Multi Skill Oriented Spatial Crowdsourcing

Mr. A.S. Dalvi, Prof. P. N. Kalavadekar

Department of Computer Engineering,
SRES SCOE, Kopargaon,
Savitribai Phule Pune University, Maharashtra.
arundalvi20@gmail.com, kprak3004@gmail.com

Abstract—Crowdsourcing is an emerging business model where tasks are accomplished by the general public; the crowd. Crowdsourcing has been used in a variety of disciplines, including information systems development, marketing and operationalization. It has been shown to be a successful model in recommendation systems, multimedia design and evaluation, database design, and search engine evaluation. Despite the increasing academic and industrial interest in crowdsourcing, there is still a high degree of diversity in the interpretation and the application of the concept. Nowadays there is fast development in smartphone devices with crowd sourcing platforms. So, attention from the database community towards spatial crowdsourcing is more. Particularly, the spatial crowdsourcing sending requests to worker for their tasks using their current live positions. In this system, Admin have to take part and assume a spatial crowd sourcing system and each worker have some special qualified set of skills for spatial task like building a house, painting a wall, roof, and performing live shows for an events which is having limited time and budget constraints and qualified skill set. System will provide solution to the problem of multi-skill spatial crowd sourcing (MS-SC), it will find an important beneficial solution to worker and task assignment methodology, so that users of the system are able to match the skills of worker with the user defined tasks. By using this approach workers as well as task user will get more benefits which is maximized with budget constraint. Hence, it proves that this problem is NP-hard. So that propose a system or providing solution to the given problem with three effective approaches, with greedy, g-divide and conquer and cost-model-based adaptive algorithms to assign qualified skilled worker for user task which is beneficial for workers as well as crowds. Through this extensive experiments with crowds and worker dataset which includes there whole information i.e. skill set with respected worker and crowd with their profile, so we are going to give the efficient and effective solution to our given problem for that we will use real as well as synthetic datasets.

Index Terms—Crowdsourcing, Cost-model based adaptive algorithm, g-divide-and-conquer algorithm, Multi Skill Spatial Crowdsourcing.

I. INTRODUCTION

Consider, a spatial crowd sourcing scenario, in which each worker has a set of qualified skills, whereas each spatial task (e.g., repairing a house, decorating a room, and performing entertainment shows for a ceremony) is time-constrained, under the budget constraint and required a set of skills. Under this scenario, there is an important problem, namely called as multi skill spatial crowd sourcing (MS-SC), which finds an optimal worker and task assignment strategy, such that skills between workers and tasks match with each other and workers benefits are maximized under the budget constraint. System prove that the MS-SC problem is NP-hard and intractable. Therefore, system propose three effective heuristic approaches, including greedy, g-divide and conquer and cost-model-based adaptive algorithms to get worker-and-task assignments. Through extensive experiments, system demonstrate the efficiency and effectiveness of our MS-SC processing approaches on both real and synthetic data sets. In the system, user login into the application they have to update their profile in complete way along with their professional working data. They also select the field of the working category which helps the application to select user based on the task. Every user has the ability in the application to post any specific work on any location as per their requirement. These data are captured from the user and its been used for processing in the server side and the task allocation on android device based on the location we get to know the person available with in that location or the nearest person located in that location with skill set of same work category as required in there. In task confirmation once user is found then the user gets a notification in the application to accept the task request. After the acceptance task person is decided and the details are shared so that they can further go ahead. After the task, if the user is not satisfied or any one has any sort of complaint then they can feed in the complaint inside the application along with the task person name. In the system Admin can login from their web server and can view all the user looking for their task to be done and who is been allocated to whom. Also if any user gets a work complain of more than five times then that user is reflected in the admin side and then admin can block that user so that user cant login into the system again.

With the progress of mobile devices and wireless broadband, a new eMarket platform termed spatial crowdsourcing is emerging, which enables workers (aka crowd) to perform a set of spatial tasks (i.e., tasks related to a geographical location and time) posted by a requester. Smart devices such as smartphones and tablet computers have become much more popular and widespread over the past decade. They are now fundamental communication and computing devices in people’s everyday lives. Moreover, they can embody various sensors, including, but not limited to, microphones, cameras, GPS, gyroscope, accelerometer, magnetometer, barometer, temperature sensor, humidity sensor, and ambient light sensor. Through the Internet connection capability of smart devices, such embedded sensors have the potential to form mobile

IJETT | ISSN: 2455-0124 (E) | 2350-0808 (P) | (IF: 0.456) | Volume 4 | Special Issue July - 2017 | 8039
© IJETT
www.ijett.in
and wireless sensor networks. This potential has led to the emergence of mobile crowdsensing (MCS) [4], which can also be called sensing as a service (S2aaS) [4].

In previous work, task assignment using spatial crowdsourcing without considering location-based information to achieve better accuracy, and prior works, they studied how to select a proper worker set for a particular task. They studied crowdsourcing problems, according to location-based information as a parameter they distribute tasks to workers. In these kinds of problems, workers are not required to accomplish tasks on sites. In this MS-SC problem, my focus is on finding an assignment such that the spatial (e.g., maximum moving distances of worker) and temporal (e.g., the arrival deadlines of tasks) conditions can be met, according to skills required by the tasks can be supported by workers, and the assignment score is maximized. Thus the existing methods like this cannot be directly applied so that some spatial crowdsourcing platform requires workers to physically move to some specific locations [3], and perform the requested services by task user, such as taking photos/videos, waiting in line at shopping malls, and decorating a room. As an example, some previous works studied the small-scale or specified campaigns benefiting from participatory sensing techniques, which utilize smart devices (equipped by workers) to sense/collect data for real applications. Kazemi and Shahabi [7] classified the spatial crowdsourcing systems from two perspectives: peoples motivation and publishing models. From the perspective of peoples motivation, the spatial crowdsourcing can be categorized into two groups: reward-based, in which workers can receive rewards according to the services they supplied, and self-incentivized, in which workers conduct tasks voluntarily. In this work, studying MS-SC problem based on the reward based model, where workers are paid for doing tasks. However, with a different goal, this MS-SC problem targets at assigning workers to tasks by using our proposed algorithms, such that the required skills of tasks can be covered, and the total reward budget (i.e., flexible budget can be maximized) [1].

According to the publishing modes of spatial tasks, the spatial crowdsourcing can be also classified into two categories: worker selected tasks (WST) and server assigned tasks (SAT). In particular, for the WST mode, spatial tasks are broadcast to all workers, and workers can select any tasks by themselves. In contrast, for the SAT mode, the spatial crowdsourcing server will directly assign tasks to workers, based on location information of tasks/workers [7]. Some prior works [3], on the WST mode allowed workers to select available tasks, based on their personal preferences. However, for the SAT mode, previous works focused on assigning available workers to tasks in the system, such that the number of assigned tasks on the server side, the number of workers self-selected tasks on the client side, or the reliability-and-diversity score of assignments [5] is maximized. In particular, Cheng et al. [5] tackles the problem of reliable diversity based spatial crowdsourcing (RDB-SC), which finds a worker-and-task assignment strategy that maximizes the assignment score (w.r.t. spatial/temporal diversity and reliability of tasks). In contrast, our MS-SC problem has a different, yet more general, goal, which involves multi-skilled workers and complex tasks with the required skills studied before, and aims to maximize a different assignment score (i.e., flexible budget, given by the total budget of the completed tasks minus the total travelling cost of workers). In addition, this MS-SC problem also needs to consider several constraints, such as skill-covering, budget, time, and distance constraints, which make this problem more challenging. Due to different assignment goals between RDB-SC [5] and MS-SC, we cannot directly borrow previous techniques such as to tackle the MS-SC problem. For instance, the greedy algorithm should design effective approach to find one assignment each time with the highest increase of flexible budget in our MS-SC problem (rather than highest reliability and diversity as discussed in RDB-SC [5]); for g-DC, we propose an effective cost model to determine the best g value to maximize the MS-SC performance most importantly, we also propose a novel cost model-based adaptive algorithm, which combines the greedy and g-DivideConquer algorithms based on this cost model that can adaptively estimate the stopping level of the recursive division, minimizing the total computation cost, which have not been studied by previous works. In contrast, focus of this system is to obtain a global optimal solution to maximize the score of assignment. Different from SCP(Set Cover Problem) and its variants that cover only one universe set, this MS-SC problem is targeting to cover multiple sets, such that the assignment score is maximized. Furthermore, this MS-SC problem is also constrained by budget, time, and distance, which is much more challenging than SCP.

A. Existing System

In existing system works on spatial crowd sourcing focused on assigning workers to tasks to maximize the total number of completed tasks, the number of performed tasks for a worker with an optimal schedule or the reliability-and-diversity score of assignments. However, they did not take into account multiskill covering of complex spatial tasks, and the assignment score with respect to task budgets and workers salaries (excluding the travelling cost). Thus, we cannot directly apply prior solutions to solve our MS-SC problem.

Recently Micro-task crowdsourcing has become a popular approach to effectively tackle complex data management problems such as data linkage, missing values, or schema matching. However, the backend crowdsourced operators of crowd-powered systems typically yield higher latencies than the machine rocessable operators, this is mainly due to inherent efficiency differences between humans and machines. This problem can be further exacerbated by the lack of workers on the target crowdsourcing platform, or when the workers are shared unequally among a number of competing requesters; including the concurrent users from the same organization who execute crowdsourced queries with different types, priorities and prices. Under such conditions,
a crowd-powered system acts mostly as a proxy to the crowdsourcing platform, and hence it is very difficult to provide efficiency guarantees to its end-users.

B. The Proposed System

In this system, user login into the application they have to update their profile in complete way along with their professional working data. They also select the field of the working category which helps the application to select user based on the task. Every user has the ability in the application to post any specific work on any location as per their requirement. These data are captured from the user and its been used for processing in the server side and the task allocation on android device based on the location we get to know the person available within that location or the nearest person located in that location with skill set of same work category as required in there. In task confirmation once user is found then the user gets a notification in the application to accept the task request. After the acceptance task person is decided and the details are shared so that they can further go ahead. After the task, if the user is not satisfied or any one has any sort of complaint then they can feed in the compliant inside the application along with the task person name. In the system Admin can login from their web server and can view all the user looking for their task to be done and who is been allocated to whom. Also if any user gets a work complain of more than five times then that user is reflected in the admin side and then admin can block that user so that user cant login into the system again.

1) Solution suggested in proposed system:

- Three effective heuristic approaches, including greedy, g-divide and conquer and cost-model-based adaptive algorithms to get worker-and-task assignments, Through extensive experiments.

- We demonstrate the efficiency and effectiveness of our MS-SC processing approaches on both real and synthetic datasets.

- System will provide solution to users task as worker who is having better score.

II. System Architecture / System Overview

Fig.1 shows the architecture of proposed system which consist of following modules.

A. Dataset Collection

It includes all information about worker and user who is having some task with their attributes. Data inserted in database when workers doing their registration or filling information with their skills and characteristics then this information is added into the database, like workers user information is also added into the database into the number format and retrieval is in normal format.
H. Admin View

Admin can login from their web server and can view all
the user looking for their task to be done and who is been
allocated to whom. Also if any user gets a work complain of
more than five times then that user is reflected in the admin
side and then admin can block that user so that user cant login
into the system again.

I. Greedy Algorithm

The greedy algorithm finds one worker with highest score
for task assignment. algorithm: Processing of algorithm based
on event ids [] Step 1 :Select max count of solution.

1) Choose event\_d, by setting event\_d = group\_d, where
user\_d = group\_d are chosen by calculating max count
from dataset.

Step 2 :If check given solution is equal to input then get final
result and goto step 5. Step 3:Then check other max count of
solution but is less than previous max count solution and goto
step 1. Step 4:Then check given solution is equal to input then
got final result goto step 5. Step 5:Stop.

J. \textit{g-Divide and Conquer Algorithm}

In this algorithm it recursively divides the original problem
into sub problems and gives solution to each sub problem and
then merge that solution by resolving conflicts. The algorithm
is as shown below: Processing of algorithm based on dividing
and merging of best group. [] Step 1 :Estimation of best
number of Group g. Step 2 :Decompose or divide the MS-SC
problem into sub problems.

1) Decompose problem into sub problem(m/g)involving
spatial task .

2) If sub problems contains only one task then apply greedy
algorithm.

Step 3: Merging sub problem by resolving conflicts.

1) If we obtain assignment i.e. worker set for each sub
problem then merge them into single worker and task
assignment set.

K. Cost-Model Based Algorithm

Until the size of task group become one, till this algorithm
will not divide the problem into sub problem like previous one.
Processing of algorithm based on estimation of best number
of group. [] Step 1 :Estimation of best number of Group.
Step 2 :Total cost of solving MS-Sc problem in g-Divide and
Conquer approach is minimized.

1) Cost of Fd.
2) Decomposition of sub problem Fc.
3) Merging sub problem by resolving conflicts Fm.

L. Mathematical Model

When we solve problem we need to analyse the difficulty
level of our problem. There are three types of classes provided
for that. These are as follows:
1. P Class.
2. NP-hard Class.

3. NP-Complete Class
1. P Class( polynomial class problem ):
Class p is a class of decision problems that can be solved in
polynomial time by (deterministic) algorithms. This class of
problem is called polynomial.
2. NP-complete( Non deterministic Polynomial complete
problem ):
Class NP is the class of decision problems that can be solved
by non deterministic polynomial algorithm. This class of
problems is called non deterministic polynomial.
3. NP Hard ( Non deterministic polynomial time hard
problem ):
A problem L is NP-hard if and only if satisfiability reduces
to L. Only a decision problem can be NP complete however
optimization problem may be NP hard.

We used a set theory concept to represent the mathematical
model for the system, which consist of different sets for
input, process, rule and output. This system that is MS-SC
problem is NP-hard.

1) Set Theory: Let,

\[ S = \{I, P, R, O\} \]

Where,
\( S \): MS-SC system
\( I \): Set of inputs
\( P \): Set of processes
\( R \): Rules or constraints
\( O \): Set of outputs/Final output

1) \( I = \{i_1, i_2\} \)
Where,
\( i_1 \): Authentication.
\( i_2 \): Posting Task.
2) \( P = \{p_1, p_2, p_3, p_4, p_5\} \)
Where,
\( p_1 \): Registration of user and worker.
\( p_2 \): Log In.
\( p_3 \): Task Allocation to worker.
\( p_4 \): Task Confirmation.
\( p_5 \): Working Complaint from user.

3) \( R = \{r_1\} \)
Where,
\( r_1 \): Only valid user can log in into system.
4) \( O = \{o_1\} \)
Where,
\( o_1 \): Task Assignment.

2) Venn Diagram: Venn diagram shows the mapping of the
input, process and output relation of the system. It also represent
the interaction between different processes along with input
and output.
As system performing operation on synthetic dataset and considering some limited records for operation. System selecting event and searching for an user id who is participating in number of events with different groups, if the person or user is present in dataset then system displays the user id with max count. If such record is not present in dataset then it displays empty table or error message to the user. Partial result of greedy algorithm for some records shown in following table.

### TABLE I

<table>
<thead>
<tr>
<th>Id</th>
<th>Select Event id</th>
<th>Max Count</th>
<th>Best User id</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>80201</td>
<td>3</td>
<td>185</td>
</tr>
<tr>
<td>2</td>
<td>80200</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>80202</td>
<td>2</td>
<td>187</td>
</tr>
<tr>
<td>4</td>
<td>80101</td>
<td>2</td>
<td>188</td>
</tr>
</tbody>
</table>

#### B. Result Analysis

The performance of the system can be measured in terms of its recall and precision. Recall measures the ability of the system to retrieve all the workers that are relevant, while precision measures the ability of the system to retrieve only the workers that are relevant. They are defined as:

\[
\text{Precision} = \frac{\text{Number of relevant workers retrieved}}{\text{Total number of retrieved}}
\]

\[
\text{Precision} = \frac{A}{A + B}
\]

\[
\text{Recall} = \frac{\text{Number of relevant workers retrieved}}{\text{Total number of relevant workers}}
\]

\[
\text{Recall} = \frac{A}{A + C}
\]

Where A represent the number of relevant workers that are retrieved, B, the number of irrelevant workers and the C, number of relevant workers those were not retrieved. The number of relevant workers retrieved is the number of the returned workers that are similar to the query worker in this case. The total number of workers retrieved is the number of workers that are returned.

### TABLE II

<table>
<thead>
<tr>
<th>Iteration</th>
<th>Number of relevant workers Retrieved</th>
<th>Total number of irrelevant workers Retrieved</th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>17</td>
<td>10</td>
<td>0.62</td>
<td>0.41</td>
</tr>
<tr>
<td>2</td>
<td>9</td>
<td>7</td>
<td>1</td>
<td>0.33</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>10</td>
<td>0.62</td>
<td>0.41</td>
</tr>
<tr>
<td>4</td>
<td>11</td>
<td>9</td>
<td>0.66</td>
<td>0.4</td>
</tr>
<tr>
<td>5</td>
<td>19</td>
<td>17</td>
<td>0.54</td>
<td>0.46</td>
</tr>
</tbody>
</table>

Partial result for above iteration in terms of Precision and Recall is shown in the following table.
IV. CONCLUSION

In this paper, we are providing a solution to the problem of (MS-SC) multi-skill oriented spatial crowdsourcing, in which we are going to assign the time constrained and multi-skill-required spatial tasks with dynamically moving workers from one location to another. So that the skills of that worker will apply on the present task. This task can be covered by skill set of workers so that the task assignment score with worker skill is maximized by using the heuristic approaches that are the algorithm which we are implementing in this system. So that this system will give us a skilled worker set for an appropriate task with user requirements with time constrained as well as budget constraints, so that user can further choose the skilled worker for their task by their score. We are proving that the processing of the MS-SC problem is NP-hard, and thus we propose three approximation approaches (i.e., greedy, g-DC, and cost model-based adaptive algorithms), which can efficiently give MS-SC solution.

REFERENCES

[1] Peng Cheng, Xiang Lian, Lei Chen, Member, IEEE, Jinsong Han, Member, IEEE, and Jizhong Zhao, Member, IEEE “Task Assignment on Multi-Skill Oriented Spatial Crowdsourcing”, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 28, NO. 8, AUGUST 2016.


[8] http://arxiv.org/abs/1510.03149.pdf This is reference website, where the Research paper is available.


The Foursquare app helps you discover new places, with recommendations from a community you trust. Find a better experience, anywhere in the world.


