Chapter 14
Study on the Personal Information Anonymization Method for the Releasing of Navigation Data

Jiannan Gao, Rendong Ying, Peilin Liu and Wenxian Yu

Abstract In this paper we give a quantitative calculation of personal information anonymization in the released navigation data based on the information theory. Individual personal information privacy index is defined and calculated based on the Markov chain model. The simulation of special state model shows that the proposed algorithm can be used to evaluate the ambiguity of different individuals from the navigation data. The simulation is based on Markov chain model, while the proposed personal information privacy metric algorithm does not depend on the motion model, one can apply it to other more accurate individual motion models and provide a quantitative basis for the personal information anonymization for the releasing of navigation data.

Keywords Navigation · Information privacy metri · Entropy · Information theory · Markov model

14.1 Introduction

Along with the development of mobile communication technology, satellite navigation technology and data mining theory, the mobile services based on the location information develop rapidly in recent years. Personal location information is used in different areas and provides customers with convenient and efficient information services including traffic analysis and optimization, mall pedestrian flow analysis, advertising, etc.

J. Gao (✉) · R. Ying (✉) · P. Liu · W. Yu
Shanghai Jiao Tong University, Shanghai 200240, China
e-mail: jiannan.gao@gmail.com

R. Ying
e-mail: rdying@sjtu.edu.cn

J. Sun et al. (eds.), *China Satellite Navigation Conference (CSNC) 2014 Proceedings: Volume I*, Lecture Notes in Electrical Engineering 303,
DOI: 10.1007/978-3-642-54737-9_14,

However, numerous utilizations of location-based services (LBS) also bring privacy and security issues. Once the personal location information is stolen, it is likely to cause the leakage of behavioral patterns, hobbies, living habits and other information, more likely to cause personal security threats. Therefore, LBS information privacy protection is becoming the hotspot and trends of the theoretical study of LBS currently. Most of the existing algorithms based on k-anonymous technologies meet the needs of location information privacy protection of single query users, while for continuous query users, those technologies do not work well because from the analysis of the trajectory, some privacy can still be revealed. When additional ancillary information is provided, it is readily to identify an individual. Therefore, resulting in the leakage of personal privacy.

There are already some trajectory information anonymization technologies including personalized k-anonymous technology [1], the silent period technology [2], the PRIVE method [3] and MOBIHIDE method [4]. SP. Li puts forward the anonymity measurement based on entropy theory [5].

In this paper, a new anonymity measurement is proposed, which is based on the changing of entropy of trace data with time. Besides by utilizing the Markov motion model, personal information anonymization problem during the navigation trajectory information dissemination is also studied in this article.

14.2 Problem Description

The scenes of this study are described as follows. Given a known initial location data of target A, and a target movement trajectory, how to avoid associating target A and the moving trajectory?

An example of application is that a large shopping mall, manager of the mall needs to optimize the layout of the shops by the analysis of the moving trajectory of customers in the store. But the customer does not want the shopping mall to associate their personal information with their trajectory, namely: the shopping mall can get the trajectory data, but is not able to identify the customers.

From a practical point of view, at some specific locations of a shopping mall, such as the elevators with video camera or the cashier section, the identity and location of customer can be known simultaneously. Therefore, to avoid this kind of information binding, the navigation information at these locations should be masked. Suppose that at time 0, the location of individual target is C_0, in order to avoid the association between the published trajectory data and the target, one need to eliminate the trajectory data between the time 0 and T. The question is: what's the minimal value of T. In the following section we will discuss the mathematic model that defined the ambiguity index with the change of time T.

14.3 Mathematic Model

We define the ambiguity of trajectory data and individual goals from the perspective of probability. Assume that at time T, there are M people ($M > 1$) at the position u. the probability distribution of the identity of the person appearing at this moment and this position can be used to infer his identity. This probability is a function of T, $\mathbf{u}$ and the individual identity m, as:

$$p(m; T, \mathbf{u}) \tag{3.1}$$

where $m = 1, 2, \ldots, M$ represent the identity of each person and u is the location an identity can appear. So for different individual m, the variation or flatness of its probability $p(m; T, \mathbf{u})$ can be used to evaluate the ambiguity of each individual. In view of the information theory, the ambiguity can be evaluated by the entropy, namely:

$$h(T; \mathbf{u}) = -\sum\nolimits_m p(m; T, \mathbf{u}) \log p(m; T, \mathbf{u}) \tag{3.2}$$

The entropy $h(T; u)$ reaches the maximum when $p(m; T, u)$ is constant for different m, i.e. each identity looks the same from the point of view of probability. Since the entropy $h(T; u)$ is a function of time T, therefore, it represents the level of personal information anonymization during the navigation trajectory

Considering the actual situation, the customers will not always linger in one area, and with the increase of T they will eventually leave the area, which means that for large enough T,

$$\sum\nolimits_m p(m; T, \mathbf{u}) < 1 \tag{3.3}$$

In order to make the definition of ambiguity in the formula (3.1) meaningful, it is necessary to "normalize" entropy, which gives the definition of the "information privacy metric" as below:

$$g(T; \mathbf{u}) = -\frac{1}{a}\sum\nolimits_m p(m; T, \mathbf{u}) \log \frac{p(m; T, \mathbf{u})}{a} \tag{3.4}$$

where $a = \sum_m p(m; T, \mathbf{u})$ is the normalization factor. For in the scenarios of M individuals, the maximum of $g(T; \mathbf{u})$ is $\log M$.

In order to find the relationship between $g(T; \mathbf{u})$ and T or $\mathbf{u}$, we need to calculate the probability $p(m; T, \mathbf{u})$. In order to define the following motion model we first discretize the time and denote each time step by integers ($n = 0, 1, 2, \ldots$). Then the first-order Markov chain model is selected as the simplified motion model for individuals. An example of the model is shown in Fig. 14.1.

In the state transition model shown above, each state represent a location, therefor 5 locations (u_k, $k = 1, 2, 3, 4, 5$) are considered in the model. Among these locations, u_5 is special, which represents the case that customers leave the mall and don't come back, namely "absorbing state". In this model the position vector of all the identite $\mathbf{u}(n)$ at any moment only depends on the position at

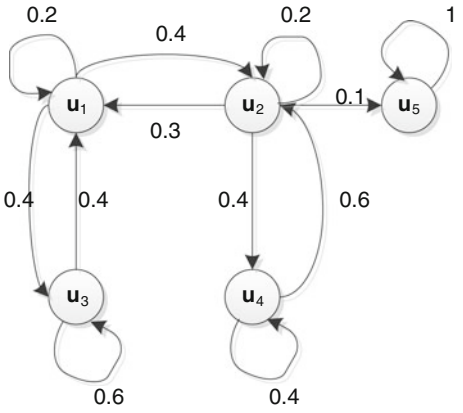

Fig. 14.1 Markov model of personal motion position

former moment $\mathbf{u}(n-1)$ with the transfer probability matrix of the Markov chain model. Numbers on the arrows of the figure represents the probability from one location to another location, which is also the state transition probability. Consider the case when there is only one person in the store, denote $\mathbf{p}(n)$ the probability that this person at time n in each position in the map, thus resulting in the state transition equation below:

$$\mathbf{p}(n) = \mathbf{p}(n-1)\mathbf{H} \tag{3.5}$$

where the matrix $\mathbf{H}$ is the state transition probability matrix. Corresponding to the example shown in Fig. 14.1, the value of $\mathbf{H}$ is given by:

$$\mathbf{H} = \begin{bmatrix} 0.2 & 0.4 & 0.4 & 0 & 0 \\ 0.3 & 0.2 & 0 & 0.4 & 0.1 \\ 0.4 & 0 & 0.6 & 0 & 0 \\ 0 & 0.6 & 0 & 0.4 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{3.6}$$

Based on state transfer formula, and the probability distribution of the initial position $\mathbf{p}(0)$, the probability $\mathbf{p}(n)$ of every individual's potion at any time n can be achieved from the above model, namely:

$$\mathbf{p}(n) = \mathbf{p}(0)\mathbf{H}^n \tag{3.7}$$

Denote

$$\mathbf{p}(0) = [0 \quad \cdots \quad \underset{i\text{th}}{0} \quad 1 \quad 0 \quad \cdots \quad 0] \tag{3.8}$$

the initial probability vector that individual m in position u_i at the initial moment, thus the jth element of $\mathbf{p}(n)$ representing the probability of individual m appear at position u_j at the time n, namely $p(m, n, u_j)$ in the formula 3.1. So calculate the

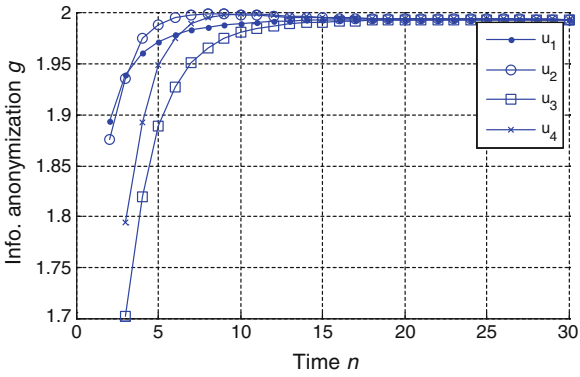

Fig. 14.2 Change of $g(n, u_i)$ with time n. At the initial time four persons are in the positions u_1–u_4, respectively

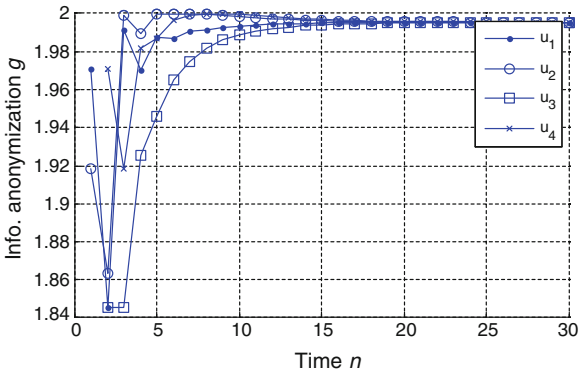

Fig. 14.3 Change of $g(n, u_i)$ with time n. At the initial time, two persons are in u_1 and the other two are in u_2

formula 3.4 and get the personal information anonymization degree function $g(n, u_i)$, and decide whether to release trajectory information at time n according to the proximity of the function and the maximum possible values.

14.4 Model Simulation

The simulation is based on Markov motion model assumptions and state transition matrix. Here is the simulation. Suppose that there are four people at the initial moment, respectively in the four position u_1–u_4. With the passage of time, the information privacy metric is showed in Fig. 14.2.

The "information privacy metric" function in the figure increases with the increasing time n, reflecting the phenomenon of the gradually rising of the difficulty to judge corresponding individual from the location information.

The three diagrams (Figs. 14.3, 14.4, 14.5) are given respectively: (1) At initial time two persons are in u_1 and the other two are in u_2; (2) At initial time one person is in u_1 and the other three are in u_2; (3) At initial time four persons are all

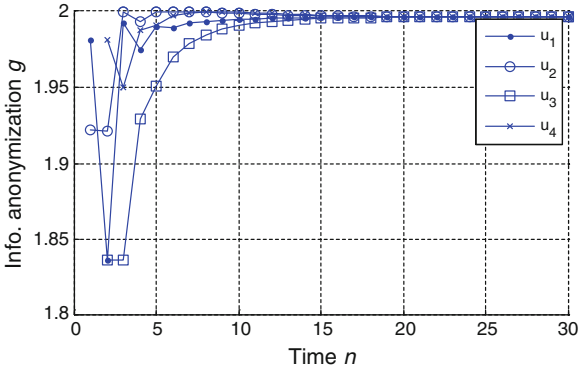

Fig. 14.4 Change of $g(n, u_i)$ with time n. At initial time one person is in u_1 and the others in u_2

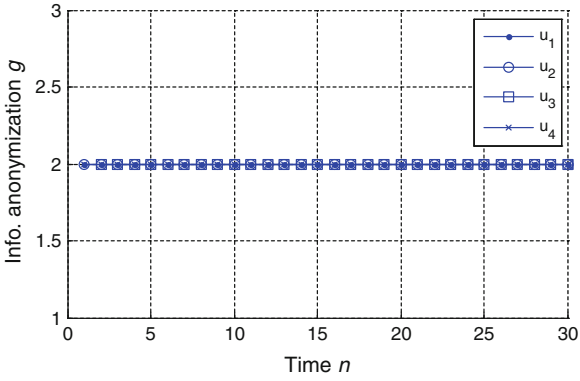

Fig. 14.5 Change of $g(n, u_i)$ with time n. At initial time all the persons are in u_2

in u_2. The diagram shows that "privacy metric" function of g increases with the concentration at initial time, which reflects the fact that at the initial moment the more concentrated the people are, the more difficult to distinguish different individuals by the trajectory.

14.5 Conclusion

In this paper we discuss the method to evaluate the extension of personal information anonymization during the navigation trajectory based on Markov chain model. Besides, this paper defines a function of "personal information privacy metric" based on the information entropy. With this function, the degree of association between data and personal information can be evaluated quantitatively, and the degree of personal information privacy is shown to be function of time and can be calculated with the person motion model. A Markov chain motion model based simulation is given in this paper, which shows the change of ambiguity of person with time and initial locations. The Markov chain model in the simulation

is a simplified model, for more accurate individual motion models, it is necessary to analysis large amount of trajectory informations, which will be the future work of this paper. The calculation method of "personal information privacy metric" proposed in this paper can be readily extended to other motion models to get more accurately evaluation result, and it provides a quantitative basis for the personal information anonymization in the publication of navigation data.

Acknowledgments The research work has been jointly funded by Beidou Navigation Satellite System Management Office (BDS office) and the Science and Technology Commission of Shanghai Municipality; the funding project number is BDZX005.

References

1. Gedik B, Liu L (2008) Protecting location privacy with personalized k-anonymity: architecture and algorithms. IEEE Trans Mob Comput 7(1):1–18
2. Huang L, Matsuura K, Yamane H (2005) Enhancing wireless location privacy using silent period. In: Proceedings of the IEEE wireless communications and networking conference, 2005, pp 1187–1192
3. Gabrial G, Panos K, Spiros S (2007) Prive: anonymous location-based queries in distributed mobile system. In: Proceedings of the 16th international conference on World Wide Web, 2007, pp 371–380
4. Gabrial Q, Panos K, Spiros S (2007) MobiHide: a mobile peer-to-peer system for anonymous location-based queries. In: Proceedings of the 10th international symposium on advances in spatial and temporal databases, 2007, pp 221–238
5. Lin X, Li SP, Yang ZH (2009) Attacking algorithms against continuous queries in LBS and anonymity measurement. J Softw 20(4):1058–1068. http://www.jos.org.cn/1000-9825/3428.htm