System for Detecting Potential Lost Person based on Conditional Random Field

R S Kusuma [1], G A P Saptawati[2]

Schools of Electrical Engineering and Informatics, Bandung Institute of Technology, Bandung, Indonesia

Email: [1] riadysastrak@gmail.com, [2]putri@informatika.org

Abstract. Global Positioning System (GPS) technology has been used widely in transsportation industry to help company in managing taxis. The most popular GPS utilization for taxi company is to identify the position of taxis and monitor theirs the mobility. Nowdays, data collected from GPS tracker is combined with data from taxi meter are analyzed to provide region information regarding potential passengers. Zicheng Liao's proposed a system based on GPS taxi data to detect anomalous area/region which was then interpreted as region with to predict rare passengers. The system was developed based on conditional random field (CRF) method and features position, velocity, passenger loading information. Our research was aimed to develop tool based on GPS data to detect potential lost person. We motivated by Liao research and modified the algorithms and features of CRF. Our experiments showed that the system has precision of 98.86% and recall of 87.478%.

1. Introduction

Nowdays, data collected from GPS can be analyzed to detect the position of a person to know the whereabouts of a person so that if the person is experiencing problems, supervisors are informed of his current position and able to provide help. However, with the limitations of GPS, supervisors do not know when the supervised experience problems. This issue has been addressed by a GPS tracker device. GPS tracker now has an emergency button that the supervised can press to tell the supervisor that he was in danger so that supervisors can immediately go to the scene to deal with the problem. However, this is still not effective regarding children or old men.

A case in point is the case of children and old man with Alzheimer. Old men with Alzheimer often wander aimlessly and reach unfamiliar areas unconsciously. Kids are also prone to getting lost when they find things that make them curious. This is very dangerous situation because they might be kidnapped. When they wander around initially, they do not realize that they are in danger, so they do not press the emergency button on the GPS tracker.

Therefore, old men with Alzheimer and kids need supervision constantly. It is impossible for supervisors to supervise them all the time since they may have other matters to attend to. To handle such cases, we propose a monitoring systems to detect potential lost person based on his/her daily habits.

2. Related Works

Zicheng Liao developed a system for detecting anomalies in GPS data from visual analysis [1]. The system was built based on machine learning with Conditional Random Fields (CRF) method for detecting anomalies in GPS data. In addition to building models for machine learning, the system has a function to display maps with the locations of potential anomalies detected. This system has been tested at a taxi company and the results are quite good since company are able to know where the passengers are rare in each region.

The system displays a map along with the potential anomaly in each region. The system required a model created by the GPS data collection on all taxi. In modelling, it takes labeled training data. The labeling process itself is carried out by the supervisory monitoring system that displays the entire taxi service. Supervisors will provide labels on certain areas in accordance to his knowledge about urban traffic monitoring whether or not the area is an anomaly.

GPS anomaly detection system requires a taxi GPS data. GPS data will be used by the system to analyze behavior patterns by observing taxis' movements. The model is based on training data that has been labeled by the supervisor. It was of a general nature, which means that the model uses the data of all existing taxis and that the model will also be used to detect anomalous areas for all taxis. It was made using CRF method with features that represents the characteristics movement patterns of taxis in general. The features cover speed, time, location, and passenger loading information.

3. Proposed System for Detecting Potential Lost Person

This section describes the development of potential lost person detection system.

3.1. Training data labelling

Training data is needed to build a model that will be used to classify whether or not a person is potentially get lost. Each user of the system (in this case, they could be kids or old man with Alzheimer has their own individual training data to build model based on their own habits and behavior.

Training data were prepared by labeling the GPS data which has already been collected using density-based outlier detection methods. Density-based outlier detection method can detect outliers in spatial data, so that it can be used to classify normal data and outlier data based on density. The normal data will be labeled as normal, while the outlier data will be labeled as an anomaly.

Density-based outlier detection is an anomaly detection based on density data. Density-based outlier detection is an approach to detect the anomaly data based on local data density around the data. Figure 1 illustrates how density-based outlier detection works. In Figure 1, there are two sets of points, C1 and C2, which have different densities. Density-based outlier detection detects the points on C1 and C2 as normal because the points in C1 and C2 have relatively uniform distance to the points around it, creating two areas with a certain density. On the contrary, point o1 and o2 are detected as anomalies, since they have different distances with the points around it.

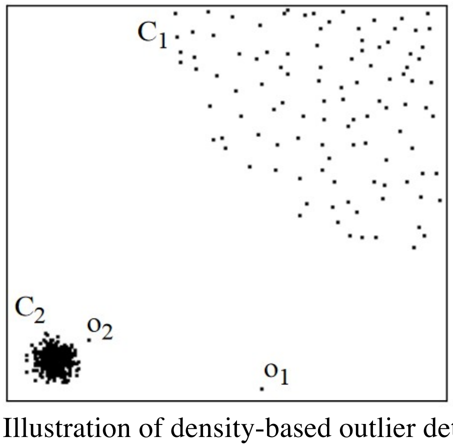

Figure 1 Illustration of density-based outlier detection

Density-based outlier detection clustered data based on the similarity of local reachable density (LRD) of any data. LRD can be determined by calculating the distance of a point with some of the closest points in the dataset using K-Nearest Neighbors (KNN). Below are the main steps for density-based outlier detection:

- Find the K-nearest neighbors of each point in dataset
- Call the maximum distance to K-nearest points that is found in previous step as K-distance. For example, if K=3 and the three nearest neighbors have distances of 1.2, 2.5 and 6.4. Then, the K-distance for this point will be 6.4.
- Next, for certain number of points (MinPts), the reach-distance is calculated.
- Then the LRD of each point is calculated using below formula:
- Finally, LOF scores is calculated using below formula:

3.2. CRF on lost person detection

CRF method is used to detect the anomaly of a pattern. Zicheng Liao used CRF method for GPS anomaly detection. CRF receives sequential data as the input. In lost person detection, CRF receives GPS data streams. The data should be split into smaller sequential data so that it can be effectively processed by CRF.

CRF method requires some feature function to conduct training and classification process on GPS data. Feature function must be defined in accordance with the needs of lost person habits so CRF method can work effectively to detect lost person. Feature function receives the input sequence of data (s), the position of the data in the sequence of data (i), the data label at position i (li), and the data label at position i-1 (li-1). It then produces a real number as the output, although the output is usually either 0 or 1. Output 1 indicates that the label at that position is correct and the output of 0 signifies that the label in that position is wrong. Features that are used to detect the lost will be discussed in the following sections:

3.2.1. Position feature. In GPS anomaly detection system, there is a position feature. If the position is often found in a certain area, the system will detect the data as normal and conclude that person is not lost. The positions on the GPS data sequences is processed by IQR (interquartile range) outlier detection methods that can detect outliers in the input dataset. Outliers are detected by calculating the IQR on the latitude and longitude separately. If the data on the current position is not an outlier in its latitude or longitude, then this feature detects that the person is not lost. But if the data on the current position is an outlier in its latitude or longitude, then this feature detects that the person is lost. Data on the current position can also be determined from the previous label. If the data on the current position is not an outlier in its latitude or longitude, then the position feature states that the current label is same as the previous label. But if the data on the current position is an outlier, then this feature states the current label is opposite to the previous label.

3.2.2. Direction feature The direction of movement can determine whether someone is getting lost or not. Direction is obtained from the difference between current GPS data and the previous GPS data with an assumption that the people who are lost have a tendency to walk with an uncertain direction. If the directions on the GPS data sequences have nearly the same amount and shows a uniform distribution, that means the person does not walk at straight directions and that the person is being lost.

3.2.3. Speed feature The speed feature is a supporting feature in this system. Walking faster usually suggests that the person is outside his usual area. In addition, faster movement also increases the emergency of receiving treatment if the person is lost. If the current data is observed to have a greater speed than the average, then this feature will detect that the person is being lost. Otherwise, if the current data shows a speed less than the average, then this feature will state that the current label is same as the previous label

3.3. System Implementation

Lost person detector use density-based outlier detection and CRF as the main method to detect whether the person is being lost or not. Lost person detector have 2 main phases: training phase and classification phase. In training phase, lost person detector obtain sequences of GPS data collected previously and then label GPS data with as normal or anomaly with density-based outlier detection method to build training dataset. Training datasets are used by CRF method to build a classification model. In classification phase, lost person detector receives sequences of GPS data and then label it accordingly using the model produced from training phase. The system flow can be seen in Figure 2

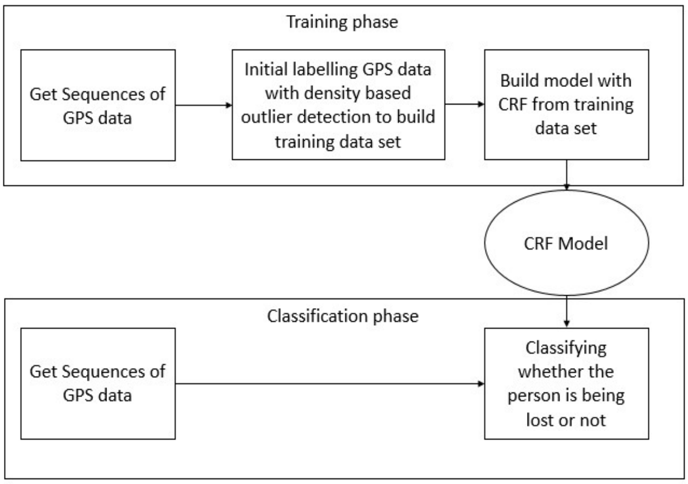

Figure 2 Lost person detector system flow

Figure 3 shows a GPS data visualization of the normal area of a supervised person's usual activity in the past 6 days. Figure 4 shows GPS data visualization of the area where the supervised person is being lost. Lost person detector use GPS data in Figure 3 and Figure 4 as training data. Figure 5 is the output of the classification phase using 10-fold cross-validation. The red markers show normal GPS data, while the purple markers show that supervised person is being lost in that area.

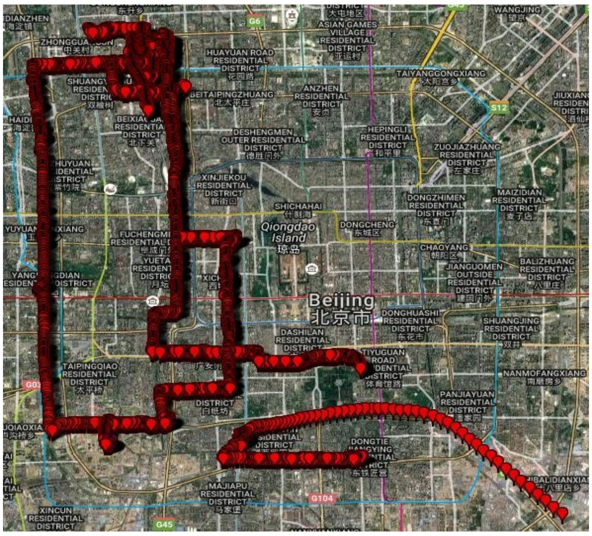

Figure 3. Normal area of supervised person’s usual activity.

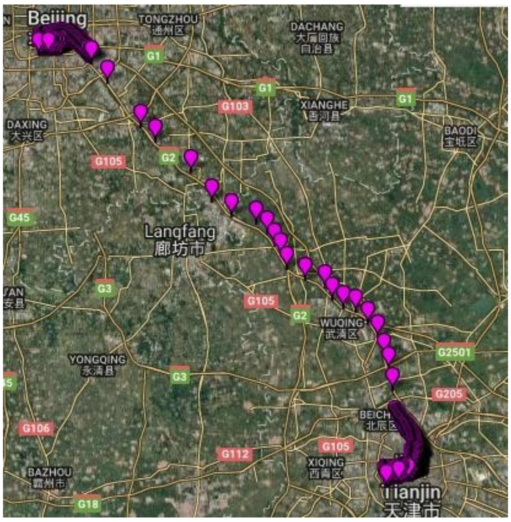

Figure 4. Area that supervised person is being lost.

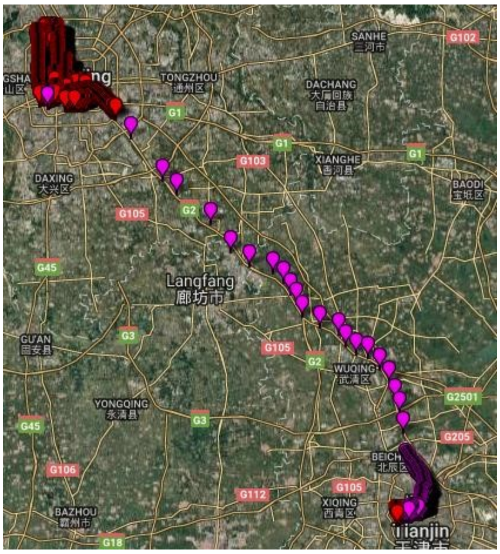

Figure 5. Output of lost person detector.

4. Experimental Result

GPS data used for experiment obtained from the collected data by Zheng et al in 2009 [9]. There are four scenarios to perform this evaluation. For each scenario, the data is collected from the same person. For scenarios 1 and 2, the normal GPS data being used is a collection of GPS data in a duration of six days. For scenarios 3 and 4, the normal GPS data being used is a collection of GPS data in a duration of three days. Anomalies that are used for each scenario are the GPS data retrieved when the person was being lost in one day.

Table 1 Lost detector experiment

Data	Training data set labelling	Precision	Recall
GPS Data collected in 6 days	Manual labelling	98.86786%	87.47855%
	Density- based outlier detection labelling	94.93670%	77.18696%
GPS Data collected in 3 days	Manual labelling	98.74367%	89.526421%
	Density-based outlier detection labelling	81.50087%	76.80921%

In comparison, the precision and recall of the classification process with training data labeled using outlier detection method is smaller than those with manually labeled training data. This is because the manually labeled training data incorporates expert insights which makes it more accurate than labeling training data using outlier detection method.

However, the difference is insignificant, and it is still considered effective to detect lost people using outlier detection method in labeling the training data since the recall is higher that 75%. The only major disadvantage is that the notification might be delayed for a while and there is slightly more probability of false alarms, where the system detects a person as lost while actually he is not.

When comparing the experiment results with different duration of training dataset, very little differences in precision and recall are seen. When outlier detection is used, the precision resulted from using three days GPS data decreased more greatly compared to that obtained from using six days GPS data. This is because the models have not saved a lot of movement patterns and habits so there are more false alarms. However, when the data that is labeled manually, the recall when using three days GPS data is more than that when using six days GPS data. This is because the three days GPS data has already formed a good enough model of the behavior patterns of the supervised person. If the training dataset contains too many normal data than necessary, over time CRF models will grow to be inflexible.

5. Conclusion

It can be concluded that the adjustments to the GPS anomaly detection system to detect lost person should be done by making changes to the training data labeling method and a change of model building. Labeling use density-based outlier detection method changes raw GPS to labeled training data that will be used to create the model. The model is created specifically for each person using the CRF. CRF method uses three features which are position, speed and direction of movement

References

[1] Liao Z 2010 Anomaly Detection in GPS Data Based on Visual Analytics *IEEE Symp. on Visual Analytics Science and Technology* p 51 (Salt Lake City: IEEE)

[2] Breunig M M, Kriegel H-P, Ng R T and Sander J 2000 LOF: Identifying Density-Based Local Outliers. SIGMOD '00 *Proceedings of the 2000 ACM SIGMOD intl. conf. on Management of data* pp 93-104 (New York: ACM)

[3] Dunning T and Friedman E 2014 Practical Machine Learning A New Look at Anomaly Detection (Sebastopol: O'Reilly Media)
[4] Ester M, Kriegel H-P, Sander J and Wu X 1996 A Density-Based Algorithm for Discovering Clusters *KDD-96 Proc*. pp 226231 (Munich: University of Munich)
[5] Kriegel H-P 2010 Outlier Detection Techniques *16th ACM SIGKDD Conf. on Knowledge Discovery and Data Mining*
[6] Lafferty J, McCallum A and Pereira F 2001 Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data *Intl. Conf. on Machine Learning* pp 282-289 (Pittsburgh: Carnegie Mellon University)
[7] National Academy of Public Administration 1995 The Global Positioning System: A Shared National Asset (Washington D.C: National Academy Press)
[8] Upton G and Cook I 1996 Understanding Statistics (New York: Oxford University Press)
[9] Zheng Y, Zhang L, Xie X and Ma W-Y 2009 Mining interesting locations and travel sequences from GPS trajectories World Wild Web pp 791-800 (Madrid: ACM Press)