A High Quality Task Assignment Mechanism in Vehicle-based Crowdsourcing using Predictable Mobility based on Markov

BING JIA, HAOTIAN XU, SHUAI LIU and WUYUNGERILE LI.
Inner Mongolia A.R. Key Laboratory of Wireless Networking and Mobile Computing, Hohhot 010021, China
College of Computer Science, Inner Mongolia University, Hohhot 010021, China
Corresponding author: Shuai Liu (e-mail: cs.liu.shuai@gmail.com)

ABSTRACT
In recent years, crowdsourcing has become a research hotspot. How to formulate a reasonable task allocation mechanism to recruit the most suitable participant for the current perceptual task, and maximize the benefits of the platform has become a problem that most researchers focus on. Great efforts have been invested on task assignment mechanisms from the perspective of the platform or requesters, i.e. quality-sensitive, budget-sensitive, time-sensitive and location-sensitive. Especially for the location-sensitive task assignment mechanism, many studies motivate users to participate by some coverage estimation methods, i.e. minimizing the traveling distance. Most existing methods statically estimate the distance between the current location of the participant and the task destination, without giving any consideration about the movement track of the participant, which may result in the failure of task for the misallocation. In this paper, we propose a location-sensitive task assignment mechanism using predictable mobility based on Markov model for the vehicle-based crowdsourcing platform. Specially, we present a location transfer prediction model based on Markov model named Markov-TPM by analyzing the positional regularity of task participants during a period of time, to predict where the participant will appear in the next time period firstly. Additionally, we propose a task assignment mechanism based on Markov-TPM, which is helpful for the platform to select the most suitable participant to complete the task. Finally, experiments are carried out by using the dataset about the taxi trajectory which is collected in Shanghai, and it is shown that the proposed algorithm can improve the accuracy of the task-delivered, which is evidently superior to two algorithms compared, i.e. random prediction algorithm and prediction algorithm based on neighbor relation.

INDEX TERMS crowdsourcing, Markov, movement track, task assignment, participant

I. INTRODUCTION
Crowdsourcing is the process of obtaining needed services, ideas, or content by soliciting contributions from a large number of people, which is a hot topic in both academia and industries in the past decades [1]–[6]. In a vehicle-based crowdsourcing market, there are usually three parties, i.e. requesters, participants and the platform. Each party of the crowdsourcing market wants to obtain more benefits, so different mechanisms of task assignment need to be provided respectively to meet the different needs of the three parties [7]. For the platform, how to formulate a reasonable task allocation mechanism to recruit the most suitable participant for the current perceptual task, and maximize the benefits of the platform has become a problem that most researchers focus on [5].

Different approaches have different focuses, which is about single task allocation, multiple task allocation, low-cost task allocation, and quality-enhanced task allocation [8]. Great efforts have been invested on task assignment mechanisms from the perspective of the platform or requesters, i.e. quality-sensitive, budget-sensitive, time-sensitive and location-sensitive, which have tried to select the suitable participant to perform the tasks under the constraint of QoS, budget, completion time, and spatial relationship [8]–[12].

In the vehicle-based crowdsourcing platform, most tasks depend on the location and the movement. So, for the
location-sensitive task assignment mechanism in the vehicle-based crowdsourcing platform, the task participants need to report their current location information and the platform will make a recruitment decision based on this. With the location information, the recruitment problem can be expressed as classical spatial coverage problems. Many studies motivate users to participate by some coverage estimation methods, i.e. minimizing the traveling distance and maximizing the number of assigned tasks during the given time interval. From the past research process, participants have a single choice, which often resulting in inefficiency due to excessive manpower and material losses. The completion of the vehicle-based crowdsourcing task mainly relies on the participation of mobile devices, i.e. mobile phone and vehicle device. Besides, the participant’s mobility rules will restrict their ability to participate in the completion of the crowdsourcing task, and will also affect the quality of the crowdsourcing data [13].

Therefore, a new selection mechanism must be proposed to select the participant node based on the predictable mobility for the participant’s vehicle. Most existing methods statically estimate the distance between the current location of the participant and the task location, without giving any consideration about the movement track of the participant, which may result in the failure of task for the misallocation of tasks. We focus on the mobility characteristics of the node, then build a model based on the participant’s movement trajectory. By inferring the participant’s positional information on the next phase, we judge and match the participant who is the most suitable for the current task. Specifically, by analyzing the positional regularity of task participants during a period of time, and using the Markov model, we present a location transfer prediction model named Markov-TPM to predict where the participant will appear in the next time period firstly. Additionally, we propose a task assignment mechanism based on Markov-TPM, which is useful to help the platform to select the most suitable participant to complete the task. Finally, we conducted GPS information collection experiments on 45 taxis in different time periods within 20 days which is collected in Shanghai, and divided different GPS positioning areas. For each car, we can give the trajectory of the car within 20 days. We link the regional trajectories in different time periods to the Markov probability model. Establishing a Markov probability matrix for different vehicles. Based on this matrix we can infer where the vehicle is most likely to be in the next time period, then pick out the most appropriate participant to match the current task. At the same time, we verified the feasible of the trajectory estimation probability model by the accurate test of the experimental records.

The rest of this paper is organized as follows. Section II briefly reviews the literature in relation to task assignment in vehicle-based crowdsourcing market. Section III establishes a location transfer prediction model named Markov-TPM based on Markov. Section IV proposes a task assignment mechanism based on Markov-TPM. Section V reports the experimental results in comparison with the random prediction algorithm (Random) and the prediction algorithm based on the neighbor relationship of the current location (BON).

II. RELATED WORK

There are many ways to distribute tasks for different purposes in the crowdsourcing market [14]–[15], such as quality-sensitive, budget-sensitive, time-sensitive and location-sensitive, which have tried to select the suitable participant to perform the tasks under the constraint of QoS, budget, completion time, and spatial relationship.

In particular, for the time-sensitive task, Wu et al. formulated the bounded task allocation problem as an integer programming problem and give a constant approximated ratio for it [9]; for the task that involve selfish participants, Li et al. proposed the max-weight best response policy for strong information scenario and the proportional allocation policy for weak information scenario [12]; for the team task, Dissanayake et al. studied how the allocation of members’ social and intellectual capital within a virtual team affects team performance in online crowdsourcing contests, and confirm that a team leader’s social capital and a team expert’s intellectual capital significantly may influence team performance [10]; for the limited budget task, it is important to assign tasks to suitable participants within a limited budget and find an optimal budget allocation that maximizes the benefits of the platform; for the quality constrained task, it is important to improve the expected quality of the results, whose solving process is similar to MCSP problem [16], [17] and the GE problem [18] [19], and which is generally done by encouraging more high quality volunteers to participate in.

However, if the participants does not get paid or only receives a small reward after completing the task, the enthusiasm of participants to participate in the task again will be greatly reduced. Therefore, for the quality constrained and budget limited task, it is important to assign tasks to suitable participants within a limited budget and with the goal that maximizes the expected quality of the collective result provided by the selected participants. For example, Yu et al. jointly consider the participant’s reputation and proximity to the task locations to maximize the expected quality of the results within a limited budget [11]; for the location-based task, which depends on the location information of the participant in the crowdsourcing market, many studies motivate users to participate by some coverage estimation methods, i.e. minimizing the traveling distance and maximizing the number of assigned tasks during the given time intervals [20]–[22]. But they only statically estimate the distance between the current location of the participant and the task location, without giving any consideration about the movement track of the participant, which may result in the failure of task for the misallocation of tasks. Without location transfer prediction, the participant recruitment becomes a stochastic problem, in which the recruiter can only use probability distribution to determine which participants to recruit [23]–[27].

In this paper, we focus on the mobility characteristics
of the participant, then build a location transfer prediction model based on the participant’s movement trajectory and Markov model. By inferring the participant’s positional information on the next phase, we judge and match the participant who is the most suitable for the current task to assign the task.

III. THE PROPOSED LOCATION PREDICTION MODEL BASED ON MARKOV

The research purpose of this paper is to formulate a reasonable participant recruitment strategy from the perspective of the vehicle-based crowdsourcing platform to improve the accuracy of the task-delivered, and then reduce the cost of human and resource. Using the method of speculating the movement trajectory is helpful to assign the task to the most suitable participant. Therefore, how to infer the movement trajectory to complete the task requirements become an urgent problem to be solved.

Markov decision process is a kind of random process, whose original model is based on the Markov model and was proposed by Russian mathematician Markov, and which is a statistical model which is widely used in speech recognition, automatic part-of-speech tagging, phonological conversion, and probabilistic grammar. In general, a Markov decision process means that the transfer of each state in the process depends on the previous states, where the number of state transitions. This process is called a n-order Markov model, and the first-order Markov model is the simplest one. It is mainly defined by the following three parts: state, initial vector, and state transition matrix. Every transition to its state depends on the previous state. In this paper, we choose the first-order Markov model to construct a location transfer prediction model.

Firstly, we must calculate the regional location information of different vehicles at different stages. Then we derive a trajectory sequence in the vehicle based on these positional information. Finally, we combine the obtained trajectory sequence with Markov model to establish the trajectory state transition probability matrix of vehicles. Using this probability matrix, we can obtain a location transfer prediction model under the goal of optimizing the task assignment of nodes, which is called Markov-TPM.

Specifically, we use the following method to build a probability matrix model. Firstly, we divide all areas of the car’s movement into several areas, taking five areas as examples for illustration, which is numbered as area1, area2, area3, area4 and area5 respectively. In addition, we form a sequence of moving trajectory based on the GPS information of each car. Assume that the trajectory of the vehicle in one day is: area1 – area2 – area3 – area5 – area3 – area2 – area3 – area2 – area1 – area5 – area1 – area2 – area3 – area5. From the movement trajectory we can see that the current position of the vehicle is area1, and we can calculate that the number of times the vehicle passes through area2 immediately after area1 is 3, and the number of times the vehicle passes through area5 immediately after area1 is 1. Then we can calculate that if the participant’s current position is area1, in the next period, the probability of the destination being area2 is 3/4, and the probability of the destination being area5 is 1/4, and the probability of other area is 0. Similarly, we can calculate the corresponding probability of other regions. Based on this, we construct the trajectory transition probability matrix P of the vehicle as follows:

\[
P = \begin{pmatrix}
area1 & area2 & area3 & area4 & area5 \\
0 & 3/4 & 0 & 0 & 1/4 \\
1/4 & 0 & 1/2 & 0 & 1/4 \\
1/3 & 1/3 & 0 & 0 & 1/3 \\
0 & 3/4 & 0 & 0 & 1/4 \\
0 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 & 0
\end{pmatrix}
\]  

(1)

Where \( P_{ij} \) in the transition matrix represents the probability of transition from area\( i \) to area\( j \), and satisfies the following conditions:

\[
\sum_{j=1}^{5} P_{kj} = 1, (k = 1, 2, 3, 4, 5)
\]

The forecasting process of Markov-TPM for one vehicle is shown in Fig. III. It can be seen that the process consists of four main stages, that is, querying the current location of the vehicle, finding each possible target region and its transfer probability in the next period in the transition matrix \( P \) of the vehicle, selecting the target region which is the maximum probability as the most likely place to arrive next tim, and returning the forecast results to the vehicle-based crowdsourcing platform.

IV. TASK ASSIGNMENT BASED ON MARKOV-TPM

In real life, we need to take into account that the task participants are always moving. In order to find the most suitable task participant, we consider here only the participants closest to the origin of the task are the most appropriate participants. Therefore, if we can analyze the motion rule and trajectory of participants, we can achieve this goal.

For the vehicle-based crowdsourcing platform, the participant is the vehicle, whose most typical feature is motion. If the platform assign the task only considering whether the current location of the vehicle is the task destination,
or applying a random assignment, most tasks cannot be completed better because that the vehicle which receive the task is likely to leave the task destination at the next period. Therefore, not considering the mobility of the vehicle will result in the misassignment of the task, and will decrease the benefits of the crowdsourcing platform by increasing the cost.

Applying the Markov-TPM to predict the vehicle’s location in the next period in the platform can help to solve the problem of misassignment. Specifically, by analyzing the historical trajectory of vehicles based on Markov-TPM, we can get the movement rules of vehicles and their location prediction, which is helpful for the platform to assign the task to the vehicle most likely to get to the task destination. The process of applying the Markov-TPM for the platform for the task assignment is shown in Fig. 2. As can be seen from Fig. 2, for each vehicle, the platform apply the Markov-TPM to predict the vehicle’s location and the corresponding probability in the next period, and the platform using the prediction result to assign the task to the vehicle whose location in the next period is the task destination and whose corresponding probability is maximum.

V. EXPERIMENTS AND RESULTS
In order to verify the performance of the proposed assignment algorithm using Markov-TPM in vehicle-based crowdsourcing, extensive experiments are conducted and the experimental results are reported in comparison with the two algorithms, i.e. random prediction algorithm (Random) and the prediction algorithm based on the neighbor relationship of the current location (BON).

A. SET UP
We adopt the GPS positioning data of the taxis of Shanghai JNJ Taxi Company in April 2015, which were sampled every 10-30 seconds during the taxi driving, and recorded information including the ID number of the vehicle equipment, data receiving time, GPS measurement time, longitude, latitude and other data fields.

The main information collected is longitude and latitude of taxis located at different time periods. Take one of the records as shown in Table 1:

<table>
<thead>
<tr>
<th>id</th>
<th>taxi.id</th>
<th>time</th>
<th>longitude</th>
<th>latitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>00010</td>
<td>06:00:01</td>
<td>121.367027</td>
<td>31.303680</td>
</tr>
</tbody>
</table>

In order to ensure the authenticity and reliability of the experiment, we only collect the data of the 45 taxis from 6am to 10pm. By collecting the GPS location of the vehicle every 10 seconds, we obtain thousands of location information records. As mentioned earlier, we have obtained the latitude and longitude information of vehicles at different time periods of the day. Next, we need to classify these GPS information. We also need to classify all the activity areas of the taxi within a day. The specific approach is as follows: firstly traverse all the GPS informations in the database table, in order to find the four extreme values of GPS information, i.e., top left value, top right value, bottom left value, and bottom right value. Here, we approximate the range of all activities of the vehicle is an irregular quadrilateral. For example, we extract one of the vehicle records to get four extreme value vertices with their addresses which is shown in Fig. 3.

Additionally, we can establish a two-dimensional plane coordinates system based on these four extremum points, which uses the extremum point at the bottom left as the origin of the coordinates. The positive direction of the X-axis is pointing to right, and the positive direction of the Y-axis is pointing to the top. The maximum values of the X-axis and Y-axis are defined as the extreme values of the top left and bottom right, respectively. Then we divide the entire area coordinates system into multiple small areas and label them according to the two proportional coefficients. Each small area with its label is used to indicate the location of the participant’s vehicle. In this way, the coordinates system of the vehicle GPS is shown as Fig. 4.

Next, we use these positional information which are represented by the numbered areas to form a sequence of vehicles’ trajectories. For example, the location of a vehicle within one minute is shown in the following Table 2.

Then we can generate a trajectory for the car with id = 00010 within one minute: \(\text{area}_{234} - \text{area}_{379} - \text{area}_{123} - \text{area}_{187} - \text{area}_{125}\). So the real moving track of the car is: \(\text{area}_{234} - \text{area}_{379} - \text{area}_{123} - \text{area}_{187} - \text{area}_{125}\). In this method, we get the trajectories of these vehicles within 20 days and the relationship between the current location and the next period’s location of the vehicle, which is shown in Fig. 5.
After the previous processes, we can get the moving trajectory sequence of the participating vehicle. We use these trajectory sequences to create the Markov-TPM model for each vehicle. Then take an example to illustrate our method. Suppose the current vehicle’s moving trajectory sequence: \(a_{13} \rightarrow a_{31} \rightarrow a_{41} \rightarrow a_{43} \rightarrow a_{41}\). We can see that the vehicle has three active areas, namely \(a_{1}, a_{3}\) and \(a_{4}\). Then we can calculate that \(P_{13}=1/2, P_{41}=1/2, P_{31}=1, P_{41}=1\) and the probability of other area is 0. After obtaining these, the transition matrix based on Markov for the vehicle is formed. Additionally, we use a two-dimensional matrix to store this probabilistic model in the database. Through this probability matrix, we choose the region with the highest probability value that may appear in the next period as the most suitable participant vehicle for the current task based on Markov-TPM.

Finally, we compare the proposed algorithm based on Markov-TPM model with the random algorithm and BON algorithm. For the random algorithm, we randomly select a vehicle to assign a task whose destination is randomly selected. If the next period’s location is destination area, the task assignment is successful. Otherwise, the task assignment fails. For the BON algorithm, it is a location prediction method based on the neighbor relationship in the current region of the vehicle. We consider the neighbor area of the current location as the possible area of the vehicle in the next period, and the task destination is randomly selected. If the task destination is one of the neighbor area of the vehicle in the next period, the task assignment is successful; otherwise, the task assignment fails.

B. PERFORMANCE COMPARISON

The performance of the task assignment for the three algorithms associated with different vehicles and different total number of vehicles are reported in Fig. 6 and Fig. 7 respectively.

From Fig. 6, we can see that the accuracy of the proposed Markov-TPM algorithm is higher than the other two comparison algorithms for most vehicles. Besides, from Fig. 7, as the number of cars increases, the accuracy of the task assignment of the proposed Markov-TPM algorithm is superior to the two comparison algorithms, which indicates that our model has a positive impact on crowdsourcing task assignment. By predicting participating vehicle’s moving trajectories to assign the tasks to the most appropriate participating vehicle, the effectiveness of crowdsourcing platform is greatly improved.

VI. CONCLUSIONS

This paper proposes a new method of task assignment for the vehicle-based crowdsourcing platform. By establishing the
Markov-TPM model, we can predict the geographical position of the participants in the next period, it is helpful for the crowdsourcing platform to accomplish task assignment and reduce unnecessary human and resource cost in an optimized way. However, the model we proposed is a probabilistic model, which is not an absolutely reliable model, because the trajectory of participants also involves many other considerations. Therefore, our further research is to predict the trajectory of participants by considering other factors, i.e. preference and social relation, to improve the efficiency of the platform further.

VII. ACKNOWLEDGEMENT

This work is supported by Chinese National Natural Science Foundation of China (No.41761086, 61502254, 61761035), Program for Young Talents of Science and Technology in Universities of Inner Mongolia Autonomous Region (No.NJY18F101), and Open Funds of Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education (No.93K17018K017).

REFERENCES

BING JIA received the B.E. degree and the M.S. degree from Northeast Normal University, Changchun, China in 2006 and 2010 respectively, and the Ph.D. degree from Jilin University(JLU), Changchun, China in 2013. She is currently a lecturer in the College of computer science, Inner Mongolia University (IMU). Her research interests include wireless networks, indoor localization, crowdsourcing and mobile computing. Dr. Jia has received many research grants as a principal investigator from the National Science Foundation of China and Inner Mongolia.

HAOTIAN XU received the Bachelor degree in Tianjin Polytechnic University, majoring in computer science and technology. Now studying for a master’s degree in computer science at Inner Mongolia University. Main research direction: crowdsourcing task allocation

SHUAI LIU was born in Oct.27,1982. He received his PhD from Jilin University in 2011. Now he is an professor in College of Computer Science, Inner Mongolia University. He published more than 20 papers in Elsevier and Springer. His interesting research domains contain fractal application and image processing, computer vision, and machine learning.

WUYUNGERILE LI received the B.E. degree from Qinghai University, Xining, China in 2004, and the M.E. and Ph.D. degree from Shizuoka University, Hamamatsu, Japan in 2010 and 2013 respectively. She is currently a lecturer in the school of computer science, IMU. Her research interests include indoor WSN and mobile computing. Dr. Li has received many research grants as a principal investigator from the National Science Foundation of China and Inner Mongolia.