s(x)$ is close to 0, indicating that the possibility of abnormal data is lower.

3.1.2. FD algorithm

The FD algorithm is composed of two parts: Factorization Machines (FM) and DNN, as shown in Figure 2.

The dense embedding layer compresses the input data into low-dimensional dense vectors to solve the problems of data sparsity and excessive dimension. The FM layer obtains the combination relations of first-order and second-order features, and the DNN layer determines the combination relations of high-order features. The FM layer and DNN layer both use vector features compressed by a dense embedding layer as input and are trained simultaneously. The final output result is shown:

$$\hat{y} = \text{Sigmoid}(y_{FM} + y_{DNN}) \quad (3.4)$$

where $\hat{y} \in (0, 1)$ is the output of the whole algorithm, which is transformed into a binary identifier of the pickpocketing individual (0 or 1). y_{FM} is the output of the FM layer, and y_{DNN} is the output of the DNN layer.

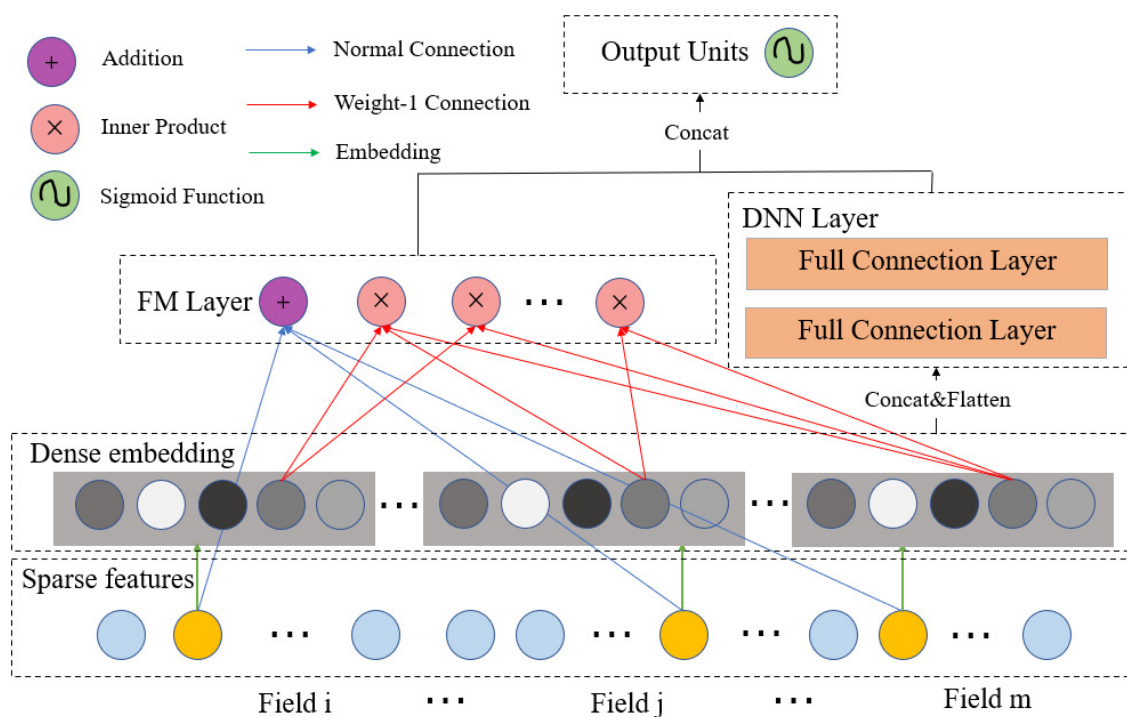


Figure 2. The architecture diagram of FD algorithm.

(1) FM algorithm. The FM algorithm solves the problem of feature combination of sparse data [39]. The algorithm can not only obtain the first-order features, but also capture the second-order features better through the inner product of vector features. The formula for this algorithm is shown below:

$$y_{FM} = \langle w_i, x_i \rangle + \sum_{i=1}^d \sum_{j=i+1}^d \langle v_i, v_j \rangle x_i x_j \quad (3.5)$$

where $w_i \in R^d$, $v_i \in R^k$. d is the feature number and k is the vector dimension. $\langle w_i, x_i \rangle$ reflects the importance of first-order features, and $\sum_{i=1}^d \sum_{j=i+1}^d \langle v_i, v_j \rangle x_i x_j$ represents the second-order feature interactions.

(2) DNN algorithm. The DNN algorithm is used to learn the interactions of high-order features. Before entering the hidden layer, the algorithm uses the dense embedding layer to compress the input vector into low-dimensional dense vectors for training [40]. The output of the dense embedding layer is:

$$a^{(0)} = [e_1, e_2, \dots, e_m] \quad (3.6)$$

$$e_i = v_{field_i} \cdot x_{field_i} \quad (3.7)$$

Then, input the value $a^{(0)}$ into the DNN algorithm, and its forward process can be expressed as:

$$a^{(h+1)} = \sigma(W^{(h)} a^{(h)} + b^{(h)}) \quad (3.8)$$

where h is the number of layers and σ is the activation function. $a^{(h)}$ is the output of layer h . $W^{(h)}$ and $b^{(h)}$ are the weight and bias of the algorithm, respectively. After passing through H hidden layers, the output of the DNN algorithm is as follows:

$$y_{DNN} = \sigma(W^{|H|+1} \cdot a^H + b^{|H|+1}) \quad (3.9)$$

3.2. CRS-Louvain pickpocketing group identification algorithm

Through the set of detected pickpocketing individuals, this paper proposes a CRS-Louvain pickpocketing group identification algorithm. Based on the idea of crowdsensing, we measure the similarity of temporal, spatial, social and identity features among the pickpocketing individuals. To represent the network relationships among individuals, we use the weighted combination similarity as the edge weight and the pickpocketing individuals as nodes to construct the pickpocketing association graph. Further, the CRS-Louvain algorithm improves the modularity of the Louvain algorithm to overcome the limitation that small-scale communities cannot be found.

3.2.1. Similarity measurement

(1) Similarity of the temporal feature $tSim(a, b)$. To reflect the difference between the probability of traveling between pickpockets a and b in a certain period of time, we constructed two histograms H_m^a and H_m^b to represent the visited frequencies v_i^a and v_j^b of pickpockets a and b in time periods t_i^a and t_j^b .

$$\begin{aligned} H_m^a &= \{t_i^a, v_i^a\} (0 < i \leq n) \\ H_m^b &= \{t_j^b, v_j^b\} (0 < j \leq n) \end{aligned} \quad (3.10)$$

Using the EMD distance of two frequency distribution histograms [41], we measure the difference between the two distributions and describe the temporal similarity $tSim(a, b)$ between pickpockets a and b . The formula of $tSim(a, b)$ is as follows:

$$\begin{cases} \sum_{i,j} f_{ij} = \min \{ \sum_i v_i^a, \sum_j v_j^b \}, f_{ij} \geq 0 \\ d_{ij} = |t_i^a - t_j^b| \\ tSim(a, b) = e^{-\sum_{i,j} f_{ij} d_{ij}} \end{cases} \quad (3.11)$$

(2) Similarity of the spatial feature $sSim(a, b)$. We use the weighted cosine similarity [42] to measure the spatial similarity of pickpockets a and b . The formula of $sSim(a, b)$ is as follows:

$$sSim(a, b) = \sum_{n_i \in S_a \cap S_b} \frac{w_i^a w_i^b}{\sqrt{\sum_{n_i \in S_a} (w_i^a)^2} \sqrt{\sum_{n_i \in S_b} (w_i^b)^2}} \quad (3.12)$$

where S_a and S_b are the sites visited by pickpocketing individuals a and b . w_i^a and w_i^b are the frequencies of the site n_i visited by individuals a and b . If the common visited site is empty, the value of $sSim(a, b)$ is 0; otherwise the visited sites of a and b are the same, and the value of $sSim(a, b)$ is 1.

(3) Similarity of the social feature $cSim(a, b)$. We construct the sequence m_a and m_b corresponding to the social features of pickpockets a and b . The negative exponent of the normalized Euclidean distance $cDis(a, b)$ between two feature sequences is used to measure the similarity of individuals a and b in social features. The formulas of $cDis(a, b)$ and $cSim(a, b)$ are as follows:

$$\begin{cases} cDis(a, b) = \sqrt{\sum_{n=1}^3 (m_n^a - m_n^b)^2} \\ cSim(a, b) = e^{-Nor(cDis(a,b))-\sigma} \\ \text{where } \sigma = \frac{\sum_{a,b=1}^{N(a,b)} cDis(a,b)}{2N(a,b)} \end{cases} \quad (3.13)$$

where $N(a, b)$ represents the number of pickpocketing individual pairs.

(4) Similarity of the identity feature $iSim(a, b)$. As pickpocketing groups are mostly organized as fellow villagers or criminals, this paper uses the Jaccard similarity coefficient [43] to measure the similarity between the household registration and criminal record. The calculation formula of $iSim(a, b)$ is as follows:

$$iSim(a, b) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \quad (3.14)$$

where $A = \{a_1, a_2\}$, a_1 and a_2 are the household registration and criminal record of pickpocketing individual a .

(5) Weighted combination similarity $WSim(a, b)$. $WSim(a, b)$ is obtained by weighted combination of the similarity of $tSim(a, b)$, $sSim(a, b)$, $cSim(a, b)$ and $iSim(a, b)$. The calculation formula of $WSim(a, b)$ is as follows:

$$WSim(a, b) = I_1 \times tSim(a, b) + I_2 \times sSim(a, b) + I_3 \times cSim(a, b) + I_4 \times iSim(a, b) \quad (3.15)$$

where I_1, I_2, I_3 and I_4 are the weight coefficients of $tSim(a, b)$, $sSim(a, b)$, $cSim(a, b)$, $iSim(a, b)$, respectively.

$$I_n = \frac{\sum_r P_{rn}}{\sum_n \sum_r P_{rn}} \quad (3.16)$$

where P_{rn} represents the value of pearson correlation coefficient [44] between the similarities r and n . $\sum_r P_{rn}$ represents the correlation between similarities n and other indicators. $\sum_n \sum_r P_{rn}$ is the correlation between any two similarities.

In this paper, the association between pickpockets is regarded as an undirected weighted association graph $G = (V, E)$, where V is the node set of undirected graph, which represents the set of pickpocketing individuals. E is the set of edges, which is the association of pickpockets a and b . The weighted combination similarity $WSim(a, b)$ represents the edge weights of temporal, spatial, social and identity features.

3.2.2. CRS-Louvain algorithm

This paper proposes a CRS-Louvain algorithm for identifying pickpocketing groups. By improving the modularity used in the traditional Louvain algorithm [45], the CRS-Louvain algorithm effectively overcomes the limitation that small-scale communities cannot be identified. The formula of the traditional modularity Q is as follows:

$$Q = \frac{1}{2m} \sum_{i,j} (A_{ij} - \frac{k_i k_j}{2m}) \delta(C_i, C_j) \quad (3.17)$$

where A_{ij} represents the edge weight between nodes i and j , $m = \frac{1}{2} \sum_{i,j} A_{ij}$. k_i represents the degree of node i , $k_i = \sum_j A_{ij}$. C_i represents the community of node i . If C_i is equal to C_j , then δ is 1; otherwise, δ is 0. This formula can be simplified as:

$$Q = \sum_x (\frac{w_x}{m} - (\frac{d_x}{2m})^2) \quad (3.18)$$

where w_x represents the total edge weights in community x , and d_x is the sum of edge weights of all nodes of community x .

This paper introduces a new weighted modularity, which considers the CRS. The formula of CRS weighted modularity Q_λ is as follows:

$$Q_\lambda = \sum_x \left(\frac{w_x \lambda_x}{m} - \left(\frac{d_x}{2m} \right)^2 \right) \quad (3.19)$$

$$\text{where } \lambda_x = \frac{2l_x}{n_x(n_x - 1)}$$

where n_x is the total number of nodes in community x . l_x is the total number of edges in community x . λ_x is the community relationship strength, which represents the ratio of the actual number of edges in community x to the number of edges in the ideal community where all points are fully connected.

The CRS-Louvain algorithm for pickpocketing group identification is as follows:

Step 1: Consider each node in the pickpocketing association graph as a community. We traverse all nodes, find the neighbor that maximizes the CRS weighted modularity of each node, and merge each node with the neighbor. Until the communities of all nodes do not change, a layer of community division is formed.

Step 2: We compress all nodes in each community into one node. The weight of nodes in the community is transformed into the self-weight of new node, and the weight between communities is transformed into the weight of new node edges. We iterate Step 1 to obtain a higher level of community division.

Step 3: We iterate Step 2 until the algorithm is stable. $G = (V, E)$ is divided into n disjoint communities $G_x = (V_x, E_x)$, $x = 1, 2, 3, \dots, n$. Each community only contains individuals with similar temporal, spatial, social and identity features. Thus the pickpocketing groups are identified.

4. Experiments

In this section, we first preprocessed the data, then extracted the data features, and conducted experiments according to the proposed model and algorithm. The experiment includes two parts: pickpocketing individual detection and pickpocketing group identification. On the one hand, we compared IForest-FD pickpocketing individual detection algorithm proposed in this paper with LR, SVM, XGBoost, FD, IForest + LR, IForest + SVM, IForest + XGBoost. On the other hand, we compared CRS-Louvain pickpocketing group identification proposed in this paper with DBSCAN algorithm, the Girvan-Newman algorithm, the Louvain algorithm.

4.1. Analysis of experimental data

4.1.1. Data preprocessing

(1) Ticketing data. The data used in this paper is based on real data in J province and construct the simulated data according to the ratio of the combination of various features. We simulate the ticketing data of passengers in J province from June 1, 2022 to July 5, 2022. The data set has approximately 6.5 million records, mainly includes name, ID number, ride time, number, etc. and the ID number is used as identification ID.

(2) Geographic information data. This data mainly includes the network and running mileage

between stations in J Province. Furthermore, the data of the network includes the routes, stations and location coordinates.

(3) Personnel identity data. This data mainly includes gender, age, place of residence, education, current address and criminal record. Approximately 1% of the data are incomplete and have missing values, and we zeroed the missing values.

4.1.2. Extraction of data features

(1) Temporal feature. The temporal feature mainly consists of four aspects: travel time, riding frequency, night travel ratio and travel regularity. We use entropy to calculate the travel regularity, and the formula is as follows:

$$E_v = - \sum_{i=1}^n P_v(T = t_i) \log P_v(T = t_i) \quad (4.1)$$

where t_i is the time interval in hours, and $P_v(T = t_i)$ represents the probability of each passenger riding within a given time interval $t_i \in T$. Passengers who have a higher night travel ratio and a lower travel regularity are more likely to be pickpockets.

(2) Spatial feature. The spatial feature represents the regularity of stations in passengers' travel behavior. The formula of the regularity of stations is as follows:

$$E_u = - \sum_{l \in l_u} P_u(l) \log P_u(l) \quad (4.2)$$

where l_u is all the stations visited by passenger u within a given time interval. $P_u(l)$ is the probability of visiting a particular station. To reduce costs and avoid being caught, the proportion of key stations and short trips in the pickpocket's record is high.

(3) Social feature. The social feature represents the average stay time of a passenger at one station. As shown in Figure 3, the normal passengers will stay for a long time to finish some work (such as school, travel, work, etc.) after arriving at a place, and the pickpockets desperately need to make his next trip. Therefore, the shorter the stay time is, the higher the abnormality of the passengers.

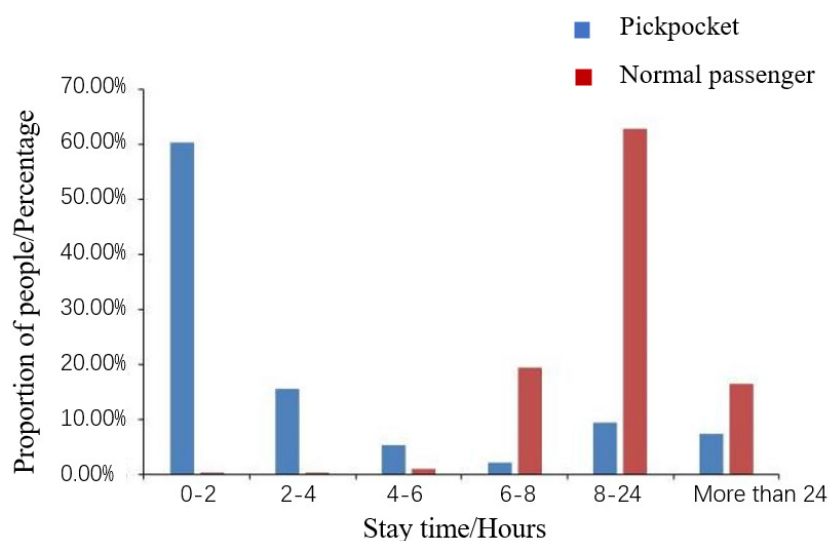


Figure 3. Description diagram of social feature.

(4) Identity feature. The identity feature mainly consist of five aspects: gender, age, place of residence, education and criminal record. Pickpockets and normal passengers have significant differences in the above features, which can improve the accuracy of pickpocketing individual detection.

4.2. Experimental results

After preprocessing the original data, we are left with 4,729,325 records of ticketing data that involve 748,681 passengers. The passengers include 748,513 normal passengers and 168 pickpockets.

4.2.1. Experiment of IForest-FD pickpocketing individual detection

First, we extract temporal, spatial, social and identity features of each passenger, and filter the passengers by the IForest algorithm. Second, we use the LR algorithm [19], SVM algorithm [20], XGBoost algorithm [21] and FD algorithm to detect the pickpockets based on the filtered passengers by the IForest algorithm. We divide all datasets using the 5-fold cross-validation method, and take 10 times to calculate the average value as the results.

(1) Evaluation indicators. To evaluate the performances of different methods, we introduce three evaluation indicators: Precision, Recall, F1score [46]. The calculation formulas are as follows:

$$Precision = \frac{TP}{TP + FP} \quad (4.3)$$

$$Recall = Sensitivity = \frac{TP}{TP + FN} \quad (4.4)$$

$$F1score = \frac{2 * Precision * Recall}{Precision + Recall} \quad (4.5)$$

where TP and FP are defined as the sample number of pickpocketing individuals detected correctly and incorrectly, respectively. TN and FN are defined as the sample numbers of normal passengers detected correctly and incorrectly, respectively.

(2) Result analysis. The experimental results of pickpocketing individual detection are shown in Table 1. We can see that the IForest algorithm can significantly improve the accuracy of the pickpocketing individual detection model. We compared the experimental results of LR, SVM, XGBoost and FD with or without the IForest algorithm. The Precision of IForest + LR, IForest + SVM, IForest + XGBoost and IForest + FD improved by 3.32, 7.41, 5.49 and 6.18%, respectively. The Recall improved by 1.82, 2.68, 6.13 and 5.13%, and the F1score improved by 3.62, 4.72, 5.79 and 5.40%, respectively.

Furthermore, we compare the experimental results of four algorithms, which are IForest + LR, IForest + SVM, IForest + XGBoost and IForest + FD. The Precision of IForest + FD is 13.92, 7.82 and 6.42% higher than that of IForest + LR, IForest + SVM, IForest + XGBoost. The Recall improved by 12.69, 8.94 and 4.39%, and the F1score improved by 11.19, 9.94 and 2.43%, respectively. It can be seen that the FD algorithm can better capture the interactive information between temporal, spatial and social features, as well as the internal information of multiple classification features in identity features, so as to reduce the probability of individual pickpocket missing detection. We also note that the IForest + FD algorithm detected 166 pickpocketing individuals.

Table 1. Experimental results of pickpocketing individual detection.

| Algorithm | Precision (%) | Recall (%) | F1score (%) |
|-------------------|---------------|------------|-------------|
| LR | 81.32 | 84.34 | 83.56 |
| IForest + LR | 84.64 | 86.16 | 87.18 |
| SVM | 83.33 | 87.23 | 83.71 |
| IForest + SVM | 90.74 | 89.91 | 88.43 |
| XGBoost | 86.65 | 88.33 | 90.15 |
| IForest + XGBoost | 92.14 | 94.46 | 95.94 |
| FD | 92.38 | 93.72 | 92.97 |
| IForest + FD | 98.56 | 98.85 | 98.37 |

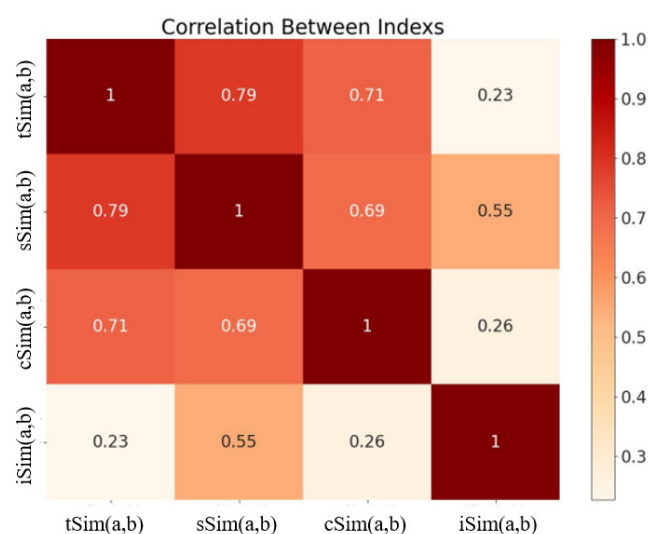
4.2.2. Experiment of CRS-Louvain pickpocketing group identification

(1) Weighted combination similarity. By analyzing the data of 168 pickpocketing individuals, we calculate the similarity of temporal feature $tSim(a, b)$, spatial feature $sSim(a, b)$, social feature and $cSim(a, b)$ identity feature $iSim(a, b)$ between any two pickpocketing individuals. The values of the four similarities and their correlations are shown in Figure 4.

According to the theory of pearson correlation coefficient, we calculate the weight values corresponding to $tSim(a, b)$, $sSim(a, b)$, $cSim(a, b)$ and $iSim(a, b)$ are 0.27, 0.31, 0.26 and 0.16 respectively. Then,

$$WSim(a, b) = 0.27tSim(a, b) + 0.31sSim(a, b) + 0.26cSim(a, b) + 0.16iSim(a, b) \quad (4.6)$$

Furthermore, we delete the pickpocketing pairs with five similarity indices that are less than 0.1, because the probability of such pickpocketing pairs in the same group is low. Then, we identify 210 pickpocketing pairs and 126 pickpocketing individuals.

**Figure 4.** The correlations of the four similarities using pearson correlation coefficient.

(2) Pickpocketing group identification. First, we use the weighted combination similarity $WSim(a, b)$ as the edge weight to construct a pickpocketing association graph, as shown in Figure 5.

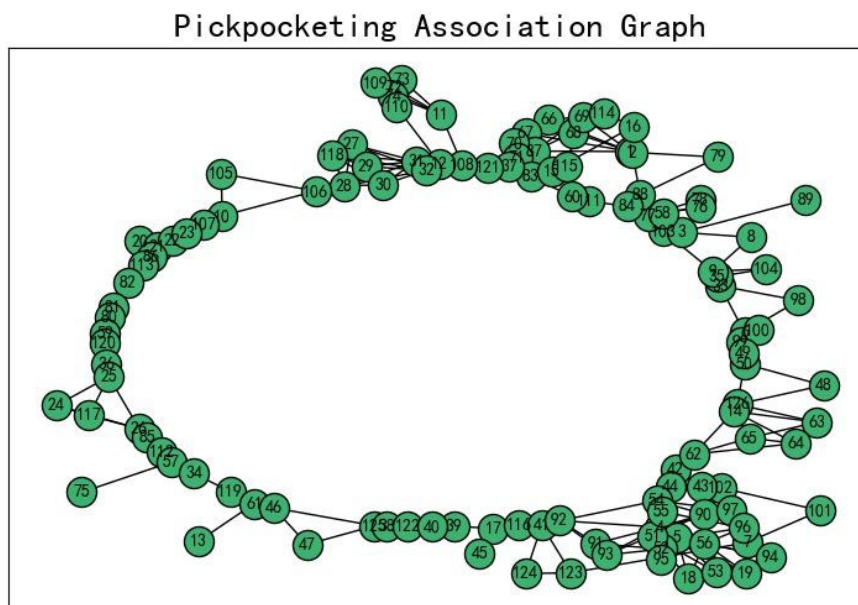


Figure 5. Pickpocketing association graph with 126 nodes and 210 edges.

The pickpocketing association graph consists of 126 nodes and 210 edges, and the aggregated points divide the association graph into multiple pickpocketing groups. Figures 6(a)–(d) respectively show some part of the detection results of the association graph obtained by the DBSCAN algorithm [32], Girvan-Newman algorithm [36], Louvain algorithm [37] and our CRS-Louvain algorithm. Figures 6(a)–(d) include a 10-person group, an 8-person group and two 3-person groups. The nodes of the detected same pickpocketing groups are marked with the same color and connected with solid lines, while the nodes of the detected different pickpocketing groups are marked with the different color and connected with dotted lines.

To compare the CRS-Louvain pickpocketing group identification algorithm with other algorithms, we introduce normalized mutual information (NMI) as an evaluation indicator [47]. The definition of NMI is as follows:

$$NMI_{x,y} = \frac{2I_{X,Y}}{H_X + H_Y} \quad (4.7)$$

where X and Y represent the community labels of nodes in x and y respectively. $I_{X,Y}$ represents the mutual information between X and Y . H_X and H_Y are the entropies of X and Y . The range of $NMI_{X,Y}$ is 0 to 1. The closer the value of $NMI_{X,Y}$ is to 1, the higher the accuracy of community division.

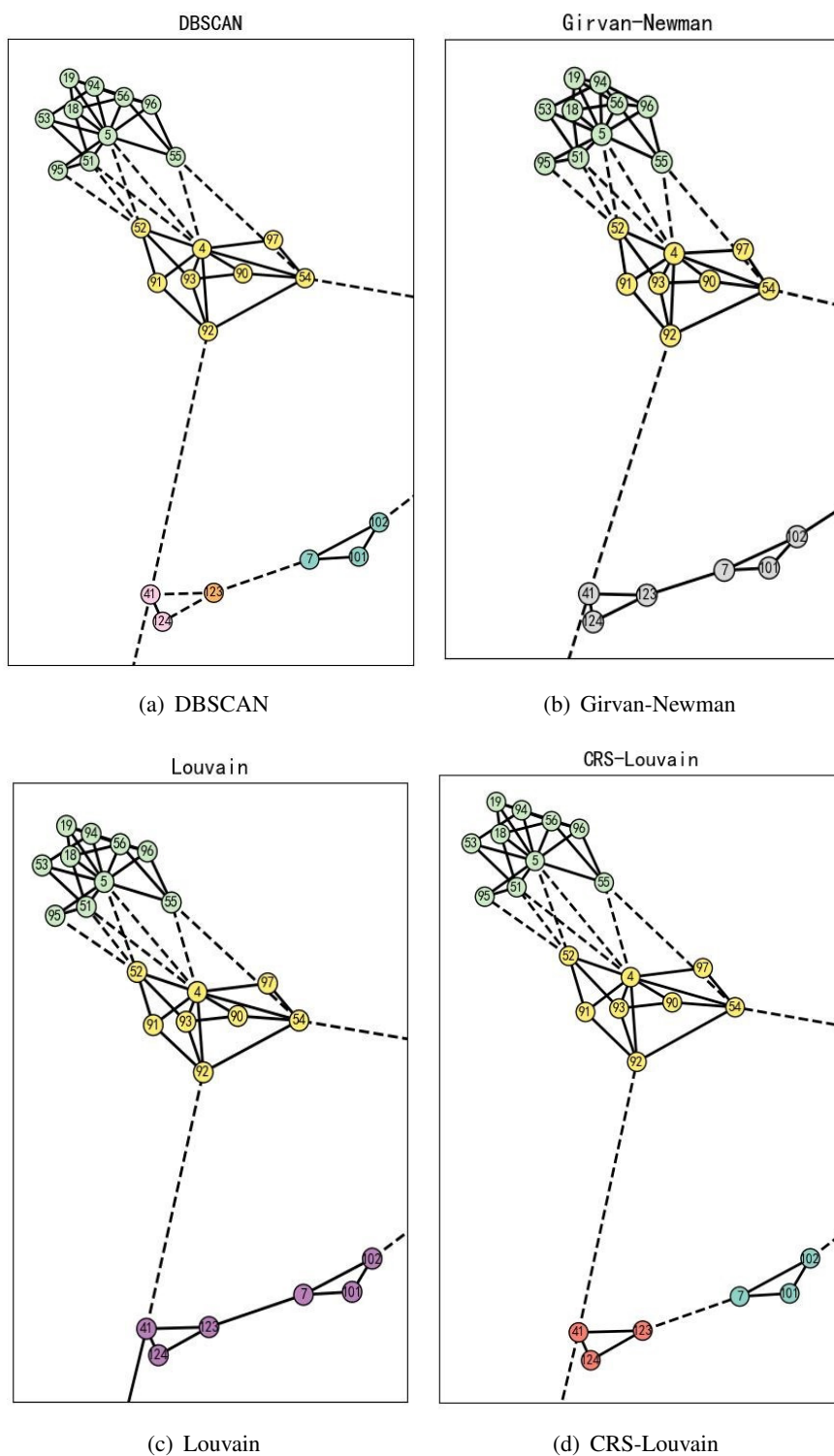


Figure 6. The detection results of each algorithm.

We compare the CRS-Louvain pickpocketing group identification algorithm with the DBSCAN algorithm, Girvan-Newman algorithm and Louvain algorithm and the NMI results are shown in Figure 7.

In the Figure 7, the abscissa is the number of members in each pickpocketing group. The ordinate

is the NMI value, which represents the difference in distribution between the actual pickpocket group and the identification group. The greater the value of NMI, the more accurately pickpocketing groups are detected. When identifying large pickpocketing groups, the four algorithms have shown good performance. However, the Girvan-Newman algorithm and Louvain algorithm do not have strong abilities to identify small groups. When identifying the pickpocketing groups composed of 2 and 3 pickpockets, the NMI values of the CRS-Louvain algorithm are 0.67 and 0.71 respectively, which is better than the DBSCAN algorithm, the Girvan-Newman algorithm and the Louvain algorithm.

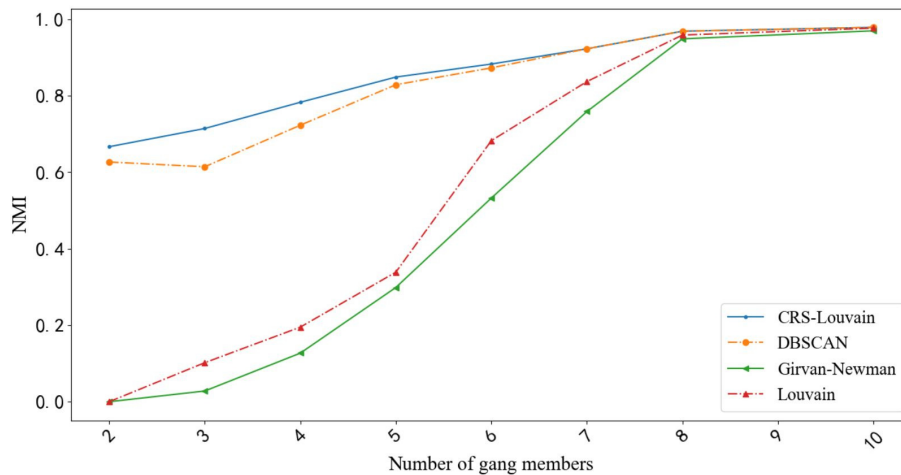


Figure 7. NMI values of the four algorithms for group division.

5. Conclusions

In recent years, the application of informatization, digitalization and intelligence in smart cities has been continuously improved. IOT technology has become one of the core technologies for collecting, storing, managing, analyzing and sharing massive data, and has been widely used in various information systems. The further integration of the IOT with big data effectively solves the problem of accurately judging passenger information with wide coverage and massive historical data for law enforcement agencies. Through intelligent analysis and automatic perception, law enforcement agencies have realized the accurate identification of pickpocketing groups in areas, which has brought new opportunities for the establishment of a decision-making system dominated by information.

This paper includes two research aspects: pickpocketing individual detection and pickpocketing group identification. First, we propose an IForest-FD pickpocketing individual detection algorithm. The IForest algorithm is used to filter the abnormal individuals of each feature extracted from ticketing data and geographic information data. Using the filtered results, the FD algorithm learns the combination relationship between low-order and high-order features to improve the accuracy of identifying pickpockets. Second, we propose a CRS-Louvain pickpocketing group identification algorithm. Based on the idea of crowdsensing, we measure the similarity of temporal, spatial, social and identity features among the pickpocketing individuals, and then use the weighted combination similarity as the edge weight to construct the pickpocketing association graph. Furthermore, the CRS-Louvain algorithm improves the modularity of the Louvain algorithm to overcome the limitation that small-scale communities cannot be identified.

The method proposed in this paper is applicable to the identification of active pickpocketing individuals and their groups. However, there are many kinds of pickpockets operating in the city, and the realization of unified modeling and identification of various criminal behaviors will be a key direction in our future research. After law enforcement officers identify the groups through the designed algorithms, their feedback will help us to identify the groups more accurately. Reinforcement learning can then continuously improve the accuracy of the algorithms through "trial and error". Therefore, how to better combine supervised learning with reinforcement learning to imitate expert decision-making in the real world and realize the automatic adjustment of the model will be another future research direction.

Use of AI tools declaration

The authors have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (Grant No.61902311, 51904226, 52274137, 62172330), in part by the Post doctoral Research Foundation of China (Grant No. 2019M663801), the Key R & D Foundation of Shaanxi Province(Grant No.2021SF-479), and the Scientific Research Project of Shaanxi Provincial Education Department (No.22JK0459).

Conflict of interest

The authors declare that they have no conflicts of interest.

References

1. A. Falahati, E. Shafiee, Improve safety and security of intelligent railway transportation system based on balise using machine learning algorithm and fuzzy system, *Int. J. Intell. Trans. Syst. Res.*, **20** (2021), 117–131. <https://doi.org/10.1007/s13177-021-00274-1>
2. K. Harshad, A Scalable, Reinforcement Learning Algorithm for Scheduling Railway Lines, *IEEE Trans. Intell. Trans. Syst.*, **20** (2019), 727–736. <https://doi.org/10.1109/TITS.2018.2829165>
3. P. Dass, S. Misra, C. Roy, T-Safe: Trustworthy Service Provisioning for IoT-Based Intelligent Transport Systems, *IEEE Trans. Veh. Technol.*, **69** (2020), 9509–9517. <https://doi.org/10.1109/TVT.2020.3004047>
4. J. Srinivas, A. K. Das, M. Wazid, A. V. Vasilakos, Designing Secure User Authentication Protocol for Big Data Collection in IoT-Based Intelligent Transportation System, *IEEE Int. Things J.*, **8** (2021), 7727–7744. <https://doi.org/10.1109/JIOT.2020.3040938>
5. C. Lin, G. J. Han, J. X. Du, T. T. Xu, L. Shu, Z. H. Lv, Spatiotemporal Congestion-Aware Path Planning Toward Intelligent Transportation Systems in Software-Defined Smart City IoT, *IEEE Int. Things J.*, **7** (2020), 8012–8024. <https://doi.org/10.1109/JIOT.2020.2994963>
6. S. Chavhan, D. Gupta, B. N. Chandana, A. Khanna, J. J. P. C. Rodrigues, IoT-Based Context-Aware Intelligent Public Transport System in a Metropolitan Area, *IEEE Int. Things J.*, **7** (2020), 6023–6034. <https://doi.org/10.1109/JIOT.2019.2955102>

7. Y. Liu, T. Feng, M. G. Peng, Z. B. Jiang, Z. Y. Xu, J. F. Guan, et al., COMP: Online Control Mechanism for Profit Maximization in Privacy-Preserving Crowdsensing, *IEEE J. Select. Areas Commun.*, **38** (2020), 1614–1628. <https://doi.org/10.1109/JSAC.2020.2999697>
8. C. Wang, Z. Y. Xie, L. Shao, Z. Z. Zhang, M. C. Zhou, Estimating travel speed of a road section through sparse crowdsensing data, *IEEE Trans. Intell. Transp. Syst.*, **20** (2019), 3486–3495. <https://doi.org/10.1109/TITS.2018.2877059>
9. J. K. Ho, B. P. Jin, H. J. Young, Fuzzy filter for nonlinear sampled-data systems: Intelligent digital redesign approach, *Int. J. Control Autom. Syst.*, **15** (2017), 603–610. <https://doi.org/10.1007/s12555-015-0437-9>
10. A. Capponi, C. Fiandrino, B. Kantarci, L. Foschini, D. Kliazovich, P. Bouvry, A survey on mobile crowdsensing systems: Challenges, solutions and opportunities, *IEEE Commun. Surv. Tutor.*, **21** (2019), 2419–2465. <https://doi.org/10.1109/COMST.2019.2914030>
11. Z. Y. Zhou, J. H. Feng, B. Gu, B. Ai, S. Mumtaz, J. Rodriguez, et al., When mobile crowd sensing meets UAV: Energy-efficient task assignment and route planning, *IEEE Trans. Commun.*, **66** (2018), 5526–5538. <https://doi.org/10.1109/TCOMM.2018.2857461>
12. K. K. Zhao, J. Zhang, L. F. Zhang, C. P. Li, H. Chen, CDSFM: A circular distributed SGLD-based factorization machines, *Database Syst. Adv. Appl.*, **10828** (2018), 701–707. https://doi.org/10.1007/978-3-319-91458-9_43
13. M. Kahng, P. Y. Andrews, A. Kalro, D. H. Chau, ActiVis: Visual exploration of industry-scale deep neural network models, *IEEE Trans. Vis. Comput. Graph.*, **24** (2018), 88–97. <https://doi.org/10.1109/TVCG.2017.2744718>
14. N. Liu, X. L. Yu, L. Yao, X. J. Zhao, Mapping the cortical network arising from up-regulated amygdaloidal activation using Louvain algorithm, *IEEE Trans. Neural Syst. Rehab. Eng.*, **26** (2018), 1169–1177. <https://doi.org/10.1109/TNSRE.2018.2838075>
15. S. Ramachandran, L. H. Palivela, C. Giridharan, An intelligent system to detect human suspicious activity using deep neural networks, *J. Intell. Fuzzy Syst.*, **36** (2019), 4507–4518. <http://dx.doi.org/10.3233/JIFS-179003>
16. D. Tsiktisiris, N. Dimitriou, A. Lalas, M. Dasygenis, K. Votis, D. Tzovaras, Real-time abnormal event detection for enhanced security in autonomous shuttles mobility infrastructures, *Sensors*, **20** (2020), 4943. <https://doi.org/10.3390/s20174943>
17. E. Selvi, M. Adimoolam, G. Karthi, K. Thinakaran, N. M. Balamurugan, R. Kannadasan, et al., Suspicious actions detection system using enhanced CNN and surveillance video, *Electronics*, **24** (2022), 4210. <https://doi.org/10.3390/electronics11244210>
18. D. Pascale, G. Cascavilla, M. Sangiovanni, D. Tamburri, W. J. Heuvel, Internet-of-things architectures for secure cyber–physical spaces: The VISOR experience report, *J. Softw. Evol. Proc.*, 2022. <https://doi.org/10.1002/smr.2511>
19. P. Chen, J. Kurland, S. C. Shi, Predicting repeat offenders with machine learning: A case study of Beijing thieves and burglars, in *2019 IEEE 4th International Conference on Big Data Analytics (ICBDA)*, (2019), 118–122. <https://doi.org/10.1109/ICBDA.2019.8713192>

20. B. Du, C. R. Liu, W. J. Zhou, Z. S. Hou, H. Xiong, Catch me if you can: Detecting pickpocket suspects from large-scale transit records, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (2016), 87–96. <https://doi.org/10.1145/2939672.2939687>
21. A. Ogunleye, Q. G. Wang, XGBoost model for chronic kidney disease diagnosis, *ACM Trans. Comput. Biol. Bioinform.*, **17** (2020), 2131–2140. <https://doi.org/10.1109/TCBB.2019.2911071>
22. H. Gu, Y. Guo, H. Yang, P. Chen, M. Yao, J. Hou, Detecting pickpocketing offenders by analyzing Beijing metro subway data, in *2019 IEEE 4th International Conference on Big Data Analytics (ICBDA)*, (2019), 62–66. <https://doi.org/10.1109/ICBDA.2019.8712833>
23. S. A. Chun, V. A. Paturu, S. C Yuan, R. Pathak, V. Atluri, R. A. Nabil, Crime prediction model using deep neural networks, in *Proceedings of the 20th Annual International Conference on Digital Government Research*, (2019), 512–514. <https://doi.org/10.1145/3325112.3328221>
24. Z. H. Lu, X. F. Hu, L. F. Qiu, Analysis on hazard degree of cases of encroaching on property based on machine learning, *J. Safe. Sci. Technol.*, **15** (2019), 29–35. [10.11731/j.issn.1673-193x.2019.12.005](https://doi.org/10.11731/j.issn.1673-193x.2019.12.005)
25. G. Xue, S. Liu, D. Gong, Identifying abnormal riding behavior in urban rail transit: A survey on “in-out” in the same subway station, *IEEE Trans. Intell. Transp. Syst.*, **23** (2022), 3201–3213. <https://doi.org/10.1109/TITS.2020.3032843>
26. I. Pradhan, K. Potika, M. Eirinaki, P. Potikas, Exploratory data analysis and crime prediction for smart cities, in *Proceedings of the 23rd International Database Applications & Engineering Symposium*, **4** (2019), 1–9. <https://doi.org/10.1145/3331076.3331114>
27. S. W. Xu, J. N. Zhu, J. Z. Jiang, P. L. Shui, Sea-surface floating small target detection by multifeature detector based on isolation forest, *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, **14** (2021), 704–715. <https://doi.org/10.1109/JSTARS.2020.3033063>
28. F. S. Zhang, B. H. Jin, T. J. Ge, Q. Ji, Y. L. Cui, Who are my familiar strangers? Revealing hidden friend relations and common interests from smart card data, in *Proceedings of the 25th ACM International Conference on Information and Knowledge Management*, (2016), 619–628. <https://doi.org/10.1145/2983323.2983804>
29. S. Liu, T. Yamamoto, E. Yao, T. Nakamura, Exploring travel pattern variability of public transport users through smart card data: Role of gender and age, *IEEE Trans. Intell. Transp. Syst.*, **23** (2020), 4247–4256. <https://doi.org/10.1109/TITS.2020.3043021>
30. X. Zhao, Y. Zhang, Y. L. Hu, Z. S. Qian, H. Liu, B. C. Yin, Modeling relation proximity of passengers using public transit smart card data, in *IEEE Intell. Transp. Syst. Mag.*, **14** (2020), 163–172. <https://doi.org/10.1109/MITS.2019.2962145>
31. J. Gravel, B. Allison, J. West-Fagan, M. McBride, G. E. Tita, Birds of a feather fight together: Status-enhancing violence, social distance and the emergence of homogenous gangs, *J. Quant. Criminol.*, **34** (2018), 189–219. <https://doi.org/10.1007/s10940-016-9331-8>
32. L. Wang, Y. Zhang, X. Zhao, H. Liu, K. Zhang, Irregular travel groups detection based on cascade clustering in urban subway, *IEEE Trans. Intell. Transp. Syst.*, **21** (2020), 2216–2225. <https://doi.org/10.1109/TITS.2019.2933497>

33. F. Troncoso, R. Weber, A novel approach to detect associations in criminal networks, *Decis. Support Syst.*, **128** (2020), 113159. <https://doi.org/10.1016/j.dss.2019.113159>
34. M. Lim, A. Abdullah, N. Jhanjhi, M. K. Khan, Situation-aware deep reinforcement learning link prediction model for evolving criminal networks, *IEEE Access*, **8** (2020), 16550–16559. <https://doi.org/10.1109/ACCESS.2019.2961805>
35. C. Ma, Q. Lin, Y. Lin, X. Ma, Identification of multi-layer networks community by fusing nonnegative matrix factorization and topological structural information, *Knowl.-Based Syst.*, **213** (2021), 1–14. <https://doi.org/10.1016/j.knosys.2020.106666>
36. M. A. Tayebi, H. Y. Shahir, U. Glasser, P. L. Brantingham, Spatial patterns of offender groups, in *2018 IEEE International Conference on Intelligence and Security Informatics (ISI)*, (2018), 13–18. <https://doi.org/10.1109/ISI.2018.8587365>
37. X. Zhao, Y. Zhang, H. Liu, S. F. Wang, Z. Qian, Y. L. Hu, et al., Detecting pickpocketing gangs on buses with smart card data, *IEEE Intell. Transp. Syst. Mag.*, **11** (2019), 181–199. <https://doi.org/10.1109/MITS.2019.2919525>
38. H. C. Qu, Z. L. Li, J. J. Wu, Integrated learning method for anomaly detection combining KLSH and isolation principles, in *2020 IEEE Congress on Evolutionary Computation (CEC)*, (2020), 1–6. <https://doi.org/10.1109/CEC48606.2020.9185626>
39. Z. Zhang, Z. Hu, H. Q. Yang, R. Zhu, D. C. Zuo, Factorization machines and deep views-based co-training for improving answer quality prediction in online health expert question-answering services, *J. Biomed. Inf.*, **87** (2018), 21–36. <https://doi.org/10.1016/j.jbi.2018.09.011>
40. L. He, G. H. Wang, Z. Y. Hu, Learning depth from single images with deep neural network embedding focal length, *IEEE Trans. Image Process.*, **27** (2018), 4676–4689. <https://doi.org/10.1109/TIP.2018.2832296>
41. J. Yang, G. Y. Wang, Q. H. Zhang, H. M. Wang, Knowledge distance measure for the multigranularity rough approximations of a fuzzy concept, *IEEE Trans. Fuzzy Syst.*, **28** (2020), 706–717. <https://doi.org/10.1109/TFUZZ.2019.2914622>
42. A. u. Rehman, A. Bermak, M. Hamdi, Shuffled frog-leaping and weighted cosine similarity for drift correction in gas sensors, *IEEE Sensors J.*, **19** (2019), 12126–12136. <https://doi.org/10.1109/JSEN.2019.2936602>
43. M. Besta, R. Kanakagiri, H. Mustafa, M. Karasikov, G. Rätsch, T. Hoefler, et al., Communication-efficient Jaccard similarity for high-performance distributed genome comparisons, in *2020 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, (2020), 1122–1132. <https://doi.org/10.1109/IPDPS47924.2020.00118>
44. G. Z. Li, A. N. Zhang, Q. Z. Zhang, D. Wu, C. J. Zhan, Pearson correlation coefficient-based performance enhancement of broad learning system for stock price prediction, *IEEE Trans. Circuits Syst. II*, **69** (2022), 2413–2417. <https://doi.org/10.1109/TCSII.2022.3160266>
45. M. Seifkar, S. Farzi, M. Barati, C-Blondel: An efficient Louvain-based dynamic community detection algorithm, *IEEE Trans. Comput. Social Syst.*, **7** (2020), 308–318. <https://doi.org/10.1109/TCSS.2020.2964197>

-
46. M. Jiang, L. Song, Y. F. Wang, Z. Y. Li, H. B. Song, Fusion of the YOLOv4 network model and visual attention mechanism to detect low-quality young apples in a complex environment, *Precis. Agric.*, **23** (2021), 559–577. <https://doi.org/10.1007/s11119-021-09849-0>
 47. J. Cho, L. Schmalen, P. J. Winzer, Normalized generalized mutual information as a forward error correction threshold for probabilistically shaped QAM, in *2017 European Conference on Optical Communication (ECOC)*, (2017), 1–3. <https://doi.org/10.1109/ECOC.2017.8345872>



AIMS Press

©2023 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)