

Towards Classification Exploration in Spatial Crowdsourcing Domain: A Systematic Literature Review

Shirley Salimun¹, Masnida Hussin², Rohaya Latif³, Shamala K. Subramaniam⁴

^{1,2,3,4}Department of Communication Technology and Network, Faculty of Computer Science and Information Technology, University Putra Malaysia (UPM), Selangor, Malaysia
²masnida@upm.edu.my

Article History: Received: 10 November 2020; Revised: 12 January 2021; Accepted: 27 January 2021;
Published online: 05 April 2021

Abstract: Today, spatial crowdsourcing concept has been widely applied in various fields. The increasing of mobile user and adoption of social network has catalyst spatial crowdsourcing growth. It has made various types of data to be easily collected and transmitted from different geographical location. However, the massive amounts of task in spatial area bring challenges for the online system to manage especially when the task is heterogeneous, and the interactions are dynamic. Such scenario has alerted the researchers to understand different types of information in order to make task assignment reliable and efficient. This study investigates current state of task assignment for spatial crowdsourcing. It basically, aims to identify several issues like trend in publication and crowd computing areas that studies task assignment in crowdsourcing. We used Systematic Literature Review (SLR) method for analysing the trends and significance of task classification for better dynamic crowd-computing.

Keywords: Systematic Literature Review (SLR), task assignment, crowdsourcing, spatial task.

1. Introduction

Spatial crowdsourcing concept has been widely applied in various fields such as transportation and commercial food delivery. The reasons of increasing in usages is because it is easily been applied due to low expenditure and distributed faster compare to traditional business model. It is a concept where worker must travel to task location to complete the task. In addition, with the advancement of mobile technology and the emergence of Internet of Thing (IoT) catalysts the concept growth rapidly. Spatial crowdsourcing consists of three main elements that play a crucial role in the process which are requester, system platform, and crowd worker. The requester outsourced the task (i.e., taking photo in certain place) and the platform will assign the task to the crowd worker. The system platform plays an essential role in ensuring the overall process is success. In spatial crowdsourcing system, the tasks and crowd workers are heterogeneous and has dynamic interactions. The arrival and departure of the tasks and crowd workers in the system are uncertain. They might leave the system at any time due to various factors such as tasks expiry or due to malicious behaviour. Consequently, uncertain conditions lead to risking incomplete task assignment. The incomplete assignment will affect the system reliability, hence reducing the number of requestors. It further affected the whole crowdsourcing system.

To optimize the task assignment, it is also important to ensure that the tasks are assigned to suitable (reliable and cost-effective) crowd worker. For example, the location of the task and worker plays an essential role in allocation decision-making. It is because it affects the willingness of the worker to travel to the task location. If the task location is too far from the worker location, the task might be left unattended. In addition, human behaviour factor such as malicious behaviour might affect the reliabilities of task assignment. Hence, it is important for the spatial crowdsourcing system to study which worker to be assigned to which spatial tasks. Therefore, the objective of this study is to investigate the current state of task assignment in the spatial crowdsourcing in regards computing scope. To conduct this review, we use a systematic literature review (SLR) method. The finding of this study gives an insight about the current trend in spatial crowdsourcing, further brings viewpoints for other researchers to work on future works.

2. Research Methods

To conduct this study, we adapted Systematic literature review (SLR) method to guide us evaluating and interpreting available research on the crowdsourcing. It is useful to develop supporting evidences and eliminate bias during the research process (Ali, 2018). The SLR method consists of three main phases which are planning review, conducting review, and reporting review.

2.1 Phase 1: Planning the review

In the planning review, we formulate research questions to narrow down the articles search results. To formulate the research questions, we used Population, Intervention, Comparison, Outcomes, and Context (PICOC) by referring to the authors in (Petticrew and Roberts, 2006). It able to help us for structuring the research questions. As concluded, we structured the research questions as follows; “*what are the focus issues in the task assignment of spatial crowdsourcing?*”.

2.2 Phase 2: Conducting the review

In this phase, there are few strategies that have been employed. The first strategy is to derive keywords from the research questions and reconstruct it into search string to find relevant articles in the digital libraries databases. The search string used in this study is (“spatial” OR “spatial crowdsourcing” OR “crowdsourcing” AND “assign” OR “assignment”). The second strategy is to select digital libraries for retrieving comprehensive published studies. The digital libraries selected in this study are ACM Digital Library, IEEE Explore, Google Scholar, Mendeley, and Science Direct. The selected digital libraries are subscribed by the University Putra Malaysia's (UPM) Library. Meanwhile Google Scholar and Mendeley are considered as a coverage across the boundaries of individual database that been used also by the authors in (Geiger and Schader, 2014). Some of the articles are also retrieved using ResearchGate webpages and backward references search. During the searching process, the articles are searched based on its relevancy towards the topic. The range date of the articles published is not limited. Some features in the digital libraries such as advance search setting and search keywords within abstract were used to narrow down the search process.

The third strategy is to screen and analyse the selected articles using the inclusion and exclusion criteria. It means to filter out articles that did not meet the study requirements. The inclusion criteria are included all articles that published in English, within the sort of relevance search, the type of publication (i.e., journal and conference proceeding). The articles must focus on the task assignment in spatial computing scope. In the other hand, the exclusion criteria include the articles that are not published in English language, published in other than journal and conference proceeding, has less than 3 pages. Such articles discussing the task assignment but not in specific for spatial scope. The qualities of the filtered selected articles were then assessed using the quality study assessment criteria. Table 1 shows the quality assessment criteria used in evaluating each article. The quality assessment consists of four questions (Q1- Q4), where each of the question given a score: Yes =1; partially = 0.5; No = 0. Based on the score given, the articles will be rated from 0 (very poor) to 4 (very good).

Table 1. The study's quality assessment

No	Assessment criteria	Answer
Q1	Are the objectives of the research is clearly explained?	Yes / No / Partially
Q2	Is the research of the article coherent?	Yes / No / Partially
Q3	Is the research supported with a primary data?	Yes / No / Partially
Q4	Is the research approach or method clearly explained?	Yes / No / Partially

2.3 Phase 3: Reporting the review

The reporting process includes synthesizing the data and reporting the finding which were further discussed in next section (Result and Discussion). The data synthesis is the process of extracting the information from selected articles that been answered according to our identified research question. To extract and synthesis the retrieved articles, we used Mendeley version 1.61.1 thus it records the references details for each identified crowdsourcing scopes. We categories the articles based on the year of publication, type of the articles, methods, abstracts, and the scopes.

3. Results and Discussion

The findings and arguments of the investigation work is explicitly described and illustrated. There is given figures and tables as evidence to support the prior investigation process. Overall, there are 198 articles that were deemed to be relevant to the research question. The retrieved articles are then screened and analysed based on their titles and abstracts. The results shown that there are 87 articles closely relevant to our scope. Later, those articles are filtered based on the inclusion and exclusion criteria. During the process all the duplicated articles, irrelevant articles or articles which did not meet the exact requirement are excluded from the article selection process. The rest of the articles were then evaluated using the study quality assessment criteria. After the filtering process, we come at 38 articles. All the selected articles are classified as good and very good articles. There are 35 articles which score

very good quality (92.1%) and three (3) papers at good quality (7.9%). Table 2 illustrates the filtering results of the quality assessment criteria for all final screening articles.

Table 2. The quality assessment result

Quality Scale	Poor (0 <1)	Fair (1-<2)	Good (2-<3)	Very Good (3-4)	Total
Number of studies	0	0	3	35	38
Percentage (%)	0	0	7.9	92.1	100

Then, those 38 selected articles are synthesized as a supporting evidence to address the research questions. In this study, the publication types of articles are selected from journal and conference proceedings. Hence, the chapter, eBook, patents etc. were excluded during the process selection. Based on the results, the conference publication for task assignment in crowdsourcing is started in 2012. General, the result (in Figure 2) shows that there are three (3) conferences proceeding published in 2012, four (4) articles published in 2015 and three (3) articles published in 2016. There is exponential increased of conference articles published in between 2016 and 2019. In other hand, the journal publication shows that it is started published in 2014. The amounts of journal published were then significantly increased in 2017 where five (5) articles published, also it reached five (5) journals published in 2019. It is probably extension work from the conference articles in the previous year.

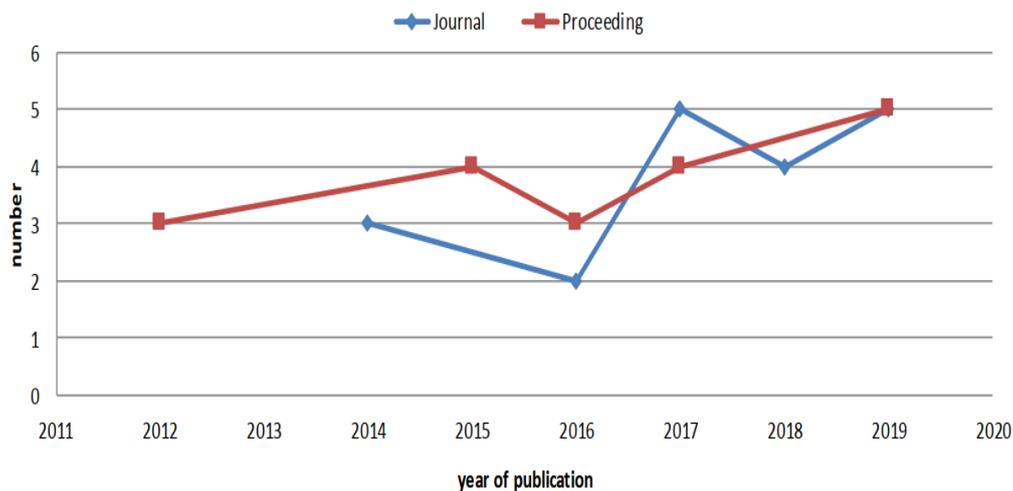


Figure 1.Publication articles

The increasing publication articles might be influenced by the advancement of the Internet, mobile technologies, and a rapid evolvement of social media. The advancement of these technologies had catalyzed massive amount of data in the networks which contains spatial information such as geolocation, photos, video, and etc (Chi et al., 2017). The vast amount of heterogeneous data and dynamic interaction within the environments brings challenges for the crowdsourcing system to understand and differentiate the data behaviour. In addition, the data collected or stored often inflected with noisy signals and repetitive waves hence it is difficult to produce clear and smooth data (Hassan & Curry, 2016). Consequently, it is affecting the task assignment decision-making process. It leads to unreliable and inefficient computing process. Therefore, further investigation on spatial task criteria needed to further study. From our prior investigation on the publication trends, it reveals that the researchers are studied within the same scope of issues. Surprisingly, most of the articles are mentioned on the issue of understanding the crowdsourcing task assignment through classification matter. It is identified from our synthesized articles there are subject of task classification that been studied for (i) privacy, (ii) semantic, (iii) sensing, (iv) location, (v) scheduling and (vi) software-testing domain. Figure 2 illustrates the matter of crowdsourcing task assignment gained from the prior studied. For privacy and sensing subject matters there are ten (10) and eight (8) articles, respectively. Some articles in the matters are inter-related to each other that makes it influenced in the reading. Meanwhile, there are nine (9) articles focused on scheduling while five (5) articles relating to location matters. Meanwhile, three (3) articles focused on each software testing domain and semantic, respectively

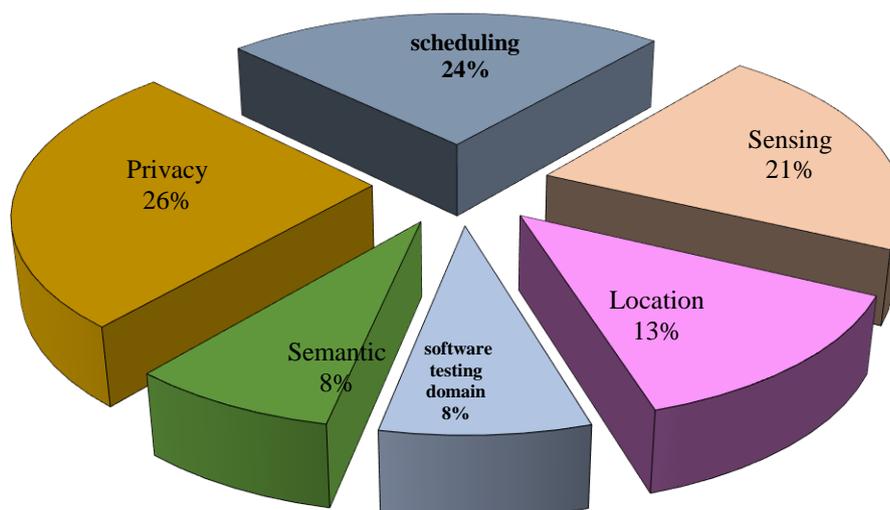


Figure 2. Subject focus of spatial crowdsourcing task assignment

Based on the results, we found that the focus of the task classification in the spatial crowdsourcing is different from the traditional crowdsourcing (Hassan & Curry, 2016). The task assignment approach in spatial crowdsourcing is different due to the spatio-temporal nature of tasks that makes the tasks take longer duration to be completed compared to the traditional crowdsourcing. In traditional crowdsourcing, the task classification is more focused on optimizing the quality of result by the crowd workers (Dekel & Sridharan, 2012; Ho et al., 2013). Most of the literatures focused on enhancing the crowdsourcing system abilities to predict and analyse accuracy of output/outcome provided by the crowd worker (Alabduljabbar, R. et. al, 2019; Hussin, M., 2018). In another example, Hassan and Curry (Hassan et al., 2013) used Bayesian approach to predict the worker performance on the new assignment tasks. Meanwhile, in Kazemi and Shahabi (2012) they proposed the maximum task assignment where later extended to maximum score assignment for optimizing the number of assignment and travel distance. The authors in (Cheng et al, 2014) also focuses on optimizing the task reliabilities while catering task diversity. In a similar research direction, Cui et.al (2017) used the agents to learn and adopt from human task allocation strategies into their task classification scheme. From the other perspective of the task classification in the spatial crowdsourcing, privacy issue becomes one of the major concerned. It is not a surprise because the crowd-computing applications that collected and stored the tasks can be reside at everywhere. It also allows the requesters and workers to engage in a wide range of interaction. This situation makes the requesters (users) or even the workers at risk of serious privacy threats (Zhang, X. et.al, 2019). Some of the researchers are focused on protecting the users' locations from being exposes. For example, Park et.al (2017) used Entropy-Maximizing Observation Function and identification algorithm to protect the user identity. Ma et al., (2017) proposed APPLLET frameworks for encryption, and prediction the recommendation results to the user. Aside from the privacy of the data, there are many researchers focused on the semantic perspective in spatial crowdsourcing in order to understand the data or description provided by the user. For example, the authors in (Vasardani et al., 2012) used task classification scheme to improve the interaction between users and services by examines the use of preposition "at" in a set of crowdsourced place descriptions. Leung and Newsam (2012) emphasised on extracting geographic information semantically for land-use classification. The researchers in (Richter et al., 2012; Hussain et al., 2017; Ramírez, J. et al, 2019) focused on understanding the types of place descriptions and Donget.al (2017) focused on probabilistic location information. Meanwhile in the sensor perspective, the researchers focused on classifying different type of sensors (Chi et al., 2017 and Mudavath et al., 2020) as a medium for collecting the data. Boutsis et.al (2016) developed the social sensing system that achieved sampling through mobile social sensors in order to accurately detect the real-time state of emergency events. On the other hand, for the location area perspective, most of the researchers are studied on how to improve the search query. For example, Shim et.al (2018) focused on enhancing the locations and the friendship between the requesters and workers. The authors in (Monteiro et.al, 2017) used task classification to improve the contextual information in order to help disaster monitoring and responding. Last but not least, we also realized that the crowdsourced has related with the software testing domain. In this area most of the studies using the crowd as the subject matter for software testing. The matter really gains the benefit from the crowdsourcing platform due to overwhelming of testing requirements which required various documentation while maximizing the number of respondents for gaining the accurate results. For example, Feng et.al (2016) proposed the Spatial

PyramidMatching (SPM) technique and natural-language processing technique for collecting the output and views from the public users and the experts that they integrated in the crowdsourcing platform. It helps for effective, quick and reliable communication.

4. Conclusion

Spatial crowdsourcing concept has opened service opportunities to engage and utilize a high-volume number of potential resources or workers. Its operation has significantly contribute toward the organisation success. However, despite the abundance of its advantage, it also brings daunting challenges which could affect organization as whole. Due to the spatial crowdsourcing environment itself, it is quite challenging to fulfill those service requirements. There are massive amounts of demand/task request which is heterogeneous and has dynamic interaction between the agents. In this study, we have conducted a systematic literature review to gain a better understanding about current state of how the resources assigned and classified accordingly and what are the focus has been highlighted in the previous studies. The finding of our study shows that there are an increasing study has been conducted emphasized on the classification since 2012 until 2019. Based on the finding, we also identified several significant subject matters of the spatial crowdsourcing which is (i) privacy, (ii) semantic, (iii) sensing, (iv) location, (v) scheduling and (vi) software-testing domain. By understanding the subject matter in spatial crowdsourcing, we hope that it could help for in the future where the spatial crowdsourcing can be utilized and further improved.

References

1. Alabduljabbar, R., & Al-Dossari, H. (2019). A dynamic selection approach for quality control mechanisms in crowdsourcing. *IEEE Access*, 7, 38644-38656.
2. Ali, M. B., Wood-Harper, T., & Mohamad, M. (2018). Benefits and challenges of cloud computing adoption and usage in higher education: a systematic literature review. *International Journal of Enterprise Information Systems (IJEIS)*, 14(4), 64-77. <https://doi.org/10.4018/IJEIS.2018100105>
3. Boutsis, I., Kalogeraki, V., & Gunopulos, D. (2016). State Detection Using Adaptive Human Sensor Sampling. *Fourth AAAI Conference on Human Computation and Crowdsourcing*.
4. Cheng, P., Lian, X., Chen, Z., Fu, R., Chen, L., Han, J., & Zhao, J. (2014). Reliable Diversity-Based Spatial Crowdsourcing by Moving Workers. *Proceedings of the VLDB Endowment*, 8(10), 1022-1033. <https://doi.org/10.14778/2794367.2794372>.
5. Chi, M., Sun, Z., Qin, Y., Shen, J., & Benediktsson, J. A. (2017). A Novel Methodology to Label Urban Remote Sensing Images Based on Location-Based Social Media Photos. *Proceedings of the IEEE*, 105(10), 1926-1936. <https://doi.org/10.1109/JPROC.2017.2730585>
6. Cui, L., Zhao, X., Liu, L., Yu, H., & Miao, Y. (2017). Complex crowdsourcing task allocation strategies employing supervised and reinforcement learning. *International Journal of Crowd Science*, 1(2), 146-160. <https://doi.org/10.1108/IJCS-08-2017-0011>.
7. Dekel, O., & Sridharan, K. (2012). Selective Sampling and Active Learning from Single and Multiple Teachers. *Journal of Machine Learning Research*, 13, 2655-2697.
8. Dong, T., Ishikawa, Y., Xiao, C., & Zhao, J. (2017). k-Expected Nearest Neighbor Search over Gaussian Objects. *Journal of Computers*, 12(2), 105-115. <https://doi.org/10.17706/jcp.12.2.105-115>
9. Feng, Y., Jones, J. A., Chen, Z., & Fang, C. (2016). Multi-objective test report prioritization using image understanding. *Proceedings of the 31st IEEE/ACM International Conference on Automated Software Engineering - ASE 2016*, 202-213. <https://doi.org/10.1145/2970276.2970367>
10. Geiger, D., & Schader, M. (2014). Personalized task recommendation in crowdsourcing information systems - Current state of the art. *Decision Support Systems*, 65(C), 3-16. <https://doi.org/10.1016/j.dss.2014.05.007>
11. Hussain, A., Mkpojiogu, E.O.C., Nawi, M.N.M. (2017). Capturing customer satisfaction and dissatisfaction in software requirements elicitation for features in proposed software systems. *Journal of Engineering and Applied Sciences*, 12 (21), pp. 5590-5597.
12. Hussin, M., & Rozlan, N. A. (2018, September). Assessing workers reliability in crowd-sourced computing using Bayesian rules. In *AIP Conference Proceedings* (Vol. 2016, No. 1, p. 020055). AIP Publishing LLC.
- Ho, C.-J., Jabbari, S., & Vaughan, J. W. (2013). Adaptive Task Assignment for Crowdsourced Classification. *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, 28,

- 534–542.
http://machinelearning.wustl.edu/mlpapers/papers/ICML2013_ho13.
13. Kazemi, L., & Shahabi, C. (2012). GeoCrowd: Enabling Query Answering with Spatial Crowdsourcing. *Proceedings of the 20th International Conference on Advances in Geographic Information Systems - SIGSPATIAL '12*, 189. <https://doi.org/10.1145/2424321.2424346>.
 14. Leung, D., & Newsam, S. (2012). Can off-the-shelf object detectors be used to extract geographic information from georeferenced social multimedia? *Proceedings of the 5th International Workshop on Location-Based Social Networks- LBSN '12*, 12. <https://doi.org/10.1145/2442796.2442801>
 15. Ma, X., Li, H., Ma, J., Jiang, Q., Gao, S., Xi, N., & Lu, D. (2017). APPLLET: a privacy-preserving framework for location-aware recommender system. *Science China Information Sciences*, 60(9).
 16. Monteiro, J., Kitamoto, A., & Martins, B. (2017). Situational Awareness from Social Media Photographs Using Automated Image Captioning. *2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, 203–211. <https://doi.org/10.1109/DSAA.2017.59>
 17. Mudavath, M., Kishore, K.H., Hussain, A., Boopathi, C.S. (2020). Design and analysis of CMOS RF receiver front-end of LNA for wireless applications. *Microprocessors and Microsystems*, 75, art. no. 102999.
 18. Park, J. D., Seglem, E., Lin, E., & Züfle, A. (2017). Protecting User Privacy: Obfuscating Discriminative Spatio-Temporal Footprints. *Proceedings of the 1st ACM SIGSPATIAL Workshop on Recommendations for Location-Based Services and Social Networks*, 2:1–2:4. <https://doi.org/10.1145/3148150.3148152>
 19. Petticrew, M., & Roberts, H. (2006). *Systematic Reviews in the Social Sciences: A Practical Guide*. In *Cebma.Org*. <https://doi.org/10.1027/1016-9040.11.3.244>.
 20. Ramírez, J., Baez, M., Casati, F., & Benatallah, B. (2019, October). Understanding the impact of text highlighting in crowdsourcing tasks. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing (Vol.7, No. 1, pp. 144-152)*.
 21. Richter, D., Winter, S., Richter, K.-F., & Stirling, L. (2012). How People Describe their Place: Identifying Predominant Types of Place Descriptions. *Proceedings of the 1st ACM SIGSPATIAL International Workshop on Crowdsourced and Volunteered Geographic Information*, 30–37. <https://doi.org/10.1145/2442952.2442959>
 22. Shim, C., Kim, W., Heo, W., Yi, S., & Chung, Y. D. (2018). Nearest close friend search in geo-social networks. *Information Sciences*, 423, 235–256. <https://doi.org/10.1016/j.ins.2017.09.049>
 23. Ul Hassan, U., & Curry, E. (2016). Efficient task assignment for spatial crowdsourcing: A combinatorial fractional optimization approach with semi-bandit learning. *Expert Systems with Applications*, 58, 36–56. <https://doi.org/10.1016/j.eswa.2016.03.022>
 24. ul Hassan, U., Curry, E., Hassan, U., & Curry, E. (2013). A Capability Requirements Approach for Predicting Worker Performance in Crowdsourcing. *Proceedings of the 9th IEEE International Conference on Collaborative Computing: Networking, Applications and Worksharing*, 429–437. <https://doi.org/10.4108/icst.collaboratecom.2013.254181>
 25. Vasardani, M., Winter, S., Richter, K.-F., Stirling, L., & Richter, D. (2012). Spatial interpretations of preposition “at.” *Proceedings of the 1st ACM SIGSPATIAL International Workshop on Crowdsourced and Volunteered Geographic Information - GEOCROWD '12*, 46. <https://doi.org/10.1145/2442952.2442961>
 26. Zhang, X., Wu, Y., Huang, L., Ji, H., & Cao, G. (2019). Expertise-aware truth analysis and task allocation in mobile crowdsourcing. *IEEE Transactions on Mobile Computing*