Multi-skill collaboration-based task assignment in spatial crowdsourcing

Zhejun Liang[a], Wenan Tan*[a,b], Jiu Liu[a], Kai Ding[a]

[a]College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu, China; [b]College of Computer and Information Engineering, Shanghai Polytechnic University, Shanghai, China

ABSTRACT

In recent years, with the rapid development of mobile terminal devices and the Internet, spatial crowdsourcing has received widespread attention. The spatial crowdsourcing problem is characterized by the location information contained in the attributes of workers and tasks, the crowdsourcing platform can assign reasonable spatial tasks to workers based on their current location, and the execution of tasks will be accompanied by dynamic changes in their physical location of the workers, so task assignment is an important research content of spatial crowdsourcing problem, and the quality of task assignment methods can affect the development of its crowdsourcing platform. For the spatial crowdsourcing problem that requires a group of workers with relevant professional skills to work together to complete special application scenarios (e.g., performance-type tasks), a cost-based greedy approach is proposed to minimize platform costs by matching a suitable team of workers for spatial tasks under the constraints of workers and tasks. Extensive experiments have been conducted on synthetic data sets to demonstrate the effectiveness and efficiency of the proposed approach.

Keywords: Spatial crowdsourcing, task assignment, multi-skill, worker collaboration

1. INTRODUCTION

In recent years, with the widespread use of smartphones with geographic information sensors, wireless and mobile networks, traditional crowdsourcing has undergone tremendous changes. A new type of crowdsourcing, the emergence of spatial crowdsourcing, has attracted widespread attention from academia and industry. The spatial crowdsourcing system is composed of task requesters, crowdsourcing platforms (such as Amazon Mechanical Turk), and workers. The task requester publishes tasks with geographic location and time requirements to the crowdsourcing platform, and the crowdsourcing platform makes reasonable assignments based on the location of workers and tasks. The worker needs to move to the designated position of the task to complete it, and the worker will be paid accordingly after the task is completed. The completion of crowdsourcing tasks is inseparable from the good collaboration between the task requester, the crowdsourcing platform, and the workers.

Task assignment is one of the key research issues in spatial crowdsourcing. The quality of the task assignment method not only determines the quantity and quality of tasks completed but also determines the interests of the crowdsourcing platform and workers. Spatial tasks can be classified into simple and complex tasks. For simple tasks (e.g., street view of Google Maps) can be performed by a single worker alone with the help of equipment, but for complex tasks (e.g., interior decoration), multiple workers with specialized skills are required to collaborate. From the perspective of professional skills, the skills of the workers assigned by the platform should include one or more of the skills required for the task, so that the selected workers can participate in the task, which can effectively avoid the waste of worker resources. At the same time, the skill set of the assigned worker should include all the skills required by the task, to ensure the completion of the task. Secondly, the intimacy between workers is one of the important factors that affect whether workers can collaborate well. High intimacy will make them communicate and collaborate smoothly, and good communication can achieve the effect of one plus one more than two, reducing team communication costs while completing the assigned tasks with high quality.

The previous research on task assignment only considered the collaboration between workers[1-2] or the assignment of skills[3-4] alone and did not consider the collaborative assignment of workers based on multiple skills. Therefore, we

*wtan@foxmail.com

cannot use the existing methods to solve the problem of multi-skill cooperative assignment. To solve the above proposed problem, this paper proposes a cost-based greedy approach. Under the constraints related to workers and tasks are satisfied, suitable teams of workers are matched for spatial tasks to ensure high-quality completion of tasks while minimizing platform costs.

2. RELATED WORK

The main job of crowdsourcing platforms is to assign tasks to the right workers, so task assignment has been one of the research hotspots of spatial crowdsourcing. From the existing studies, the task assignment methods of spatial crowdsourcing can be divided into two types, worker selected tasks (WST) and server assigned tasks (SAT). Although these two assignment approaches address different problem scenarios, most research works have maximized the number of tasks assigned[5-6], maximizing the system revenue[3,7], or minimizing the platform cost[8] as the main optimization objectives. For WST, reference[9] allows workers to browse and choose which tasks to participate in according to their wishes. Reference[10] finds a time scheduling table for workers given a set of workers and tasks such that the number of completed tasks is maximized. This way of assigning tasks can protect workers' privacy to a certain extent while making it possible to improve workers' motivation, but it can result in certain tasks in remote locations not being selected by workers for a long time.

For SAT, reference[6] proposed a task assignment method based on time preference. With the aid of historical data and context matrix, it calculates the preferences of different types of workers in different periods. Give priority to workers who are more interested in tasks to maximize the number of tasks assigned; A task assignment approach based on platform profit is proposed in reference[7], which first establishes a task pricing model and then uses a tree decomposition algorithm for task assignment to maximize system revenue. Reference[11] is the first to consider task location as a factor that affects task assignment. It studies the online task assignment of three types of objects in a spatial crowdsourcing environment and proposes a random-threshold-based algorithm for task assignment.

In addition, if the task assigned is complex and difficult, multiple workers with specialized skills are required to work together to complete it. A multi-skill-oriented task assignment approach is proposed in reference[3], where workers who satisfy time and spatial constraints will be assigned to spatial tasks with time constraints using greedy algorithms, partitioning algorithms, and adaptive algorithms based on cost models, making the flexible budget maximized, i.e., maximizing the number of tasks completed while minimizing the cost of worker movement. The study only considers the multi-skills of workers and ignores the collaborative relationship between workers. Reference[2] proposed a heuristic task framework FCC-SA, which uses feedback from workers and requesters to accurately assess worker skills and intimacy, and then allocate workers to collaborative tasks to maximize the number of tasks assigned. The study considers the collaboration between workers, but for the skills of workers, only single skills rather than multiple skills are considered. Reference[1] only considers the collaborative relationship between workers when assigning tasks, ignoring the skill requirements of workers.

3. PROBLEM DEFINITION

Definition 1 (skills set): Let $D = \{d_1, d_2, \cdots, d_n\}$ is a skill set that contains different k skills, and skills can be painting walls, repairing roofs, replacing pipes, etc. Each worker has at least one skill, while multiple skills are required for the completion of each task.

Definition 2 (multi-skilled workers set): Let $W_\varphi = \{w_1, w_2, \cdots, w_n\}$ is workers set that contains n multi-skilled workers at timestamp φ. For each worker $w_i \in W_\varphi$, has a skills set $D_i (\subseteq D)$, is located at a location $l_i(\varphi)$ at timestamp φ, can move with speed v_i, has a maximum moving distance r_i, and has the intimacy a_{ik} with worker w_k.

The platform will maintain the intimacy matrix. Firstly, the platform can initialize the intimacy matrix according to the social relationship α between workers, α is calculated using simple social attributes, such as region, age, gender, and more complex features, and normalized to a decimal of 0 to 1. Then, as the number of task assignment rounds increases, we can dynamically update the intimacy matrix, as shown in equation (1).

$$a_{ik} = \alpha + (1 - \alpha) \cdot |T_i \cap T_j| \cdot (|T_i| \cdot |T_j|)^{-1/2} \tag{1}$$

where $|T_k|$ represents the number of tasks completed by worker w_i, $|T_k|$ represents the number of tasks completed by worker w_k, and $|T_i \cap T_j|$ represents the number of tasks completed by worker w_i and worker w_k.

The intimacy between workers determines the smoothness of communication between workers. Therefore, this paper converts intimacy into the cost of communication. The higher the intimacy, the lower the cost of communication. The cost of communication is calculated as shown in equation (2).

$$I_{(w_j)} = \begin{cases} 0 & |W_j| = 0,1 \\ K_I \cdot \left(1 - \sum_{w_i \in W_j} \sum_{w_k \in W_j \wedge k \neq i} a_{ik} \cdot \left(|W_j| \cdot (|W_j| - 1)\right)^{-1}\right) & |W_j| > 1 \end{cases} \tag{2}$$

where W_j represents the set of workers assigned to the task t_j, and K_I is a constant and represents the hourly call subsidy.

Definition 3 (Complex spatial tasks): Let $T_\varphi = \{t_1, t_2, \cdots, t_m\}$ is tasks set that that contains m complex spatial tasks at timestamp φ. For each task $t_j \in T_\varphi$, needs a skill set for its completion, is located at a location l_j at timestamp φ, and has a deadline e_j and a budget B_j.

Definition 4 (valid worker-task pairs): At the timestamp φ, given workers set W_φ and tasks set T_φ, a valid worker-task pair $\langle w_i, t_j \rangle$. is constructed by worker w_i and task t_j while satisfying constraints on distance, time, skill, travel cost, etc.

Constraint 1 (distance constraint): This paper uses Euclidean distance as distance function. The distance between the current location $l_i(\varphi)$ of worker w_i and the location l_j of task t_j is not greater than the maximum moving distance r_i of worker w_i, that is $dist(l_i(\varphi), l_j) \leq r_i$.

Constraint 2 (time constraint)：The worker w_i need to arrive at the task location l_j before deadline e_j to perform the task, that is $dist(l_i(\varphi), l_j) \cdot v^{-1} \leq e_j - \varphi$.

Constraint 3 (travel cost constraint): The worker w_i need to physically move to the location l_j of task t_j and the travel cost c_{ij} of worker w_i is not greater than the budget B_j of task t_j, that is $c_{ij} = K_c \cdot dist(l_i, l_j) \leq B_j$, where K_C is a constant, representing the subsidy per kilometer.

Constraint 4 (skills constraint): The assigned worker w_i needs to provide the required skills for the task t_j that is $D_i \cap D_j \neq \emptyset$.

An assignment instance I refers to a set of valid worker-task pairs.

Definition 5 (MSC-CS problem): At timestamp φ, given workers set W_φ and tasks set T_φ, The MSC-CS problem is to assign the currently available workers to the spatial tasks and minimize platform total cost $S(T)$:

$$S(T) = \sum_{t_j \in T} S(W_j) = \sum_{t_j \in T} C(W_j) + I(W_j) \tag{3}$$

$$C(W_j) = \sum_{w_i \in W_j} c_{ij} \tag{4}$$

where $C(W_j)$ represents the travel cost of task t_j and $I(W_j)$ is the communication cost of the task t_j defined by equation (2).

4. TASK ASSIGNMENT METHOD

4.1 The total cost increase

Firstly, this paper defines the increase of the total cost. At timestamp φ, when a new worker w_i is assigned to the task t_j, the increase of total cost is calculated as shown in equation (5).

$$\Delta S_\varphi = S(W_j) - S(W_j - \{w_i\}) = \left(c_{ij} + \Delta I(W_j)\right) \cdot \left(|D_i \cap (D_j - D_j')|\right)^{-1} \tag{5}$$

$$\Delta I(W_j) = I(W_j) - I(W_j - \{w_i\}) \tag{6}$$

where W_j is the set of workers assigned to the task t_j (including w_i), and $\Delta I(W_j)$ is the change in communication cost of the task t_j after worker w_i is assigned.

4.2 Task priority calculation

If an assignment conflict is encountered during task assignment, this paper adopts the strategy of assigning the higher priority first. The priority is calculated as shown in equation (7).

$$p_j = \beta \cdot exp\left(-\left|D_j\right| \cdot B_j^{-1}\right) + (1 - \beta) \cdot exp\left(-(e_i - \varphi) \cdot \left|D_j\right|^{-1}\right) \tag{7}$$

where β represents the weight of task benefit and task urgency.

4.3 Cost-based greedy approach

In this section, we propose a cost-based greedy approach, as shown in Algorithm 1.

Algorithm 1 Cost-based greedy approach

Input: A timestamp φ

Output: An assignment instance I_φ

1: **initialize**: $I_\varphi \leftarrow \emptyset$;

2: retrieve all available workers and tasks then add to W_φ and T_φ

3: compute all valid worker-and-task pairs $\langle w_i, t_j \rangle$ from W_φ and T_φ

4: **while** $W_\varphi \neq \emptyset$ and $T_\varphi \neq \emptyset$ **do**

5:     obtain the best pairs $\langle w_i, t_j \rangle$ with the lowest total cost increase $\Delta S\varphi$

6:     **if** obtain more than one best pair **then**

7:         select the pair $\langle w_i, t_j \rangle$ with the highest task priority

8:         $I_\varphi \leftarrow I_\varphi \cup \{\langle w_i, t_j \rangle\}$

9:         $W_\varphi \leftarrow W_\varphi - \{w_i\}$

10:     **else**

11:         $I_\varphi \leftarrow I_\varphi \cup \{\langle w_i, t_j \rangle\}$

12:         $W_\varphi \leftarrow W_\varphi - \{w_i\}$

13:     **end if**

14:     **if** $D_j \subseteq \bigcup_{w_i \in W_j} D_i$ **then**

15:         $T_\varphi \leftarrow T_\varphi - \{t_j\}$

16:     **end if**

17: **end while**

18: **return** I_φ;

At timestamp φ, this paper first retrieves all available workers and tasks (lines 2), then for each worker w_i a range query is executed using spatial indexing methods (e.g., R-Tree) to determine which tasks are reachable by the worker w_i. The time constraint, travel cost constraint, and skill constraint are further used to determine which tasks that worker w_i can participate in, while constructing a valid worker-task pair about worker w_i. Finally, all valid worker-task pairs at the current timestamp φ can be obtained (line 3). Lines 4 to 18 are iteratively select the worker-task pair $\langle w_i, t_j \rangle$ with the smallest increase in total cost, if only one pair is found, add it directly to the assignment instance I_φ. If several pairs are found, select pairs with the highest priority of conflicting tasks, add them to the assignment instance I_φ, and then

remove worker w_i from the current workers set W_φ(lines 5 to 13). If the skills set of the workers set W_j and the skills set required by the task t_j satisfy $D_j \subseteq \bigcup_{w_i \in W_j} D_i$, then task t_j can be removed from the current tasks set T_φ(lines 14 to 16). Finally, the task assignment instance I_φ is returned (line 18) and the selected worker w_i is notified to perform the corresponding task t_j.

5. EXPERIMENTAL STUDY

5.1 Experimental setup

Data sets: For the location of workers and tasks, in a 2D data space, generated following a skewed distribution, First locate 80% of them in a Gaussian cluster (center at (0.5, 0.5), variance = 0.2^2), and the rest are uniformly distributed in the 2D data space. For each worker's moving speed v and maximum moving distance r, and each task's budget B, following Gaussian distribution, generated in the range $[0.2, 0.3]$, $[r-, r+]$, $[1, 5]$, for the deadline e of each task, it is the current timestamp plus 1 to 3 random values. Randomly select a user and a group in the Meetup data set[12] to associate to a worker and a task, then use user and group tags as the worker and task skills, then use the number of events that users participate in to initialize intimacy, that is $a_{ik} = 0.5 * 0.5 + 0.5 * |N_{i \setminus cap N_k}| \cdot (|N_i| \cdot |N_k|)^{-1/2}$, where N_i, N_k respectively represent the set of events in which worker w_i and worker w_k participate. For Gaussian distribution $N(0, 0.2^2)$, Using linear mapping to map data samples within $[-1,1]$ to the target range. For the other parameters, the number of tasks is 100, the value of K_I and K_C is 1, and the value of β is 0.5.

Comparative Experiment: By changing the number and the moving distance of workers, the method proposed in this paper is compared with Top_1 Greedy[13] and the random method, and analyzes the difference between the total cost and running time. During the experiment, these methods will repeat several times and the average value will be calculated to obtain the result.

5.2 Result analysis

As shown in Figures 1 and 2, for the platform cost, the increase of the number of workers or the distance workers can move will lead to the increase of the workers available for tasks, and eventually, the number of completed tasks will also increase while the corresponding total platform cost will increase, after, due to the limitation of the number of tasks or the movement speed of workers, the number of completed tasks will stabilize, and the total cost of the platform will also stabilize. Therefore, the curves of the three methods increase first and then flatten. The random method does not consider anything when assigning, so its total platform cost is the highest, Top_1 Greedy considers both costs and benefits when assigning, so its total platform cost is second, Cost-based Greedy selects the worker-task pair with the lowest total cost increase in each round, so it has the lowest total cost of the platform. For the running time, when performing task assignments, the number of workers available for tasks increases will also lead to an increase in system overhead, so the running time of all three algorithms shows an increasing trend. The random method is the simplest and it has the least running time. Cost-based Greedy only needs to calculate the travel cost and communication cost of the workers when assigning, so its running time is secondly, Top_1 Greedy needs to perform additional calculation and comparison of the profit ratio when assigning, and its running time is the highest.

According to the above experimental results, by changing the two parameters, the method in this paper generally performs better than the other two methods in terms of total platform cost, and the overall performance is stable. Because this paper is a static offline scenario, the time complexity of the method is not high, so in terms of running time, the method of this paper can meet the requirements.
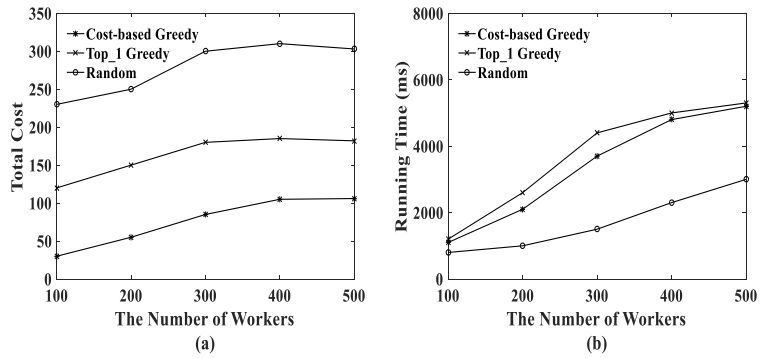
Figure 1. The impact of the number of workers on total cost and running time.
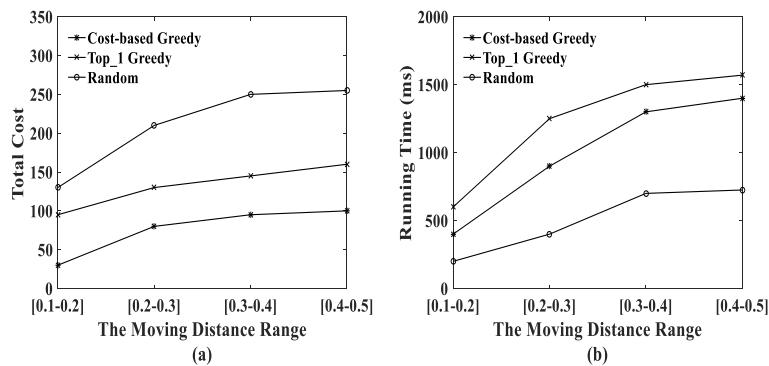

Figure 2. The impact of the moving distance of workers on total cost and running time.

6. SUMMARY AND OUTLOOK

For the multi-skill collaboration-based task assignment problem, whose optimization objective is to minimize the total platform cost, this paper transforms the intimacy between workers into the communication cost. Based on the travel and communication costs, a cost-based greedy approach is proposed to iteratively select the worker task pair that minimizes the total cost growth and matches the task with the right team of workers to minimize the total cost of the platform. Finally, the method proposed in this paper is validated on a synthetic dataset. The next step of the study will no longer be limited to the local optimum of the current timestamp but will be to achieve the global optimum of a period by accurate prediction.

REFERENCES

[1] Cheng, P., Chen, L. and Ye, J. P., "Cooperation-aware task assignment in spatial crowdsourcing," Proc. of the 35th IEEE Int. Conf. on Data Engineering (Macau), 1442-53 (2019).

[2] Qiao, L., Tang, F. L. and Liu, J. C., "Feedback based high-quality task assignment in collaborative crowdsourcing," Proc. of the 32nd Int. Conf. on Advanced Information Networking and Applications (Krakow), 1139-46 (2018).

[3] Cheng, P., Lian, X., Chen, L., Han, J. S. and Zhao, J. Z., "Task assignment on multi-skill oriented spatial crowdsourcing," IEEE Transactions on Knowledge and Data Engineering, 28, 2201-15 (2016).

[4] Mavridis, P., Gross-Amblard, D. and Miklos, Z., "Using hierarchical skills for optimized task assignment in knowledge-intensive crowdsourcing," Proc. of the 25th Int. Conf. on World Wide Web (Montreal), 843-853 (2016).

[5] Tran, L., To, H., Fan, L. Y. and Shahabi, C., "A real-time framework for task assignment in hyperlocal spatial crowdsourcing," ACM Trans. Intell. Syst. Technol., 9, 1-26 (2018).

[6] Zhao, Y., Xia, J. F., Liu, G. F., Sun, H., Lian, D. F., Shang, S. and Zheng, K., "Preference-aware task assignment in spatial crowdsourcing," The Thirty-Third AAAI Conf. on Artificial Intelligence, 2629-36 (2019).

[7] Xia, J. F., Zhao, Y., Liu, G. F., Xu, J. J., Zhang, M. and Zheng, K., "Profit-driven task assignment in spatial crowdsourcing," Proc. of the Twenty-Eighth Int. Joint Conf. on Artificial Intelligence, 1914-20 (2019).

[8] Liu, Y., Guo, B., Wang, Y., Wu, W. L., Yu, Z. W. and Zhang, D. Q., "TaskMe: Multi-task allocation in mobile crowd sensing," Proc. of the 2016 ACM Int. Joint Conf. on Pervasive and Ubiquitous Computing, 403-414 (2016).

[9] Deng, D. X., Shahabi, C., Demiryurek, U. and Zhu, L. H., "Task selection in spatial crowdsourcing from worker's perspective," GeoInformatica, 20, 529-568 (2016).

[10] Deng, D. X. and Shahabi, C., "Maximizing the number of worker's self-selected tasks in spatial crowdsourcing," Proc. of the 21st SIGSPATIAL Int. Conf. on Advances in Geographic Information Systems, 314-323 (2013).

[11] Song, T. S., Tong, Y. X., Wang, L. B. and Xu, K., "Online task assignment for three types of objects under spatial crowdsourcing environment," Journal of Software, 28, 611-630 (2017).

[12] Liu, X. J., He, Q., Tian, Y. Y., Lee, W. C., McPherson, J. and Han, J. W., "Event-based social networks: Linking the online and offline social worlds," Proc. of the 18th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 1032-40 (2012).

[13] Gao, D., Tong, Y. X., She, J. Y., Song, T. S., Chen, L. and Xu, K., "Top-k team recommendation and its variants in spatial crowdsourcing," *Data Sci. Eng.*, 2, 136-150 (2017).