Front. Comput. Sci., 2023, 17(3): 173607
https://doi.org/10.1007/s11704-022-2090-9

LETTER

Group relational privacy protection on time-constrained point of interests

Bo NING (✉)[1], Xiaonan LI[1], Fan YANG[1], Yunhao SUN[2], Guanyu LI[1], George Y. YUAN[3]

1 Faculty of Information Science and Technology, Dalian Maritime University, Dalian 116026, China
2 Faculty of Software, Dalian University of Foreign Languages, Dalian 116044, China
3 Thinvent Digital Technology Co., Ltd., Nanchang 330029, China

1 Introduction

With the rapid development of mobile networks, location-based services has become popular in the daily lives of people. The service providers can recommend the profitable services to persons through mining the frequent interests or places of persons. However, one aspect is that the historical data on Internet can easily cause the leakage of user-relationship privacy, another aspect is that the historical interests of person are always bound to time. Therefore, this paper devotes to study a privacy protection method on time-constrained point of interests (PoIs) based on the group relationships of users.

Existing technologies provide different strategies for the privacy protection problem, that can be roughly divided into two categories. First category [1–3] is to protect the private data by replacing part of the original data by noisy data. A representative strategy of this category is generalization-based method, that integrates a series of lower-level concepts into a higher-level similar ones. Second category [4–6] is to classify and divide data based on the attributes of data, the type of service provided by third-party applications, and the type of privacy protection. The method usually divides the data into sensitive and non-sensitive data. A more detailed division can be employed in the graph-based data, including nodes, attributes and edges, etc.

In practical applications, the above two classified methods can be combined to protect the privacy of user PoIs. Our contributions are shown as following. Firstly, we mine the inter-user relationships, to extract the common PoIs of multiple users within the different time dimensions. Secondly, we propose an algorithm of relational privacy protection based on the above mined inter-user relationships to hide the PoIs of users. Finally, extensive empirical studies on real and synthetic graphs demonstrate that our techniques outperform the state-of-the-art algorithms.

2 Our model

The model of $k^{m,n}$-anonymity is proposed to protect the time-constrained PoIs of users, denoted in Definition 1.

Received February 13, 2022; accepted June 2, 2022

E-mail: ningbo@dlmu.edu.cn

Definition 1 ($k^{m,n}$-anonymity). Given a time-constrained data set $D(U,I)$ of users, a $k^{m,n}$-anonymity is used to make the number of each vulnerable PoIs in D not less than k, where m and n denote the number of common PoIs of two-user and multi-user as the prior knowledge known by attackers, respectively.

Where, a time-constrained data set $D(U,I)$ contains a set of users U and time-constrained PoIs $= \langle I,T\rangle$, and each point of interest a at t time-stamp is denoted as $\langle a,t\rangle \in \langle I,T\rangle$.

3 Privacy protection of user relationship

The privacy protection of user relationship is classified as the anonymity of single-relationship and group-relationship within the prior knowledge of attackers.

3.1 Anonymity of single-relationship

Assumed that an attacker already knows that two users contain m common PoIs, the anonymity of single-relationship is to anonymize the vulnerable PoIs, defined as k^m_{single}, such that the number of anonymized PoIs is more than k at each timestamp.

Considering a time-constrained data set of user PoIs in Table 1, that contains a set of users $U = \{u_1, u_2, u_3, u_4, u_5\}$, time $T = \{t_1, t_2, t_3, t_4, t_5, t_6\}$ and PoIs $I = \{a, b, c, d, e, f\}$. Regarding the prior knowledge $m = 3$ and $k = 2$, the single-relationship between users u_3 and u_5 is calculated as a set $k^m_{single}(u_3, u_5) = \{\langle t_1, d\rangle, \langle t_2, d\rangle, \langle t_3, c\rangle, \langle t_5, e\rangle\}$, The anonymity of $R_{single}(u_3, u_5)$ is to anonymize the common PoIs, such that the number of common PoIs is not more than m and the number of column-sorted PoIs at each timestamp is not less than k. Thus, the way of replacing the one of common PoIs d, c and e can satisfy the constraint of m (= 3) prior knowledge. The anonymity of $k^m_{single}(u_3, u_5)$ can replace d to a at timestamps t_1 and t_2 or replace c to b at timestamp t_3 in user u_3. If e is replaced to c at timestamp t_5, the number of e is 1, that is less than 2, thus the replacement of e to c at timestamp t_5 does not satisfy the constrain of k (= 2) prior knowledge.

Further, the selected PoIs are filtered based on the distance of PoIs. The distance should not exceed a threshold τ, that is used to calculate the value of pairwise PoIs by two users. The larger the value is, the easier the attackers can be perceived,

Table 1 Time-constrained data set of user PoIs

U	T					
	t_1	t_2	t_3	t_4	t_5	t_6
u_1	a	a	b	b	c	e
u_2	a	a	c	b	c	f
u_3	d	d	c	b	e	f
u_4	d	d	b	c	c	f
u_5	d	d	c	c	e	e

thus the a smaller value is employed to anonymize the vulnerable PoIs of user. Considering an assumed distance matrix of PoIs in Table 2, there exists $\tau = 10$, the distance from d to a (= 13) is more than τ (= 10) and the distance from c to b (= 3.9) is less than τ (= 10) in single-relationship $k^m_{single}(u_2, u_5)$, thus c is only replaced to b in the limits of anonymous distance in Table 2.

3.2 Anonymity of group-relationship

Assumed that an attacker already knows that multiple users contain n common PoIs, the anonymity of group-relationship is to anonymize the vulnerable PoIs, defined as k^n_{group}, such that the number of common anonymized PoIs is not less than k at each timestamp.

The anonymity of group-relationship is similar to the one of single-relationship. Considering a time-constrained data set of user PoIs in Table 1 and prior knowledge $n = k = 2$, the group relationship of users u_3, u_4 and u_5 is calculated as a set $k^n_{group}(u_3, u_4, u_5) = \{\langle t_1, d\rangle, \langle t_2, d\rangle, \langle t_3, c\rangle\}$, thus the PoI d or c of one user can be replaced by a or b to destroy the n prior knowledge of group-relationship in the limits of k prior knowledge and anonymous distance.

Not that, if n is bigger than m, the m prior knowledge is meaningless in group relationship. Since the anonymity of single-relationship should ensure that the number of common PoIs is not more than m, there dose not exist a group prior knowledge of n, such that $m < n$.

3.3 Anonymous algorithm of user relationship

The anonymous algorithm of user relationship is denoted in Algorithm 1. The inputs are the time-constrained data set of user PoIs $D(U, I)$, prior knowledge m, n, k and a distance matrix of PoIs M. The output is an anonymous data set $\mathcal{D}$ from D.

Algorithm 1 consists of two subroutines. The first subroutine is to construct the user relationships as a tree (Lines 1−5). The tree is rooted by the single-relationship R^m_{single} and takes the group-relationship R^n_{group} as node, where the branches

Table 2 Distance matrix of PoIs

Pols	a	b	c	d	e	f	g
a	0	14	10.5	13	13.2	9.7	21.9
b	14	0	3.9	1.2	0.75	4.4	13.1
c	10.5	3.9	0	2.9	3	1	15.6
d	13	1.2	2.9	0	1	3.4	13.6
e	13.2	0.75	3	1	0	3.6	13.8
f	9.7	4.4	1	3.4	3.6	0	16.1
g	21.9	13.1	15.6	13.6	13.8	16.1	0

Algorithm 1 Anonymous algorithm of user relation-ship

Input: time-constrained data set of user PoIs $D(U, I)$, prior knowledge m, n, k, distance matrix M
Output: anonymous data set $\mathcal{D}$

1 **for** $u, u' \in U$ **do**
2 $\quad R^m_{single} \leftarrow R^m_{single}(u, u')$;
3 **for** $u \in U$ *and* $r \in R^m_{single}$ **do**
4 $\quad R^n_{group} \leftarrow R^n_{group}(u, u')$;
5 $\quad Tree \leftarrow Root(r) \cup Node(u, r)$;
6 **if** $m \leqslant n$ **then**
7 $\quad$ anonymize R^m_{single} on D limited by M;
8 **else**
9 $\quad$ **for** $r \in R^m_{single}$ **do**
10 $\quad\quad$ anonymize r on D limited by M;
11 $\quad\quad$ iteratively update and anonymize nodes in $Tree(r)$ limited by M;

denote a user linking single-relationship and group-relationship (Line 5). The second subroutine is to anonymize the user PoIs in time-constrained data set D (Lines 6−11). According to Section 3.2, the m prior knowledge is meaningless in group relationship if n is bigger than m. Therefore, the user PoIs of single-relationship are only anonymized if $m \leqslant n$ (Lines 6−7), otherwise both user PoIs of single-relationship and group-relationship are anonymized (Lines 9−11). The anonymous strategy is executed in a tree, iteratively (Line 11). If one node have been anonymized, its descendant nodes are verified whether there still exists group-relationship or not. If one descendant node still satisfies the group-relationship, the node is anonymized and verifies its descendant nodes.

4 Experimental evaluation

Experiments were conducted on a machine with an Intel i3 3.80GHz CPU and 4GB memory. Smart Driving Dataset[1)] is a real user trajectory data set, which contains the trajectory data of 6,180 users, and the data of each user is stored in a single text file. Shanghai PoIs Dataset[1)] records the location preference of users in the city of shanghai, whose geographic location is in the range of latitude 120.85°−122.13° and longitude 30.67°−31.87°.

Experimental parameters: error rate of query (ERQ), denotes the error rate of querying a user PoI at a certain time; clustering rate of change (CRC), denotes a ratio between original PoI and its anonymity; clustering average rate of change (CARC), defines the average of CRC; change of attributes (CA), denotes a ratio between original PoI and its anonymity at a certain time.

The experimental evaluations of parameters on the different scales of data tables are shown in Fig. 1 and its settings are denoted in Table 3. Data sets of smart driving and shanghai PoI are divided by time and location, respectively. Figure 1 shown that the values of CARC, ERQ and ACA appear an overall upward trend as the increase of m and the unmodified number of points of interest NO has been shown a downward

[1)] See GitHub website

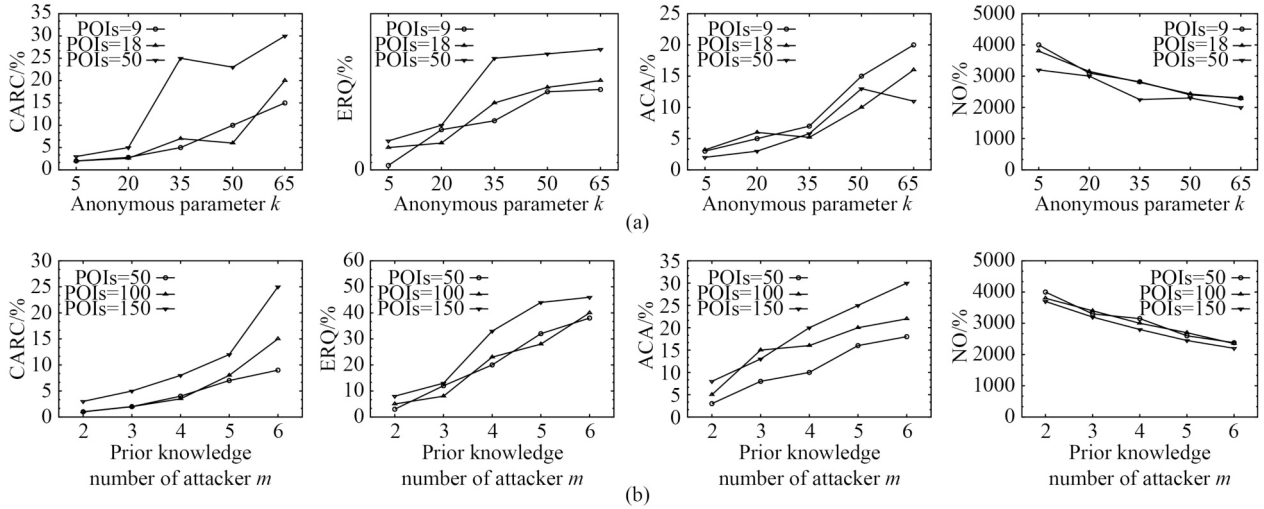

Fig. 1 Experiments of CARC, ERQ, average CA(ACA) and number of changing PoIs. (a) Dataset of smart driving; (b) dataset of Shanghai PoIs

Table 3 Parameter settings

Dataset	m	n	k	|PoIs|	Partition
Smart driving	3	2	3	{9, 18, 50}	$T = 96$
Shanghai PoI	6	4	3	{50, 100, 150}	$p = 156$

trend. This upward trend mean that the anonymous data has a worse data and query availability and decreased number of NO refers to a more laborious anonymous operations. In general, the larger the value of m is, the worse the data availability of anonymous data is.

5 Conclusions

This paper designed an algorithm to protect the privacy relationship exposed by the user PoIs. Through the experimental evaluation, our method can protect the time-constrained PoIs under the hypothetical prior knowledge of attackers.

Acknowledgements This work was supported by the National Natural Science Foundation of China (Grant Nos. 61976032 and 62002039), and the General Scientific Research Project of Liaoning (No.LJKZ0063).

References

1. Yuan Y, Wang G, Xu J Y, Chen L. Efficient distributed subgraph similarity matching. The VLDB Journal, 2015, 24(3): 369–394
2. Ge Y F, Cao J, Wang H, Chen Z, Zhang Y. Set-based adaptive distributed differential evolution for anonymity-driven database fragmentation. Data Science and Engineering, 2021, 6(4): 380–391
3. Xu J, Zhao J, Zhou R, Liu C, Zhao P, Zhao L. Predicting destinations by a deep learning based approach. IEEE Transactions on Knowledge and Data Engineering, 2021, 33(2): 651–666
4. Cai T, Li J, Mian A, Li R H, Sellis T, Yu J X. Target-aware holistic influence maximization in spatial social networks. IEEE Transactions on Knowledge and Data Engineering, 2022, 34(4): 1993–2007
5. Tong Y, Zeng Y, Ding B, Wang L, Chen L. Two-sided online micro-task assignment in spatial crowdsourcing. IEEE Transactions on Knowledge and Data Engineering, 2021, 33(5): 2295–2309
6. Xiao X, Tao Y. Anatomy: simple and effective privacy preservation. In: Proceedings of the 32nd International Conference on Very Large Data Bases. 2006, 139–150