



QITA: Quality Inference Based Task Assignment in Mobile Crowdsensing

Chenlin Liu, Xiaofeng Gao^(✉), Fan Wu, and Guihai Chen

Shanghai Key Laboratory of Scalable Computing and Systems,
Department of Computer Science and Engineering, Shanghai Jiao Tong University,
Shanghai 200240, China

ChanningLiu@sjtu.edu.cn, {gao-xf, fwu, gchen}@cs.sjtu.edu.cn

Abstract. With the rapid proliferation of mobile devices, Mobile Crowdsensing (MCS) has become an efficient way to ubiquitously sense and share environment data. Due to the openness of MCS, sensors and workers are of different qualities. Low quality sensors and workers may yield low sensing quality. Thus it is important to infer workers' qualities and seek a valid task assignment with enough total qualities for MCS. To solve the quality inference problem, we adopt *truth inference* methods to iteratively infer workers' qualities. This paper also proposes an task assignment problem called *quality-bounded task assignment with redundancy constraint* (QTAR) based on truth inference. We prove that QTAR is NP-complete and propose a $(2 + \epsilon)$ - approximation algorithm QTA for QTAR. Finally, experiments conducted on real dataset prove the efficiency and effectiveness of the algorithms.

Keywords: Mobile Crowdsensing · Task assignment · Truth inference

1 Introduction

In recent years, the rapid proliferation of mobile devices has changed people's lives. *Mobile Crowdsensing* (MCS) [2] has thus become an efficient way to ubiquitously sense and share environment data with mobile devices. One of the significant advantage of MCS comparing with traditional sensor networks comes from its active involvement of workers to collect and share sensing data. For a typical MCS, one of the most important problem is how to appropriately assign tasks to workers, which has recently been widely studied by researchers [4, 6, 7].

Due to the openness of MCS, sensors and workers are of different qualities. Low quality sensors and workers may yield low sensing quality. Low quality devices may collect noisy, even inaccurate sensing data. Moreover, low quality workers may randomly collect data ignoring the location constraints in order to deceive payment. To overcome the quality inference problem, this paper introduce a method called *truth inference* to simultaneously infer the workers' qualities and the truth of each task. Most existing truth inference methods are redundancy-based, which means that one task is assigned to multiple workers.

By guaranteeing a certain amount of redundancy for each tasks, this paper investigates the quality-aware task assignment problem with budget constraint and propose quality bounded task assignment with redundancy constraint (QTAR). In QTAR, the total quality of selected workers exceeds the quality bound while the overall cost is minimized in the task assignment process. Different from tradition task assignment problem, we add redundancy constraint to satisfy the preliminaries of truth inference, which requires that each task should be assigned a certain or more amount of workers.

Therefore, our method is divided into two steps. Firstly, a truth inference method is adopted to infer workers' qualities based on previously aggregated sensing data. Next, we solve QTAR with the inferred workers' qualities in step 1. After workers finish their assigned tasks, the sensing data are aggregated and prepared to be used in the following quality inference step. This method, which is called Quality Inference Based Task Assignment (QITA), is an effective way to improve the overall sensing quality of MCS.

This paper proves that QTAR problem is NP-complete and propose a $(2 + \epsilon)$ approximation algorithm QTA for QTAR. Finally, we evaluate our algorithms by conducting a series of experiments on both synthesis data and real dataset.

2 Related Work

In this paper, we focus on minimizing monetary cost for task assignment while satisfying the quality constraint. In order to infer workers' qualities under the lack of real data, we adopt truth inference methods [1, 9], which has been widely studied in existing crowdsourcing works. Based on workers' answers, truth inference methods iteratively infer the truth and workers' qualities. Different from online crowdsourcing, in MCS, the aggregated sensing data are numeric data such as air quality and WiFi signal strength. A few researches studied truth inference methods for numeric tasks. [5] assumed that the answers of workers follows certain biases and variance and the paper adopted EM algorithm to iteratively infer the truth and workers' qualities. Moreover, [3] considered the confidence of each worker and proposed a confidence-aware truth discovery method to infer truth by considering the confidence interval of the inference.

3 Problem Formulation

We consider an MCS consists of a cloud server, multiple sensing tasks and multiple workers with mobile devices. When a worker is assigned a sensing task, her/he moves to the specified location to sense data. Based on the moving distance, cloud server allocates monetary reward to the worker. Let C_{ij} denote the monetary cost for assigning worker j to task i .

Suppose that there are m sensing tasks and n workers. Let $T = \{t_1, t_2, \dots, t_m\}$ denote the set of tasks and $W = \{w_1, w_2, \dots, w_n\}$ be the set of workers. The quality evaluation of worker j is represented by $q_j \in \mathbb{R}^+$, the larger q_j is, the more accurate the collected data of worker j will be. Let

$Q = \{q_1, q_2, \dots, q_n\}$ denote the set of workers' qualities. We will further discuss how to compute Q in Sect. 4. In order to satisfy the redundancy requirements R of truth inference methods, for each task, the number of assigned workers is no less than R .

Since we prefer to assign tasks to workers with higher qualities, this paper aims to find a subset of worker set W with one task assigned to each worker, such that the total worker quality is no less than the quality bound Q_B and the overall cost is minimized. For each task, the number of assigned workers is greater or equal to the redundancy requirement R . This problem is defined as *quality bounded task assignment with redundancy constraint* (QTAR).

To solve the problem, we propose a method called *Quality Inference Based Task Assignment* (QITA). There are two processes in QITA as follows:

Quality Inference: Based on the previously aggregated sensing data, QITA adopts truth inference method to iteratively infer both the truth of tasks and workers' qualities. Workers who submit sensing data that are close to the truth will be assigned higher qualities.

Task Assignment: With worker set W , Task set T and quality set Q , QITA selects a subset $S \in W$ as the selected worker set, then assigns one task for each worker $w \in S$. After the workers finish their tasks, the aggregated sensing data, which is denoted as D , will be used for next quality inference process.

Figure 1 shows the workflow of QITA.

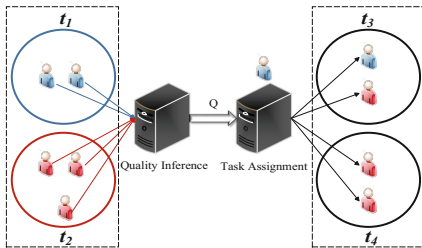


Fig. 1. The Workflow of QITA

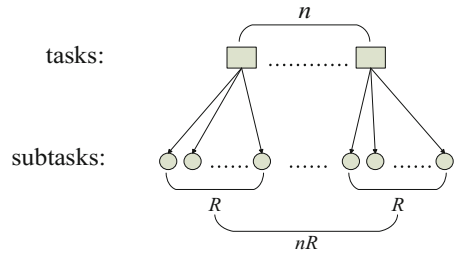


Fig. 2. An example of tasks dividing into subtasks

4 Quality Inference

Considering that the sensing data in MCS are mostly numeric numbers (e.g., WiFi signal strength, air quality index, etc.), we adopt quality inference algorithm based on truth inference method called LFC_N [5]. In this section we introduce the modeling and quality inference algorithm of this method.

For the i^{th} task $t_i \in T$, let $S_i = \{w_1^i, w_2^i, \dots, w_{m_i}^i\}$ denote the worker set assigned to task t_i . Let $d_j^i = \{d_{j1}^i, d_{j2}^i, \dots, d_{j\lambda}^i\}$ be the aggregated sensing data of worker w_j^i for task t_i . Assume that d^i is the truth (real sensing data) of task

t_i , our model is that the worker provides a noisy version of the actual true value d^i . For worker w_j^i we assume a Gaussian with mean d^i (the truth) and inverse-variance (quality) q_j , that is, $\Pr(d_j^i | d^i, q_j) = \mathcal{N}(d_j^i | d^i, 1/q_j)$. To obtain the exact methods of truth inference step and quality inference step, the LFC_N method adopts a *Maximum-Likelihood Estimator* as follows:

Assume the tasks are independent, for task t_i , we suppose that a worker need to sense for λ times of data to complete this task. Let $\mathbf{d}^i = \{y_1^i, y_2^i, \dots, y_\lambda^i\}$ denote the inferred truth for task t_i each time the worker senses the data. $\mathbf{q} = \{q_1, q_2, \dots, q_m\}$ denotes the inferred quality of each worker. The likelihood of the parameter $\theta = (d^i, \mathbf{q})$ given the observations \mathcal{D}_i can be factored as $\Pr(\mathcal{D}_i | \theta) = \Pr(d_1^i, d_2^i, \dots, d_{m_i}^i | \theta)$. By maximizing the log-likelihood, we obtain the update equation for the inferred quality and inferred truth

$$\frac{1}{\hat{q}_j} = \frac{\sum_{k=1}^{\lambda} (d_{jk}^i - \hat{y}_k^i)^2}{\lambda} \quad \hat{y}_k^i = \frac{\sum_{j=1}^{m_i} \hat{q}_j d_{jk}^i}{\sum_{j=1}^{m_i} \hat{q}_j} \tag{1}$$

Since the two parameters \mathbf{q} and \mathbf{d}^i are coupled together, by using the equations in Eq. (1), we can iteratively infer the qualities and truth until convergence. Therefore we can reasonably infer the workers' qualities by adopting the truth inference based method. The inferred quality set will be used as an input of the task assignment process of QITA.

5 Quality Bounded Minimum Cost Task Assignment

5.1 QTA: An Approximation Algorithm for QTAR

The goal of our assignment is to minimize the overall cost within the quality bound and the redundancy constraint, which is formally formulated as QTAR problem in Sect. 2. Let the redundancy constraint $R = 0$ and suppose the number of workers $m = 1$, the problem can be reduced to a 0-1 knapsack problem and thus QTAR is NP-complete.

To satisfy the redundancy constraint, we divide each task t_i into R subtasks. As is shown in Fig. 2, the subtasks have the same locations with the tasks they are born from. Completing the original tasks with redundancy constraint is equivalent to assigning more than one workers to these subtasks.

This paper proposes a $(2 + \epsilon)$ -approximation algorithm called QTA to solve QTAR. This algorithm successively solves two related problems, then combines the results of both problems as the final result.

Minimum Worker Matching (MWM): Assign exactly one worker to each task and the total cost is minimized.

We formulate this problem as a *Minimum Weighted Complete Matching* problem. Let task set T and worker set W be the disjoint sets of nodes in a bipartite graph. The edge set E where each edge has one endpoint in each of T and W denotes the assignment between workers and tasks; let $B = (T, W, E)$ denotes

such a bipartite graph. If edge $(t_i, w_j) \in E$, then task t_i is assigned to worker w_j . We assume that the weight of each edge $(t_i, w_j) \in E$ is the cost C_{ij} .

By adopting a negative weight $\tilde{C}_{ij} = -C_{ij}$ for each edge, we convert the problem into a *Maximum Weighted Complete Matching* problem, which can be solved in polynomial time by the Hungarian method (also been known as the Kuhn-Munkres algorithm or Munkres assignment algorithm).

Quality Bounded Minimum Assignment (QBMA): An assignment satisfying that the total qualities of workers exceeds the quality bound and the overall cost is minimized. Each worker can only be assigned one task.

QBMA can be reduced from 0-1 knapsack problem when $m = 1$. Due to the NP-completeness of QBMA, this paper proposes a dynamic programming based FPTAS $(1 + \epsilon)$ - approximation algorithm by scaling the cost down enough such that the costs of all assignments are polynomially bounded in n . Let C_{max} denote the maximum cost among workers and tasks, the algorithm is as follows:

Algorithm 1. FPTAS Approximation Algorithm for QBMA

- 1 Given $\epsilon > 0$, let $K = \frac{\epsilon C_{max}}{n}$;
 - 2 For each possible assignment a_{ij} , define $C'_{ij} = \lceil \frac{C_{ij}}{K} \rceil$;
 - 3 Using the dynamic programming to find the minimum cost assignment S' with these as the costs of assignments ;
 - 4 Output S' ;
-

Suppose that a_{ij} represents an assignment between worker w_j and task t_i , then we introduce our approximation algorithm QTA for QTAR.

Algorithm 2. QTA: An Approximation Algorithm for QTAR

Input: task set T , worker set W , quality set Q and quality bound Q_B

Output: assignment set S

- 1 Divide the QTAR into an MWM and a QBMA ;
 - 2 Solve the corresponding MWM and QBMA, the output is A_1 and A_2 ;
 - 3 Initially $S = \emptyset$;
 - 4 **foreach** $1 \leq j \leq n$ **do**
 - 5 **foreach** $1 \leq i \leq mR$ **do**
 - 6 **if** $a_{ij} \in A_1$ **then** $S \leftarrow a_{ij}$; **break** ;
 - 7 **else if** $a_{ij} \in A_2$ **then** $S \leftarrow a_{ij}$;
 - 8 **return** S ;
-

In QTA, we divide the QTAR into an MWM and a QBMA. For the MWM, we construct a bipartite graph $B = (T, W, E_1)$. Let T be the task set of mR subtasks and W denote the worker set. The cost C_{ij} denotes the edge weight between t_i and w_j . For the QBMA, we find an assignment for task set T and worker set W . Different from QTAR, each worker can only be assigned one task while each task can be assigned to multiple or zero workers.

5.2 Algorithm Analysis

We define C_1 , C_2 and C_S as the total cost of solutions of MWM, QBMA and QTAR respectively. Then we define OPT_2 and OPT_S as the total cost of optimal solutions of QBMA and QTAR. We claim that QTA is a $(2 + \epsilon)$ -approximation. To prove it, we first prove the following two lemmas.

We define C_1 , C_2 and C_S as the total cost of solutions of MWM, QBMA and QTAR respectively. Then we define OPT_2 and OPT_S as the total cost of optimal solutions of QBMA and QTAR.

Theorem 1. *QTA is a $(2 + \epsilon)$ -approximation algorithm.*

Proof. In MWM, we find a complete match for tasks with minimum total cost C_1 . Assume each task is only assigned to one worker in QTAR, the result of QTAR is at most as good as MWM, which means the total cost of QTAR can not be smaller than MWM. Thus $C_1 \leq OPT_S$.

Comparing with QTAR, we ignore the redundancy constraint in QBMA, thus the result of QTAR is at most as good as QBMA. Then we have $C_2 \leq (1 + \epsilon)OPT_2 \leq (1 + \epsilon)OPT_S$.

In the combining step of QTA, we discard some assignment to ensure that one worker is assigned exactly one task. The total cost decreases in this step, which means $C_S \leq C_1 + C_2$. Finally, we have

$$C_S \leq C_1 + C_2 \leq OPT_S + (1 + \epsilon)OPT_S = (2 + \epsilon)OPT_S,$$

which completes the proof.

6 Experiments

In this Section, we evaluate the performance of QITA by conducting a series of experiment on both synthesis data and real dataset. From the results of our experiments, the effectiveness and efficiency of QITA is proved.

6.1 Quality Inference Simulation Experiments

For each worker, we randomly generate Gaussian distribution data as the sensing data. Suppose the truth is d and the quality of worker w_j is q_j , then the generated data D_j satisfies the Gaussian distribution with d as the mean and $1/q_j$ as the variance. Initially, we have three kinds of workers: (1) bad workers with $q = 1$; (2) normal workers with $q = 10$; (3) good workers with $q = 100$. For totally 30 workers, workers with ID (1–10) are bad workers, the next 10 workers are normal workers and the last 10 workers are good workers. The three kinds of workers should be clearly distinguished in the result of quality inference.

As is shown in the following figures, for all workers, the average error between inferred qualities and real qualities is less than 10%. Therefore, we can draw the conclusion that the quality inference is accurate and effective (Figs. 3, 4 and 5).

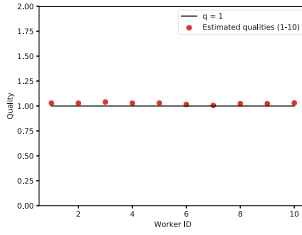


Fig. 3. Worker (1–10)

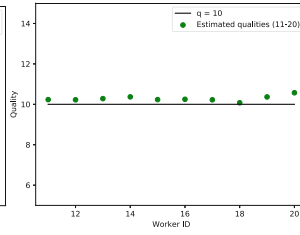


Fig. 4. Worker (11–20)

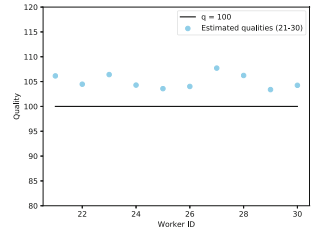


Fig. 5. Worker (21–30)

6.2 Task Assignment Experiments

We evaluate our task assignment algorithms by conducting experiments on real dataset TSMC2014 [8]. This dataset includes long-term (about 10 months) check-in data in New York city and contains 227,428 check-ins. We randomly select check-ins of different users as our worker set. The qualities of workers are generated by Gaussian Distribution with different means. Unless specified otherwise, we suppose that the positions of tasks are randomly placed in the sensing area. For each set of experiment parameter choices, we run the experiment 100 times with randomized qualities for workers and locations for tasks then task the average as the result. Figure 6 provides the evaluation results.

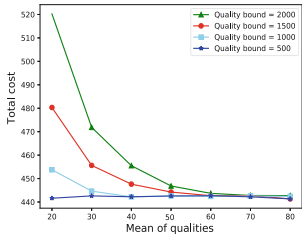


Fig. 6. Performance of QTA

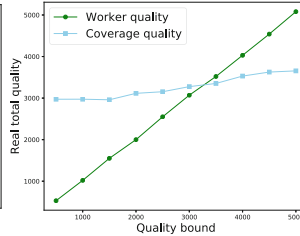


Fig. 7. Quality comparison

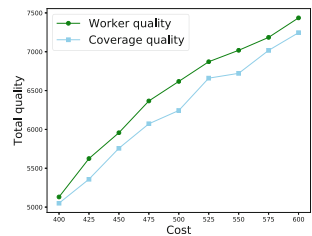


Fig. 8. Cost comparison

The total cost decreases with the increase of quality means. Higher mean of qualities means higher average value of qualities, which makes it easier for QTA to reach the quality bound. Moreover, with the decrease of quality bound, the total cost also declines due to the decreasing demand for workers.

This paper evaluates QTA by implementing a benchmark: considering the coverage quality. The coverage quality has been considered as an optimization goal to improve sensing quality in many previous researches [4, 7], which is defined as the number of sensor readings in each MCS sensing cycle. The larger amount of sensor readings in MCS, the higher the coverage quality.

Figure 7 shows the comparison result of QTA and benchmark under different quality bounds. The total quality of benchmark's result becomes lower than QTA with the increase of quality bound. The comparison result under different total cost is illustrated in Fig. 8. The total quality of QTA's result is larger than benchmark when the amounts of total cost are close for both algorithms, which proves the effectiveness and efficiency of QTA.

7 Conclusion

Motivated by the truth that low quality and malicious workers may yield low sensing quality in MCS, we studied quality-aware redundancy-based task assignment problem in MCS. We first adopted truth inference methods to iteratively infer the truth and qualities based on the aggregated sensing data. By taking the quality inference result as an input, we proposed *quality bounded task assignment with redundancy constraint* (QTAR). We proved that QTAR is NP-complete and proposed $(2+\epsilon)$ -approximation algorithm QTA for QTAR. By conducting experiment on both synthesis data and real dataset, we compared the performance between our algorithms and benchmarks. Experiment results showed that our algorithms is efficient and effective.

Acknowledgement. This work is supported by the program of International S&T Cooperation (2016YFE0100300), the China 973 project (2014CB340303), the National Natural Science Foundation of China (Grant number 61472252, 61672353), the Shanghai Science and Technology Fund (Grant number 17510740200), and CCF-Tencent Open Research Fund (RAGR20170114).

References

1. Dong, X.L., Berti-Équille, L., Srivastava, D.: Integrating conflicting data: the role of source dependence. *PVLDB* **2**(1), 550–561 (2009)
2. Ganti, R.K., Ye, F., Lei, H.: Mobile crowdsensing: current state and future challenges. *IEEE Commun. Mag.* **49**(11), 32–39 (2011)
3. Li, Q., et al.: A confidence-aware approach for truth discovery on long-tail data. *PVLDB* **8**(4), 425–436 (2014)
4. Philipp, D., Stachowiak, J., Alt, P., Dürr, F., Rothermel, K.: Drops: model-driven optimization for public sensing systems. In: *PerCom*, pp. 185–192 (2013)
5. Raykar, V.C., et al.: Learning from crowds. *JMLR* **11**, 1297–1322 (2010)
6. Wu, S., Gao, X., Wu, F., Chen, G.: A constant-factor approximation for bounded task allocation problem in crowdsourcing. In: *GLOBECOM*, Singapore, pp. 1–6, 4–8 December 2017
7. Xiong, H., Zhang, D., Chen, G., Wang, L., Gauthier, V.: Crowdtasker: maximizing coverage quality in piggyback crowdsensing under budget constraint. In: *PerCom*, pp. 55–62 (2015)
8. Yang, D., Zhang, D., Zheng, V.W., Yu, Z.: Modeling user activity preference by leveraging user spatial temporal characteristics in lbsns. *IEEE Trans. Syst. Man Cybern.: Syst.* **45**(1), 129–142 (2015)
9. Zheng, Y., Li, G., Li, Y., Shan, C., Cheng, R.: Truth inference in crowdsourcing: is the problem solved? *PVLDB* **10**(5), 541–552 (2017)