

Research Article

Participant Recruitment Method Aiming at Service Quality in Mobile Crowd Sensing

Weijin Jiang,^{1,2} Junpeng Chen ,^{1,2} Xiaoliang Liu,^{1,2} Yuehua Liu,^{1,2} and Sijian Lv^{1,2}

¹Key Laboratory of Hunan Province for New Retail Virtual Reality Technology, Changsha 410205, China

²College of Computer and Information Engineering, Hunan University of Technology and Business, Changsha 410205, China

Correspondence should be addressed to Junpeng Chen; jjpchen@163.com

Received 11 December 2020; Revised 16 February 2021; Accepted 9 March 2021; Published 19 April 2021

Academic Editor: Yingjie Wang

Copyright © 2021 Weijin Jiang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the rapid popularization and application of smart sensing devices, mobile crowd sensing (MCS) has made rapid development. MCS mobilizes personnel with various sensing devices to collect data. Task distribution as the key point and difficulty in the field of MCS has attracted wide attention from scholars. However, the current research on participant selection methods whose main goal is data quality is not deep enough. Different from most of these previous studies, this paper studies the participant selection scheme on the multitask condition in MCS. According to the tasks completed by the participants in the past, the accumulated reputation and willingness of participants are used to construct a quality of service model (QoS). On the basis of maximizing QoS, two heuristic greedy algorithms are used to solve participation; two options are proposed: task-centric and user-centric. The distance constraint factor, integrity constraint factor, and reputation constraint factor are introduced into our algorithms. The purpose is to select the most suitable set of participants on the premise of ensuring the QoS, as far as possible to improve the platform's final revenue and the benefits of participants. We used a real data set and generated a simulation data set to evaluate the feasibility and effectiveness of the two algorithms. Detailedly compared our algorithms with the existing algorithms in terms of the number of participants selected, moving distance, and data quality. During the experiment, we established a step data pricing model to quantitatively compare the quality of data uploaded by participants. Experimental results show that two algorithms proposed in this paper have achieved better results in task quality than existing algorithms.

1. Introduction

The rapid development of smart sensing technology and the widespread popularity of mobile smart devices have made it possible for the holder of each mobile device to become a sensing unit [1] which has led to the rapid development of mobile crowd sensing (MCS) [2, 3]. The use of mobile devices to build an interactive and participatory sensor network allows ordinary users to participate in the data collection process, which makes the data collection technology under the big data environment highly developed. Compared with traditional static sensing technology, MCS utilizes existing sensing equipment and communication infrastructures, saving the expense of building additional sensing equipment [4]. At the same time, MCS has the advantages of high mobility and a wide range of potential participants, especially for sudden and unpredictable events, and MCS provides unprece-

dent time and space coverage conditions [5]. Nowadays, MCS has been widely used in public safety [6], environmental monitoring [7], smart transportation [8, 9], etc., which brings a lot of convenience to our daily life while also improving our quality of life. Compared with traditional wireless perception technology, MCS pays more attention to and emphasizes the participation process of participants in the collection process and the decisive role of perception data [10].

The main theoretical research of MCS includes three major aspects: participant selection [11], task distribution [12], and incentive mechanism [13–15]. The key of the research is how to set up an excellent incentive mechanism to recruit suitable users for perception tasks, so as to meet the time requirements, quality requirements, and cost requirements of the tasks, so that the data platform can obtain considerable benefits [16].

The problem of participant selection refers to how to effectively select suitable participants from a large user group to perform various perception tasks under certain constraints [17]. On the one hand, the data platforms hope to use less expenditure to obtain the desired data in order to maximize the benefits of them; on the other hand, the participants also hope to make as much profit as possible in the perception task that is to increase each the number of tasks undertaken by participants. There are gaps in the data that different participants can collect, and the returns they demand from the platform are also different [18, 19]. How to achieve a game equilibrium between platforms and users is an important direction of current academic research. In addition to the above two main goals, the number of participants in each task, the number of tasks assigned by each participant, the quality of the data submitted by participants, and the total distance a participant needs to move to complete tasks are all in the selection process, which need to be considered [20]. Participant selection, as the core key issue in the field of MCS, is the focus and difficulty of current research. The current related research mainly stays in single-objective optimization, such as platform-centered seeking to maximize profits and user-centered to maximize user benefits [21]. Single-objective optimization often produces various problems. For example, maximizing the benefits of the data platform will damage the benefits of participants, thereby reducing the willingness of users to participate, and ultimately will affect the benefits of the platform [22]. In reality, it is necessary to comprehensively consider various impact conditions in order to ensure the benefits of the data platform and participants. Therefore, this paper comprehensively considers a variety of factors and conducts an in-depth study on participant selection methods under multitask conditions.

The main contributions of this article include the following:

- (1) First, sort out and explain the research on participant selection in the theory of MCS. Among them, the MultiTasker method proposed by Liu et al. provides constructive ideas for the two selection methods proposed in this paper
- (2) Research on the multitask distribution model of MCS. Based on the historical task completion of participants, we use the reputation and willingness of participants to establish a service quality model and propose a participant selection plan based on service quality
- (3) Aiming at the task-centered and user-centered participant selection problems, the heuristic greedy algorithm is used to solve the problem, and a participant selection algorithm is proposed, respectively, and the platform is realized under the conditions of meeting the service quality constraints and distance constraints the requirements for minimum expenditure and maximum user benefits
- (4) Verify the feasibility and effectiveness of the selection scheme through simulation experiments on real data

and simulated data, construct a data value evaluation index system, and find the most suitable distance constraint value. Compare it with existing algorithms in terms of the number of participants, the number of tasks, and the data quality

2. Related Work

Now, a train of scholars have conducted extensive research and exploration on the selection of participants in the MCS system.

Some scholars' research mainly focuses on designing excellent incentive mechanisms to encourage participants to perform perceptual tasks. Jin et al. [23] proposed a new type of MCS system framework, which integrates data aggregation, incentives, and disturbance mechanisms. Its incentive mechanism selects users who are more likely to provide reliable data and compensates for their perceived cost; its data aggregation mechanism combines the reliability of users to generate high-precision aggregation results; its data perturbation mechanism weakens users' privacy. Leaking concerns improve participant satisfaction and the desired accuracy of the final disturbance result. Hu et al. [24] focused on the location-based MCS system and proposed a demand-based dynamic incentive mechanism. The mechanism can dynamically change the reward of each task as needed to balance the popularity between tasks. Propose a solution based on optimal backtracking for opportunistic scenarios to help each user choose a task, while maximizing their profits, and better achieve the balance between participants and tasks. Xu et al. [25] proposed two models, MCT-M and MCT-S, for the purpose of minimizing social costs. These two solutions use the real relationship of social networks to group participants, so that each collaborative task can be completed by a group of compatible users. Experiments show that the proposed model can further reduce social costs while reducing grouping time.

At the same time, some researchers pay attention to the research on the selection mechanism of excellent participants. Guo et al. [26] mainly studied the problem of participant selection in the task types oriented to diversity. Starting from the microscopic and macroscopic visual task types, respectively, they designed the UtiPay visual group perception framework, which greatly improved the quality of group intelligence perception data. Zhou et al. [27] defined the coverage model of "t-Sweep k-Coverage" and proposed a selection method based on linear programming and a selection method based on greedy strategy. Through solving on real data sets, they verified the feasibility and effectiveness of the two selection methods. However, this method only solves the situation when the participant's moving time, location, and other attributes are known and does not predict the future location of the user's historical movement law. Estrada et al. [28] proposed a framework for selecting participants in order, which comprehensively considers the reputation and payment model of participants and maximizes the quality of service under certain time constraints. However, this method does not collect the actual movement speed of the participant and randomly generates the movement speed of

the participant, and it also does not consider the movement trajectory of the user. Pu et al. [29] proposed another sequence selection scheme, which models the nature of tasks, user capabilities, and real-time performance to predict the quality of services provided by participants. The plan considers the importance of keywords to maximize the possibility of participants accepting and completing tasks in a timely manner but does not consider the reputation of participants and the expenses caused by moving distance. Azzam et al. [30] proposed a group-based selection scheme, which considers maximizing coverage, sampling frequency, and battery life and greatly improves data quality. However, using this model to select participants requires a lot of running time, and the model does not consider the cost of the task platform and the movement trajectory of participants. Yang et al. [31] proposed a multiarmed slot machine user selection model with a budget mechanism, which solves the cost problem of effectively learning to select each user without prior knowledge and reduces the selection to a large extent. The accumulated regrets are considered in this process; but the time correlation of user environment information has not been studied in depth. Wang et al. [32] divided users into two groups, formulated different pricing plans, proposed a semi-Markov model to determine the user's interest points distribution, and proposed a prediction-based participant preference Method to minimize the cost of uploading data. Jiang et al. [16] proposed an optimal participant decision model based on voting mechanism. In the participant selection stage, considering the platform benefits, they designed a participant decision model based on the reverse auction model, which is similar to the traditional reverse auction model. In contrast, this method increases the parameter of the number of effective perception tasks and effectively reduces the redundancy of participants. However, this method does not fully consider the geographic location information of participants. For large-scale surveillance systems, Li et al. [33] introduced the caching mechanism to mobile group intelligence perception for the first time. They proposed a dynamic participant selection problem with heterogeneous perception tasks and designed offline and online algorithms to solve this problem. Simulations on real data sets verify the effectiveness and efficiency of the proposed algorithm. Liu [34] et al. proposed three algorithms: T-Random, T-Most, and PT-Most, for multitask participant selection from task-centric and user-centric. The number of participants, the number of tasks assigned by each user, and the moving distance were compared through experiments, and the most appropriate participant selection strategy under different indicators was selected. However, this method only uses the number of users, tasks, and movement distance as the main optimization indicators and does not consider the service quality of participants.

In summary, most of the current research methods characterize the selection problem as a target optimization problem on certain constraints. In this process, it considers the reputation of participants, geographic location, moving distance, and incentive mechanism and converts multiobjective constraints into single-objective constraints. However, there are few studies on using participant historical information

to estimate the task quality of participants. Some methods ignore the platform's need for task quality when considering the number of participants or moving distances. Other methods take into account the task quality requirements but ignore the actual moving speed or willingness of participants. In reality, the different speeds of participants and willingness have a great influence on the task completion time and task quality. This paper takes into account the historical tasks of participants and establishes a service quality model. In the task-centered and user-centered selection algorithms, many factors such as travel distance, reputation, and task completion degree are comprehensively considered to find the best option under the best service quality. Finally, we use real data sets and simulated data sets to carry out simulation experiments on the two proposed algorithms and analyze and study the participant task set and running time in terms of distance weakening.

3. Multitask Participant Selection Model Based on QoS

The MCS system consists of three parts: task release, user selection, and data collection. For various needs, it is often necessary to collect relevant information at a certain time in certain regions and in real life. Collecting data by themselves often consumes a lot of time and financial resources. Nowadays, the common solution is that the data requester uploads the task requirements to a certain task distribution platform; then the cloud server receives the request information and distributes the task to some practitioners for completion. Participants receive tasks and perform data perception according to the task requirements, then upload the data to the cloud. Cloud cleans the data and packages it to the requester. In this process, it is undoubtedly crucial to select the right participants for the perception task, and the quality of participants directly determines the quality of the collected data. Aiming at task-centered and user-centered, under the premise of maximizing service quality, this paper proposes two multitask distribution participant selection methods, from the two aspects of maximizing the benefits of the data platform and participants optimization.

The task platform tends to issue multiple tasks at one time to form a task set $T = \{t_1, t_2, t_3, \dots, t_n\}$. At the same time, the data platform will have certain restrictions on the task completion time. Late data is often worthless; this paper agree that the completion time of each task set T should not exceed h hours. There are also certain differences in the urgency and importance of tasks. In order to ensure the completeness of the collected data as much as possible and consider the emergencies of the participants and other factors, we consider that each task requires multiple participants to complete, and the number of people s_i required to complete each task t_i is different. For ease analysis, the completion time of each task is assumed to be 5 minutes. Suppose there are $U = \{u_1, u_2, u_3, \dots, u_m\}$ participants in the task platform, and the moving speed of each participant is different, $V = \{v_1, v_2, v_3, \dots, v_m\}$. The set of tasks that each participant needs to complete is represented by $TU_i = \{t_{i1}, t_{i2}, t_{i3}, \dots\}$. The sum

of the time taken by the participants to complete tasks and spent on the trip is lower than required by the platform. The constraint conditions are defined as follows: $D(TU_i)$ is the total distance the participant needs to move to complete the task set:

$$TU_i \times 5 + \frac{D(TU_i)}{v_i} \leq h \times 60, \quad 1 \leq i \leq m. \quad (1)$$

This paper integrates the willingness to participate and the reputation of the participants in completing tasks in the past, constructing the service quality model of the participants.

The quality of data uploaded by participants is positively related to their willingness. For example, the higher the willingness of a participant, the more actively participant will collect data and the higher the quality the data uploaded; if a participant adopts a negative attitude to collect data, the quality will also be low. However, it is not enough to only consider the impact of participants' willingness on data quality; the impact of participants' objective perception on data quality is also crucial. We have established evaluation indicators for participation willingness and data quality. In order to reflect the willingness to participate w_i and data quality D_i are equally important, we restrict them to the range of [0,1], namely,

$$\begin{aligned} w_i &= \begin{cases} \min(w_i, 1), \\ \max(0, w_i), \end{cases} \\ D_i &= \begin{cases} \min(D_i, 1), \\ \max(0, D_i). \end{cases} \end{aligned} \quad (2)$$

Currently, there are studies that use distance and other conditions to define the willingness of participants. Different participants' willingness to participate in a task is often different and difficult to consider comprehensively. Therefore, this paper defines participant willingness from another perspective. In normal circumstances, the longer it takes a participant to receive the task from the data platform to the confirmation of acceptance of a task, the lower the participant's willingness to participate in the perception task.

The time from when the task is distributed to a participant to when the participant accepts the task is called hesitation time. And willingness is defined as a function related to hesitation time. Inspired by social principles [35], the longer the participants hesitate, the less value they contribute to the platform. We take the average hesitation time of all participants as the critical value. If a participant's hesitation time is equal to it, his willingness to participate is neutral, and his willingness value is 0.5.

Based on the above discussion, participation willingness and hesitation time are modeled as the following functions:

$$w_i = 1 - \max \left[0, \min \left(\frac{1}{2} \log_t t_i, 1 \right) \right], \quad (3)$$

where w_i' represents the willingness of the participant, t_i is the hesitation time, and \bar{t} represents the average hesitation time. It can be seen from the above formula that the longer the participants hesitate, the lower their willingness to participate is.

Participants' untrustworthiness can be obtained from the historical data of participating in the perception task. No record of untrustworthiness indicates that participants have a higher sense of responsibility for participation; the quality of the collected data will be higher.

For the sake of simplicity, we use an improved reputation model [36] to describe the reputation of participants, and the calculation formula as follows:

$$R = \frac{T + 1}{T + F + 2} \times \xi \quad (0 \leq R < 1, T > 0, F > 0). \quad (4)$$

The closer the value of R is to 1, the higher the reputation value of participants is. T and F are the records of "true" and "false" in the participant's historical task. ξ is the weighting factor of the participants' malicious events; the calculation formula as follows:

$$\begin{cases} \xi = \lambda^k, & 0 \leq k < K, \\ \xi = 0, & k \geq K. \end{cases} \quad (5)$$

In this formula, k is the malicious event of some perceived participants, and K is the malicious event threshold set by the data platform. Once the threshold is exceeded, the reputation score of the participant will be set to 0. Considering that the platform's tolerance for malicious events is extraordinarily low, participants who uploaded the wrong information three times can be considered malicious participants, and $K = 3$ by default during the experiment. λ is the attenuation factor within the threshold range, and the default value of λ is set to 0.8. The reputation of participants can be infinitely close but not equal to 1. It can be seen from (4) and (5) that the reputation of a participant who performed the perception task for the first time is 0.5. With the number of perception tasks increasing, there is a large gap between the reputations of participants with no wrong records and those with wrong records. At the same time, the credibility of participants who submit correct data each time should be different. Participants who have 10 correct records must be more credible than those who have only one correct record. When selecting participants, participants with bad reputations are often eliminated. Participant reputation constraints δ are established; if $R \in (0, \delta)$, we do not consider choosing this participant for perception tasks.

Combining the aforementioned reputation model and participation willingness model, we establish the participant's quality of service (QoS), which is calculated as follows:

$$QoS = w \times R. \quad (6)$$

The configuration or brand of the mobile devices carried by participants is different; even the sensing data collected by the devices in the same location may be different. The quality of sensing device may have a certain impact on the data

Input: task set T , user set U , integrity constraint set C , distance constraint θ , reputation constraint δ .
Output: the participant set P and the completed task set TU .

1. Calculate the reputation R of all candidates and the data integrity indicator D
2. Delete participants with substandard reputation
3. Select the task with the most people in the task set to be completed as the initial node $t_{ij}(j=1)$
4. Calculate the Qos of the participants who meet the condition $D \geq \varepsilon_j$ within the bound distance θ from the initial task t_{ij}
5. Select the participant $p_i(i \geq 1)$ with the highest Qos in θ range
6. Select the task $t_{i(j+1)}(j \geq 1)$ that meet $\varepsilon_{(j+1)} \leq D$ and closest to the task t_{ij} as the new central task node
7. Repeat 4~6, and stop the loop when the time for the participant p_i to complete these tasks exceeds the time constraint
8. Output task set $TU_i = \{t_{i1}, t_{i2}, t_{i3}, \dots\}$ of p_i
9. Repeat 3~8 until all tasks are completed
10. Output participant set $P = \{p_1, p_2, p_3, \dots\}$ and completed task set $TU = TU_1 \cup TU_2 \cup \dots \cup TU_i$
11. End

ALGORITHM 1

quality. The complete performance of the data characterizes the perception of participants. Quantitative research on the perception of participants' equipment and professional conditions obviously requires consideration of a variety of positive and negative related factors; the degree of difference between the data submitted by participants and the expected data on the platform can characterize the perception of the participant.

Evaluate the historical submitted data of participants, calculate the gap between it and the expected value of the platform, and establish data integrity function to characterize data quality D_i . The calculation formula is as follows:

$$D_i = \begin{cases} 0.8, & N = 0, \\ \frac{1}{N} \sum_{t=1}^N P_t, & N \geq 1. \end{cases} \quad (7)$$

P_t is the completeness of the data uploaded by a participant at historical time t , $P_t \in (0, 1)$, N is the number of historical tasks completed by the participant, and the larger the value of D , the more complete the information uploaded by the participant high. For the participant who performs the perception task for the first time, platforms are always willing to give him more opportunities to complete the task, and the tolerance for novices is often higher, so we expect the complete performance of the participant to submit the task to 0.8. Similarly, we establish the integrity constraint index ε . When $D \in (0, \varepsilon)$, the task completion quality of the participant cannot meet the current task completion requirements, so this participant will not be considered for the perception task.

The platform's tolerance for data quality is often higher than untrustworthy participants. Some tasks are only a rough perception that can meet the needs of the platform; that is, the requirements for the quality of participation in each task may be different. When the platform publishes the task set $T = \{t_1, t_2, t_3, \dots, t_n\}$, it often gives the corresponding minimum integrity data set $C = \{\varepsilon_1, \varepsilon_2, \varepsilon_3, \dots, \varepsilon_n\}$. As long as the completeness of the tasks submitted by participants is higher than this standard, the needs of the platform can be met.

The optimization goal of the participant selection method proposed in this paper is to minimize the movement

distance of participants under the condition of ensuring service quality to maximize the benefits of the data platform.

4. Participant Selection Method

Algorithm 1 proposed in this paper is improved on the basis of the T-Most algorithm. It is task-centered to select participants and maximizes the service quality of participants as the main optimization goal. At the same time, it also takes into account the minimum number of participants; each participant should complete multiple tasks in the specified time and the smallest possible moving distance, while meeting the data quality requirements of the data platform.

The specific process of participant selection can be considered as follows: select the task which needs the largest number of people among the tasks to be completed as the initial task, and select the participant with the highest service quality within a certain distance from the initial task to complete the task. Next, select the task that is closest to the initial task point and that the participant is eligible to complete as the second task to be completed. When the second task is completed, take this task as the initial node, and select the most recent task and the participant is eligible to complete as the third task to be completed. Repeat this, and follow the above process until the task set completed by this participant in h hours is selected. Eliminate the participant in the above process, and reduce the number of people required for each task in the task set that they have participated in by one. According to the above method, continue to select participants and the corresponding task set until all tasks are completed.

Algorithm 1 is task-centered for participant selection, and its main goal is to maximize the Qos of participants, thereby improving the benefits of data platform. This paper proposes a participant-centric selection scheme based on the PT-Most algorithm. Different from Algorithm 1, the main optimization goal of Algorithm 2 is to maximize the number of tasks completed by participants and to maximize the benefits of participants on the premise of meeting the data quality requirements of the platform.

The specific process of participant selection is considered as follows: randomly select participants in the user set as candidate, and select the nearest task that satisfies the integrity

Input: task set T , user set U , integrity constraint set C , distance constraint θ , reputation constraint δ .
Output: the participant set P and the completed task set TU .

1. Calculate the reputation R of all candidate participants and the data integrity indicator D
2. Delete participants with substandard reputation
3. Randomly select a user in the participant set as the task candidate p_{ji}
4. Select the task $t_{ik} (k=1)$ that satisfies the condition $\varepsilon_k \leq D_i$ closest to the user p_{ji} within the constraint condition as the initial task
5. Select the closest task $t_{i(k+1)}$ that satisfies $\varepsilon_{k+1} \leq D_i$ under the distance constraint from task t_{ik} as the next task
6. Repeat 5~6, and stop the loop when the time for the participant p_{ji} to complete these tasks exceeds the time constraint
7. Output task set $TU_{ji} = \{t_{i1}, t_{i2}, t_{i3}, \dots\}$ of candidate p_{ji}
8. Execute 3~8 in a loop to determine the task collection of each candidate p_{ji} within the time constraint TU_{ji}
9. Choose the participant p_j with the largest task set to complete the task set TU_j
10. Repeat 3~10 until the task set of each participant who meets the constraints is determined
11. Output participant set $P = \{p_1, p_2, p_3, \dots\}$ and completed task set $TU = TU_1 \cup TU_2 \cup \dots \cup TU_i$
12. End

ALGORITHM 2

constraint within the range of this candidate θ as the initial task. Next, select the task that is closest to the initial task and the candidate is eligible to complete as the next task. This is repeated until the task set for each participant within the agreed time is selected. Choose the candidate with the most tasks as the first participant. Eliminate the participants in the above process, and reduce the number of people required for each task in the task set by 1. According to the above method, continue to select the participants and the corresponding task set until all tasks are completed.

Algorithm 1 is a task-centric algorithm. The time complexity of the algorithm is directly related to the number of tasks. The time complexity of Algorithm 1 is $n!$. Algorithm 2 is a user-centered selection algorithm. As the number of tasks increases, the number of participants x required increases. The value of x will affect the amount of calculation of Algorithm 2. The time complexity of Algorithm 2 is $m!/(m-x)!$, where m is a participant whose credibility meets the standard.

5. Experimental Evaluation

5.1. Data Set and Experimental Settings

5.1.1. Data Set. This paper selects a real data set and a simulated data set to evaluate the participants' preferred schemes in multitasking conditions.

- (1) Real data set: this paper uses the crowdsourced task allocation data set of the Chinese Society of Industrial and Applied Mathematics. The data set contains 835 tasks and 1877 participants in the four cities of Guangzhou, Shenzhen, Foshan, and Dongguan. On the basis of this data set, we randomly assign a task complete index ε for each task to be constrained, and at the same time, assign a certain speed value to each participant randomly to represent their true moving speed; the value is 50-1000 m/min. The data set also provides the participant's reputation score R ; we map it to $[0,1]$. Table 1 summarizes the parameter settings for this data set.

TABLE 1: Real data set parameter settings.

| Parameter | Description | Parameter value |
|---------------|------------------------------------|-----------------|
| m | Number of perception tasks | 835 |
| n | Optional participants | 1877 |
| R | Participant reputation | $[0,1]$ |
| ε | Minimum task integrity requirement | $[0,1]$ |
| v | Participant moving speed | 50-1000 m/min |

- (2) Simulation data set: this paper follows the existing data generation method to generate a simulation data set. Evenly distribute the perception task and the location of the participants in a rectangular plane of 20 km \times 20 km. In addition, set a reputation parameter R within the range of $[0,1]$ for each participant. In the process of the experiment, we consider task sets and participant sets of different sizes, from $\{100,200,500,800\}$ and $\{50,100,150,200\}$; select m perception tasks and n participants to combine. And specify the same other parameters as the real data set for the simulated data set. Table 2 summarizes the parameter settings of the simulation data set.

5.1.2. Experimental Setup. Since the real data set used in this paper has given the task location and the participant's current location, the generated simulation data set is also randomly distributed on the participant and task location; we will not take other considerations of location information. However, in real life, participants and tasks are often not connected in a straight line. Considering the buildings and traffic routes is more in line with the real situation. Choosing a suitable path obviously requires a special study. Since the content of this paper is the preferred plan of the participants, for the convenience of the experiment, we use the Euclidean distance between two points to represent the distance that needs to be moved.

TABLE 2: Simulation data set parameter settings.

| Parameter | Description | Parameter value |
|---------------|------------------------------------|-------------------|
| m | Number of perception tasks | {100,200,500,800} |
| n | Optional participants | {50,100,200,300} |
| R | Participant reputation | [0,1] |
| ε | Minimum task integrity requirement | [0,1] |
| v | Participant moving speed | 50-1000 m/min |

In the subsequent experiments, in order to facilitate the comparison of algorithm performance, we default that each task requires 5 users to complete without special instructions. At the same time, set the completion time of each task to 5 minutes, and the completion time of each task set to 2 hours.

This paper considers experimental design from two aspects. The first is to study the parameters of the algorithm and compare the experimental results to select the most suitable distance constraint θ . At the same time, in the case of changes of the perception task and the number of participants, the performance of the algorithm is considered; the second is to consider the quality of the collected data and evaluate the effect of the algorithm.

5.2. Experimental Results

5.2.1. Selection of Suitable Distance Constraint θ . This experiment is conducted on the real data set and the simulated data set. For the real data set, we determine the appropriate θ through experimental comparison; for the simulated data set, on the basis of changing the number of perception tasks and the number of participants available for selection, we conducted multiple experiments to verify the effect of θ changes in different task densities and participant densities on the platform's revenue and then chose the appropriate θ restrictions.

The larger the range of θ , the higher the QoS of the selected participant will undoubtedly be. However, when pursuing higher QoS, the time required for the participant to complete the task will be higher, which increases the cost of the platform. We define participant's revenue as task revenue and travel revenue and establish a pricing mechanism for the participant's revenue time. The longer the participant takes to reach a certain task, the higher the participant's travel revenue will be, and the platform's variable expenses will be larger. For simplicity, we establish a linear time expenditure function to characterize the variable expenditure of the platform. Its function can be set as follows:

$$P(u) = kt. \quad (8)$$

$P(u)$ is the variable expenditure of the platform, t is the time required for the participant to reach the next task point, and k is the movement cost per unit time. For simplicity, set the value of k to 1.

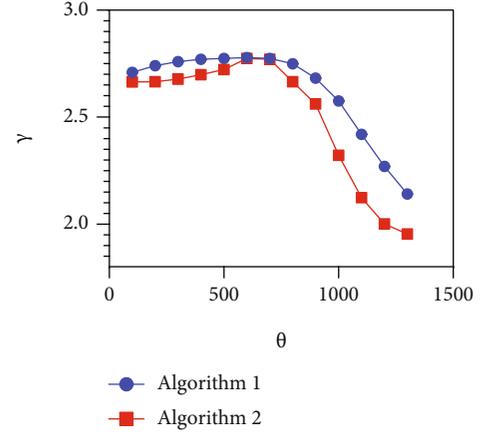


FIGURE 1: The influence of the change of θ value on the real data set γ .

At the same time, pricing is based on the data collected by the participants. The higher the service quality of the participants, the greater the revenue of the data platform. We establish a step-type pricing model. The pricing model is as follows:

$$G(u) = \begin{cases} 10 \text{ QoS} : 0 \leq \text{QoS} < 0.2, \\ 2 + 10(\text{QoS} - 0.2) : 0.2 \leq \text{QoS} < 0.5, \\ 7 + 10(\text{QoS} - 0.5) : 0.5 \leq \text{QoS} < 0.8, \\ 10 + 10(\text{QoS} - 0.8) : 0.8 \leq \text{QoS} < 1. \end{cases} \quad (9)$$

Among them, $G(u)$ is the income that the platform can finally obtain with the improvement of QoS.

And set the objective function γ to represent the final return of the platform as the distance changes and consider the impact of changes in θ on the platform's return. The γ function is expressed as follows:

$$\gamma = G(u) - P(u). \quad (10)$$

Figure 1 is the result of the experiment on the real data set. It can be clearly seen that with the relaxation of the

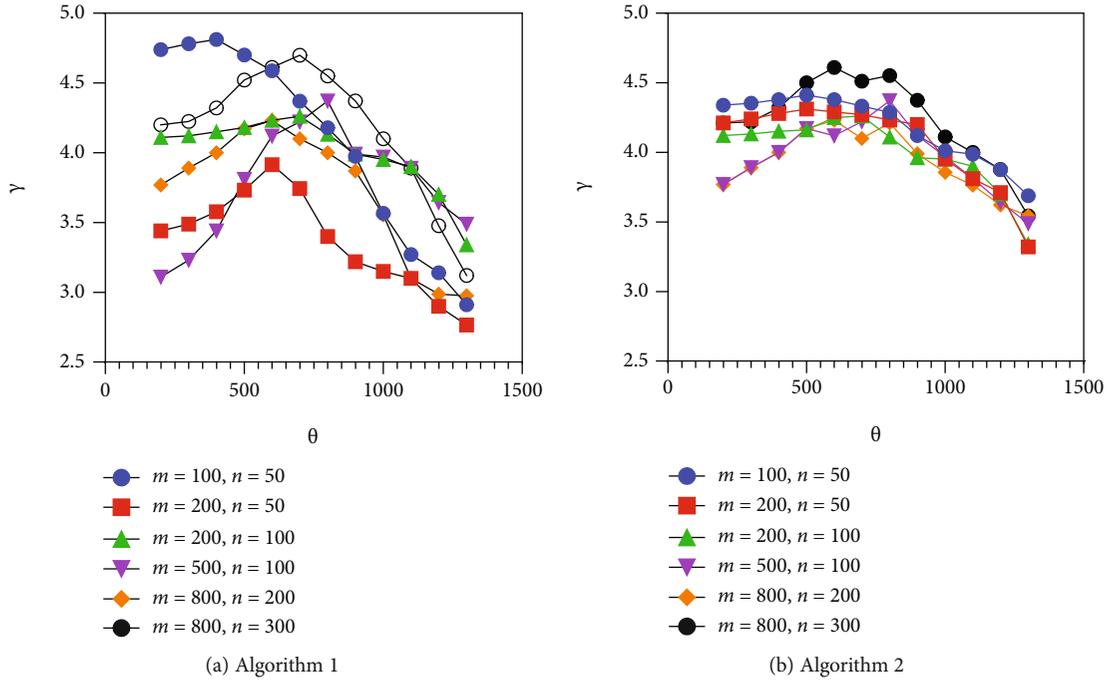


FIGURE 2: The influence of the change of θ value on the simulated data set γ .

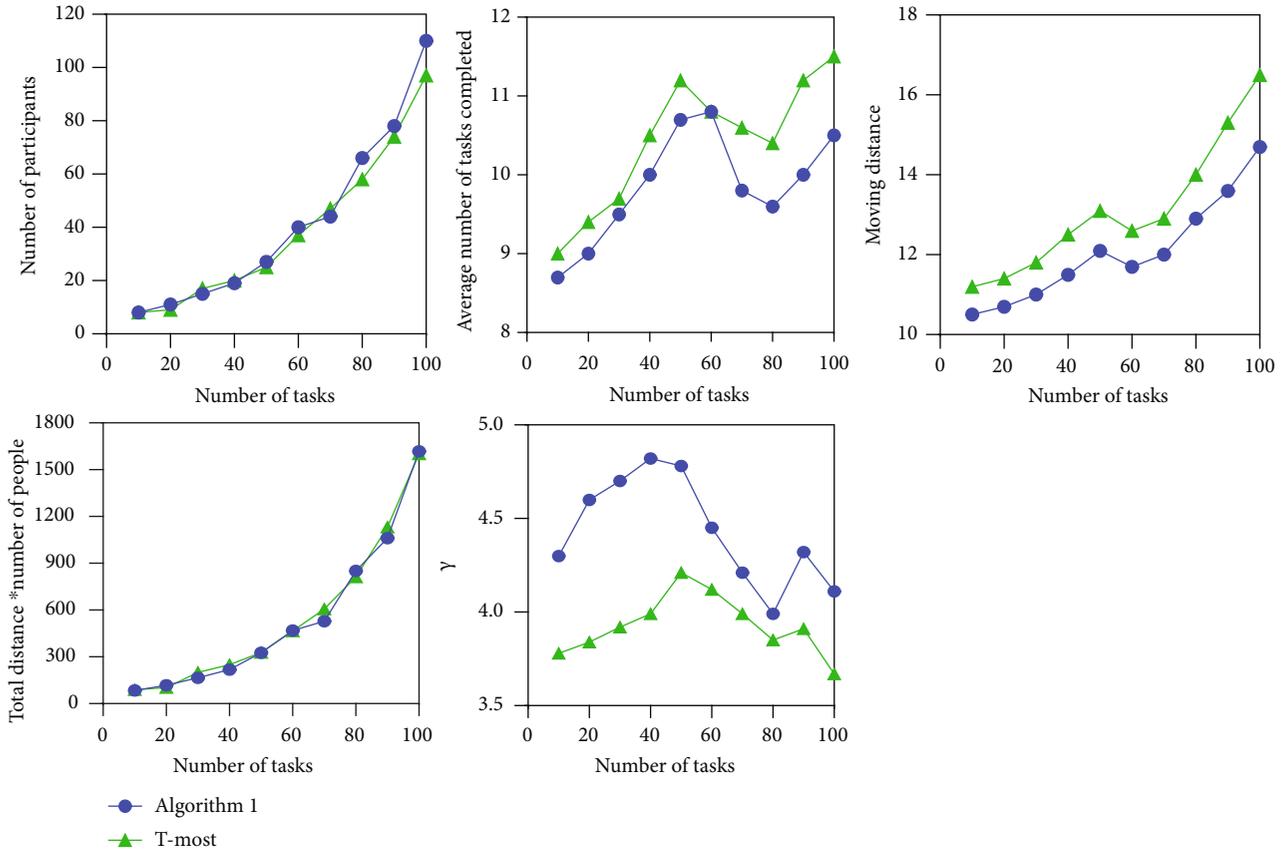


FIGURE 3: The impact of changes in the number of tasks on algorithm performance.

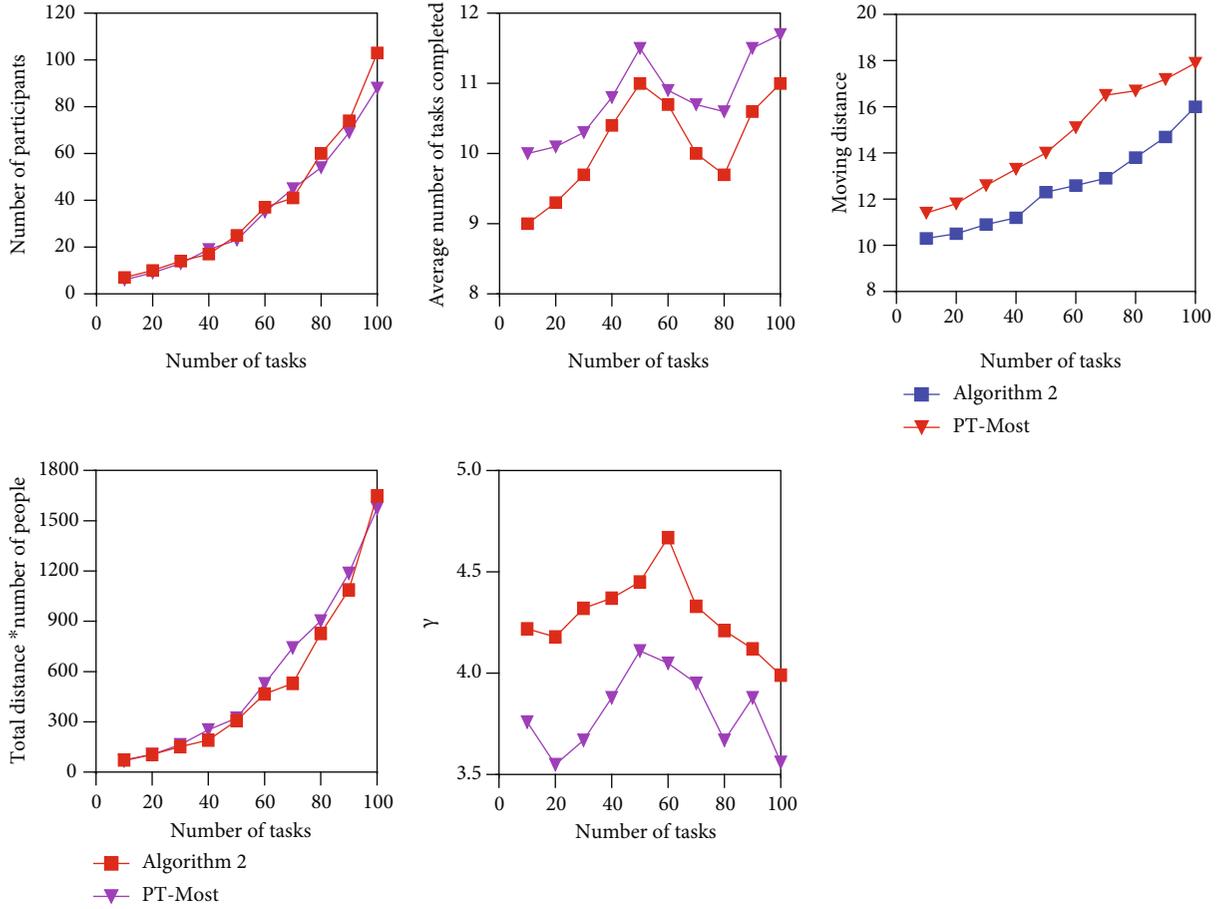


FIGURE 4: The impact of changes in the number of tasks on algorithm performance.

constraints, the final revenue of the platform will first increase and then decrease and, when the constraint distance is greater than 900 m, the revenue of the platform will significantly decrease; this should be consistent with the actual situation. Too short a distance constraint is difficult to select participants with better service quality, which will make the platform’s revenue less than the optimal level; and the value of θ is selected too large; although participants with better QoS can be selected, the distance the users needs to move will increase, which will undoubtedly increase the additional expenses of the platform.

Figure 2 is the experimental result of the dynamic combination of the number of tasks and the number of participants on the simulated data set. It can be seen that as the number of tasks distributed and the number of participants changes, the distance when γ reaches the maximum value also changes greatly. Too high or too low θ will make the γ value lower. Therefore, it is not easy to select the most appropriate value of θ for different data, but it can be seen from Figure 2 that in view of the changes in the number of tasks and participants, most data will achieve better results in the condition of $\theta \in [500,800]$. In subsequent experiments, if there are no special instructions, we default $\theta = 600$ as the distance constraint value.

5.2.2. The Impact of Changes in the Number of Tasks on the Algorithm. The number of tasks issued by the platform is not constant. This experiment considers changing the number of tasks distributed each time when other factors are determined to make the selection of participants. Figure 3 shows the results of a comparison experiment between Algorithm 1 and T-Most algorithm. As the number of tasks increases, the gap between the number of participants selected by Algorithm 1 and the T-Most algorithm is not too large. Even in some certain tasks, Algorithm 1 will select fewer users. Regarding the number of tasks assigned to users on average, there is not much difference between Algorithm 1 and T-Most. In terms of moving distance, because Algorithm 1 weakens the distance constraint, the moving distance of participants in the T-Most is less than that in Algorithm 1, but in the index of “total distance \times number of people,” in the two algorithms, the performance effect is not much different. The optimization goal of the T-Most is to minimize the distance and reduce the cost. Algorithm 1 is not so strict in the distance requirement but pays more attention to the service quality of participants. So on the platform’s final profit evaluation index γ , Algorithm 1 achieves significantly better results than T-Most, which is

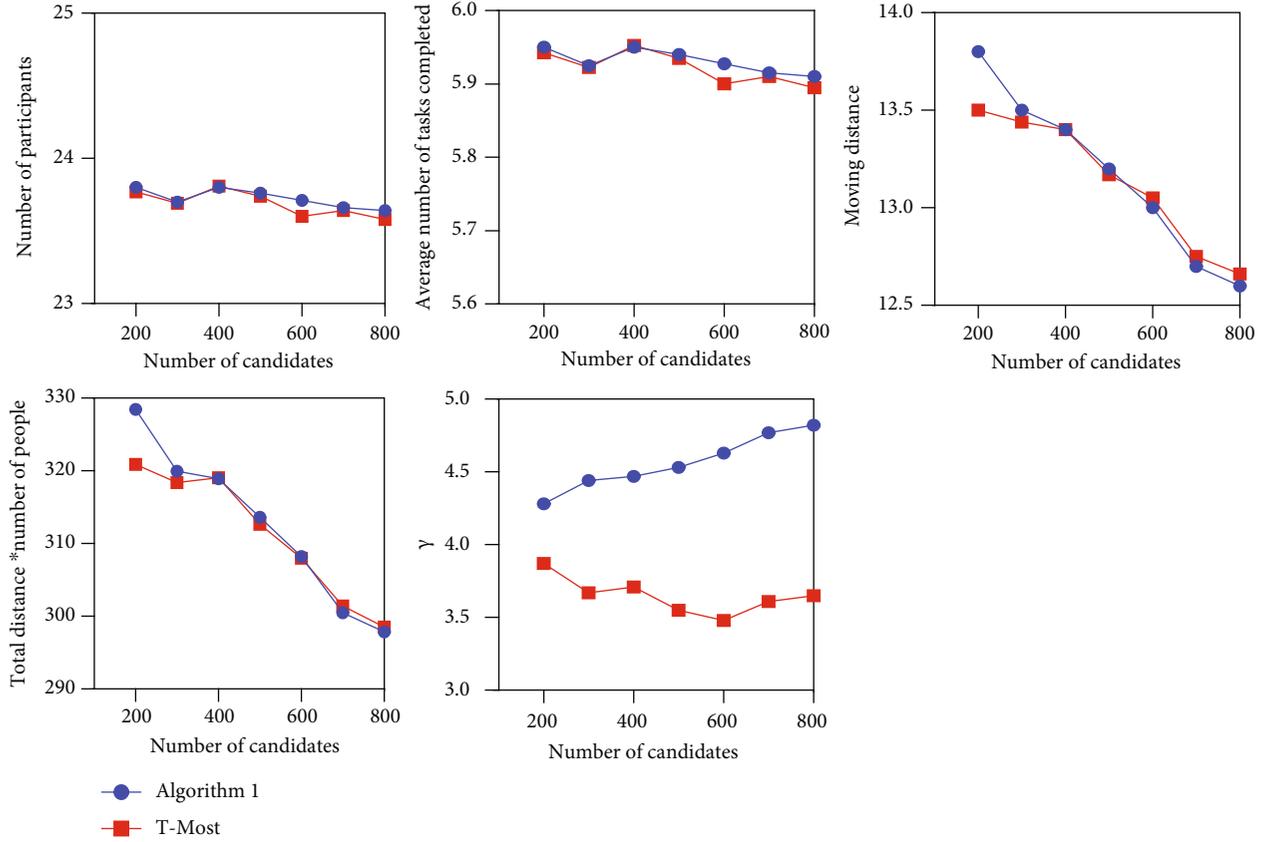


FIGURE 5: The impact of changes in the number of candidates on algorithm performance.

consistent with what we envisioned when proposing the algorithm.

Figure 4 shows the results of a comparison experiment between Algorithm 2 and PT-Most. Like the results in Figure 3, when the gap between other indicators is not obvious, the effect of our proposed algorithm 2 on the evaluation index γ is obviously better than that of PT-Most.

5.2.3. The Impact of Changes in the Number of Candidates on the Algorithm. In the process of participant selection, in addition to the number of tasks to be completed that has a great influence on the selection, the number of participants to be selected will also have an impact on the selection effect to a large extent. The more candidates there are, the closer the selected participant will be to the task, and the quality of service for participants will be further improved. This experiment changes the number of candidate participants in the task area when the task to be completed is constant. The evaluation indicators are the same as the previous experiment. It can be seen from Figures 5 and 6 that as the number of candidates in the area increases, the number of participants selected by the four algorithms and the average number of tasks completed by each participant did not change much, and there was not much difference in quantitative indicators. However, as the number of candidates increases, the total distance that participants need to move and distance \times

number of people indicators will drop rapidly. The experimental results are in line with our expectations. As the number of candidates increases, the closer the selected participant will be to the task, the corresponding movement distance will decrease. In terms of task completion quality, it is obvious that the two algorithms mentioned in this paper are better than T-Most and PT-Most. As the number of candidates increases, the performance of the two algorithms in this paper will get better and better on the γ index, while T-Most and PT-Most have no obvious changes.

5.2.4. Perceive Time Changes. The specified completion time of task is also a major factor affecting the performance of the algorithm. This experiment changes the limited time to 1 h, 2 h, and 3 h for experimental comparison and analysis of various situations when the number of tasks remains unchanged. It can be seen from Figure 7 that with the relaxation of the time limit, the number of tasks performed will continue to decrease and the number of tasks that each user needs to complete increases accordingly. The moving distance increases slightly, and the index of distance \times number of people also shows a downward trend. Compared with the comparison algorithm, the two algorithms in this paper have no obvious difference in the above four indicators. In most cases, the algorithms in this paper have similar effects to the comparison algorithm. On the γ index, the algorithm

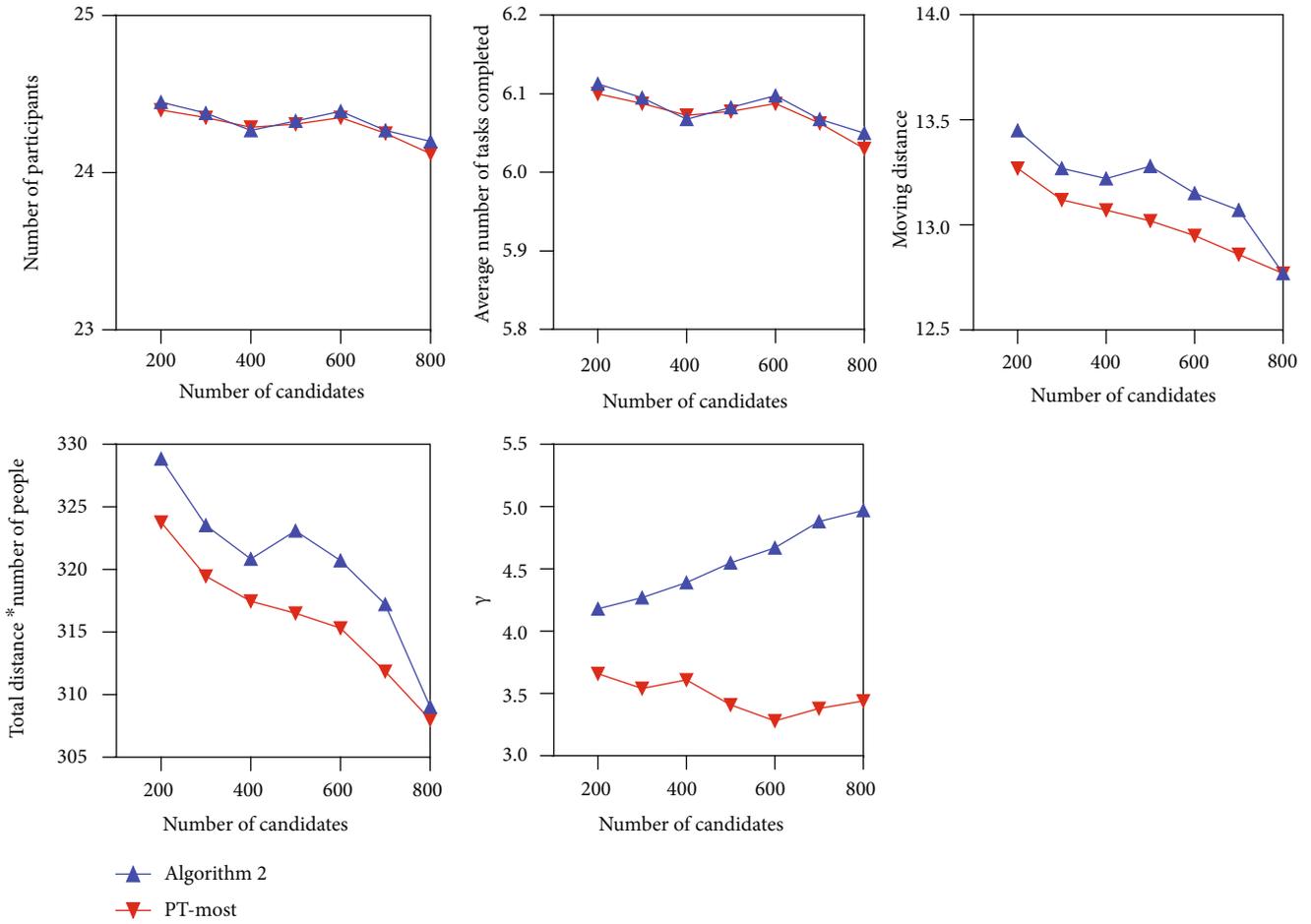


FIGURE 6: The impact of changes in the number of candidates on algorithm performance.

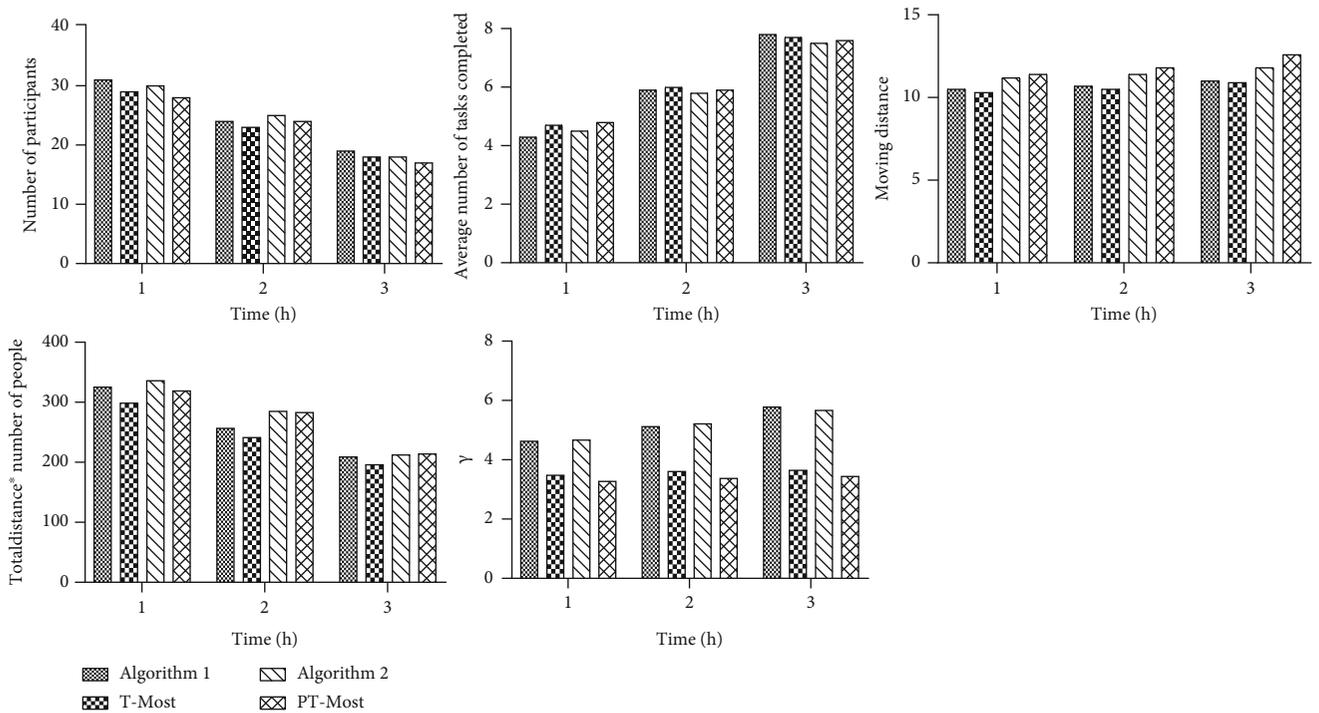


FIGURE 7: The impact of perceptual time changes on the algorithm.

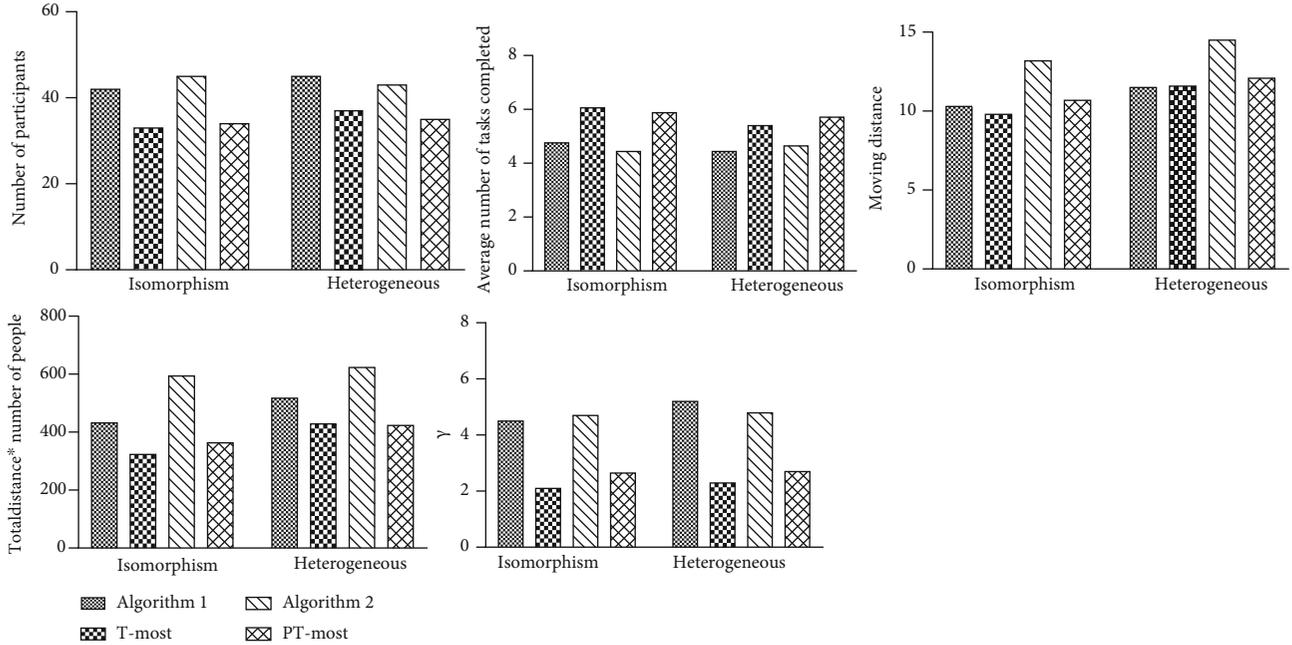


FIGURE 8: The influence of homogeneous and heterogeneous conditions on algorithms.

proposed in this paper is significantly better than the two comparison algorithms, and with the relaxation of the time limit, the γ index of our algorithms has a gradual upward trend.

5.2.5. Change the Number of People Required for Each Task.

The above experiments are conducted under the premise that each task requires 5 participants to complete. But in real life, due to the difficulty level of each task, the different requirements of the data requester for data quality, and the difference in the budget for each task, the number of participants required for each task is not the same. Therefore, this experiment will study the changes of the five comparative indicators in the above experiment under the condition that the task requires the same and different participants.

In order to facilitate the experiment, we set the number of tasks to be completed to 20. Under homogeneous conditions, each task requires 10 participants to complete. Under heterogeneous conditions, the number of people required to set tasks is 5, 10, and 15 and accounted for 25%, 50%, and 25% of the total number of tasks; that is, 5 tasks require 5 participants to complete, 10 tasks require 10 participants to complete, and 5 tasks require 15 participants to complete. We generate two 3 km \times 3 km rectangular areas to simulate the real environment and make the participants and tasks evenly distributed on the rectangular plane. In the two scenarios, except for the differences in the participants required for the task, the other parameter settings are roughly the same.

It can be seen from Figure 8 that there is a big difference when the participants required by the task are different and when the participants required by the task are completely consistent. Algorithms 1 and 2 have certain weaknesses in the first four comparison indicators, which is consistent with

our expected situation. However, comparing Algorithms 1 and 2 longitudinally, we find that in terms of the number of participants and the average number of tasks completed, Algorithm 1 has achieved better results under the condition of task isomorphism. Under the condition of heterogeneous tasks, Algorithm 2 performs better. In terms of moving distance, Algorithm 2 has moved a longer distance compared to Algorithm 1 under both conditions, because the main goal of Algorithm 2 is to maximize the completion of tasks for each participant, this result is also in line with expectations. On the index of distance \times people, Algorithm 1 is superior to Algorithm 2. On the γ factor, the two algorithms proposed in this paper are undoubtedly superior to the existing algorithms.

6. Conclusion

This paper studies the participant selection method in the multitask situation in MCS. Based on the participant's historical task completion, the reputation and willingness of participants are used to establish a service quality model, aiming at task-centered and participant-oriented for the center; we establish a preferred plan for participants based on service quality. Under the condition of satisfying service quality constraints and distance constraints, the requirements of minimizing platform expenses and maximizing user benefits can be achieved. We establish a data quality pricing function and quantitatively evaluate the data uploaded by participants. In terms of experimental evaluation, we have compared multiple index factors with the T-Most and PT-Most. The experimental results show that the two algorithms proposed in this paper have no obvious gap with the comparison algorithm in terms of participant selection evaluation indicators. However, in terms of data quality,

our algorithms are significantly better than the control algorithm. Our participant selection program based on QoS can effectively select better participants, thereby greatly increasing platform revenue.

In the follow-up research, we plan to study the heterogeneity of users and the law of historical movement to further improve the quality of service.

Data Availability

The data used to support the findings of this study have not been made available because our organization has confidentiality measures.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work has been supported by the National Natural Science Foundation of China (61472136 and 61772196), the Natural Science Foundation of Hunan Province (2020JJ4249), the Social Science Foundation of Hunan Province (2016ZDB006), the Key Project Social Science Achievement Review Committee of Hunan Province (XSP19ZD1005), the Degree and Graduate Education Reform Research Project of Hunan Province (2020JGYB234), the and Scientific Research Project of Hunan Provincial Department of Education (20A131).

References

- [1] X. Liu, M. S. Obaidat, C. Lin, T. Wang, and A. Liu, "Movement-based solutions to energy limitation in wireless sensor networks: state of the art and future trends," *IEEE Network*, vol. 34, pp. 1–6, 2020.
- [2] Y. Liu, "Crowd sensing computing," *Communications of the ccf*, vol. 8, pp. 38–41, 2012.
- [3] Y. Ruiyun, W. Pengfei, B. Zhihong, and W. Xingwei, "Participatory sensing: people-centric smart sensing and computing," *Journal of computer research and Development*, vol. 54, no. 3, pp. 457–473, 2017.
- [4] T. Huang, J. Liu, S. Wang, C. Zhang, and R. J. Liu, "Survey of the future network technology and trend," *Journal on Communications*, vol. 42, no. 2, pp. 1–23, 2021.
- [5] T. Wang, H. Luo, X. Zeng, Z. Yu, A. Liu, and A. K. Sangaiah, "Mobility based trust evaluation for heterogeneous electric vehicles network in smart cities," *IEEE Transactions on Intelligent Transportation Systems*, vol. 99, pp. 1–10, 2020.
- [6] Y. Liu, T. Feng, M. Peng et al., "COMP: online control mechanism for profit maximization in privacy-preserving crowdsensing," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 7, pp. 1614–1628, 2020.
- [7] X. Zhu, Y. Luo, A. Liu, W. Tang, and M. Z. A. Bhuiyan, "A deep learning-based mobile crowdsensing scheme by predicting vehicle mobility," *IEEE Transactions on Intelligent Transportation Systems*, vol. 99, pp. 1–12, 2020.
- [8] H. Xiong, D. Zhang, L. Wang, and G. Chen, "CrowdRecruiter: selecting participants for Piggyback Crowdsensing under probabilistic coverage constraint," in *Acm International Joint Conference on Pervasive & Ubiquitous Computing*, pp. 703–714, September 2014.
- [9] R. K. Ganti, F. Ye, and H. Lei, "Mobile crowdsensing: current state and future challenges," *IEEE Communications Magazine*, vol. 49, no. 11, pp. 32–39, 2011.
- [10] J. Huang, L. Kong, H. N. Dai et al., "Blockchain based mobile crowd sensing in industrial systems," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 10, pp. 6553–6563, 2020.
- [11] F. Restuccia, N. Ghosh, S. Bhattacharjee, S. K. Das, and T. Melodia, "Quality of information in mobile crowdsensing," *ACM Transactions on Sensor Networks (TOSN)*, vol. 13, no. 4, pp. 1–43, 2017.
- [12] J. An, Z. Peng, X. Gui, and L.-L. Xiang, "Research on task distribution mechanism based on public transit system in crowd sensing," *Chinese Journal of Computers*, vol. 42, no. 2, pp. 65–78, 2019.
- [13] W. J. Jiang and X. L. Liu, "Research on anti-greedy incentive mechanism for mobile user recruitment in crowd sensing," *Control and Decision*, vol. 36, pp. 1–10, 2021.
- [14] Y. Wu, J. R. Zeng, H. Peng, H. Chen, and C. P. Li, "Survey on incentive mechanisms for crowd sensing," *Journal of Software*, vol. 27, no. 8, pp. 2025–2047, 2016.
- [15] H. Jin, L. Su, and K. Nahrstedt, "CENTURION: incentivizing multi-requester mobile crowd sensing," in *IEEE INFOCOM 2017-IEEE conference on Computer Communications*, pp. 1–9, Atlanta, GA, USA, May 2017.
- [16] N. Jiang, D. Xu, J. Zhou, H. Yan, T. Wan, and J. Zheng, "Toward optimal participant decisions with voting-based incentive model for crowd sensing," *Information Sciences*, vol. 512, pp. 1–17, 2020.
- [17] Y. Wen, J. Shi, Q. Zhang et al., "Quality-driven auction-based incentive mechanism for mobile crowd sensing," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 9, pp. 4203–4214, 2015.
- [18] Z. Sheng, C. Mahapatra, C. Zhu, and V. C. M. Leung, "Recent advances in industrial wireless sensor networks toward efficient management in IoT," *IEEE access*, vol. 3, pp. 622–637, 2015.
- [19] Y. Sei and A. Ohsuga, "Differentially private mobile crowd sensing considering sensing errors," *Sensors*, vol. 20, no. 10, p. 2785, 2020.
- [20] B. Tian, Y. Yuan, H. Zhou, and Z. Yang, "Pavement management utilizing mobile crowd sensing," *Advances in Civil Engineering*, vol. 2020, 16 pages, 2020.
- [21] J.-S. Lee and B. Hoh, "Sell your experiences: a market mechanism based incentive for participatory sensing," in *2010 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pp. 60–68, Mannheim, Germany, April 2010.
- [22] M. Xiao, J. Wu, L. Huang, Y. Wang, and C. Liu, "Multi-task assignment for crowdsensing in mobile social networks," in *2015 IEEE Conference on Computer Communications (INFOCOM)*, pp. 2227–2235, Hong Kong, China, April 2015.
- [23] H. Jin, L. Su, H. Xiao, and K. Nahrstedt, "Incentive mechanism for privacy-aware data aggregation in mobile crowd sensing systems," *IEEE/ACM Transactions on Networking*, vol. 26, no. 5, pp. 2019–2032, 2018.
- [24] J. Hu, Z. Wang, J. Wei et al., "Towards demand-driven dynamic incentive for mobile crowdsensing systems," *IEEE*

- Transactions on Wireless Communications*, vol. 19, no. 7, pp. 4907–4918, 2020.
- [25] J. Xu, Z. Rao, L. Xu, D. Yang, and T. Li, “Incentive mechanism for multiple cooperative tasks with compatible users in mobile crowd sensing via online communities,” *IEEE Transactions on Mobile Computing*, vol. 19, no. 7, pp. 1618–1633, 2019.
 - [26] B. Guo, H. Chen, Q. Han, Z. Yu, D. Zhang, and Y. Wang, “Worker-contributed data utility measurement for visual crowdsensing systems,” *IEEE Transactions on Mobile Computing*, vol. 16, no. 8, pp. 2379–2391, 2016.
 - [27] J. Zhou, Z. Y. Yu, W. Z. Guo, L. K. Guo, and W. P. Zhu, “Participant selection algorithm for t-sweep k-coverage crowd sensing tasks,” *Computer Science*, vol. 45, no. 2, pp. 157–164, 2018.
 - [28] R. Estrada, R. Mizouni, H. Otrok, A. Ouali, and J. Bentahar, “A crowd-sensing framework for allocation of time-constrained and location-based tasks,” *IEEE Transactions on Services Computing*, vol. 13, no. 5, pp. 769–785, 2017.
 - [29] L. Pu, X. Chen, J. Xu, and X. Fu, “Crowd foraging: A QoS-oriented self-organized mobile crowdsourcing framework over opportunistic networks,” *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 4, pp. 848–862, 2017.
 - [30] R. Azzam, R. Mizouni, H. Otrok, S. Singh, and A. Ouali, “A stability-based group recruitment system for continuous mobile crowd sensing,” *Computer Communications*, vol. 119, pp. 1–14, 2018.
 - [31] S. Yang, F. Wu, and G. Chen, “On designing most informative user selection methods for mobile crowdsensing,” *Chinese Journal of Computers*, vol. 43, no. 3, pp. 409–422, 2020.
 - [32] E. Wang, Y. Yang, J. Wu, W. Liu, and X. Wang, “An efficient prediction-based user recruitment for mobile crowdsensing,” *IEEE Transactions on Mobile Computing*, vol. 17, no. 1, pp. 16–28, 2017.
 - [33] H. Li, T. Li, W. Wang, and Y. Wang, “Dynamic participant selection for large-scale mobile crowd sensing,” *IEEE Transactions on Mobile Computing*, vol. 18, no. 12, pp. 2842–2844, 2018.
 - [34] Y. Liu, B. Guo, W. Wu, Z. Yu, and D. Zhang, “Multitask-oriented participant selection in mobile crowd sensing,” *Chinese Journal of Computers*, vol. 40, no. 8, pp. 1872–1887, 2017.
 - [35] D. J. Hardisty, K. C. Appelt, and E. U. Weber, “Good or bad, we want it now: fixed-cost present bias for gains and losses explains magnitude asymmetries in intertemporal choice,” *Journal of Behavioral Decision Making*, vol. 26, no. 4, pp. 348–361, 2013.
 - [36] C. Miao, H. Yu, Z. Shen, and C. Leung, “Balancing quality and budget considerations in mobile crowdsourcing,” *Decision Support Systems*, vol. 90, pp. 56–64, 2016.