모바일 크라우드센싱 시스템을 위한 진실 탐지 응용 동향 분석

장얀 [1], 백우호 [2], 이명 [3] 서승현 [4]
[1] 한양대학교 전자공학과 석사과정
[2] 한양대학교 전자공학과 박사과정
[3] 허난사범대학 컴퓨터정보공학과 부교수
[4] 한양대학교 ERICA 전자공학부 교수

z2021189899@hanyang.ac.kr, byh2018@hanyang.ac.kr, liming@htu.edu.cn, seosh77@hanyang.ac.kr

Survey on Truth Discovery in Mobile Crowdsensing and Its Application

Yan Zhang[1], Yuhao Bai[2], Ming Li[3], Seung-Hyun Seo[4]
[1,2]Dept. of Electronic & Electrical Engineering, Hanyang University
[3]Dept. of Computer and Information Engineering, Henan Normal University
[4]Dept. of Electrical Engineering, Hanyang University ERICA

Abstract

The mobile crowdsensing platform obtains sensing data from mobile users, and the involvement of the public increases the untrustworthy of collected data. In order to distinguish factual data from inaccurate data provided by untrustworthy users, the truth discovery method has been introduced for accurate data aggregation in mobile crowdsensing (MCS). To explore the application of truth discovery in mobile crowdsensing, we overview the general concepts of truth discovery algorithms. Finally, we summarize the main existing application prospects of truth discovery in mobile crowdsensing.

1. Introduction

In mobile crowdsensing networks, mobile users use mobile smart devices instead of sensors to collect a large amount of sensing data, which can meet more complex task requirements. Crowdsensing is a technology that uses mobile devices to collect data on people and their environment, analyzing their service and activity patterns. By mining this data, hidden information about user behavior, community structure, and service-related attributes can be revealed. This ultimately provides useful information and services to end users and is a significant departure from traditional fixed sensor networks. However, the quality and confidence of the sensing data cannot be guaranteed due to the user's subjective will, the mobility of the device, the instability of the communication environment, and the diversity of sensing tasks [1]. At present, more researchers have paid attention to the problem of anomaly detection and information extraction for a large number of multi-source sensing data in mobile crowdsensing. In [2], temporal stability and spatial correlation of data are used for outlier data detection. In addition, truth discovery methods and machine learning methods such as clustering are also often used in data aggregation and outlier detection [3]. Among them, truth discovery algorithms are potential data integration solutions that can more accurately identify real information from noisy data. In this paper, to effectively apply truth discovery algorithms to extract the truth of data, we summarize the application prospects of truth discovery algorithms in mobile crowdsensing in recent years.

2. Truth Discovery

In traditional crowdsensing, workers complete sensing tasks and upload data and are then rewarded by the task publisher. However, the publisher cannot always guarantee that the received data meets their requirements. In order to tackle the data quality problem, Yin. et al. [4] introduced the concept of truth discovery. The general principle of truth discovery is to judge the reliability of data by evaluating the credibility of users who provide data. If the user has higher credibility, the data collected by them could be regarded as relatively reliable. By reasonably evaluating the trustworthiness of users utilizing the truth discovery mechanism, the mobile crowdsensing platform could

efficiently analyze sensing data collected by the untrustful public. Thus the platform could detect outlier data, manage users' reputations, and evaluate the data quality. Here we introduce the overall flowchart of the truth discovery algorithm in mobile crowdsensing scenarios. And Table 1 gives the symbolic explanation of the truth discovery algorithm.

Table 1 Symbolic explanation in truth discovery algorithm.

Symbol	Explanation
$\mathcal{U} = \{1,2,\dots,n\}$	The set of mobile users
$\mathcal{U}_j = \{i \mid i \epsilon \mathcal{U}, \tau_j \epsilon \mathcal{T}_i\}$	The set of users sensing data for τ_j
$\mathcal{T} = \{\mathcal{T}_1, \mathcal{T}_2, \dots, \mathcal{T}_m\}$	Task set
$\mathcal{D}_i = \{(d_j^i, t_j^i) \mid \tau_j \epsilon \mathcal{T}_i\}$	Dataset accomplished by user i
$\tau_j \epsilon \mathcal{T}_i$	Task done by user i
d_j^i	Data collected by user i for task τ_j
d_j	The estimated truth for task τ_j
w_i	The weight of user i

Let us consider an MCS model consisting of a centralized platform and a group of mobile users $\mathcal{U} = \{1,2,\dots,n\}$. Firstly, a set of sensing tasks $\mathcal{T} = \{\mathcal{T}_1, \mathcal{T}_2, \dots, \mathcal{T}_m\}$ is published in platform, each sensing task can be a task in the specific sensing region, such as monitoring temperature, humidity, and PM2.5 in different areas. Once user i accomplished several tasks, the sensing data will be packed as a set $\mathcal{D}_i = \{(d_j^i, t_j^i) \mid \tau_j \epsilon \mathcal{T}_i\}$ and submit to the platform, where d_j^i is the sensing data for task τ_j, and t_j^i is the corresponding timestamp. When all the sensing data from all users have been collected, for each task $\tau_j \epsilon \mathcal{T}_i$, the platform could get an aggregated result d_j, which is the estimated truth for this task, and the user's weight will be evaluated based on it.

A general truth discovery algorithm is an iterative estimation process of user weight and truth, which consists of two phases: weight evaluation and truth estimation. Algorithm 1 gives an overview of this iterative estimation process.

Algorithm 1 Truth discovery algorithm

Input: Uploaded data $\mathcal{D} = \{\mathcal{D}_i \mid i \in \mathcal{U}\}$;
Output: Estimated truth $\{d_j \mid \tau_j \in \mathcal{T}\}$;

1. Randomly initialize the ground truth of each task;
2. **repeat**
   // User weight evaluation
3. **for each** $i \in \mathcal{U}_j$ **do**
4. Use equation (1) to update the user's weight;
5. **end**
   // Truth evaluation
6. **for each** $\tau_j \in \mathcal{T}$ **do**
7. Use the equation (2) to update the truth of the collected data;
8. **end**
9. **until** the estimated truth of collected data converges;
10. **return** estimated truth $\{d_j \mid \tau_j \in \mathcal{T}\}$。

In the beginning, for each task, a random truth for this task is guessed by the algorithm, and then the algorithm iteratively updates each user's weights and the estimated fact until it is convergence.

Let $\mathcal{U}_j = \{i \mid i \epsilon \mathcal{U}, \tau_j \epsilon \mathcal{T}_i\}$ be the set of users who participated in the task $\tau_j \epsilon \mathcal{T}_i$. For each user i, given the data collected by him and the estimated truth of the task, the weight w_i of user i is calculated as follows:

$$w_i = \mathcal{W}\left(\sum_{\tau_j \epsilon \mathcal{T}_i} \mathcal{D}(d_j^i, d_j)\right) \quad (1)$$

where $\mathcal{W}(\cdot)$ is a monotonically decreasing function, and $\mathcal{D}(\cdot)$ is a function evaluating the difference between the estimated truth d_j and user's data d_j^i.

Afterward, the estimated truth d_j for task τ_j is updated as follows:

$$d_j = \frac{\Sigma_{i \epsilon \mathcal{U}_j} w_i \, d_j^i}{\Sigma_{i \epsilon \mathcal{U}_j} w_i} \quad (2)$$

where w_i is the weight of user i, and d_j^i is the data collected by user i for task τ_j.

3. Application

Truth discovery is generally applied to information extraction, outlier data detection, data quality assessment, and user reliability assessment in mobile crowdsensing scenarios. This section analyzes the application of truth discovery in mobile crowdsensing from two aspects of security aggregation and privacy-aware truth discovery.

Security aggregation: There are various attacks in the application scenarios of mobile crowdsensing to affect the accuracy of data truth discovery. The most common one is a data poisoning attack, where attackers upload a large amount of malicious data to affect the accuracy of the final truth discovery. In order to solve the impact of such attacks on truth discovery, some algorithms can be introduced to combine with the truth discovery algorithm. In [5], the combination of a similar account grouping algorithm and truth discovery algorithm reduces the impact of malicious data on the accuracy of truth discovery in the presence of Sybil attacks. However, when the malicious data is too large or too much, it still has a significant impact on the accuracy of the truth discovery algorithm. Furthermore, [6] proposed a robust truth discovery algorithm to resist data poisoning attacks by additional source evaluation and source filtering before data aggregation. Therefore, combining truth discovery algorithms with other outlier data detection algorithms is beneficial for more accurate data aggregation in the presence of attackers.

Privacy-aware truth discovery: In addition to ensuring the accuracy of the truth discovery algorithm. In mobile crowdsensing scenarios, the privacy protection of data and ground truth is also crucial. In [7] and [8], the combination of truth discovery and encryption algorithm prevents the privacy leakage of data and ground truth. It should be noted that combining encryption algorithms and aggregating large amounts of data will affect the performance of truth discovery.

4. Conclusion

In this paper, we introduce general truth discovery algorithms and summarize recent applications of truth discovery in mobile crowdsensing in terms of security aggregation and privacy-aware truth discovery. Inspired by the summarization of our work, how to keep the accuracy and efficiency of truth discovery schemes under different threats is a challenging topic, which requires us to put more interest. In the future, we will study more concrete truth discovery approaches for energy-limited devices and complex network security environments.

5. Acknowledgments

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2023-2018-0-01417) supervised by the IITP(Institute for Information & Communications Technology Planning & Evaluation).

References

[1] Ganti R K, Ye F, Lei H. Mobile crowdsensing: current state and future challenges[J]. IEEE communications Magazine, 49(11): 32-39, 2011.

[2] Huang J, Kong L, Dai H N, et al. Blockchain-based mobile crowd sensing in industrial systems[J]. IEEE Transactions on Industrial Informatics, 16(10): 6553-6563, 2020.

[3] An J, Liang D, Gui X, et al. Crowdsensing quality control and grading evaluation based on a two-consensus blockchain[J]. IEEE Internet of Things Journal, 6(3): 4711-4718, 2018.

[4] Yin X, Han J, Yu P S. Truth discovery with multiple conflicting information providers on the web[C]. Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining. 2007: 1048-1052.

[5] Wang E, Cai J, Yang Y, et al. Trustworthy and efficient crowdsensed data trading on sharding blockchain[J]. IEEE Journal on Selected Areas in Communications, 40(12): 3547-3561, 2022.

[6] Huang Z, Pan M, Gong Y. Robust truth discovery against data poisoning in mobile crowdsensing[C]. 2019 IEEE Global Communications Conference (GLOBECOM). IEEE, 2019: 1-6.

[7] Wu H, Wang L, Cheng K, et al. Privacy-enhanced and practical truth discovery in two-server mobile crowdsensing[J]. IEEE Transactions on Network Science and Engineering, 9(3): 1740-1755, 2022.

[8] Liu Y, Liu F, Wu H T, et al. RPTD: Reliability-enhanced Privacy-preserving Truth Discovery for Mobile Crowdsensing[J]. Journal of Network and Computer Applications, 207: 103484, 2022.