International Journal of Engineering Research And Management (IJERM)
ISSN: 2349- 2058, Volume-05, Issue-10, October 2018

A Survey of Participant Selection Methods Based on Data Quality in Mobile Crowd Sensing Tasks

Kai Zhang

Abstract—**Mobile Crowd Sensing is an emerging sensing model that accomplishes a perceptual task by perceiving specific environments, collecting perceptual data, and providing information and analysis. This paper focuses on the choice of participants in the mobile crowd sensing perception task under the premise of prioritizing data quality. The task publisher publishes the perceived task recruiting worker and gives a certain reward. After the worker completes the sensing task, the worker feedbacks the sensory data to the task publisher, the quality of the feedback data is crucial for the result of the entire sensing task. This paper proposes to study the existing task allocation methods, and compares the current most advanced methods. Finally, it is found that Budget-TASC is the current method with the highest data quality, and it can control the budget within a certain range in the group intelligence perception task. While maximizing data quality, this method studies the reputation of the worker and the distance between the worker and the task position. While controlling the cost to a certain extent, the quality of the expected result is maximized.**

Index Terms—**Mobile crowd sensing; data quality; participant selection.**

I. INTRODUCTION

Smart devices (including smartphones, tablets, etc.) can not only communicate as a mobile device for everyday communication, but also because of its embedded sensors, such as acceleration sensors, digital compasses, gyroscopes, GPS, microphones, Camera, etc., and use it as a powerful sensing unit. The use of these sensors makes it possible to recruit ordinary people to mobile phones and share sensory data. "Mobile group perception" is a new research field based on the development of this emerging application [1]. Figure 1 shows the user participation in the mobile crowd sensing task flow.

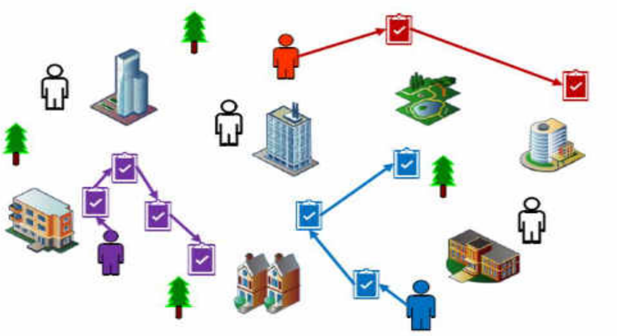

Fig 1 the user participation in the mobile crowd sensing task flow.

Manuscript received October 09, 2018
Kai Zhang, School of Computer Science & Software Engineering, Tianjin Polytechnic University, Tianjin, 300387, China

The crowd sensing network is a perceptual network composed of mobile devices (such as mobile phones) that are widely used by ordinary users and integrated with a large number of sensors [2]. Ordinary users' mobile devices (such as mobile phones, tablets, etc.) are used as basic sensing units to conduct conscious or unconscious cooperation through the Internet (such as WiFi, cellular networks, and wired networks) to achieve perceptual task distribution and sensory data collection. Large-scale, complex social perception tasks [3].

The current application of crowd sensing network is mainly focused on the perception of geospatial information. This application is mainly used to predict some terrorist events, geological disasters and special behaviors [4]. For example, the CommonSense[5] system uses a portable handheld air quality sensing device to connect to the user's mobile phone to monitor the air quality of the environment through a mobile-aware network; NoiseTube[6] and Ear-Phone[7] use the microphone of the mobile phone to measure environmental noise. And collect a large number of users' perceptual data to construct a city's environmental noise map. The completion of the crowd sensing task is inseparable from the establishment of the crowdsourcing platform. There are already many online crowdsourcing platforms that provide commercial services, such as China Mobile Online Crowdsourcing Platform, Baidu Crowdsourcing Platform, and the world famous platform with Amazon's Mechanical Turk[8].

II. THE SIGNIFICANCE OF COLLECTING DATA BY MOBILE CROWD SENSING TASKS

1) Enhance the temporal and spatial coverage of perception
Crowd sensing uses the public's own smartphone to perceive information anytime and anywhere, greatly expanding the time and space coverage of information perception. Since there is no need to manually deploy sensors, the perceived cost is low, and group intelligence overcomes the costly problem of traditional sensing methods, and will become an effective sensing mode for large-scale information sensing.
2) Extended perception perspective
Traditionally aware devices are typically deployed statically and are limited in cost and are generally not densely deployed. These features will make the perceived perspective single, the data scarcity, and the inability to perceive objects. With the mobility and wisdom of people, group intelligence can not only obtain comprehensive sensory data from multiple angles, but also artificial labeling and evaluation can greatly reduce the risk caused by the machine's misunderstanding of data.
3) Reduce the complexity of the sensing system

The traditional sensing mode requires manual deployment of a large number of sensor nodes, which is not only costly, but also increases in system complexity as the scale increases. In particular, the performance of data collection protocols (including a priori and on-demand routing protocols) in large-scale wireless sensor networks cannot meet the actual needs, mainly due to high control cost and low data transmission success rate. Opportunistic data collection uses social behavior context information (including trajectory, mobility, physical environment, interest preferences, reputation, etc.) to collect data in the form of opportunistic routing. The networking mode is more flexible and flexible, and the control load is small and extended. Strong.

4) Increase user engagement and provide a foundation for regional data convergence

Collecting data in the form of opportunistic routing does not require that the perceived data be uploaded in real time, but rather the right time to forward the data to the appropriate repeater, which means that the perceived participants do not need to upload via a paid network (eg GPRS, 4G). Data, but distribute data (such as Wi-Fi, Bluetooth) on a free close-range communication network. On the other hand, most group perception systems have incentives. This opportunistic data collection not only increases user engagement, but also enables regional data fusion through opportunity transmission to further improve data quality [9].

III. Typical Participant Selection Method

3.1 *Factors affecting data quality*

Sensing data quality is affected by many factors [10], including:

1) The type of sensor device that the user is using. For example, sensors with expensive high-end phones are generally more accurate than those with low-cost phones.

2) The environment and manner in which users collect data. For example, the quality of data collected in the hands of the mobile phone to collect environmental noise is higher than the quality of the data collected in the clothes pocket or handbag to collect environmental noise.

3) User's subjective cognitive ability. For example, an image search application based on mobile group intelligence perception relies on the user's ability to recognize images, and different users may have different perceptions of the same image.

4) User engagement attitude. For example, some users will collect data strictly according to requirements, while some users will be more casual, and even some malicious users will upload fake and forged data.

3.2 Participant selection in crowd sensing tasks

All of the above factors can cause the quality of sensing data to be uneven. Below, we first introduce the methods of selecting participants in several typical crowdsourcing tasks based on the type of perceived task or object:

3.2.1 Participant selection in traditional crowd sensing tasks

Because workers in crowdsourcing situations have different abilities and behavioral tendencies, research on finding credibility information for each worker to find efficient task distribution solutions has become a research hotspot in traditional crowdsourcing tasks. The crowdsourcing task assignment in the traditional mode, because the distance between the worker and the task location is not considered, the task assignment is not mobile and time-sensitive, and the quality of the final obtained data is also affected. In this field, the most important aspect to be studied is that the task publisher pre-determines the number of workers based on which to maximize the quality of the data that is expected to be obtained. In the article [18-24], the authors jointly considered the resource constraints and reputation of workers to arrive at a task allocation plan based on the network queuing theory. However, they did not consider the effects of spatial separation between workers and mission locations, or constraints on limited budgets.

The most representative method of traditional crowdsourcing tasks is CrowdBudget [19] proposed by L.Tran-Thanh et al. By analyzing the cost of the task and the quality of the expected results, CrowdBudget distributes the rewards that each worker should receive on average. This method determines the number of workers' work in advance before recruiting workers, and does not take into account information such as the distance between the worker and the person's location.

3.2.2 Participant selection in mobile crowd sensing tasks

Compared with the crowd sensing task assignment in the traditional mode, the allocation method of the mobile crowd sensing task has been widely studied in recent years. A more representative method is the GeoTruCrowd [20] method proposed by L. Kazemi et al. The author of GeoTruCrowd proposed a heuristic-based efficient algorithm that combines the characteristics of mobile crowd sensing tasks to achieve a task-allocation solution that is close to the optimal solution. GeoTruCrowd tends to require workers to have a minimum total distance to move solution, but they focus on worker-based volunteering (that is, workers don't expect to pay only to volunteer to complete tasks) and do not consider budgetary constraints.

IV. Current Best Participant Selection Method

In this section, we highlight the current state-of-the-art participant selection method, Budget-TASC [21].

The core of this method is to discuss the credibility of workers(Formula 1).

$$c_j^{T_i} = r_j \cdot \delta\left(l_j, l^{T_i}\right) \tag{1}$$

Among them, we use i to represent the task publisher, T_i represents the sensing task, and task publisher i publishes the perceived task T_i. T_i contains a lot of information, namely l^{T_i} , R^{T_i} , B^{T_i} , $p_H^{T_i}$, $p_M^{T_i}$. l^{T_i} represents the specific location of the perceived task. This position is specifically represented by longitude and latitude. L. Kazemi used this method in his article [11]. R_i means that the recruited worker is farthest from the task position and cannot exceed this range. B^{T_i} represents the total budget paid by the task issuer to the worker participating in the task T_i.

The fees we pay to workers of different credibility will also be different. $p_H^{T_i}$ represents the remuneration paid to workers with high reputation, and $p_M^{T_i}$ represents the remuneration paid to workers with medium reputation.

In our model, the worker's reputation value is determined by its previous performance in other crowd sensing tasks. For the sake of discussion, we divide the reputation value of the worker into three levels, which are high reputation value, medium reputation value and low reputation value. In the actual task, the task publisher can further subdivide the worker's reputation value. We use the mTurk method [8] proposed in the P.G. Ipeirotis article here, assuming that the task publisher only wants to recruit high-credit and medium-reputation workers to participate in the perceived task in order to ensure data quality.

We have divided the reputation of workers into three levels. Now we need to use Th_{HM} and Th_{ML} to divide the worker's reputation value ($0 \le Th_{ML} \le Th_{HM} \le 1$). Th_{HM} is used to divide high-reputation workers and medium-reputation workers, Th_{ML} is used to divide middle-reputation workers and low-reputation workers. At the same time, we stipulate that the credibility value of workers is from 0 to 1. The workers have a medium reputation when they have not participated in any perceived tasks. We denote worker, worker j; use r_j to represent the credibility of worker j ($0 \le r_j \le 1$).

Budget-TASC defines the issue of participant selection for mobile crowd sensing tasks as a multi-choice backpacking issue. Essentially, given a group of workers, we aim to determine the final set of workers to be as high as possible, and to ensure that the total budget is not exceeded, the sum of the credibility of the workers needs to be as high as possible.

V. CONCLUSION

The results of the three algorithms are shown in Table 1. The e represents the final average error rate, B(-) represents the average budget utilization, and D represents the average distance traveled by the worker. The smaller the values of the three indicators, the better. Budget-TASC's average error rate is 45.0% lower than GeoTru-Crowd and 81.0% lower than CrowdBudget. Budget-TASC's average budget utilization is 17.1% less than GeoTru-Crowd, and CrowdBudget can reduce its budget by 28.8%. Budget-TASC achieves comparable performance to GeoTruCrowd. Both methods are significantly better than CrowdBudget in this respect. Thus we can conclude that Budget-TASC is the most advanced method of participant selection in current mobile crowd sensing tasks.

Table 1 Performance of the three methods in terms of average error rate, average budget utilization, and worker average moving distance

Approach	$\bar{e}$	$\overline{B(-)}$	$\overline{D}$
CrowdBudget	81.1%	96.5%	13.9 km
GeoTruCrowd	28.0%	84.8%	0.60 km
Budget-TASC	15.4%	67.6%	0.58 km

In future research, we focus on improving data quality from two aspects. The first aspect, while considering the worker's reputation value and the distance between the worker and the task location, joins the current ability of the worker (ie, whether the worker currently has the necessary equipment to complete the task); the second aspect It is in the process of selecting participants, taking into account the data provided by malicious participants and solving such situations.

REFERENCES

[1] LANE N D et al. A Survey of Mobile Phone Sensing [J]. IEEE Communications Magazines, 2010, 48(9): 140-150.

[2] He Hong, Xiang Chaoshen et al. Research Status and Development of Group Intelligence Network [J]. Journal of Jilin University, 2016, 34(3): 374-383.

[3] Liu Yunhao et al. Group Intelligence Perception Computing [J]. Chinese Computer Society Newsletter, 2012, 8(10): 38-41.

[4] Zhang Zhen, Li Peng. Data-driven group-aware task allocation algorithm [J]. Computer Applied Research, 2017, 34(8): 2376-2379.

[5] Dutta P, Aoki P, Kumar N et al. Common Sense: participatory urban sensing using a network of handheld air quality monitors [C]. In Proc. of ACM SenSys, 2009: 349-350.

[6] Stevens M, D' Hondt E. Crowdsourcing of pollution data using smartphones [C]. In Workshop on Ubiquitous Crowdsourcing, 2010.

[7] Rana R, Chou C, Kanhere S, et al. Ear-phone: an end-to-end participatory urban noise mapping system [C]. In Proc. of ACM/IEEE IPSN, 2010: 105-116.

[8] P.G. Ipeirotis, Analyzing the Amazon Mechanical Turk marketplace, XRDS: Crossroads[C]. The ACM Magazine for Students, 17 (2) (2010) 16 – 21.

[9] Chen Xiang, Xu Jia et al. Research on group intelligence sensing data collection based on social behavior analysis [C]. Computer Application Research,2015, 32(12): 3534-3541

[10] Zhao Dong, Ma Hualiang et al. Measurement and guarantee of mobile group perception quality[J]. ZTE Technologies, 2015,21(6): 1-5.

[11] L. kazemi, C. Shahabi, L Chen, GeoTruCrowd: Trustworthy query answering with spatial crowdsourcing[C]. Proceedings of the 21th International Conference on Advances in Geographic Information Systems (ACM SIGSPATIAL'13), 2013. pp. 304 - 313.

[12] H. Yu, Z. Shen, C. Miao, B. An, Challenges and opportunities for trust management in crowdsourcing[C]. Proceeding of the IEEE/WIC/ACM International Conference on Intelligent Agent Technology (IAT' 12), 2012: 486 – 493.

[13] H. Yu, Z. Shen, C. Miao, B. An, A reputation-aware decision-making approach for improving the efficiency of crowdsourcing systems, Proceedings of the 12th International Conference on Autonomous Agents and Multi-Agent Systems(AAMAS' 13), 2013: 1315-1316.

[14] H. Yu, C. Miao, B. An, C. Leung, V.R. Lesser. A reputation management model for resource constrained trustee agents[C]. Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI' 13), 2013:418-424.

[15] H. Yu, C. Miao, B. An, Z. Shen, C. Leung, Reputation-aware task allocation for human trustees[C]. Proceedings of the 13th International Conference on Autonomous Agents and Multi agent Systems (AAMAS'14), 2014: 357-364.

[16] H. Yu, X. Yu, S.F. Lim, J. Lin, Z. Shen, C. Miao, A multi-agent game for studying human decision-making, Proceedings of the 13th International Conference on Autonomous Agents and Multi-agent Systems (AAMAS' 14), 2014. pp. 1661 – 1662.

[17] H. Yu, C. Miao, Z. Shen, C. Leung, Y. Chen, Q. Yang, Efficient task sub-delegation for crowdsourcing[C]. Proceedings of the 29th AAAI Conference on Artificial Intelligence (AAAI-15), 2015:1305-1311.

[18] Z. Pan, H. Yu, C. Miao, C. Leung. Efficient collaborative crowdsourcing[C]. Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI-16), 2016: 4248-4249.

[19] L. Tran-Thanh, M. Venanzi, A. Rogers, N.R. Jennings. Efficient budget allocation with accuracy guarantees for crowdsourcing classification tasks[C]. Proceedings of the 12th International Conference on Autonomous Agents and Multi-agent Systems (AAMAS'13), 2013: 901-908.

[20] L. Kazemi, C. Shahabi, GeoCrowd: Enabling query answering with spatial crowdsoucing[C]. Proceedings of the 20th International Conference on Advances in Geographic Information Systems (ACM SIGSPATIAL' 12), 2012: 189-198.

[21] Miao Chunyan, Han Yu et al. Balancing quality and budget considerations in mobile crowdsourcing[J]. Decision Support Systems, 2016, 90(2):56-64.