Data Quality Aware Task Allocation Under A Feasible Budget in Mobile Crowdsensing

Xiaohui Wei, Yongfang Wang, Shang Gao, Yao Tang
College of Computer Science & Technology, Jilin University, China
weixh@jlu.edu.cn, yongfang16@mails.jlu.edu.cn, shanggao@jlu.edu.cn, tangyao15@mails.jlu.edu.cn

***Abstract*—Satisfying spatial-temporal coverage requirement in the interested regions while considering the quality of the sensing data with budget limitation is a major research challenge in mobile crowdsensing. Most existing research in this field focus on the number of sensor readings collected in each covered subarea and do not consider individual differences of participants for contributing to data quality improvement. In this paper, we propose a novel coverage metric, quality coverage, which considers both the spatial coverage and the quality of sensing data and then use task allocation approaches to achieve highly diverse and spatial quality coverage level within a limited budget for different application scenarios.**

I. Introduction

With the explosive popularity of wireless mobile devices, a new paradigm, mobile crowdsensing (MCS) [1], has emerged. It leverages existing communication infrastructures like 3G/4G, Wi-Fi and Bluetooth networks for large-scale and fine-grained sensing without expensive infrastructure cost. The service provider in a MCS system usually integrates sensing queries of data requesters and launches diversity-based location-aware MCS tasks that require mobile workers with different attributes (*e.g.*, arrival time, moving directions, and types of hand-held devices) to travel over specific locations to complete tasks in sensing areas. In many scenarios, data requesters ask for specific quality levels of the sensing data and diverse-spatial coverage. Such requirements impose various costs, *e.g*, energy, timing constraints and the risk of privacy disclosure. To incentivize mobile workers, data requesters having limited budget for obtaining sensing data may offer financial compensations.

The main research question is how to achieve highly diverse and spatial coverage level and obtain expected quality of sensing readings simultaneously under the budget limitation. Existing research in this field using k-th sensor readings in each point of interesting (PoI) to avoid the large deviations between the ground truth and single sensing reading [2], [5]. However, they do not consider individual differences of participants for contributions to data quality improvement. In [3], [4], authors use the data aggregation mechanism and show the relationship between reliability and the quality of aggregated results. In this work we leverage this relationship to the coverage problem, and let workers' contribution to coverage levels of the same PoI be different.

We regard a crowd of participants within a time interval (or combined with other diversity requirements) as a sensing data source. Such diversity-based location-aware tasks can be covered by multiple sources in each sensing subarea. We give a new definition of coverage for MCS tasks, and study the relationship between participants' reliability and the quality of aggregation results, and then use it to quantify coverage quality. We design efficient algorithms to solve optimization problems in two different application scenarios.

II. Problem Statement

We use $L_j, j = 1, \ldots, J$ to denote the jth PoI in the sensing area, and $w_n, n = 1, \ldots, N$ to denote the nth mobile workers in the MCS system. We regard the mth data source request as a subtask $t_m, m = 1, \ldots, M$. A unit task deployed at PoI L_j for subtask t_m is denoted as t_j^m with the weight ω_j. Each worker has a reliability level R_n obtained by directly calculating the expected deviation of workers' data from the ground truths, and it can be used to compute her execution ability e_n that measures the contribution of her per sensing data to the quality improvement of the aggregation result. Knowing their identical basic incentive b_0 and the different unit reward bonus u_n, mobile workers submit an interested unit task set $\mathbb{T}_n$ to the platform, who allocates the MCS tasks. We point that workers with higher execution ability have higher unit reward bonus. After selected workers complete all the sensing tasks, they get their compensations. Fig. 1 shows this process in our MCS system.

Quality Coverage. From our study, if S_j^m denotes the set of selected worker set for unit task t_j^m, then the expected quality of the aggregation results satisfies $Q_j^m = \sum_{n \in S_j^m} e_n$. We use Q to denote the quality requirement of per sensing results. We say a unit task t_j^m is covered with desired quality if it satisfies $Q_j^m \geq Q$. Then we can present the spatial quality coverage level of the mth data source under a selected worker set S as

$$CL^m(S) = \sum_{j=1}^{J} \omega_j \times \min \left\{ \sum_{w_n \in S, t_j^m \in \mathbb{T}_n} e_n, Q^m \right\} \qquad (1)$$

Problem Formulation. In small sensing areas like the indoor scene, the service provider can attract enough participants to form high spatial coverage level, so we try to solve the max-min fair source quality coverage (Max-minFSQC) problem by maximizing the minimum of the source spatial coverage level among all data sources. In a large-scale sensing area like urban areas, we are more concerned with the maximum weighted quality coverage (MWQC) problem that maximizes the total multi-source and spatial coverage level. The above problems can be formulated as follows.

Select a mobile worker set $S \subset W$, such that

978-1-5386-2542-2/18/$31.00

Max-minFSQC: $\max \ \min_{m} CL^m(S)$

MWQC: $\max \sum_{m=1}^{M} CL^m(S)$

subject to: $\sum_{n \in S} (bo + u_n |\mathbb{T}_n|) \leq B$

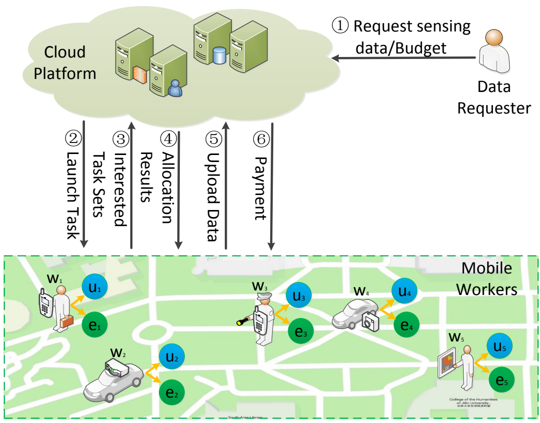

Fig. 1: Crowdsensing System Process

III. The Proposed Solution

The above two problems are NP-hard. We design algorithms, a genetic algorithm for Max-minFSQC problem and a greedy algorithm for the MWQC problem.

Genetic algorithm for Max-minFSQC problem. We use code '0' to indicate that a worker is not selected, and code '1' to indicate that the worker is selected. The length of a gene is equal to the total number of workers. The genetic algorithm is summarized as follows. 1) Generate the primary population of genes randomly; 2) Calculate the minimum source spatial coverage level of each gene in the population; 3) Perform selection, crossover, mutation operations. Different from traditional genetic algorithms, we added a penalty parameter to the crossover process by modifying the genes that exceed certain threshold ; 4) Check the termination condition: the maximum number of iteration we set has been reached or the optimal gene gets the results $\sum_{j=1}^{J} \omega_j \times Q$, otherwise repeat from step 2). Simulation studies show that the algorithm converges quickly.

Approximation algorithm for MWQC problem. We definite the coverage quality function as $QC(S) = \sum_{m=1}^{M} CL^m(S)$, and the marginal contribution per cost of a mobile worker w_n as $PMQ_n = \frac{QC(S' \bigcup w_n) - QC(S')}{c_n}$, where $c_n = b0 + u_n \times |\mathbb{T}_n|$ and S' is the set of currently selected workers. Since function $QC(S)$ is non-decreasing and submodular, we design a greedy strategy that selects workers with the maximum marginal contribution per cost in each iteration. The time complexity of the Algorithm 1 is upper bounded by $O(MJN^2)$.

We evaluate our strategy in two simulation scenarios, with $M = 4$, $N = 150$ and $J = 120$ for evaluation of genetic algorithm, and $M = 8$, $N = 2500$ and $J = 1000$ for Algorithm 1. We set $b_0 = 30$ and take the unit reward from the set $\{0.3, 0.6, 0.9, 1.2, 1.5\}$. We assume $|\mathbb{T}_n|$ ranges from 15 to 25 in the two scenarios. We compare our algorithms with k-th coverage strategies, Maxmin-KC and MKC with $k = 5$, and set the average value of e_n to be one fifth of Q. Results (Fig. 2) show that our two algorithms achieves significantly better weighted average quality of aggregation sensing results than existing approaches for all budget limits.

Algorithm 1 Approximation Algorithm for MWQC Problem

Require: W, $\mathbb{T}$, B, Q
Ensure: A set of selected mobile workers S

$S \leftarrow \emptyset$; $curentB \leftarrow B$; $W' \leftarrow W$
while $W' \neq \emptyset$ **do**
  $i \leftarrow \arg\max_{n, w_n \in W'} \frac{QC(S \bigcup w_n) - QC(S)}{c_n}$
  if $curentB - c_i \geq 0$ AND $PMQ_i > 0$ **then**
    $S \leftarrow S \bigcup w_i$; $curentB \leftarrow curentB - c_i$
  end if
  $W' \leftarrow W' \setminus w_i$
end while
return S

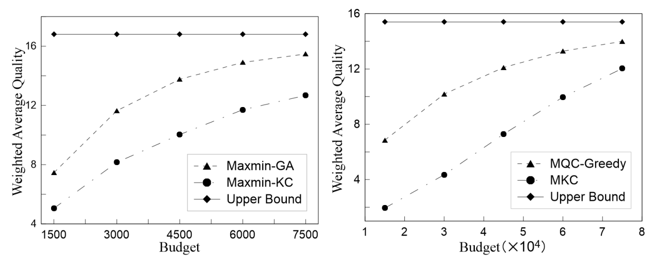

Fig. 2: Budget vs. Weighted Average Quality

IV. Conclusion

This paper designed and evaluated a new strategy to recruit workers with high execution ability to achieve high coverage level on both the multiple data sources and the sensing areas. By giving workers with high execution ability more incentive bonus, our approach can attract more reliable workers and ensure relatively high quality of sensing readings. We studied two algorithms to solve the optimization problems for different scenarios. Simulation results show that our algorithms outperform existing approaches.

Acknowledgment

This work is supported by National Natural Science Foundation of China (NSFC) (Grants No.61772228), National Key research and Development Program of China (Grant 2017YFC1502306, 2016YFB0701101), Jilin Scientific and Technological Development Program (20170520066JH).

References

[1] Ganti R K, Ye F, Lei H. Mobile crowdsensing: current state and future challenges. IEEE Communications Magazine, 2011, 49(11):32-39.
[2] Xiong H, Zhang D, et al. iCrowd: Near-Optimal Task Allocation for Piggyback Crowdsensing. IEEE TMC, 2016, 15(8):2010-2022.
[3] Jin H, Su L, Nahrstedt K. CENTURION: Incentivizing Multi-Requester Mobile Crowd Sensing. 2017.
[4] Jin H, Su L, Xiao H, et al. INCEPTION:incentivizing privacy-preserving data aggregation for mobile crowd sensing systems. ACM MobiHoc, 2016:341-350.
[5] Miao X, Liu K, Chen L, et al. Quality-Aware Online Task Assignment in Mobile Crowdsourcing. In IEEE ICMAHSC, 2016:127-135.