Improving Data Credibility for Mobile Crowdsensing with Clustering and Logical Reasoning

Tongqing Zhou, Zhiping Cai, Yueyue Chen, and Ming Xu

College of Computer, National University of Defense Technology, China
{zhoutongqing, zpcai, yueyuechen, xuming}@nudt.edu.cn

Abstract. Mobile crowdsensing is a new paradigm that tries to collect a vast amount of data with the rich set of sensors on pervasive mobile devices. However, the unpredictable intention and various capabilities of device owners expose the application to potential dishonest and malicious contributions, bringing forth the important issues of data credibility assurance. Existed works generally attempt to increase data confidence level with the guide of reputation, which is very likely to be unavailable in reality. In this work, we propose CLOR, a general scheme to ensure data credibility for typical mobile crowdsensing application without requiring reputation knowledge. By integrating data clustering with logical reasoning, CLOR is able to formally separate false and normal data, make credibility assessment, and filter out the false ingredient. Simulation results show that improved data credibility can be achieved effectively with our scheme.

Keywords: Mobile Crowdsensing; Data Credibility; Clustering Algorithm; Logical Reasoning.

1 Introduction

The past few years have witnessed the massive prevalence of human-carried computing devices equipped with a rich set of powerful embedded sensors. Such advancements have given rise to a new sensing paradigm, known as mobile crowdsensing (MCS) [1], where individuals use their own mobile devices to perform sensing task, and collect interested physical data for further analysis [2] in cloud-based platform. So far, a broad spectrum of MCS applications have been developed, including environment monitoring, city management, network measurement [3], etc.

A major challenge for the adoption of MCS is how to assure the credibility of collected sensory data [4]. Unlike the specialized sensors used in traditional wireless sensor networks (WSNs), MCS relies on individuals with unknown trustworthiness and varied capabilities. Generally, other than unintentionally false data, normal anonymous participants may tend to submit random measurements to get reward with minimal effort [5]. Further, potential malicious participants intend to mislead the data analysis process for their own profit by injecting fabricated data. For instance, leasing agents may submit false low noise readings regarding a specific region to promote the rental for their houses. An Internet Service Provider may generate fictitious measurements to degrade its competitor's performance evaluation while increasing its own profit. Therefore, it is critical to design an effective scheme to verify the data in order to derive reliable conclusion from them.

Typically, previous works have investigated credibility assurance for sensory data from unreliable sources by either building reputation systems [6][7] or utilizing false detection [8][9][10] techniques. The former category attempts to evaluate the trustworthiness of the collected data based on participants' reputation information and related provenance, such as location proximity and real-time performance [6][7]. In [8] and [9], spatial-temporal compressive sensing technique and overall reputation of clusters are applied to detect false ingredient and improve data credibility. However, these approaches all rely on the prior knowledge of participants' reputation information, which may in fact be unavailable due to anonymity [11]. To address the problem of "trust without reputation", some works attempt to increase data credibility by identifying contributions that fail to pass location verification as false [12]. Unfortunately, it implicitly ignore a common unreliable form containing false sensory data with a valid location. Alternatively, in [13], the concept of provenance logic is introduced to evaluate data trust based on extended Event Calculus and Markov Logic Network. However, [13] only focus on several special application scenarios that collect data with finite domain of state (e.g. event happens or not), making it unable to handle the more common scenarios where numerical sensory data are collected [3].

In this work, we attempt to solve the problem of "trust without reputation" for general MCS applications. A Clustering and LOgical Reasoning based scheme (CLOR) is proposed to ascertain the credibility of multi-dimension numerical sensory data without requiring any prior knowledge from the participants. Specifically, two characteristics of MCS are exploited to achieve this goal. First, as crowd-contributed data for one MCS task are spatial correlated, clustering algorithm is performed to formally distinguish false ingredient from the normal part. Second, co-located events observed within a short period of time are very likely to share logical relations with the current MCS task, so logical reasoning is introduced to assess the credibility of sensory data through identifying potential logical supports.

The main contributions of our work are three-fold:

1. A novel clustering-and-merging based translation mechanism is presented to map the numerical sensory measurements to countable discrete levels and further represent them with First-Order Logic (FOL) predicates;
2. Given pre-defined logical relations between related events and MCS task, a logical reasoning based module is proposed to assess the credibility of quantization levels corresponding to sensory data clusters of that task;
3. A general data credibility assurance algorithm is developed by jointly applying clustering algorithm and logical reasoning to filter out false ingredient

in the crowd-contributed sensory data, while no reputation information is required in this process.

The rest of this paper is organized as follows. Section II describes system model, problem formulation and introduce logical reasoning knowledge. Then we outline the key components of scheme CLOR, and introduce how CLOR facilitates better data credibility in section III. In Section IV, simulation results that indicate the effectiveness of the scheme are provided. Finally, conclusions are drawn in Section V.

2 Preliminaries

2.1 System Model

We consider a typical MCS architecture as shown in Fig. 1. It consists of a cloud-based platform and a set of participants $U = \{u_1, ..., u_N\}$ that perform sensing task T at location L. Among the three stages of a MCS application, data credibility is considered as an crucial part of the utilization stage, and data falsification threats arise as the crowd participants have various capabilities and purposes. A sensing task normally specifies multiple modalities of sensory data to be collected, so we consider that the collected data in MCS application are multi-dimensional numerical sensor readings (e.g. temperature, noise level).

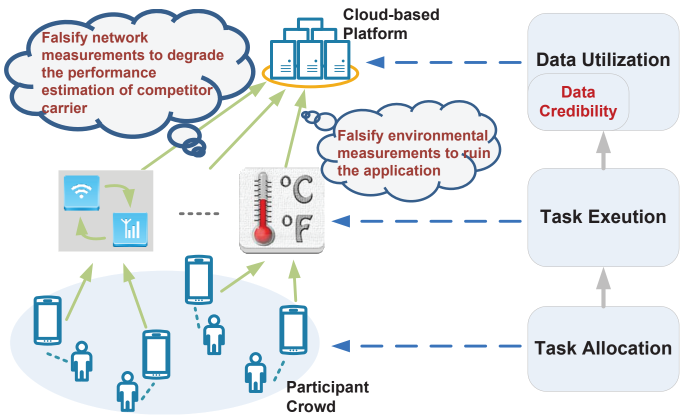


Fig. 1. Architecture of typical MCS applications. Network measurement and temperature monitoring are depicted as two illustrative example applications. Two potential credibility threats are also listed.

During the execution of task T, u_i collects a series of measurements with K different sensor types, where each can be denoted by $s_i(k)$, where $i \in [1, N]$, and $k \in [1, K]$. Sensory data are submitted together with location l_i to the platform before time deadline. The contribution from u_i forms a tuple in key-value syntax, denoted as $d_i =< l_i, s_i >$. By the end of T, the centralized platform will obtain a data set $D = \{d_1, ..., d_N\}$, based on which some aggregation function f will be performed to derive statistical conclusion.

Finally, we assume that L refers to an area of interest within certain distance of L instead of a specific spot as physical measurements are usually spatial correlated, and all sensory data are aligned on measurement features. No additional information of the participants is required.

2.2 Problem Formulation

Data collected from individuals with unknown trustworthiness are unreliable. The potential erroneous or fabricated ingredient measurements injected in the collected data would deviate the analysis result from the expected true value. Here we consider data d_i as trust if its location component l_i and sensory measurement component s_i are both valid. We classify the state space of credibility of data d_i into four categories as shown in Table 1, where symbol T (F) means the value is true (false). Note that location attestation-based schemes like [12] try to assure data credibility by picking out data with invalid location in category B and category C, ignoring possibly false data in category A, which is more common in a MCS application especially when malicious intention is considered. Unlike these schemes, we propose to improve overall credibility through identifying the group of data with invalid measurement (Category A and C) in this work, we consider data fall in category B to be normal as its value is valid.

Table 1. Space state for validity of typical MCS data

l_i \ s_i	T	F
T	Normal data	Category A
F	Category B	Category C

In view of the above-described false data forms, two types of adversary model are considered:

1. *Random Falsification.* Participants submit measurements with random value to minimize their efforts, or tamper the measurement to facilitate a misleading effect. For the latter intention, dishonest participants would try to deviate the aggregation result as much as possible.
2. *Falsification with Conspiracy Cooperation.* A group of adversaries collude with each other to intentionally induce the final aggregation result to a wrong value. Moreover, in order to avoid being identified by statistical analysis-based abnormal detection method, the dishonest group is able to fabricate and submit data obeying normal distribution.

Collusion among participants would result in a more significant deviation, and the injected artificial data cannot be easily picked out. Taking average function f_{avg} as an example of the aggregation function, if we have a crowd contributed data set $D_{eg} = \{d_n^1, ..., d_n^N, d_f^1, ..., d_f^N\}$, where $d_n^i =< L, M >$ denotes a normal measurement, and $d_f^i =< L, 2M >$ denotes a false measurement, then we will have $f_{avg}(D_{eg}) = 1.5M$ which is 1.5 times larger than the actual value M. Additionally, we do not make any assumption or set any limitation on the number of dishonest participants in U, in which situation vote-based false detection approaches are not effective any more.

2.3 Logical Reasoning

Logical reasoning is the formal manipulation of the symbols representing a collection of something known to produce representations of new ones. It generally involves ontology, basic predicates, and knowledge base (KB). The underlying ontology can be time points, events (e.g. accident), and fluents (e.g. high temperature), while a predicate represents a property of or relation between ontology that can be true or false. A KB contains general axioms describing the relations between predicates. Resolution is one of the most widely used calculi for theorem proving in logical reasoning. It proves a theorem by negating the statement to be proved and adding this negated goal to the sets of axioms that are known to be true to tell whether it leads to a contradiction.

In this work, we consider to map and translate the sensory data collected during current MCS task into FOL predicates, and use resolution rules to tell whether the predicates are satisfiable by jointly considering the co-located events and basic KB. We assume the basic KB has been pre-established given a specific application scenario, and real time computation only involves translating related events into predicates and add them to the reasoning KB.

3 Design of CLOR

3.1 Overview

CLOR scheme tries to improve the overall credibility of crowd data in MCS through identifying and discarding the corrupted part with invalid sensory measurements (i.e. the data belonging to category A and C in Table 1). Theoretically, only normal part of the collected data remains after the processing.

Framework of CLOR is illustrated in Fig. 2, which basically consists of three modules. The quantization and representation module formally translates the input numerical sensory data into predicates, wherein clustering-and-merging mechanism, projection and translation operation are carried out sequentially. The KB construction module provides reasoning KB based on co-located events and pre-established casual logical relations in regard of the MCS tasks. The filtering module adopts logical resolution to find logical supports for each cluster and assess overall credibility for them. Finally, clusters with low assessment score are filtered out and aggregation function is performed on the filtering results.

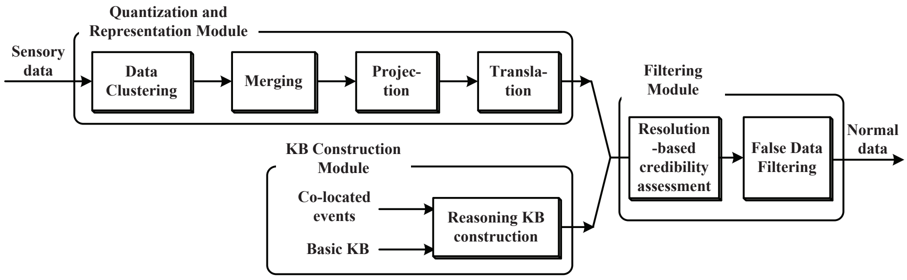


Fig. 2. Framework of CLOR

CLOR improves data credibility, and provides space for privacy preserving. Essentially, CLOR improves the quality of service provided by a MCS application to the sensory information requesters.

3.2 Data Quantization and representation

In this module, we aim to obtain several quantization levels from the sensory data and translate them into FOL predicates for further logical reasoning. The number of levels is application specific and dynamically determined by the distribution of the sensory measurements, and denoted as m. Each level represents the property of corresponding data cluster. The detail stages are introduced as follows.

1) Data Clustering: Normally, physical measurements act as signatures that characterize a place of interest, which implies that measurements for the same location are correlated with each other, while on the other hand, false data behave abnormal in the feature space. Meanwhile, the collected data are mainly exploited at a community scale which provides sufficient participant density support for clustering the data around a specific location target. Hence, a fixed width clustering algorithm is first performed on D to group similar data instances into clusters with similar property. The first data is assigned to be the centroid of the first cluster. Then for every subsequent data d_i in D, distance between centroid of each cluster and data d_i is calculated as,

$$dis(s_i, s_c) = \sqrt{\sum_{k=1}^{K} (s_i(k) - s_c(k))^2} \tag{1}$$

where s_i is the sensory measurements in d_i, and s_c is the sensory values of cluster centroid. If the distance to one cluster is less than the cluster width ω, it is added to that cluster and the centroid of that cluster is adaptively adjusted to the mean of the inner data. Otherwise a new cluster is formed with that data as the initial centroid. Here, we novelly define ω as half of the minimum expected deviation from the true aggregation result for a potential falsify behavior among the crowd data, i.e.,

$$\omega = 1/2 \cdot \min_i(\left| f(D(i)) - f(\tilde{D}) \right|) = 1/2 \cdot \sigma_{dev} \tag{2}$$

where $D(i)$ represents one of the possible collected data sets containing corrupted ingredient, f denotes the aggregation function during data analysis, $\tilde{D}$ denotes the set of normal data, and σ_{dev} denotes the minimum expected misleading degree. The value of parameter σ_{dev} is updated adaptively according to the application context, e.g., a dishonest participant may prefer to consider a deviation of at least $4°c$ as effective for a task that measures city temperature, while $10dB$ may be a meaningful value in an application of received signal strength measurement. The clustering operation generates a set of fixed width clusters $C = \{C_1, .., C_n\}$ in the feature space.

2) Clusters Merging: In order to map the sensory data into several discrete levels, we introduce a merging stage to combine similar clusters together. The similarity between clusters can be measured by their inter-cluster distance. Hence, distance $dis(s_c^i, s_c^j)$ between each cluster C_i and C_j is calculated, and a merging operation is performed between the two clusters with the minimum inter-cluster distance to generate a new cluster. The new cluster combines the data points in the two neighbor clusters and is added into C to join the next round of comparison and merging. This iterative procedure is continued until the inter-cluster distances of remaining clusters are all bigger than width ω, making these distances the m largest ones.

Note that clusters with a bigger size do not inherently indicate higher credibility as the number of malicious participants may be more than the honest one, so vote-base approach can not help here.

3) Projection: The above process provides us with m separated areas (i.e. clusters) in the K-dimension feature space. However, what we need are discrete levels, say, some points distributed in the space, so a mapping function is required to map one cluster into one single point. Generally, the centroid of a cluster can describe its property well, so we propose to use the centroid of clusters to represent them. Hence, the extracted m levels are defined as $Lev = \{s_c^1, ..., s_c^m\}$, and s_c^i equals to $f_c(i)$, which denotes the calculation of the centroid of cluster C_i.

4) Translation: Each quantization level is a K-dimension feature vector, we propose to translate them independently. Specifically, each sensory measurement in s_c^i is first converted into linguistic variables $M(s_c^i(k))$ with function $M(x)$. There has already been many works on linguistic representation such as fuzzier in fuzzy logic, so we propose to rely on the state-of-art methods. For example, the value is replaced with "WA" or "CO" (Warm or Cold) according to its scale in a temperature monitoring application. Then we introduce a FOL predicate, denoted as $HoldsAt(F, T, L)$, which means Fluent F holds at time interval T at location L, to describe the statements corresponding to the linguistic variables. Finally, the translation result regarding level i is defined as

$$T(i) = HoldsAt(M(s_c^i(1)), T, L) \wedge ... \wedge HoldsAt(M(s_c^i(K)), T, L) \qquad (3)$$

3.3 Knowledge base construction

The reasoning KB is constructed by incorporating related events into the basic KB. Physical phenomena sensed in the same region during a time period are often related, we define these phenomena as related events. Basically, the collection of related events can be obtained from two main sources: i) the reports of other MCS applications from the same MCS platform, and ii) the geo-tagged observations and information collected from social networks. For example, the degree of crowdedness in different regions of an urban area could influence the corresponding noise level, and practically, it can not only be identified through mobility-based sensing application [14], but also be detected from human observations with social sensing [15]. On the other hand, the basic KB contains a set of logical formulas representing the causal relations between events and the sensory phenomena of MCS applications, and it is application specific. The method of event collection and relation formulation are out the scope of this paper, we assume related events and basic KB are known prior, and denoted as E and KB_{basic}. Finally, the reasoning KB can be represented as $KB = E \cup KB_{basic}$.

3.4 Data Filtering

The quantization module formally map sensory data indicating different phenomenal property into discrete levels. We employ logical resolution to find evidence for each level of announced sensory phenomenon, and estimate its credibility.

Algorithm 1 Cluster credibility assessment

Input:
$M(s_c^i(k))$, KB_{basic}, E, additive increase factor r_a
Output:
A_c^i: credibility assessment score of the i-th cluster

Set $A_c^i \leftarrow 0$ for $i = 1...m$
for $i = 1 \rightarrow m$ **do**
  for $k = 1 \rightarrow K$ **do**
    $statement = \neg HoldsAt(M(s_c^i(k)), T, L)$
    for $\forall f \in KB_{basic}$ **do**
      if $resolution(E, statement, f) \Rightarrow NIL$ **then**
        $A_c^i \mathrel{+}= r_a$
      end if
    end for
  end for
  $A_c^i = \frac{A_c^i - \min(A_c^i)}{\max(A_c^i) - \min(A_c^i)}$ //Normalization
end for

The proposed algorithm is described in Algorithm 1, which will be repeated sequentially for each dimension in the K-dimensional of the m quantization levels. Initially, we introduce a variable A_c to denotes the credibility assessment score of levels in Lev. Then we adopt logical resolution to estimate A_c^i for each level i. Specifically, we first pick out the logical reasoning for one level's k-th sensory measurement and negate it to obtain a statement. We then use inference rules of resolution to iteratively perform resolution on the statement, axiom set E, and every formula in KB_{basic} to show whether this leads to a contradiction (logically, an empty clause). A contradiction means that this measurement of level i is logically supported by E, in which situation we propose to addictively increase A_c^i with a factor r_a. The rationale of the 3rd iteration (Line 5) is that with more events logically supporting the current level, it should be more reliable, while the 2nd iteration (Line 3) indicates that with more dimension of sensory measurements being supported, the current level should be more reliable. For each cluster, an estimation score would be generated with this iteration reasoning procedure. We normalize these estimation values and obtain an assessment for the credibility of each level (Line 13).

Finally, the level with the highest credibility assessment is determined to be reliable, and the corresponding data cluster regarded as the container of normal data, i.e., $C^* = \arg\max_{C_i}\{A_c^i\}$. Other contributions in the data set are regarded as false and filtered out.

4 Evaluation

4.1 Settings

In this section, we aim to test the effectiveness of CLOR with a typical MCS-based environmental monitoring application. In such applications, portable sensors are equipped with mobile participants to collect physical information. Specifically, we choose an open source temperature measurement traces obtained from the CRAWDAD data set [16], which contains 5030 measurement items from 289 active taxicabs collected around the GPS location $(41.9, 12.5)$ in Rome. In order to simulate the potential dishonest behaviors that falsifies sensory data, some items of the data set are modified. Here, we consider the adversary model of falsification with conspiracy cooperation as it is harder to detect. Measurements of these items are replaced with random values generated from a normal distribution with mean parameter μ equalling to the value of misleading target S_{err} and standard deviation parameter $\sigma = 1$. Note that the false data are fabricated to obey normal distribution to imitate smart collusion among a dishonest group. Further, the synthetic data set is divided into two sets according to the submission time of the contribution to conduct two experiments independently. Finally, as mentioned above, we assume the KB has been pre-established given our application scenario.

We consider the minimum possible deviation caused by data falsification to be $4°$, so the cluster width is set to be $2°$. Meanwhile, the falsification target S_{err} for time period 1 and time period 2 are set to $14°$ and $8°$ to effectively mislead the aggregation results. Two sets of fabricated measurements are then generated and used to replace sensory measurements of the selected data items in the original data set. We assume that all the simulations are under a closed-world assumption, i.e., all relevant events are defined in the KB. For each time period, an complete KB is defined as presented in Table 2. Here we provide 3 levels (e.g.

Cold, Warm, Hot) to quantize the collected temperature measurements, and we emphasize that other situations can be easily generalized.

We evaluate the effectiveness of CLOR using credibility metric $\Re_D$, given by

$$\Re_D = 1 - (\frac{\left|f(D) - f(\tilde{D})\right|}{\min(f(D), f(\tilde{D}))}) \tag{4}$$

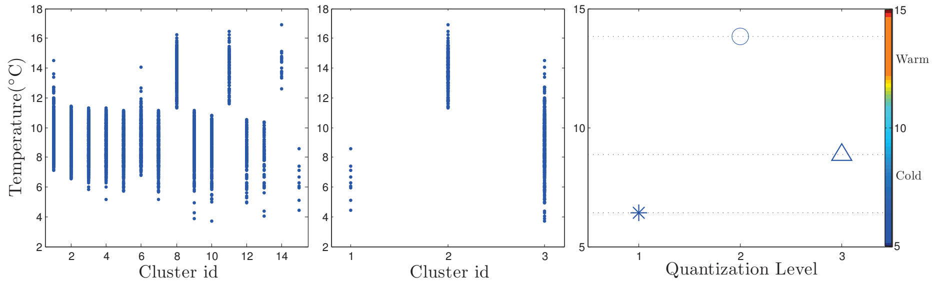
where $\tilde{D} = D - D_f$, and D_f is the set of false data. $\Re_D$ is a posterior value calculated by comparing the filtering results of CaPa with the ground truth. Obviously, the less false data in D, the more similar $f(D)$ and $f(\tilde{D})$ will be, and the higher credibility D could achieve. Without loss of generality, we adopt average function as the aggregation function f during data analysis.

Table 2. A complete knowledge base for logical reasoning

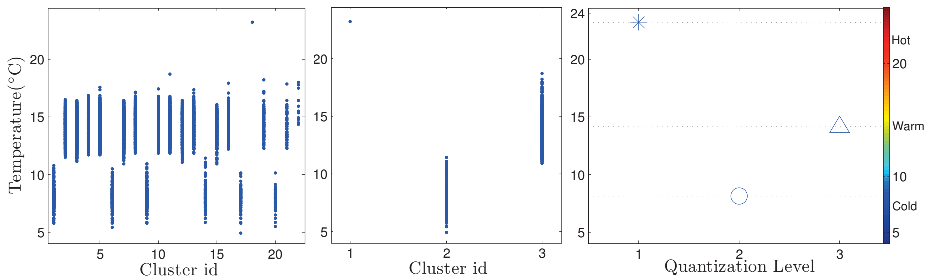
Period	(# of events, # of logical relations)		
	Cold ($< 10°C$)	Warm ($10 \sim 20°C$)	Hot ($> 20°C$)
Morning	(8,10)	(4,10)	(1,10)
Noon	(3,10)	(9,10)	(4,10)

4.2 Results

According to the scheme, collected data are first processed with the quantization and representation module to generate discrete levels for further reasoning. We evaluate the effectiveness of this procedure based on data of the two time periods. The results of each processing stage are shown in Fig. 3 (from left to right), separately. For time period 1, the data are first clustered into 15 groups, which are then merged by comparing their inter-cluster distance with 2, the cluster width, generating 3 new clusters with values distributed in the feature space. Centroid of the 3 clusters are extracted to project data cluster into discrete data point, and finally mapped to linguistic state "Cold", "Warm", "Cold", respectively. Similarly, 22 clusters are generated in time period 2, which are represented with state "Hot", "Cold", and "Warm" by merging, projection, and mapping. The number of output levels (3 for both periods) are mainly determined by cluster width (2 in our simulation). With application specific knowledge, proper width can be carefully chose to roughly separate normal and false ingredient into different groups. During the clustering phase, one data point is added to all the clusters within certain distance, so some clusters depicted in Fig. 3 are similar with each other. Note that the similarity (or redundancy) is significantly reduced with the merging phase.



(a) Result for time period 1 (from 6 o'clock to 10 o'clock)



(b) Result for time period 2 (from 11 o'clock to 15 o'clock)

Fig. 3. Data quantization and representation results for two time periods. The result of data clustering, clusters merging, and projection are given, separately

Given the pre-established KB in Table 2, credibility of the discrete levels are assessed using logical reasoning. The normalized assessment scores are presented in Fig 4. As expected, level with more supported events achieves higher score. Further, clusters are classified as normal or false based on their assessment score. Obviously, level 1 and 3 present the highest score (around 0.8) for time period 1, while the highest score of period 2 is level 3. Thus, cluster 1 and 3 generated from data of period 1 are labelled as normal, and cluster 3 is determined to be the normal container for time period 2. The remaining data points in the set are labelled as false.

Finally, we evaluate the performance of CLOR in Table 3. Since we assume all the unmodified measurements are reliable, the average temperature measurements are regarded as the ground truth value here. The falsification target is where the collusive group likes to mislead the aggregation result, and the falsification result is the real value it turns out to deviate to. Then the posterior credibility for both falsification result and CLOR output are evaluated. As illustrated in Table 3, CLOR improves the overall credibility from 0.84 and 0.83 to 0.99 for the two selected time periods.

(**Comparison**) We consider CLOR as the first attempt to assess data credibility considering possible false forms without reputation information for typical MCS applications. Hence, we emphasize that the credibility advantage CLOR

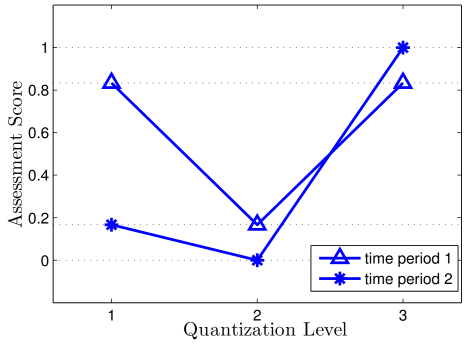


Fig. 4. Credibility assessment results of the quantization levels for the two periods

Table 3. Performance evaluation of CLOR

Metric	Ground Truth	Falsification			CLOR	
		Target	Result	Credibility	Result	Credibility
Period 1	8.85	14	10.27	0.84	8.81	0.99
Period 2	14.05	8	12.05	0.83	14.16	0.99

achieves is also relative to the location attestation-based scheme in [12], reputation-based scheme in [7] and provenance logic-based scheme in [13].

5 Conclusion

We have presented a Clustering and LOgical Reasoning based scheme CLOR to improve data credibility for typical MCS applications. In view of the potential data falsification threat, CLOR proposes to assess overall credibility of sensory data without the aid of reputation system. Clustering and logical reasoning techniques are introduced to exploit spatial correlation of data and logical relation of co-located events. We describe the corresponding processing and filtering module in detail. The simulation results show that CLOR can adequately improve the overall credibility under the cases considered.

Acknowledgment

This work is supported by National Key Basic Research Program of China (No.2012CB933504), National Natural Science Foundation of China under Grant Nos. 61379144, 61379145, 61402513.

References

1. Ganti, R.K., Ye, F., Lei, H., "Mobile crowdsensing: current state and future challenges," IEEE Communications Magazine, vol. 49, n. 11, 2011, pp. 32-39.
2. Gu, B., Sheng, V.S., Tay, K.T., Romano, W., Li, S., "Incremental Support Vector Learning for Ordinal Regression," IEEE Transactions on Neural Networks and Learning Systems, vol. 26, no. 7, 2015, pp. 1403-1416.
3. Khan, W.Z., Xiang, Y., Aalsalem, M.Y., Arshad, Q., "Mobile phone sensing systems: A survey," IEEE Communications Surveys & Tutorials, vol. 15, no. 1, 2013, pp. 402-27.
4. Reddy, S., Samanta, V., Burke, J., Estrin, D., Hansen, M., Srivastava, M., "Mobisense - mobile network services for coordinated participatory sensing," Proc. of International Symposium on Autonomous Decentralized Systems(ISADS), IEEE, 2009, pp. 1-6.
5. Downs, J.S., Holbrook, M.B., Sheng, S., Cranor, L.F., "Are your participants gaming the system? screening mechanical turk workers," Proc. of the SIGCHI Conference on Human Factors in Computing Systems, ACM, 2010, pp. 2399-2402.
6. Wang, X., Cheng, W., Mohapatra, P., Abdelzaher, T., "Artsense: Anonymous reputation and trust in participatory sensing," Proc. of INFOCOM, IEEE, 2013, pp. 2517-2525.
7. Amintoosi, H., Kanhere, S.S., "A reputation framework for social participatory sensing systems," Mobile Networks and Applications, vol. 19, n. 1, 2014, pp. 88-100.
8. Cheng, L., Kong, L., Luo, C., Niu, J., Gu, Y., He, W., Das, S.K., "False data detection and correction framework for participatory sensing," Proc. of IEEE/ACM International Symposium on Quality of Service(IWQoS), IEEE/ACM, 2015.
9. Zhou, T., Cai, Z., Xu, M., Chen, Y., "Leveraging crowd to improve data credibility for mobile crowdsensing," Proc. of the 21th IEEE Symposium on Computers and Communications(ISCC), IEEE, 2016.
10. Li, J., Li, X., Yang, B., Sun, X., "Segmentation-based Image Copy-move Forgery Detection Scheme", IEEE Transactions on Information Forensics and Security, vol. 10, no. 3, 2015, pp. 507-518.
11. Fu, Z., Ren, K., Shu, J., Sun, X., Huang, F., "Enabling Personalized Search over Encrypted Outsourced Data with Efficiency Improvement," IEEE Transactions on Parallel and Distributed Systems, DOI: 10.1109/TPDS.2015.2506573, 2015.
12. Talasila, M., Curtmola, R., Borcea, C., "Improving location reliability in crowd sensed data with minimal efforts," Proc. of the 6th Joint IFIP Wireless and Mobile Networking Conference(WMNC), IEEE, 2013, pp. 1-8.
13. Wang, X., Fu, H., Xu, C., Mohapatra, P., "Provenance logic: Enabling multi-event based trust in mobile sensing," Proc. of IEEE International Performance Computing and Communications Conference(IPCCC), IEEE, 2014, pp. 1-8.
14. Liu, S., Liu, Y., Ni, L., Li, M., Fan, J., "Detecting crowdedness spot in city transportation," IEEE Transactions on Vehicular Technology, vol. 62, n. 4, 2013, pp. 1527-1539.
15. Srivastava, M., Abdelzaher, T., Szymanski, B., "Human-centric sensing," Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, vol. 370, n. 1958, 2012, pp. 176-197.
16. Alswailim, M.A., Hassanein, H.S., Zulkernine, M., "CRAWDAD dataset queensu/crowd temperature (v. 2015-11-20)," Downloaded from `http://crawdad.org/queensu/crowd\_temperature/20151120`, 2015.