Implementation of Multi Skill Spatial Crowdsourcing Processing Approaches

Prof. Gurpreet Kaur Bhatti[1], Bhagyashree Deshmukh[2], Namrata Kumbhar[3], Sonali Naiknaware[4], Arati Shinde[5]
[1, 2, 3, 4, 5] Dept of Computer Engineering
[1, 2, 3, 4, 5]SPPU, Pune

Abstract- *Data mining is the computing process of discovering pattern in large data sets. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for future use. With the rapid development of mobile devices and crowdsourcing platforms, the spatial crowdsourcing has attracted much attention from the database community. Specifically, the spatial crowdsourcing refers to sending location-based requests to workers, based on their current positions. The system consider a spatial crowdsourcing scenario, in which each worker has a set of qualified skills, whereas each spatial task such as 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. For developing such system different methods and techniques are used such as greedy, g divide- and-conquer and cost-model-based adaptive algorithms to get worker-and-task assignments, the work should be d one within time and under budget. The system introduces a task assignment on multi-skill oriented spatial crowdsourcing to demonstrate the efficiency and effectiveness of our MS-SC processing approaches on both real and synthetic data sets.*

Keywords- Multi-skill spatial crowdsourcing, Greedy algorithm, g-Divide and Conquer algorithm, Cost-Model-based adaptive algorithm.

I. INTRODUCTION

Data mining is computing process of discovering pattern in large data set. With the popularity of GPS equipped smart and device and wireless mobile network people can easily have identified and handle.Crowd-sourcing platform assign a number of moving workers to do spatial task nearby which required worker to move some one specified location, under budget constraint set required set of skills. For developing such a system different method and techniques are used such as greedy ,g divide-and-conquer and cost-model-best adaptive algorithm to get workers and task , the work should be done within time and under budget. The system introduce a task assignment on multi-skill oriented spatial crowd-sourcing to demonstrate the efficiency and effective all also MS-SC process approaches on both real and synthetic data set. A framework called spatial crowdsourcing for employing workers to conduct spatial tasks, has emerged in both academia and industry.

In other words, these complex tasks cannot be simply accomplished by normal workers, but require the skilled workers with specific expertise (e.g., fixing roofs or setting up the stage). Inspired by the phenomenon of complex spatial tasks, in this consider an important problem in the spatial crowdsourcing system, namely multi-skill spatial crowdsourcing (MS-SC), which assigns multi-skilled workers to those complex tasks, with the matching skill sets and high scores of the worker-and-task assignments.

II. MULTISKILLED WORKERS

System contains Universe set of skills of workers and each worker has the skills which are the part of that universe set of skill. Proposed system has the universe set of workers. Universe set of skills and universe set of workers are associated with each other. Each worker has a skill that are present in universe set of skills.

III. THE MULTI-SKILL SPATIAL CROWDSOURCING PROBLEM

In multi-skill spatial crowdsourcing problem, which assigns spatial tasks to workers such that workers can cover the skills required by tasks and the assignment strategy can achieve high scores. MSSC problem considers travelling distance of workers, budget and deadline of task. Proposed system considers all these parameters to solve the MSSC problem.

IV. PRUNING STRATEGIES

Pruning is a technique in machine learning that reduces the size of list by removing the section of list that provide little power to classify instances. Pruning reduces the complexity of the final classifier, and hence improves predictive accuracy by the reduction of overfitting. There are different pruning strategies are use to reduce the list in meaning full way. This system uses following three pruning strategies.

1. If given worker – task pair is {(wa,tj) (wb,tj)} then system consider budget and feedback ,travelling cost of worker to choose a worker task pair. Worker wa has more skills than the worker wb and also less travelling cost than wb then (wb,tj) pair prune from instant set.
2. If a given worker task is {(wa,tj) (wa,tk) (wb,tk)} then system consider a worker who can do more task. Wa worker can do task tk and tj also therefore (wb,tk) pair is prunes from instant set.
3. If unassigned available worker has more travelling cost which is greater than budget of task then system not assign task to that worker.

V. PROPOSED ARCHITECTURE

There are two sections one is user section and another is a worker section.

Firstly user and worker need to give their information for registration purpose. In registration phase user also fix username and password. After registration user can login to the system by entering username and password. User can add the task and required information which specifies a set of required skills, deadline of task and budget. All these information stored in database. Worker need to mention their skills and location while registration.In system, user's requirement is matching with worker's skills when it is matched then task is assign to worker by checking status. When worker accept that task his status becomes busy.

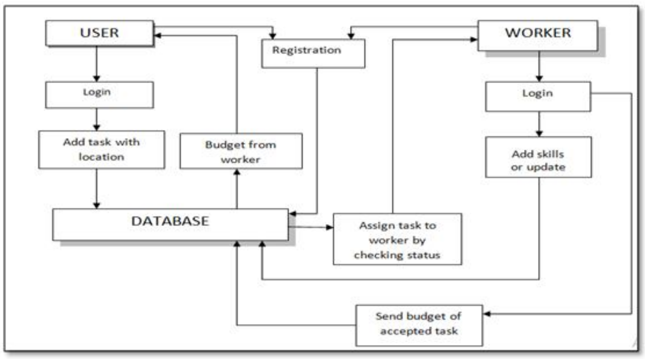

Fig.1. Proposed Architecture

VI. MATHEMATICAL MODEL

Let S be a system which assign the task to worker.
S= {…}
Identify Input as I
S= {I,}

Let I= {u,w,t}

The input will be the user database, Worker database and task assigned by worker.

Identify Output as O

S= {I, O…}

O = User assigned task will assigned to multi-skill worker according to budget and time.

Identify the processes as P

S= {I, O, P...}

P= {Ta,Tc}

Ta=task will be allocated to worker.

Tc=task will be completed by worker and add status as free.

Identify the failure as F

S= {I, O, P, F…}

F= Failure occurs when the system fails to assign the task to worker

Identify Success as s

S= {I, O, P, F, s.}

s=success occurs when task is successfully allocated to skilled worker

Identify Initial Condition as Ic

S= {I,O,P,F,s,Ic}

Ic= Registration of workers

VII.METHODOLOGY

In general, the complex tasks cannot be simply accomplished by normal workers, but it is simple when skilled workers with specific expertise which assigns multi-skilled workers to those complex tasks, with the matching skill sets.

VIII. ALGORITHMIC STRATEGY

1. Greedy Approach:

In the greedy algorithm, which iteratively assigns a worker to a spatial task that can always achieve the highest score increase shows the pseudo code of MS-SC greedy algorithm, which obtains one worker-and-task pair with the highest score increase each time, and returns a task assignment instance set.

2. g-Divide-and-Conquer Approach:

The first algorithm incrementally finds one worker-and-task assignment at a time; it may incur the problem of only achieving local optimality. Therefore, in this section, an efficient g-divide-and-conquer algorithm (g-D&C), which first divides the entire MS-SC problem into sub problems, such that each sub problem involves a smaller subgroup of spatial tasks, and then conquers the sub problems recursively. Since different numbers of the divided sub problems may incur different time costs, cost - model-based method to estimate the best value. During the recursive process combine/merge assignment results from subgroups, and obtain the assignment strategy for merged groups, by resolving the assignment conflicts among subgroups. Finally the task assignment instance set Ip, with respect to the entire worker and task sets.

MS-SC Problem Decompositions

In this section to decompose a MS-SC problem into sub problems. In order to illustrate the decomposition, first convert our original MS-SC problem into a representation of a bipartite graph. Bipartite graph representation of the MS-SC problem. Specifically, given a worker set Wp and a spatial task set Tp, denote each worker/task (i.e., wi or t j) as a vertex in the bipartite graph, where worker and task vertices have distinct vertex types. There exists an edge between a worker vertex wi and a task vertex tj, if and only if worker wi can reach spatial task tj under the constraints of skills ,time and budget and the worker-and-task assignment pair is valid, if there is an edge between vertices wi and tj in the graph. Decomposing the MS-SC problem will illustrate how to decompose the MS-SC problem, with respect to task vertices in the bipartite graph. An example of decomposing the MS-SC problem into three sub problems, where each sub problem contains a subgroup of one single spatial task, associated with its connected worker vertices. In a general case, given n workers and m spatial tasks, partition the bipartite graph into g subgroups, each of which contains spatial tasks, as well as their connecting worker presents the pseudo code of our MS-SC problem decomposition algorithm, namely MS-SC Decomposition, which returns g MS-SC sub problems, Ps, after decomposing the original MS-SC problem.

Merging conflict reconciliation

In this section, the merging conflict reconciliation procedure, which resolves the conflicts while merging assignment results of sub problems. Assume that Ip is the current assignment instance set we have merged so far. Given a new sub problem Ps with assignment set the merging algorithm, namely MS-SC Conflict Reconcile, which combines two assignment sets Ip resolving conflicts. In particular, two distinct tasks from two sub problems may be assigned with the same worker wi. Since each worker can only be assigned to one spatial task at a time, we thus need to avoid such a scenario when merging assignment instance sets of two sub problems.

3. Cost-Model-Based Adaptive Approach:

A cost-model-based adaptive approach, Initially, estimate the cost, of applying thegreedy approach over worker/task sets. Similarly, it can also estimate the best group and compute the cost, If it holds that the cost of the greedy algorithm is smaller than that of the g-D&C approach then it will use the greedy algorithm by invoking function MS-SC Greedy otherwise g-D&C algorithm are used shows the pseudo-code of our cost-model-based adaptive algorithm, namely MS-SC Adaptive. Initially estimate the cost, cost greedy, of applying the greedy approach over worker/task sets Wp and Tp estimate the best group size, g, and compute the cost, cost gd &c of using the g- D&C algorithm. If it holds that the cost of the greedy algorithm is smaller than that of the g-D&C approach, then we will use the greedy algorithm by invoking function MS-SC Greedy (due to its lower cost). Otherwise, we will apply the g-D&C algorithm, and further partition the problem into sub problems Ps Then, for each sub problem Ps, recursively call the cost-model-based adaptive algorithm, and retrieve the assignment instance set After that, merge all the assignment instance sets from sub problems by invoking function MS-SC Conflict Reconcile. Finally, we return the worker -and-task assignment instance set Ip.

Cost Model for the Stopping Condition

To determine the stopping level, when using our cost -model-based adaptive approach to recursively solve the MS-SC problem. Intuitively, at the current level k, need to estimate the costs, cost greedy and cost g-d-c, of using greedy and g-D&C algorithms, respectively, to solve the remaining MS-SC problem. If the greedy algorithm has lower cost, then will stop the divide-and-conquer, and apply the greedy algorithm for each sub problems. In the bipartite graph of valid worker-and-task pairs, denote the average degree of workers as (degw), and that of tasks as (degt), the computation of valid worker and- task pairs cost. Since there are at most n iterations, for each round apply two worker-pruning methods to at most pairs, and select pairs with the highest score increases, which need cost in total. For the cost of task-pruning, there are totally n rounds in each round, there are at most degw out of m tasks that may be potentially pruned .To check each of degw tasks, need cost. Therefore, the total cost of task-pruning is given If we cannot prune a task that was assigned with a worker in the last round, then need to update score increases of deg workers for that task.

IX. RELATED WORK

Spatial crowdsourcing without considering the location information in crowdsourcing, previous works studied the task assignment to achieve better accuracy, and prior works studied how to select a proper worker set for a particular task. Prior works like, usually studied crowdsourcing problems, which treat the location in formation as a parameter and distribute tasks to workers. In these problems, workers are not required to accomplish tasks on sites. In our MS-SC problem, we focus on finding an assignment such that the spatial (e.g., maximum moving distances of worker) and temporal (e.g., the arrival deadlines of tasks) constraints can be met, the skills required by the tasks can be supported by workers, and the assignment score is maximized. Thus the existing methods cannot be directly applied.

1) A crowdsourcing worker quality evaluation algorithm on Map reduce for Big Data Application, IEEE2015

   With the development crowdsourcing system, the data size of crowdsource, contractor and task grows rapidly. With respect to all that is applied to any critical task such as tagging, matching, filtering and other emerging application without wasting resources.

2) Reliable diversity based spatial crowdsourcing by moving worker, IEEE2015

   In this paper, consider an important spatial crowdsourcing problem, namely RDB-SC in which spatial task are constraint and worker are moving towards the same direction. RDB-SC problem is to assign worker to spatial task such as completion, reliability and spatial diversity of spatial task are maximized.

3) Crowdsource POI labelling: location-aware result inference and task assignment, IEEE2016

   In this paper, by observing that crowdsourcing is the best fit for computer hard task, crowdsourcing is improving the quality of POI labelling.

4) Spatial crowdsourcing: current state and future direction.

   In this paper, discuss unique challenges of spatial crowdsourcing, provide a comprehensive view of this new paradigm by introducing taxonomy and give future direction.

X. RESULT

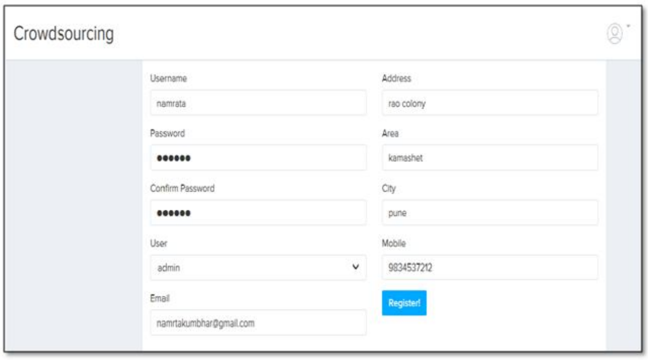

Fig.2. Registration Page

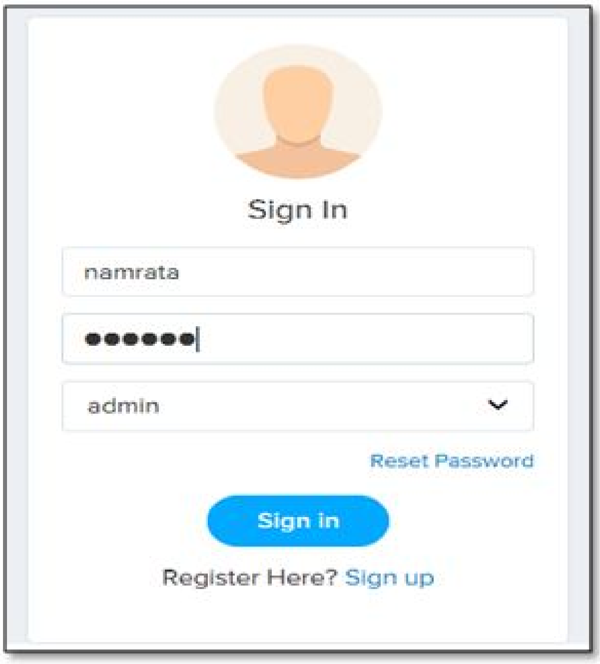

Fig.3. Login Page

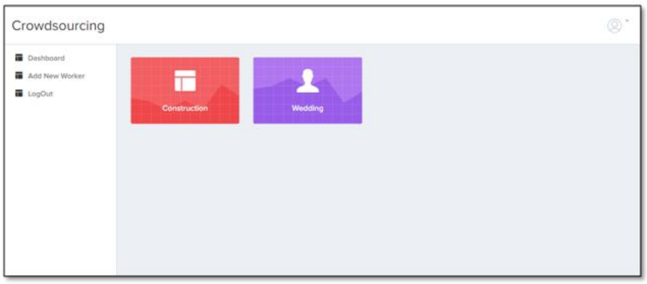

Fig.4. Dashboard

Fig.5. Constrution Plan

Fig. 6. Post Construction Work

XI. CONCLUSION

In this paper the problem of the multi-skill oriented spatial crowdsourcing (MS-SC), which assigns the time constrained and multi-skill-required spatial tasks with dynamically moving workers, such that the required skills of tasks can be covered by skills of workers and the assignment score is maximized and that the processing of the MS-SC problem is NP-hard, and thus propose three approximation approaches (i.e., greedy, g-D&C, and cost model- based adaptive algorithms), which can efficiently retrieve MS-SC answers. Extensive experiments have shown efficiency and effectiveness of our proposed MS-SC approaches on both real and synthetic data sets.

REFERENCES

[1] Zhe Yang, Zhehui Zhang, Yuting Bao, Xiaoying Gan, Xiaohua Tian, Xinbing Wang. “On Tac: Online Task Assignment for Crowd sourcing”, In IEEE 2016

[2] Habibur Rahman, SenjutiBasu Roy, Saravanan Thirumuruganathan, SihemAmer-Yahia, Gautam Das. “Task Assignment Optimization in Collaborative Crowdsourcing”, In 2015 IEEE.

[3] IoannisBoutsis, VanaKalogeraki, ” On task assignment for real-time reliable crowdsourcing”, IEEE 2014.

[4] Khanh-Hung Dang and Kim-Tuyen Cao, “Towards Reward-based Spatial Crowdsourcing”, ICCAIS 2013.

[5] IoannisBoutsis, VanaKalogeraki, “Crowdsourcing under Real-Time Constraints”, 2013 IEEE

[6] Z. B. G. L. J. F. Huiqi Hu, Yudian Zheng and R. Cheng, “Crowd-sourced poi labelling: Location-aware result inference and task assignment”, ICDE 2016.

[7] S. S. Kanhere, “Participatory sensing: Crowdsourcing data from mobile smartphones in urban spaces”, In MDM 2011.

[8] M. F. Bulut, Y. S. Yilmaz, and M. Demirbas.“Crowdsourcing location based queries”,In PERCOM Workshops 2011, pages 513518.

[9] P. Cheng, X. Lian, Z. Chen, R. Fu, L. Chen, J. Han, and J. Zhao.“Reliable diversity-based spatial crowdsourcing by moving workers”, VLDB 2015,8(10).

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

[11] P. Cheng, X. Lian, L. Chen, J. Han, and J. Zhao. Task assignment on multi-skill oriented spatial crowdsourcing (technical report). http://arxiv.org/abs/1510.03149.

[12] Google street view. https://www.google.com/maps/views/streetview. It gives location of user .

[13] C.-J. Ho and J. W. Vaughan. Online task assignment in crowdsourcing markets. In AAAI, volume 12, pages 45–51, 2012.

[14] Y. Zheng, J. Wang, G. Li, R. Cheng, and J. Feng. Qasca: A quality aware task assignment system for crowdsourcing applications. In SIGMOD-15, pages 1031–1046. ACM.

[15] K. Wang, X. Qi, L. Shu, D. J. Deng, and J.J.P.C. Rodrigues, “Toward trustworthy crowdsourcing in social internet of things”, *IEEE Wirel.Commun.*, vol. 30, no. 5, pp. 30-36, Oct. 2016.