

Specialty-Aware Task Assignment in Spatial Crowdsourcing

Tianshu Song, Feng Zhu, Ke Xu

SKLSDE Lab, School of Computer Science and Engineering
Beihang University, Beijing 100191, China

Abstract. With the rapid development of Mobile Internet, spatial crowdsourcing is gaining more and more attention from both academia and industry. In spatial crowdsourcing, spatial tasks are sent to workers based on their locations. A wide kind of tasks in spatial crowdsourcing are specialty-aware, which are complex and need to be completed by workers with different skills collaboratively. Existing studies on specialty-aware spatial crowdsourcing assume that each worker has a united charge when performing different tasks, no matter how many skills of her/him are used to complete the task, which is not fair and practical. In this paper, we study the problem of specialty-aware task assignment in spatial crowdsourcing, where each worker has fine-grained charge for each of their skills, and the goal is to maximize the total number of completed tasks based on tasks' budget and requirements on particular skills. The problem is proven to be NP-hard. Thus, we propose two efficient heuristics to solve the problem. Experiments on both synthetic and real datasets demonstrate the effectiveness and efficiency of our solutions.

1 Introduction

With the development mobile Internet and the blossom of sharing economy, all kinds of *spatial crowdsourcing (SC)* platforms become popular, where the *online crowd workers* are employed by their phones to participate in and complete *offline crowdsourcing tasks* in the physical world [9]. Typical SC platforms includes Gigwalk¹, TaskRabbit² and gMission³ [2].

One fundamental issue in SC is task assignment, namely assigning crowdsourcing tasks to suitable crowd workers. Generally speaking, there are two kinds of tasks. The first kind is micro tasks which can be completed by any single worker such as taking photos and delivering things. The second kind is specialty-aware tasks such as repairing a house and organizing a party, where crowd workers with different kinds of skills are needed to work collaboratively and finish the task. For micro-task assignment, there are many existing works and we refer the readers to [14] for more details. In this paper we focus on specialty-aware tasks assignment.

¹ www.gigwalk.com

² www.taskrabbit.com

³ gmission.github.io

Table 1: Tasks and their lists of skills

Tasks	Lists of required skills
t_1	$s_1(\text{music}), s_2(\text{drinks})$
t_2	$s_1(\text{music}), s_3(\text{barbecue}), s_4(\text{lights})$
t_3	$s_1(\text{music}), s_2(\text{drinks}), s_3(\text{barbecue}), s_4(\text{lights}), s_5(\text{stage})$

Table 2: Workers' skills and fees

Workers	Skills and fees
w_1	$(s_1, 3), (s_2, 4), (s_4, 5)$
w_2	$(s_3, 5), (s_5, 3)$
w_3	$(s_4, 2)$
w_4	$(s_1, 5), (s_5, 1)$
w_5	$(s_1, 2), (s_2, 2), (s_3, 3), (s_4, 6)$

Existing works [3,4] on specialty-aware tasks assignment formulate that each crowd worker has multiple skills and will get a united fee if s/he is employed, which is not very practical as (1) workers often have unbalanced workloads, (2) workers may be confused what they should do in a task and (3) the payment and the workload do not often match. To solve the above drawbacks, in this paper we propose the Specialty-Aware Task Assignment (SATA) problem where each crowd worker specify a fee for each of her/his skill to make the payment proportional to the workload.

We then illustrate the STAT problem by a motivation example of organizing a party.

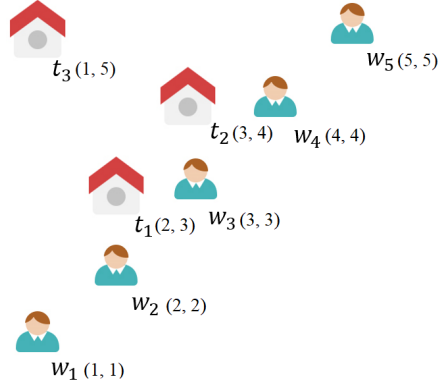


Fig. 1: Locations of tasks and workers.

Example 1. Suppose we have three tasks of throwing parties, each has different styles and thus different kinds of works need to be done. For example, party 1 is a mini one and only needs music and drinks, while party 3 is ceremonious and requires music, drinks, barbecue, lights and a stage. The skill lists of the three tasks are shown in Table 1. Besides, we have some workers shown in Table 2, each with different skills and corresponding fees. For example, if w_1 is required to finish the music job (a_1), s/he will be paid 3. Besides, each worker will get the transportation fee, which equals the distance from the worker to the assignment task times a global unit price. For example, Figure 1 shows the locations of tasks and workers, and if the global unit price is 0.5, the transportation fees for assigning w_1 to t_1 is $\sqrt{5} \times 0.5 \approx 1.12$.

Motivated by the example above, we will formalize the STAT problem, which aims to efficiently assign crowd workers to specialty-aware tasks to maximize the total utility of the assignment. Note that existing works either focus on assigning workers to micro tasks to optimize different goals, or assume that the workers have a united fee. Thus, their methods cannot be directed adopted to solve our problem.

In this paper, we first prove the NP-hardness of the SATA problem, indicating that SATA is not tractable and it is challenging to gain the optimal solution. Therefore, we propose two efficient and effective heuristics to solve it.

To summarize, we make the following contributions.

- We formally define a new task assignment problem in spatial crowdsourcing, called the Specialty-Aware Task Assignment (SATA) problem.
- We prove the SATA problem is NP-hard, and develop two efficient heuristics to solve it.
- We verify the effectiveness and efficiency of the proposed methods with extensive experiments on real and synthetic datasets.

The rest of the paper is organized as follows. We define the SATA problem and prove its NP-hardness in Section 2. Section 3 discusses extensive experiment results on both synthetic and real datasets. We review related works in Section 4 and conclude in Section 5.

2 Problem Definition

We first introduce two basic concepts, namely Task and Worker. Then, we introduce how to calculate the reward of worker. Finally, we formally give the definition of the Specialty-Aware Task Assignment (SATA) problem.

Definition 1 (Worker). A worker w is defined as $\langle L_w, S_w, P_w \rangle$, where L_w is the location of w which can be described by longitude and latitude, $S_w = \langle s_1^w, s_2^w, \dots, s_{|S_w|}^w \rangle$ is the list of skills that w masters, and $P_w = \langle p_1^w, p_2^w, \dots, p_{|S_w|}^w \rangle$ is the list of fees for each skill in S_w .

Similar to the definition of a worker, a task is formally defined as follows.

Definition 2 (Task). A task t is defined as $\langle L_t, S_t, B_t \rangle$, where L_t is the location of t which can be described by longitude and latitude, $S_t = \langle s_1^t, s_2^t, \dots, s_{|S_t|}^t \rangle$ is the list of skills that are needed to complete t collaboratively, and B_t is the total monetary budget of t .

Briefly, a worker's reward includes two parts: (1) transportation fee, which is directly proportional to the distance between the worker and the task; (2) labor fee, which is the sum of the fees for the skills used to perform a task.

Definition 3 (Reward of Worker). The reward of task w to perform task t equals $r_w = \gamma \cdot \text{dis}(L_w, L_t) + \sum_{s \in S'_w} p_s^w$, where $\text{dis}(L_w, L_t)$ is the distance between L_w and L_t , which can be Euclidean distance or road network distance, γ is a global parameter representing the unit transportation fee, and S'_w is the set of skills that w uses to perform the task.

We define the utility of a task as follows.

Definition 4 (Utility of Task). The utility of task t is defined as $u_t = B_t - \sum_{w \in W_t} r_w$, where B_t is the budget of the task and $\sum_{w \in W_t} r_w$ is the summation of rewards of workers assigned to t if t is completed. If t cannot be finished, the utility is zero.

We finally define our problem as follows.

Definition 5 (Specialty-Aware Task Assignment (SATA) Problem). Given a set of tasks T , a set of workers W and a global unit transportation fee γ , the problem is to assign workers to tasks to maximize the total utility of the completed tasks and the following constraints should be satisfied:

- **Specialty Constraint:** a task can be completed as long as the workers assigned to it can cover the required skills of the task;
- **Budget Constraint:** the total rewards of workers assigned to a task cannot exceed the task's total budget;

We then prove the hardness of SATA problem.

Theorem 1. The SATA problem is NP-hard.

Proof. We prove through a reduction from the set cover problem [20]

We first introduce the set cover problem. Given a universe $U = \{s_1, s_2, \dots, s_n\}$ and its m subsets $S_1, S_2, \dots, S_m \subseteq U$, $\cup_{i=1}^m S_i = U$. Each S_i is associated with a cost c_i . The set cover problem is to find a set $K \subseteq \{1, 2, \dots, m\}$ to minimize $\sum_{i \in K} c_i$ satisfying $\cup_{i \in K} S_i = U$.

We next show how to transform the set cover problem to an instance of our SATA problem. We only have one task t which requires skills $S_t = U$ and has infinite budget B_t . For m workers $\{w_1, w_2, \dots, w_m\}$, their required fees for skills are all zero, and we adjust their locations and γ to make their transportation fee to perform t is c_i . For this instance of our SATA problem, we aim to find a set of workers K to maximize the utility of t , which equals to minimize $\sum_{i \in K} c_i$. In this way, we reduce set cover problem to our SATA problem. As the set cover problem is known to be NP-hard [20], SATA problem is also NP-hard.

Algorithm 1: Total Budget Based Algorithm (TBA)

input : set of workers W , set of tasks T
output: Assignment M

- 1 $Q \leftarrow$ sorting tasks in T according to their budgets in descending order;
- 2 **foreach** t in Q **do**
- 3 Assign $w \in W$ to t with minimum $\frac{r_w}{|S'_w \cap S_t|}$;
- 4 Update M and W ;
- 5 $S_t \leftarrow S_t - S'_w$;
- 6 **if** S_t is \emptyset **then**
- 7 Break;
- 8 **return** M

3 Algorithms

In this section, we give two efficient heuristic algorithms to solve the SATA problem.

3.1 Total Budget Based Algorithm

Our first algorithm is called the Total Budget Based Algorithm (TBA). The main idea is that we always try to assign workers to the tasks with the largest budget. During the procedure of task assignment, we refer to the greedy algorithm to solve the set cover problem [20].

The procedure of TBA is shown in algorithm-1. The algorithm takes the set of workers W and set of tasks T as input, and return an assignment M between them as shown in lines 1-2. In line 3, the algorithm first sorts the tasks in T in descending order according to their budgets, and the sorted result is saved in Q . In lines 4-13, for each task t in Q , we refer to the greedy algorithm to solve the set cover problem [20] to assign workers. Specifically, in lines 5, we find worker with minimum $\frac{r_w}{|S'_w \cap S_t|}$. Notes that here S'_w considers all possible subsets of S_w . In lines 6-7, we update M , W and S_t . In lines 9-11, if S_t is \emptyset , which means it can be completed, we break the loop and start to assign workers for the next task.

Example 2. Back to our running example in Example 1. TBA first finds the task with the largest total budget, which is t_3 . Then it starts to assign workers for t_3 . As w_3 has the minimal $\frac{r_w}{|S'_w \cap S_t|}$ of 2, we first assign w_3 to t_3 . After assigning w_3 , t_3 's list of skills has not been covered, thus we assign w_5 to t_3 with $\frac{r_w}{|S'_w \cap S_t|}$ of $\frac{7}{3}$. We finally assign w_4 to t_3 and the total reward paid to w_3 , w_4 and w_5 is $2 + 2 + 2 + 3 + 1 + (\sqrt{2} + 4 + \sqrt{10}) \times 0.5 \approx 15$. Thus, the utility of t_3 is $30 - 15 = 15$. Similarly, we assign workers to t_2 and t_3 successively, and the final utility of TBA is 21.08.

Complexity. If we take the maximum number of skills a worker may have as a constant, the time complexity of TBA is $O(\max\{|T|\log|T|, |T||W|\})$.

Algorithm 2: Average Budget Based Algorithm (ABA)

input : set of workers W , set of tasks T
output: Assignment M

- 1 $Q \leftarrow$ sorting tasks in T according to their budgets in descending order;
- 2 **foreach** t in Q **do**
- 3 Assign $w \in W$ to t with minimum $\frac{r_w}{|S'_w \cap S_t|}$;
- 4 Update M and W ;
- 5 $S_t \leftarrow S_t - S'_w$;
- 6 **if** S_t is \emptyset **then**
- 7 Break;
- 8 **return** M

3.2 Average Budget Based Algorithm

The TBA algorithms only considers the total budget of tasks. However, a large budget may result from a large number of skills required in the task. Thus, in this subsection, we propose another algorithm, called Average Budget Based Algorithm (ABA). The main idea is that we first measure the average budget of all the tasks, and prefer to assign workers to tasks with a larger average budget.

The pseudo codes of ABA is shown in Algorithm-2. The biggest difference between TBA and ABA lies on line 3. In TBA, we first sort tasks in T based on average budget, which is defined as $\frac{B_t}{|S_t|}$. The procedure of how to assign workers to a given task is the same as TBA, which is shown in lines 4-13.

Example 3. Back to our running example in Example 1. Different from TBA, ABA first finds the task with the largest average budget, which is t_1 . Then it assigns workers to t_1 . As w_5 has the minimal $\frac{r_w}{|S'_w \cap S_t|}$ of 2, we first assign w_5 to t_1 . After assigning w_5 , we find t_1 's list of skills has been covered, thus the total utility is $20 - (2 + 2) - (\sqrt{13} \times 0.5 \approx 14.20)$. Similarly, we next assign workers for t_2 and t_3 , and the final utility of TBA is 25.78.

Complexity. If we take the maximum number of skills a worker may have as a constant, the time complexity of ABA is also $O(\max\{|T|\log|T|, |T||W|\})$.

4 Evaluation

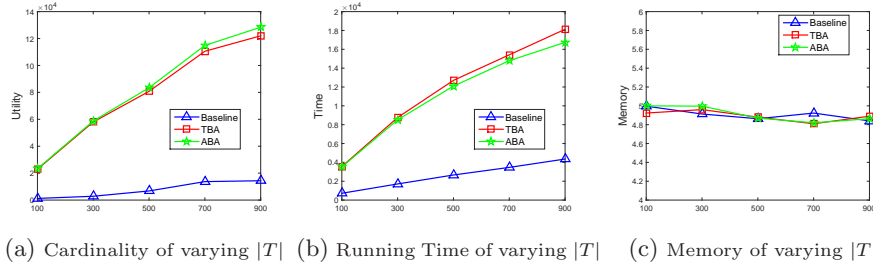
4.1 Experiment Setup

We use real and synthetic datasets to evaluate our algorithms. Real data comes from CSTO (<http://www.csto.com/>), which is an outsource task platform. In the CSTO dataset, each task is associated with a set of skills needed to complete the software development task, and each coder is associated with a set of skills and an average price which can be deduced from the history data. Since the CSTO data is not associated with location information, we generate the distance of each

Table 3: Synthetic Dataset

Factor	Setting					
$ T $	100	300	500	700	900	
$ W $	1000	3000	5000	7000	9000	
γ	0.1	0.3	0.5	0.7	0.9	
B_t	60	80	100	120	140	
P_w	10	15	20	25	30	
$ S $	10	20	30	40	50	

coder from the task following uniform distribution. For synthetic data, based on the observation from real data set, the price of skills owned by a worker and the budget of a task follow Gaussian distribution, respectively. Statistics of the synthetic data are shown in Table3, where we mark our default settings in bold font.

Fig. 2: Results on varying $|T|$.

4.2 Experiment Results

In this subsection, we test the performance of our proposed algorithms by setting different parameters. We evaluate two exact algorithm, called TBA and ABA, and a baseline algorithm in terms of total utility score, running time and memory cost, and study the effect of varying parameters on the performance of the algorithms. The baseline algorithm uses a simple random strategy, which assigns workers to tasks randomly. The algorithms are implemented in CodeBlocks16.1, and the experiments are performed on a machine with Intel(R) Core(TM) i5 2.50GHZ CPU and 8GB main memory.

Effects of the number of tasks $|T|$. The results of varying $|T|$ are presented in Fig.2a to 2c. First, we can observe that the utility increases as $|T|$ increases, which is reasonable as more tasks available. Also, we can observe that TBA algorithm and ABA algorithm are much better than baseline algorithm

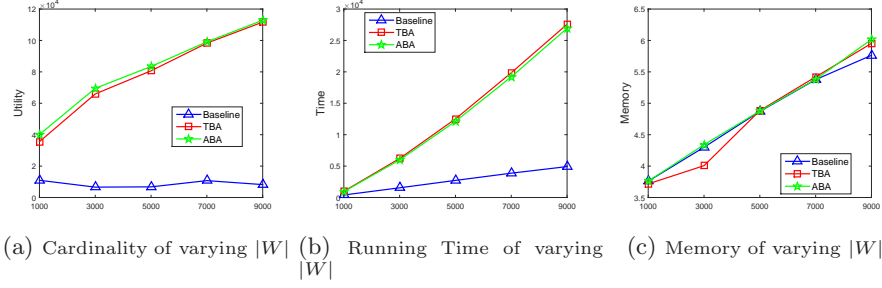


Fig. 3: Results on varying $|W|$.

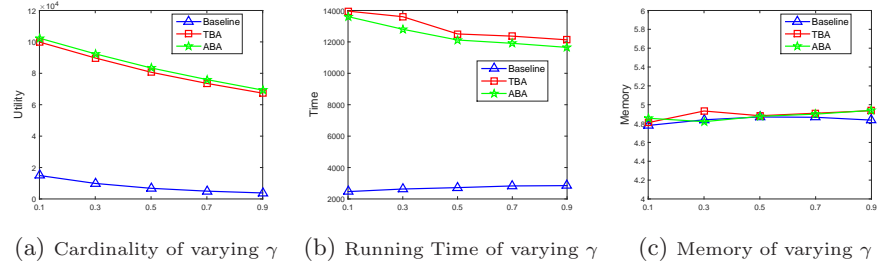


Fig. 4: Results on varying γ .

and TBA algorithm has advantages over ABA algorithm. As for running time, TBA and ABA are slower than the baseline due to sorting tasks and finding more economic schedule, and the running time is acceptable for better performance on utility. Moreover, TBA is faster than ABA for it is easier to find suitable workers for each tasks. The three algorithm do not vary much in memory consumption.

Effects of the number of workers $|W|$. The results of varying $|W|$ are presented in Fig.3a to 3c. We can observe that the utility, running time and memory consumption generally increase as $|W|$ increase, which is reasonable as more workers need to be assigned. Again, we can see that TBA are better than ABA in terms of Utility and running time.

Effects of the global unit transportation fee γ . The results of varying γ are presented in Fig.4a to 4c. We can see that the utility and running time decrease as the γ increases for higher transportation fee and less workers that could be assigned to far tasks.

Effects of the average budget of tasks B_t . The results are presented in Fig.5a to 5c. We can first see from the figure that the utility increases as the average budget increases. And there is no large differences of the running time and memory consumption between various B_t .

Effects of the variance of the price of different skills P_w . The results are presented in Fig.6a to 6c. We can see from the figures that TBA algorithm

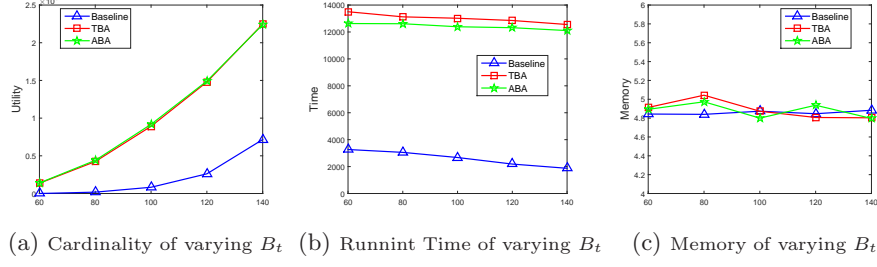


Fig. 5: Results on varying B_t .

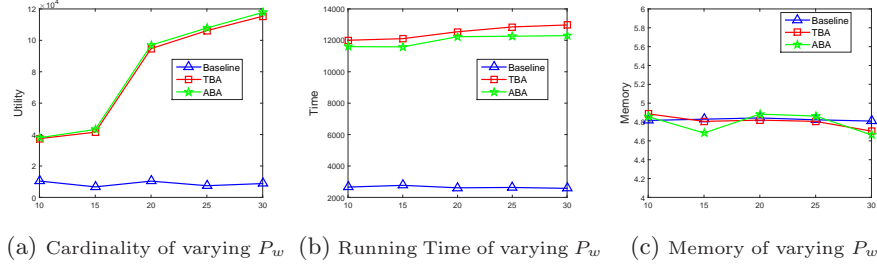


Fig. 6: Results on varying P_w .

and ABA algorithm have much better performance than baseline algorithm as the price increases. And the running time and memory consumption do not vary too much in different price.

Effects of the total number of skills $|S|$. The results are presented in Fig. 7a to 7c. First, we can observe that the utility and memory consumption do not change greatly as the number increases. Then, we can see that the running time increases as the number increases, and this is reasonable because it is much harder to find suitable workers to finish the task for more kinds of skills.

Real Dataset. The results on real dataset are shown in Fig. 8a to 9c, where we vary $|T|$ and price. We can observe similar patterns as those in Fig. 2a to 2c and Fig. 4a to 4c.

Conclusion. For Utility, TBA is better than ABA and baseline algorithm, and both TBA and ABA algorithm have a much better performance than baseline algorithm. As for running time, baseline algorithm is fastest, but the speed of TBA and ABA algorithm is acceptable for most circumstances. Moreover, TBA algorithm is faster than ABA algorithm.

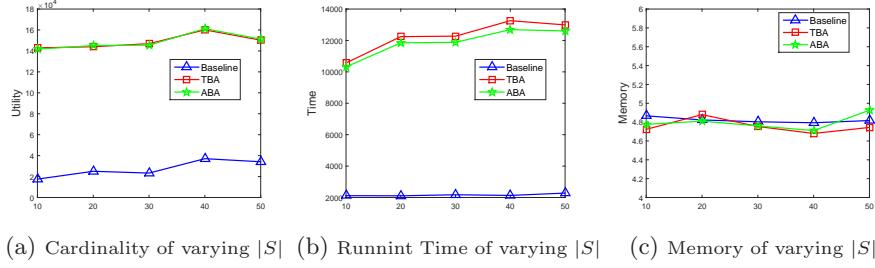


Fig. 7: Results on varying $|S|$.

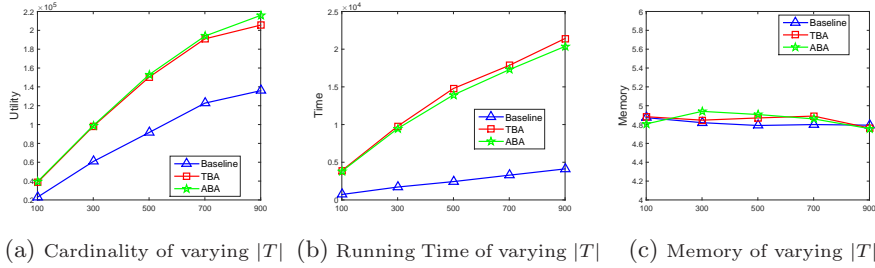


Fig. 8: Results of Real Dataset on varying $|T|$.

5 Related Work

In this section, we review related works from two categories, namely task assignment and team formation problem.

5.1 Task Assignment in Spatial Crowdsourcing

The research on task assignment in spatial crowdsourcing mainly includes two parts: micro-task assignment and specialty-aware task assignment.

Micro task refers to the spatial tasks that can be completed by any single worker. [6] is the first work on task assignment in spatial crowdsourcing, whose optimization objective is to maximize the total number of the assignment tasks. [16] is the first work focusing on the online scenario of task assignment, and studies the two-sided online task assignment problem, whose goal is to maximize the total utility score of the assignment. [10] also focuses on the online scenario and considers the influence of work space on task assignment, whose goal is to maximize the total utility score. [15] studies the problem of online minimum weighted bipartite matching, which can be used in online task assignment. [18] considers the problem of flexible online matching where workers can be scheduled if no task is assigned. [11] recommends routes dynamically for workers to deal with online tasks, and the goal is to maximize the total utility. [21] assigns tasks to

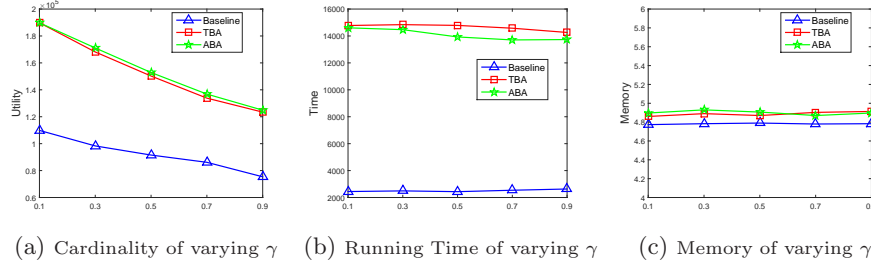


Fig. 9: Results of Real Dataset on varying γ .

workers while trading off quality and latency of task completion. [17] proposes a match-based approach to solve the dynamic pricing problem in spatial crowdsourcing. [22] takes the destinations of workers into consideration to perform task assignment. [13] considers performing online task assignment while preserving the privacy of tasks and workers under the circumstance that the server is untrusted. [12,19] proposes a real-time framework for task assignment. The difference between our work and the aforementioned works is that they focus on micro tasks which can be completed by a single worker, and we study on the assignment for specialty-aware tasks which have requirements on skills of workers and usually have to be completed by multiple workers collaboratively.

[4,5] recommend top-k teams with the minimum cost to a specialty-aware task. [3] studies assigning workers for specialty-aware tasks to maximize the total utility score. The difference between our work and [3] is that in our work workers specify fees for each of their skills, and in [3] workers only have a united fee, which is not practical.

5.2 Team Formation Problem

A closely related topic is the team formation problem [7], whose goal is to find a team of experts with the minimum cost, according to the skills and social relationships of the users. [1] studies the online version of the team formation problem, where the issue of workload balance is also considered. [8] studies another variant of the team formation problem where the capacity constraint of experts is considered. The difference between our problem and the team formation problem and its variants is that we do not consider the social relationships between users and focus on task assignment.

6 Conclusion

In this paper we study the problem of Specialty-Aware Task Assignment (SATA) in spatial crowdsourcing, where the tasks have requirements on skills, and the workers specify fees for each of their skills. The goal is to maximize the total

utility of the task assignment between tasks and workers. We prove the SATA problem is NP-hard. To solve the problem, we propose two efficient and effective heuristic algorithms. We conduct extensive experiments on both synthetic and real-world datasets to evaluate our algorithms. The experiment results show that our solutions are efficient and effective.

References

1. Anagnostopoulos, A., Becchetti, L., Castillo, C., Gionis, A., Leonardi, S.: Online team formation in social networks. In: WWW 2012. pp. 839–848
2. Chen, Z., Fu, R., Zhao, Z., Liu, Z., Xia, L., Chen, L., Cheng, P., Cao, C.C., Tong, Y., Zhang, C.J.: gmission: A general spatial crowdsourcing platform. PVLDB 7(14), 1629–1632 (2014)
3. Cheng, P., Lian, X., Chen, L., Han, J., Zhao, J.: Task assignment on multi-skill oriented spatial crowdsourcing. TKDE 28(8), 2201–2215 (2016)
4. Gao, D., Tong, Y., She, J., Song, T., Chen, L., Xu, K.: Top-k team recommendation in spatial crowdsourcing. In: WAIM 2016. pp. 191–204
5. Gao, D., Tong, Y., She, J., Song, T., Chen, L., Xu, K.: Top-k team recommendation and its variants in spatial crowdsourcing. DSE 2(2), 136–150 (2017)
6. Kazemi, L., Shahabi, C.: Geocrowd: enabling query answering with spatial crowdsourcing. In: GIS 2012. pp. 189–198
7. Lappas, T., Liu, K., Terzi, E.: Finding a team of experts in social networks. In: SIGKDD 2009. pp. 467–476
8. Majumder, A., Datta, S., Naidu, K.: Capacitated team formation problem on social networks. In: SIGKDD 2012. pp. 1005–1013
9. Musthag, M., Ganesan, D.: Labor dynamics in a mobile micro-task market. In: CHI 2013
10. Song, T., Tong, Y., Wang, L., She, J., Yao, B., Chen, L., Xu, K.: Trichromatic online matching in real-time spatial crowdsourcing. In: ICDE 2017. pp. 1009–1020
11. Tao, Q., Zeng, Y., Zhou, Z., Tong, Y., Chen, L., Xu, K.: Multi-worker-aware task planning in real-time spatial crowdsourcing. In: DASFAA 2018
12. To, H., Fan, L., Tran, L., Shahabi, C.: Real-time task assignment in hyperlocal spatial crowdsourcing under budget constraints. In: PerCom 2016. pp. 1–8
13. To, H., Shahabi, C., Xiong, L.: Privacy-preserving online task assignment in spatial crowdsourcing with untrusted server. In: ICDE 2018
14. Tong, Y., Chen, L., Shahabi, C.: Spatial crowdsourcing: Challenges, techniques, and applications. PVLDB 10(12), 1988–1991 (2017)
15. Tong, Y., She, J., Ding, B., Chen, L., Wo, T., Xu, K.: Online minimum matching in real-time spatial data: Experiments and analysis. vol. 9, pp. 1053–1064 (2016)
16. Tong, Y., She, J., Ding, B., Wang, L., Chen, L.: Online mobile micro-task allocation in spatial crowdsourcing. In: ICDE 2016. pp. 49–60
17. Tong, Y., Wang, L., Zhou, Z., Chen, L., Du, B., Ye, J.: Dynamic pricing in spatial crowdsourcing: A matching-based approach. In: SIGMOD 2018
18. Tong, Y., Wang, L., Zhou, Z., Ding, B., Chen, L., Ye, J., Xu, K.: Flexible online task assignment in real-time spatial data. PVLDB 10(11), 1334–1345 (2017)
19. Tran, L., To, H., Fan, L., Shahabi, C.: A real-time framework for task assignment in hyperlocal spatial crowdsourcing. TIST 9(3), 37 (2018)
20. Vazirani, V.V.: Approximation Algorithms. Springer Science & Business Media (2013)

21. Zeng, Y., Tong, Y., Chen, L., Zhou, Z.: Latency-oriented task completion via spatial crowdsourcing. In: ICDE 2018
22. andYang Li andYu Wang andHan Su andKai Zheng, Y.Z.: Destination-aware task assignment in spatial crowdsourcing. In: CIKM 2017. pp. 297–306