A Data Aggregation Framework based on Deep Learning for Mobile Crowd-sensing Paradigm

Junjie Pang, Zongduo Liu

Abstract—**Mobile crowd sensing is a effective sensing solution. Many IoT applications are benefit from this sensing method and has larger sensing coverage, higher quality sensing data. In MCS platform, human resource becomes the new type of sensing resources. The wearing devices and human mobility improve the IoT sensing coverage. However, the worker performance can effect the sensing quality directly. For the edge-enabled MCS platform designs, though they have brought significant success in flexible computation/storage resource management and alleviating the work-loads of end-devices and clouds, are commonly concentrate on resolving the offloading strategy and task allocation mechanism. Most of them fail to take enough consideration about introducing quality guarantee of sensing data into edge-enabled MCS system design. To fulfill our goal, we use EM algorithm to deduce the data quality in the proposed framework. Furthermore, we use deep learning techniques to filter the sensing data for further analysis. Simulation validates our method with significant results.**

Index Terms—**mobile crowd sensing, data aggregation, deep learning, data quality.**

I. Introduction

MOBILE crowd sensing(MCS), which is a hybrid human-machine sensing paradigm[1][2] that leverage the crowd intelligence to solve urban sensing problems. It originates from the concept of crowdsourcing and achieve success by leveraging human power as the extended sensing resource. With the development and proliferation of mobile devices in recent years, the mobile phones, laptops or unmanned aerial vehicles(UAVs) are equipped diverse sensing facilities and with longer battery life to help mobile device users to perform well in various sensing task of MCS applications. Past decades witnessed the MCS applications achievements in various areas, such as the noise pollution monitoring that provided by NoiseTube, the road conditions based traffic planning that provided by Nericell and PotHole, and so on. Such successful attempts make MCS a versatile platform to complement existing IoT architecture with the incorporation of human power and sensors capabilities, which is further explained by Fig.1.

Junjie Pang and Zongduo Liu are with the College of Computer Science and Software Engineering, Qingdao University, Qingdao, Shandong 266000, China.

Along with the substantial growth in MCS platforms, the scale and range of sensing tasks is increasing while the computational and energy resource of the mobile devices is still limited. Mobile sensing tasks that include multimedia information collection and text description are become the common case in MCS applications which keep the pressure on the computation and storage resources of the cloud and push up the service promotion cost. Moreover, the growing amount of data transition/preprocessing requirements between the cloud servers and end users adds substantial stress to the limited communication resources[3].

Therefore, many researchers bring edge computing concept into the MCS platform designs[4] to address the above-mentioned challenges. According to these successful attempts, edge enabled MCS platforms[5] make the sensing data possible to be preprocessed and analyzed in the edge of the network for bandwidth saving and latency reduction[6]. And it also means that the volume of sensing data can be scale down through preprocessing procedure on the edge, which can reduce the network traffic between edge and cloud severs to a large extent. Besides, edge-enabled MCS is valuable to end users since they are able to offload their tasks to the edge resources that closed to their current locations to save the energy consumptions without task abort.

However, the edge-enabled MCS platform designs, though having brought significant success in flexible computation/storage resource management and alleviating the workloads of end-devices and clouds, are commonly concentrate on resolving the offloading strategy and task allocation mechanism[7]. Most of them fail to take enough consideration about introducing quality guarantee of sensing data into edge-enabled MCS system design[8]. Generally, participants provide their sensing data with diverse quality based on different capabilities and effort levels. Meanwhile, there are also some malicious participants tend to maximize their rewards by using fake data. Such uncertainty that hide in their behaviors makes the quality guarantee becomes a real challenge. Intuitively, moving the quality estimation method to the edge and incorporated with appropriate preprocessing procedure[9] may give MCS the capability to guarantee data quality for various

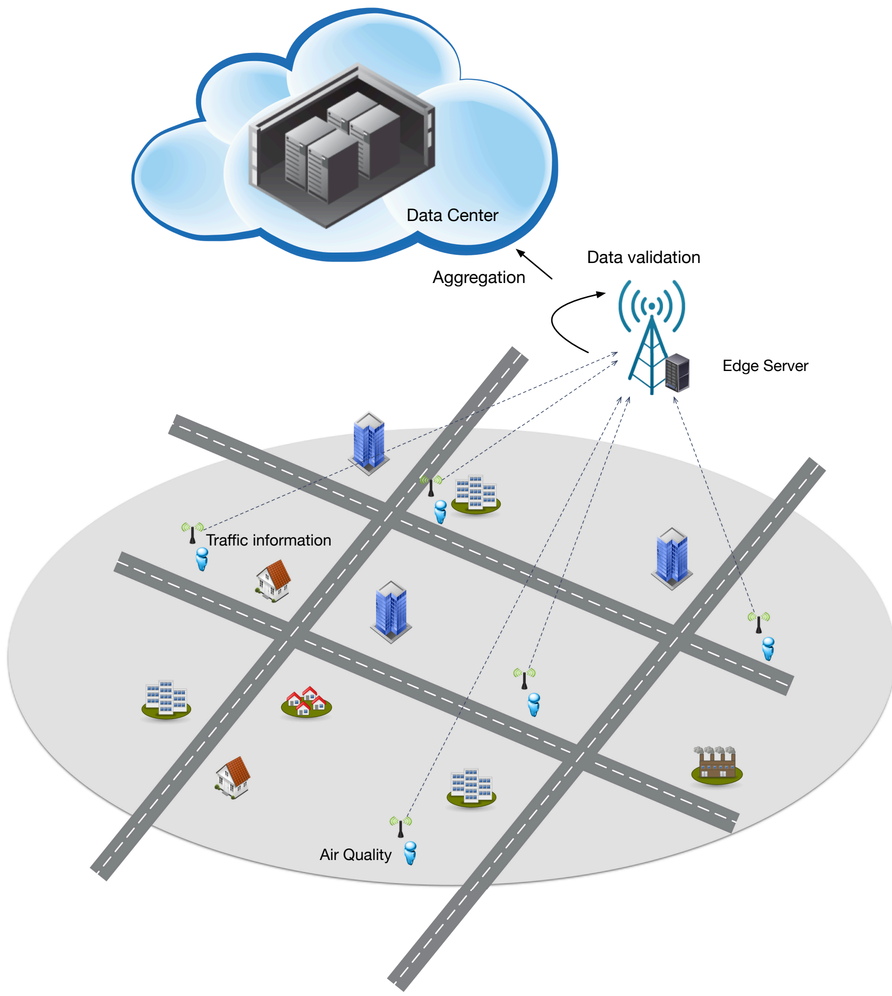

Fig. 1. An Example Application Scenario

type of sensing data effectively, which motivate us to design a quality-driven framework of MCS platforms with edge computing support. To the best of our knowledge, few existing works have investigated such problem and the primary challenges and difficulties are summarized below:

First of all, it is challenging to design a quality guarantee framework that take both structure and unstructured data into consideration. With more and more sensing tasks need participants provide multimedia information like pictures, audios and videos, such heterogeneous data required appropriate preprocessing and quality estimation method [10]to maintain the sensing data availability among different data types.

Second, most of existing sensing tasks are location awareness since they need to gather the desired data in specific locations. With the consideration of location-aware character, recently works takes great effort on design location-aware incentive mechanism[11] and task allocation strategy to meet the geography coverage requirements. Unfortunately, the quality estimation method neglect that the location variety will also affect the data quality and participant effort estimation[12]. For example, a sensing task often covered a number of area which has different environment conditions, participant groups and difficulties[13]. The significant difference makes the sensing data may drop into different probability distributions. Lacking in consideration of location diversity, existing truth discovery and quality estimation methods have limitations when deal with tasks with multiple geographic areas.

Third, it is nontrivial to bridge the gap between finding quality estimatation results and making guarantee of sensing data quality for every stage of life cycle in MCS platforms. Though quality estimation method could help us to measure participants behavior in every single task, a carefully designed participator assessment method is required to maintain a long-term observation of their behaviors and contributions among multiple tasks types[14].

And the participator selection strategy, incentive mechanism and reward distribution should based on a contribution quantification from a comprehensive behavior analysis[15]. The reason is that the participators working history may reflect the variation trend of data quality in the future. A quality driven MCS framework which make full use of the quality estimation results throughout the whole life cycle will continuous encourage the participators to keep a high effort level to make effective contributions. While the traditional MCS framework that using uniform pricing method and auction mechanism are unlikely to encourage the high-quality contribution effectively so that it fail to provide the long-term quality guarantee.

Therefore, we explore the space of quality guarantee framework designs in edge-enabled mobile crowd sensing platforms and propose a carefully designed framework to resolve above challenges by:

In the rest of the paper, we review related work in Section 2, and discuss our framework designs and motivations in Section 3. Sections 4 present the case study to better explain the practical advantages of the framework, which are evaluated in Section 5. Section 6 concludes the paper and give the future work discussions.

II. Related Work

Mobile crowd sensing[16] is a new type of urban sensing infrastructure which based on the concept of crowd sourcing. By leveraging the sensing ability of the mobile devices, the sensing paradigm empowers human participants to collect data for various urban sensing applications. As a hybrid human-machine system, mobile crowd sensing has a number of successful attempt[17], such as NoiseTube[18], SoundofTheCity[19], and UAir[20]. With the rapid development of mobile crowd sensing paradigms, many researchers realize that there are still many challenges and problems in the area of incentive mechanism, task allocation and framework designs need to be resolved well the coverage requirement and limited budget restrictions.

A. Incentive Mechanism Designs with Different Purpose

For the incentive mechanism design,. In these earlier research on incentive mechanisms, only few works took data quality into considerations in their designs, which lead to their rewards have weakly connections with how well the participant accomplish the task[21]. To fill the gaps, pay as well first proposed a quality estimation method based on EM algorithm which is inspired by quality based pricing[22]. In addition, they apply the classical Information Theory to measure the participant contribution rely on data quality and then resolved the reward distribution strategy. In quality-driven, they proposed a quality-driven auction based incentive mechanism,which use data utility to represent sensing data quality instead of working time to calculate the participant pay-off. In a case study of indoor localization system, they define the data quality by a relevance threshold and perform a series experiments to validate the efficiency and system profitable. In high quality vehicle, a quality-aware participant recruitment method is proposed for a vehicle-based crowdsourcing paradigm. In their cases, the quality requirement is related to the spatial and temporal coverage so that they used a trajectory prediction method to improve high quality crowdsouring.

Besides, most of incentive mechanism needs a effective reward settings. To keep the reward can motivate users to participant to the tasks, the reward setting needs to be changed when the former setting has graded performance. A well designed reward is based on analysis on sensing task types, target users' behaviors, and the system budget. In most of application scenarios, there is a apparent contradiction between system budget and effective reward setting. To face the challenge, some research start to find a solution instead of meeting the requirement of clients number. Sparse mobile crowd sensing is emerged. It use seed clients to gather a part of information. Then, the others sensing area use data inference to get enough sensing data. This method can alleviate the budget limitation problem and achieves well performance in recent years.

However, existing works are still limited when the sensing paradigm have a large scale of heterogeneous sensing task and various kind of sensing data. A more general quality guarantee framework is needed to provide a long-term motivation for high-quality contributions.

B. Data Quality Estimation in MCS

Different from the incentive mechanism research, data quality evaluation and estimation tasks are mainly focus on evaluating the contributed data from workers, and the primary goal is evaluating the quality of data or the trustworthiness of workers. Based on these works, MCS platform can have a deep analysis on the behaviors of workers, then it can make informed, further decisions to tell whether the worker can be trust, or improved the pricing strategy. There are many methods to assess workers performance, like reputation based solution can assess the trustworthiness of both workers and their contributions data by using centralized reputation value and individual vote-based collaborative reputation[23] values at the same time. EndorTrust system uses a trust-based worker relationship and collaborative filtering to evaluate workers behaviors and also give the prediction results, which can be a reference for future task allocation. From the existing research we can see that a well-designed data quality estimation method can help MCS platform to be robust and improve the data trustworthiness[24]. There are also another type of research use truth discovery to re-processing the original sensing data, eliminating the errors to make the sensing data more accurate to the real conditions. We called these are post-evaluation method. Most of data aggregation frame work can be benefit from these solutions.

C. Exploiting the Power of Deep Learning in Mobile Crowd Sensing Paradigms

Beyond that, there are also some efforts have been made in the area of deep learning based data processing. Firstly, deep learning methods are often used in time-series analysis issues[25]. For most IoT application scenarios, due to increasing size and source of available time-series data streams, there is a need for a powerful analysis tool with much smaller storage space and faster processing capability. Deep learning often serves as the primary step to make a deep insight in time correlation and semantic relationship construction[26], which is very important in IoT data aggregation tasks[27]. However, there are many challenges to deal with the time serial data[28]. For example, some relations in IoT data are not liner relationship, a graph may better catch their correlations. But a graph based solution is always a daunting task. For this challenge, graph based neural network is emerged and give potential solutions to resolve the difficulties. Powered by deep learning techniques, graph based neural network can process a large graph with lower computation expenses and higher performance. It provide a new solution for graph based data processing tasks.

In contrast, our work differs from existing research in the following aspect: We re-define the quality guarantee problem in a cache-enabled mobile crowd sensing paradigm, which has various kind of raw sensing data and task; To incorporated with multimedia data, we design a sensing data preprocessing procedure to provide the basis of quality estimation method and reduce the traffic between edge server and cloud.

III. Edge Computing-Enabled Mobile Crowdsensing with Data Quality Guarantee

In this section, we first give the mobile crowdsensing model as a motivating example, and then give a deep learning based method to learn the features from sensing data with multiple types and use edge computing to obtain the trade-off between efficiency and high computation cost. Finally, the formulation of contribution quantification method is conducted to make the quality guarantee in our general framework, which is the core challenge in our design.

A. System Overview

In this section, we provide an example scenario of our framework that illustrated by figure 2. The system consists of three layers:sensing data collection layer,data preprocessing layer and data aggregation layer with quality estimation.

B. Sensing Cata Collection Step

The basic function of sensing data collection layer is to collect data according to the sensing task description by participants with various sensors or devices. In most of the cases, the participants will be asked to accumulating sensing data by specific sensors and time-location requirement which can be convenient for sensing data pre-processing. Under such restrictions and rules, the task coverage restriction and the worker capability requirement may becomes a contradiction since it is often difficult to find enough people to accomplish the sensing tasks. For example, there is a sensing task has a need of sensing data coverage higher than 95 of the target area. It indicates that the whole task may be result in vain if some of the area cells do not collect enough data to meet the standards. Though the data correlation method can address the data sparsity problem, it will fall into trouble when some sensing data of key area cell is missing. Meanwhile, the limited system budget even make the things worse. Assumed that the budget is fixed, people's income will be affected when the competent participants claim a higher cost. Above all, the reality is that a MCS platform is hard to obtain well coverage and data quality at the same time. Such obstacles push existing research take great effort on sensing task assignment method. Such methods pay their hopes on finding a reasonable allocation strategy by a new-defined geographic region to save the sensing cost and reach the ideal target of sensing coverage requirement closely.

Instead of struggling with above problems and trade-offs, we propose a new view point: discover the participants' potential of themselves more than specific sensors restrictions. In real practice, people could sense the surroundings by different ways such as a text description, images and videos. Each of these data source may contains useful information. Obviously, if the MCS platform can lower the capability barriers, the sensing coverage will significantly improved since more and more people can be evolved. Motivated by such thoughts, the sensing data collection layer in XXX system design takes full consideration of heterogeneous capabilities among potential participants: everyone who have willingness can be regarded as an sensing data resource to achieve the tasks. Since every coin has its two side, such thought encounters other limitations and challenges such as the high-dimensional, data noise and heterogeneous data sources. To resolve these challenges, we leverage the power of deep learning assisted method in the sensing data aggreagation layer to integrate the sensing data and support task-critical applications. The details of the sensing data aggregation designs will be discussed in next section.

C. Sensing Data Aggregation Step

Apparently, if the MCS platform relaxed the restrictions of the requirement of specific sensors, it means the sensing data could be heterogeneous among various data formats and resources. Take an air pollution task as a example, participants can use a text describe the smell of the air in target location, or simply take a picture of the sky instead of using specific sensors comes from a home air monitor. The lower threshold means more people could be included in the sensing task so that the coverage ration would be increase. After being collected, the multi-media data and text description will be cleaned to remove noise and then the feature will be extracted by a pre-trained NLP model and image processing model to accompolish the multi-modal feature fusion process. Therefore, different from other system design that only provide data pre-processing function to achieve standardization or normalization in this layer, we follow the standard process of data aggregation in our design to filter and generate valuable information from these multiple disparate data modalities. With collabration of deep learning techniques, MCS platform has the inference ability to make use of all available sensing data. Though it is hard to obtain accurate sensor readings from multi-media data and text data, there is sill a positive effect on sensing data correlation learning and missing data inference since they provide such methods an accurate range rather than no-data.

D. Quality Estimation and Contribution Quantification Step

We consider a general case that the avaliabilty and preciseness of sensing data is significantly affects the service of various MCS applications. Intuitively, a contribution quantification that based on data quality estimation could be a long-term effective method for MCS system to calculate each participant's material rewards and enable them to provide high-quality data all the time. The quality estimation among multiple data source has a number of chanllenges: firstly, quality estimation could not rely on the ground truth because we often lack of the prior knowledge of the sensing tasks. Secondly, it is challenging to bridge the gap between quality of sensing data and heterogenous data sources. Different participant may submit different data to make their own sensing results that are not always directly comparable. Thirdly, the geographic differences will affect participants performance evaluation. For a sensing task that include a large sensing area, it is more reasonalbe to compare participants' quality within their community instead of the whole sensing area. Above all, the quality estimation should not only have punishment on poor performance and malicious behavior but also have positive reinforcement to any piece of available sensing data. To address such problem, we incorporated the consideration of sensing location variations into the quality estimation method and used EM algorithem to calculate each participants' sensing data quality and find the ground truth at the same time.

1) There is not enough sample data has positive correlation to the ground truth in most of the sensing task, especially in the new tasks;
2) When the model makes a mistake, some sensing data will be wrongly filter out and the participant will be regarded as a malicious worker. Such error would be harmful to the participation, and therefore, have negative affect to reach the coverage requirement for new tasks.
3) Beyond that, the low-quality sensing data can be combined to provide better quality results-which is the main concept of crowd wisdom. Therefore, filter out the irreverent data in a too aggressive way may not suitable in mobile crowd sensing scenario.

Consider the challenges and difficulties, we replace the data validation by designing a sensing data preprocessing scheme in a mild way to retain the valuable information among each piece of data. Edge computing support. We illustrated the proposed deep learning based data preprocessing scheme in Figure 2. From the above process, the valuable information of sensing data can extracted. For the sensing task with multimedia data, the data size can be largely reduced and more convenient for data quality estimation since most of quality estimation methods have difficulties to deal with multimedia data. Take image data as a example, Figure 3 shows the comparison of reduced data size and origin input data size. After the preprocessing, XX of data have been filtered.

IV. Quality Estimation for Edge enabled MCS

After we obtain the data after deep learning empowered pre-processing procedure, we need to evaluate the data quality and give the feedback to the learning models. Then, the model can improve itself to adapt the data aggregation requirements. In this section, we review the traditional method for quality estimation. Then, we describe our proposed solution.

A. EM based Quality Evaluation

Algorithm 1 EM based Quality Evaluation Algorithm

Input: Observation data $x = (x^1, x^2, ;x^m)$, joint distribution $p(x, z; \theta)$, conditional distribution $p(z|x; \theta)$ maximun iteration times J.

Output: Model parameters θ

1: The initial value θ^0 random initialization parameters θ model.
2: for j from 1 to J to start the EM algorithm iteration:
3: Step E:
4: Calculate the conditional probability expectation of the joint distribution:

$$Q_i(z^{(i)}) = P(z^{(i)}|x^{(i)}, \theta^j)$$

$$L(\theta, \theta^j) = \sum_{i=1}^{m} \sum_{z^i} Q_i(z(i)) log P(x^{(i)}, z^{(i)}; \theta)$$

5: Step M:
6: Maximize $L(\theta, \theta^j)$, get θ^{j+1}:

$$\theta^{j+1} = argmax_{\theta} L(\theta, \theta^j)$$

7: If θ^{j+1} has converged, the algorithm ends. Otherwise, continue to return to step 3) to perform step E iteration.

For m sample observation data, the model parameters θ of the sample $x = (x^1, x^2, \ldots, x^m)$, we use a log-likelihood function that maximizes the model distribution is as follows:

$$\theta = arg \max_{\theta} \sum_{i=1}^{m} log P(x^{(i)}; \theta)$$

If the obtained observation data has unobserved implicit data$z = (z^{(1)}, z^{(2)}, ...z^{(m)})$, then the log-likelihood of our maximization model distribution. The function is described as follows: However, it is hard to find the above formula directly θ. Therefore, some tricks are needed to resove this problem. We first scale this formula as follows: In this equation, we use the Jensen inequality as follows:

$$log \sum_j \lambda_j y_j \geq \sum_j \lambda_j log y_j \quad, \lambda_j \geq 0, \sum_j \lambda_j = 1$$

Meanwhile, if there is a need to satisfy the Jensen's inequality, we can use follows formulas:

$$\frac{P(x^{(i)}, z^{(i)}; \theta)}{Q_i(z^{(i)})} = c, cisaconstant$$

Since $Q_i(z^{(i)})$ is a probability distribution, it satisfies follows restrictions:

$$\sum_z Q_i(z^{(i)}) = 1$$

From the above two formulas, we can get: If $Q_i(z^{(i)}) = P(z^{(i)}|x^{(i)}; \theta))$, then formula is a lower bound of our log-likelihood containing hidden data. If we can maximize this lower bound, we are also trying to maximize our log-likelihood. That is, we need to maximize the following formula:

$$arg \max_{\theta} \sum_{i=1}^{m} \sum_{z^{(i)}} Q_i(z^{(i)}) log \frac{P(x^{(i)}, z^{(i)}; \theta)}{Q_i(z^{(i)})}$$

Remove the constant part in the above formula, then we need to maximize the lower limit of the log-likelihood:

$$arg \max_{\theta} \sum_{i=1}^{m} \sum_{z^{(i)}} Q_i(z^{(i)}) log P(x^{(i)}, z^{(i)}; \theta)$$

The above formula is the M step of our EM algorithm. Note that $Q_i(z^{(i)})$ is a distribution in the above formula, so

$$\sum_{z^{(i)}} Q_i(z^{(i)}) log P(x^{(i)}, z^{(i)}; \theta)$$

can be understood as the expectation of $log P(x^{(i)}, z^{(i)}; \theta)$ based on the conditional probability distribution $Q_i(z^{(i)})$.

B. Convergence Analysis of our Proposed algorithm

Convergence of EM algorithm. To prove the convergence of the EM algorithm, we need to prove that the value of our log-likelihood function has been increasing during the iteration. which is:

$$\sum_{i=1}^{m} log P(x^{(i)}; \theta^{j+1}) \geq \sum_{i=1}^{m} log P(x^{(i)}; \theta^j)$$

Since we have follow equations:

$$L(\theta, \theta^j) = \sum_{i=1}^{m} \sum_{z^{(i)}} P(z^{(i)}|x^{(i)}; \theta^j)) log P(x^{(i)}, z^{(i)}; \theta)$$

And

$$H(\theta, \theta^j) = \sum_{i=1}^{m} \sum_{z^{(i)}} P(z^{(i)}|x^{(i)}; \theta^j)) log P(z^{(i)}|x^{(i)}; \theta)$$

These two subtractions above give following features:

$$\sum_{i=1}^{m} logP(x^{(i)};\theta) = L(\theta,\theta^j) - H(\theta,\theta^j)$$

In the above formula, we take θ as θ^j and θ^{j+1} respectively, and then we can obtain following equations:

$$\sum_{i=1}^{m} logP(x^{(i)};\theta^{j+1}) - \sum_{i=1}^{m} logP(x^{(i)};\theta^j)$$
$$= [L(\theta^{j+1},\theta^j) - L(\theta^j,\theta^j)] - [H(\theta^{j+1},\theta^j) - H(\theta^j,\theta^j)]$$

To give the convergence analysis of the proposed EM based quality evaluation algorithm, we only need to prove that the right part of the above formula is non-negative. Since θ^{j+1} makes $L(\theta,\theta^j)$ extremely large, there are:

$$L(\theta^{j+1},\theta^j) - L(\theta^j,\theta^j) \geq 0$$

And for the second part, we have:

$$H(\theta^{j+1},\theta^j) - H(\theta^j,\theta^j) = \sum_{i=1}^{m}\sum_{z^{(i)}} P(z^{(i)}|x^{(i)};\theta^j)\; log\frac{P(z^{(i)}|x^{(i)};\theta^{j+1})}{P(z^{(i)}|x^{(i)};\theta^j)}$$
$$\leq \sum_{i=1}^{m} log(\sum_{z^{(i)}} P(z^{(i)}|x^{(i)};\theta^j)\; \frac{P(z^{(i)}|x^{(i)};\theta^{j+1})}{P(z^{(i)}|x^{(i)};\theta^j)})$$
$$= \sum_{i=1}^{m} log(\sum_{z^{(i)}} P(z^{(i)}|x^{(i)};\theta^{j+1})) = 0$$

Thus, we have:

$$\sum_{i=1}^{m} logP(x^{(i)};\theta^{j+1}) - \sum_{i=1}^{m} logP(x^{(i)};\theta^j) \geq 0$$

Finally, we prove the convergence of the EM algorithm.

V. Performance Evaluation

To test the proposed method, we run several experiments in a simulating environments. We randomly generate a number of sensing tasks, the data requirement including text, image and numerical data. We use a CNN based model to accomplish the data aggregation task, using data inference to deduce the semantic information hiding in these multi-model data from multiple sources. The simulation result is depicted in figure 2.

For validate our approach, we test the proposed model in terms of accuracy. We observed when number of users is increasing, all of the three approach's accuracy are growing. However, the random method and greedy method

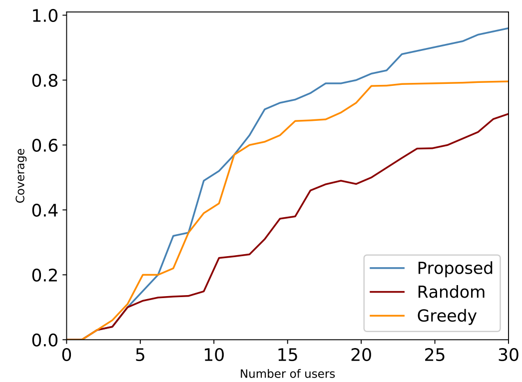

Fig. 2. Accuracy Comparison

fail to maintain the increasing trend. Our approach can maintain increase.

We also validate our approach in terms of coverage ratio and the result is depicted in figure.3. We observed when number of users is increasing, all of the three approach's coverage ratio are growing. However, the random method and greedy method end the increasing trend more quickly and our proposed method can increase the coverage ratio more quickly and maintain the stable coverage ratio for a longer period.

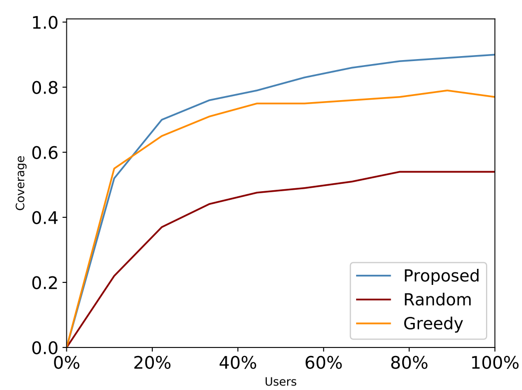

Fig. 3. Coverage ratio Comparison

VI. Conclusion

In this paper, we consider a general case that the availability and preciseness of sensing data are significantly affecting the service of various MCS applications. Intuitively, a contribution quantification based on data quality estimation could be an effective long-term method for the MCS system to calculate each participant's material rewards and enable them to provide high-quality data all the time. In existing research, the summary that quality estimation among multiple data sources has several challenges: firstly, quality estimation could not rely on the ground truth because we often lack prior knowledge of the sensing tasks. Secondly, it is challenging to bridge the gap between sensing data quality and heterogeneous data sources.

However, the different participants may submit additional data to make their sensing results that are not always directly comparable. Thirdly, the geographic differences will affect participants' performance evaluation. For a sensing task that includes a large sensing area, it is more reasonable to compare participants' quality within their community instead of the whole sensing area. Above all, the quality estimation should have punishment on poor performance and malicious behavior and have positive reinforcement to any piece of available sensing data.

To address such problem, in this paper, we incorporated the consideration of sensing location variations into the quality estimation method and used EM algorithm to calculate each participants' sensing data quality and find the ground truth at the same time. Besides the obtained achievements, there are remain many other open issues, such as the performance improvements of distributed computing framework runs in the edge, and the resource balancing of cloud platform and edge layer. We aim to address the further challenges in future work.

ACKNOWLEDGMENT

The authors would like to thank anonymous reviewers.

REFERENCES

[1] Chenguang Wang, Yangqiu Song, Dan Roth, Chi Wang, Jiawei Han, Heng Ji, and Ming Zhang. Constrained information-theoretic tripartite graph clustering to identify semantically similar relations. *IJCAI International Joint Conference on Artificial Intelligence*, 2015-Janua:3882–3889, 2015.

[2] Hua Ting Chang and Shie Yuan Wang. Using SDN technology to mitigate congestion in the OpenStack data center network. *IEEE International Conference on Communications*, 2015-Septe:401–406, 2015.

[3] R Dos Reis Fontes and C E Rothenberg. Mininet-WiFi: A platform for hybrid physical-virtual software-defined wireless networking research. *2016 ACM Conference on Special Interest Group on Data Communication, SIGCOMM 2016*, pages 607–608, 2016.

[4] Layla Pournajaf, Li Xiong, Daniel A Garcia-ulloa, Vaidy Sunderam, Layla Pournajaf, Li Xiong, Daniel A Garcia-ulloa, and Vaidy Sunderam. A Survey on Privacy in Mobile Crowd Sensing Task Management. Technical report, 2014.

[5] Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei. LINE: Large-scale Information Network Embedding. 2015.

[6] Pablo N Mendes, Max Jakob, Andrés García-silva, and Christian Bizer. DBpedia Spotlight : Shedding Light on the Web of Documents. *Proceedings of the 7th International Conference on Semantic Systems (I-Semantics).*, 95:1–8, 2011.

[7] Amar Viswanathan, New York, Geeth De Mel, and James A Hendler. Pragmatic Reformulation in Querying Heterogeneous Knowledge Graphs. *Proceedings of the 30th Conference on Artificial Intelligence (AAAI 2016)*, pages 4317–4318, 2016.

[8] Deqing Yang, Yanghua Xiao, Hanghang Tong, Wanyun Cui, and Wei Wang. Towards Topic Following in Heterogeneous Information Networks. *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2015, Paris, France, August 25 - 28, 2015*, pages 363–366, 2015.

[9] Yuan Ni, Qiong Kai Xu, Feng Cao, Yosi Mass, Dafna Sheinwald, Hui Jia Zhu, and Shao Sheng Cao. Semantic Documents Relatedness Using Concept Graph Representation. *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*, pages 635–644, 2016.

[10] Weiping Zhu, Wenzhong Guo, Zhiyong Yu, and Haoyi Xiong. Multitask Allocation to Heterogeneous Participants in Mobile Crowd Sensing. *Wireless Communications and Mobile Computing*, 2018, 2018.

[11] Yan Liu, Bin Guo, Yang Wang, Wenle Wu, Zhiwen Yu, and Daqing Zhang. TaskMe : Multi-Task Allocation in Mobile Crowd Sensing. *International Joint Conference on Pervasive and Ubiquitous Computing*, pages 403–414, 2016.

[12] V Kumar and S Madria. PIP: Privacy and Integrity Preserving Data Aggregation in Wireless Sensor Networks. *2013 IEEE 32nd International Symposium on Reliable Distributed Systems (SRDS)*, pages 10–19, 2013.

[13] Zhenyu Zhou, Haijun Liao, Bo Gu, Kazi Mohammed Saidul Huq, Shahid Mumtaz, and Jonathan Rodriguez. Robust Mobile Crowd Sensing: When Deep Learning Meets Edge Computing. *IEEE Network*, 32(4):54–60, 2018.

[14] Bin Guo, Zhiwen Yu, Xingshe Zhou, and Daqing Zhang. From participatory sensing to Mobile Crowd Sensing. *2014 IEEE International Conference on Pervasive Computing and Communication Workshops, PERCOM WORKSHOPS 2014*, pages 593–598, 2014.

[15] Thomas Ludwig, Tim Siebigteroth, and Volkmar Pipek. Crowdmonitor: Monitoring physical and digital activities of citizens during emergencies. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8852:421–428, 2015.

[16] Manfred Reichert, Jochen Herrmann, Berthold Langguth, and Winfried Schlee. Mobile Crowd Sensing in Clinical and Psychological Trials – A Case Study. pages 23–24, 2015.

[17] Ioannis Krontiris and Tassos Dimitriou. Privacy-respecting discovery of data providers in crowd-sensing applications. *Proceedings - IEEE International Conference on Distributed Computing in Sensor Systems, DCoSS 2013*, pages 249–257, 2013.

[18] Adam Koli Ski and Pawel Fajfer. Noise tube. *Environmental Science and Engineering (Subseries: Environmental Science)*, 3:617–626, 2011.

[19] Jan Höller, Vlasios Tsiatsis, Catherine Mulligan, Stamatis Karnouskos, Stefan Avesand, and David Boyle. Participatory Sensing. *From Machine-To-Machine to the Internet of Things*, pages 295–305, 2014.

[20] Yuchen Zhang, Xi Chen, Dengyong Zhou, and Michael I. Jordan. Spectral Methods meet EM: A Provably Optimal Algorithm for Crowdsourcing. 2014.

[21] Jiajun Sun and Huadong Ma. Heterogeneous-belief based incentive schemes for crowd sensing in mobile social networks. *Journal of Network and Computer Applications*, 42:189–196, 2014.

[22] Wendong Wang, Hui Gao, Chi Harold Liu, and Kin K. Leung. Credible and energy-aware participant selection with limited task budget for mobile crowd sensing. *Ad Hoc Networks*, 0:1–15, 2016.

[23] Jesús Bobadilla, Fernando Ortega, and Antonio Hernando. A collaborative filtering similarity measure based on singularities. *Information Processing and Management*, 48(2):204–217, 2012.

[24] Maryam Pouryazdan, Burak Kantarci, Tolga Soyata, Luca Foschini, and Houbing Song. Quantifying user reputation scores, data trustworthiness, and user incentives in mobile crowd-sensing. *IEEE Access*, 5:1382–1397, 2017.

[25] Xukan Ran, Haoliang Chen, Zhenming Liu, and Jiasi Chen. Delivering Deep Learning to Mobile Devices via Offloading. *Proceedings*

of the Workshop on Virtual Reality and Augmented Reality Network - VR/AR Network '17, pages 42–47, 2017.

[26] Fengxiao Tang, Bomin Mao, Zubair Md Fadlullah, Nei Kato, Osamu Akashi, Takeru Inoue, and Kimihiro Mizutani. On Removing Routing Protocol from Future Wireless Networks: A Real-time Deep Learning Approach for Intelligent Traffic Control. *IEEE Wireless Communications*, 25(1):154–160, 2018.

[27] Aleksandar Antonić, Martina Marjanović, Krešimir Pripužić, and Ivana Podnar Žarko. A mobile crowd sensing ecosystem enabled by CUPUS: Cloud-based publish/subscribe middleware for the Internet of Things. *Future Generation Computer Systems*, 56:607–622, 2014.

[28] Tuyen X. Tran, Abolfazl Hajisami, Parul Pandey, and Dario Pompili. Collaborative mobile edge computing in 5G networks: New paradigms, scenarios, and challenges. *IEEE Communications Magazine*, 55(4):54–61, 2017.