Accuracy Estimation for Real World Sensor Application Using Inference and Learning Algorithm

Mrs. R.C.Subashini
M.E II year
Dept. of computer Science and Engineering
Oxford Engineering College
Trichy

Mr. D. SelvaPandian
Assistant Professor
Dept. of Computer Science and Engineering
Oxford Engineering College
Trichy

ABSTRACT: In sensor networks, the measurement collected by the sensors is prone to noise. Due to this uncertainty sensor data that is referred as probability measurement are unprecedented wealth. Accuracy estimation of the sensor data should be done so that data can be used by the sensor application. Prior information about model data over multiple timestamp is used to infer and learn the probability measurement, which helps to estimate the accuracy of the sensor data. For accuracy estimation along with learning, inference techniques are equally viable alternative which should not be confined to their traditional use in state estimation problems. Undirected graphical models are used for inference and learning techniques.

Keywords: Sensor, Network, Noise, Uncertainty, Estimate, Accuracy

1. INTRODUCTION

Now sensor technology becomes ubiquitous. The measures of sensor are highly uncertain, because of the noise. Most of the application receives measurement with noise which cannot be true to the application. Due to the result of the noise and uncertainty, sensor measurement will have the confidence intervals or distribution variances. This uncertain measurement will be hereafter referred to as probabilistic measurements.

To overcome the presence of noise in sensor data, a lot of researches in the area of sensor networks are carried out to estimate its accuracy over the last two decades. This work is done in two classes 1) State Estimation: inference algorithm is used to estimate the state of the noisy measurement with the help of prior models 2) Parameter Estimation: here the learning techniques are used to find the uncertain measurement's parameter values which is used to generate the accuracy of sensor measurements.

This work investigates the problem that lays in-between the above classes state estimation and parameter estimation, and which will hereafter refer to as accuracy estimation. First start with a noisy probabilistic measurement (e.g. temperature value with a 95% confidence interval), and the objective is to estimate the accuracy of this measurement, with the help of inference and learning technique. It is done by finding out, how far the uncertain measurement lies from the ground truth given with certain distance metric.

It is critical to detect, when a sensor starts malfunctioning. When the accuracy of the measurements drops significantly below a certain threshold then it is known that the measurement is not true. Finally, the emergence of social sensing applications has been the challenge of estimating the truth of human participants in sensing the data. When people report some observations (say, temperature for weather forecast) it should be checked how far it is true from the ground truth.

2. PROBLEM DESCRIPTION

The presence of noise in sensor data has motivated a lot of research in areas of sensor networks, mobile tracking and vehicle tracking over the last two decades. For addressing accuracy estimation, accuracy estimation framework was proposed. It lies between the application and the sensor network.

Accuracy estimation framework contains taxonomy of four layers: pre-processing, state estimation, accuracy estimation and accuracy indexing. By passing the monitored sensor observations through those layers, the proposed framework will provide the accuracy of the reported measurements, and then reason about the accuracy of the sensor systems measurements.

2.1 Accuracy Estimation Framework

Formulate the difficulty in accuracy estimation of sensor systems in a general manner which covers a broad spectrum of sensing application. Motivate the problem in the context of pricing sensing services, ranking them if they are competing for the same user, detecting faults and establishing trustworthiness of different individual in social sensing.

Propose a framework to address the formulated accuracy estimation problem which has the following layers.

1. Pre-processing
2. State Estimation
3. Accuracy Estimation
4. Accuracy Indexing

The framework firstly activates all the coexisting sensor systems that can monitor the measurements on the queried states for a given time stamp, and the raw observed measurements pulled from the systems are firstly processed through the pre-processing layer. Then the processed measurements are forwarded to the state estimation layer, where the states of monitored measurements are estimated based on sensor observations and available prior over single timestamp [16]. The next accuracy estimation layer takes the estimated states of the monitored value together with the pre-processed sensor measurements as input, and evaluates the accuracy of measurements with the help of accuracy metric.

The estimated measurement accuracy is then forwarded to the accuracy indexing layer, where an accuracy indices is built for each sensor system, according to the attributes. Finally, the accuracy indices of the sensor systems are reported as the output of the accuracy estimation framework, which empowers the sensing application to choose the desired value of measurement depending upon the context. The framework will initiate this process periodically to keep the accuracy indices updated.

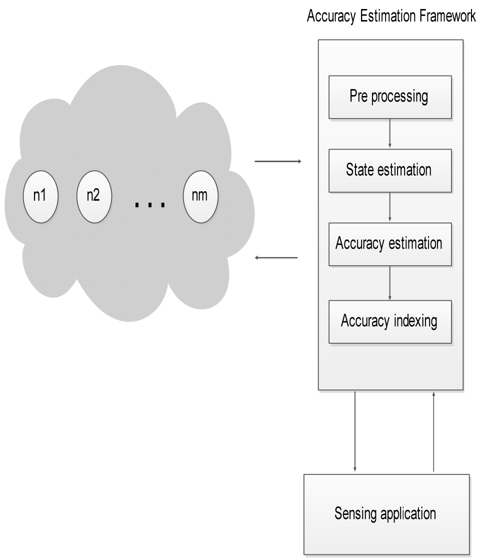


Fig 2.1 Accuracy Estimation Frameworks

Inference and learning techniques are used to estimate the state of the monitored date after pre-processing.

2.2 Inference-based approach

Depending upon the temporal information used between the states, it is divided into static and dynamic inference.

It models the monitored states, the measurements from one or more sensor systems and the available prior knowledge in a probabilistic model with a set of parameters θ known a prior. Generally θ depends on the type and structure of the model.

Static Inference

It ignores any temporal correlations between the latent states and only accesses the measurements and priors at single timestamp using likelihood of monitored data.

$$\hat{X}_t \propto p(^{x_t}/_{\Psi_t}) \prod_{m=1}^{M} p(Z_t^m/x_{t,\theta^b})$$

Dynamic Inference

Assumes that the monitored states are temporally correlated and estimates the states with all the observed measurements and priors under Markovian assumption.

$$\hat{X}_t \alpha p(Z_{1:t}^{1:M}, \Psi_{1:t}, {}^{x_t}/_{\theta}) p({}^{Z_{t+1:T}^{1:M}}/_{\Psi_{t+1:T}}, x_t, \theta)$$

2.3 Learning-based approach

Learning does not assume any prior knowledge on the model parametersθ. It starts with an estimate of the parameters and iteratively refines this estimate to be more consistent with the sensor measurements and priors. The estimated state distribution is then inferred with the learned model parameters and the observed data.

Static Learning

It tries to find the model parameters that are the most consistent with the data observed within each timestamp, given by maximum likelihood estimation. It can be computed by Expectation maximization approach.

a) E-step computes the expected log likelihood function $Q(\theta',\theta)$ of the new parameter θ'

$$Q(\theta',\theta) = \int_{x_t} p({}^{x_t}/_{Z_t^{1:M}}, \Psi_t, \theta) \log p(Z_t^{1:M}, \Psi_t, {}^{x_t}/_{\theta'}) dx_t$$

b) M-step computes the new parameters θ' that maximize the Q function

$$\theta' = \arg\max_{\theta'} Q(\theta',\theta)$$

Dynamic Learning

It assumes the hidden state varies over time and the state transitions are governed by parametersθ^a.

$$\theta = \{\theta^a, \theta^b\}$$

3. PROPOSED SYSTEM

Dynamic inference and learning are significantly better since the temporal correlations between monitored states are considered. A limitation is that directed graphical models and recursive Bayesian techniques are used to capture the correlation between the states and measurements.

Proposed system will overcome this by using undirected graphical models for inference and learning in the context of accuracy estimation. Multiple time stamps are used to incorporate the other forms of prior knowledge on state.

3.1 Objective

An undirected graphical model – Conditional Random Factor (CRF) is used to capture the temporal correlation between the monitored states and measurements for learning and inference approaches. Inference is supported by the multiple timestamps of the prior knowledge to estimate the accuracy of the sensor system.

3.2 System Architecture

The proposed framework contains five layers.

- Raw data collection
- Pre processing
- State estimation using CRF
- Accuracy estimation
- Accuracy indexing

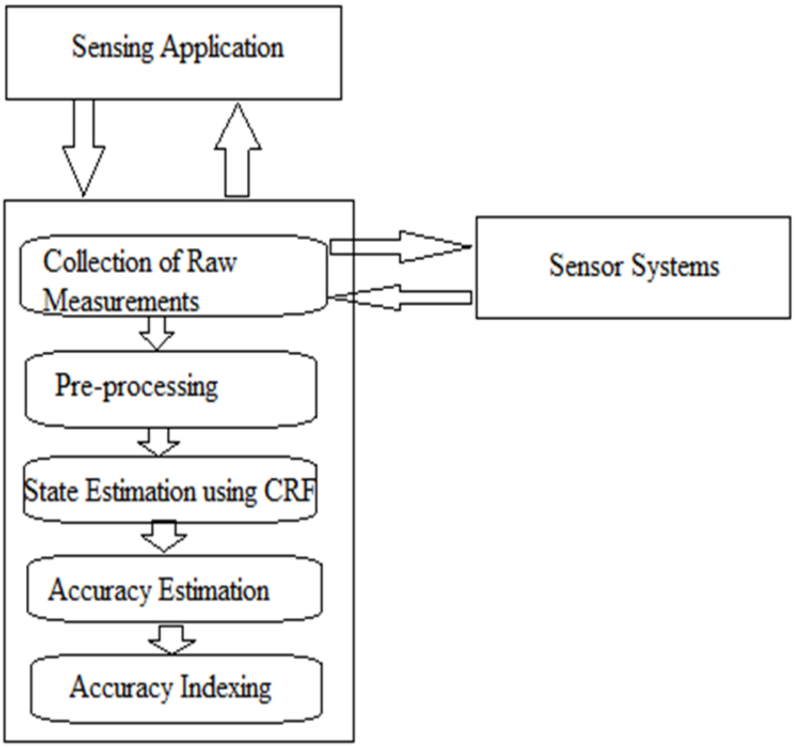


Fig.3.1 Accuracy Estimation Frame Works

3.3 Raw Data collection

Application will place a request for sensor data by passing the accuracy metric $f_e$and the attribute A. On receiving the request, system will start the entire coexisting sensors and collect the monitored value after a short duration.

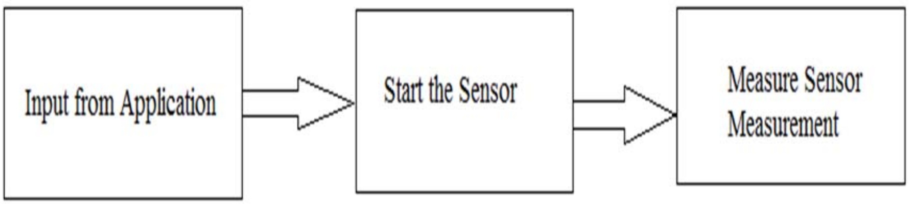


Fig.3.2 Raw data collection

3.4 Pre-processing

The pre-processing layer collects and processes the raw measurements from different sensor systems. The collected measurement Z_t^m from sensor systemSn_m, 1<=m<=M, at the given timestamp t is typically heterogeneous, and they may be in different time other than the global clock. They may be generated at different time or in different space granularities.

The raw sensor measurements are processed in the pre-process layer through two steps:

a) The synchronization step
b) The resample step

In the synchronization step, the measurements from different sensor systems are firstly re-timestamped according to the global clock provided by the proposed framework and convert the measurement into same metric if they are generated in different metric[4][6][11].

After the synchronization step, the resample step further subsamples or interpolates the raw measurements so that measurements from different sensor systems will have the same time stamp and space granularity. Gaussian process non linear regression is used to interpolate the measurement of different sensor system [10].

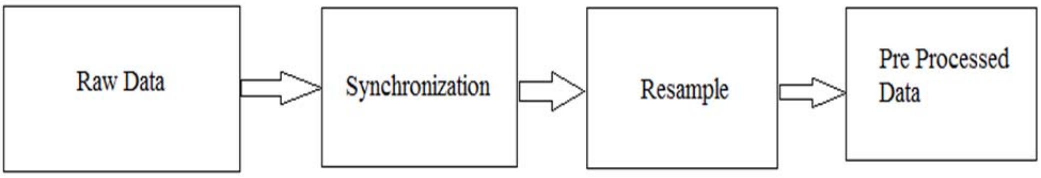


Fig.3.3 Pre-processing

3.5 State Estimation using CRF

Given the pre-processed sensor measurements and the available priors over multiple timestamp as input, the state estimation layer computes the estimated state distribution$\hat{X}_t$.

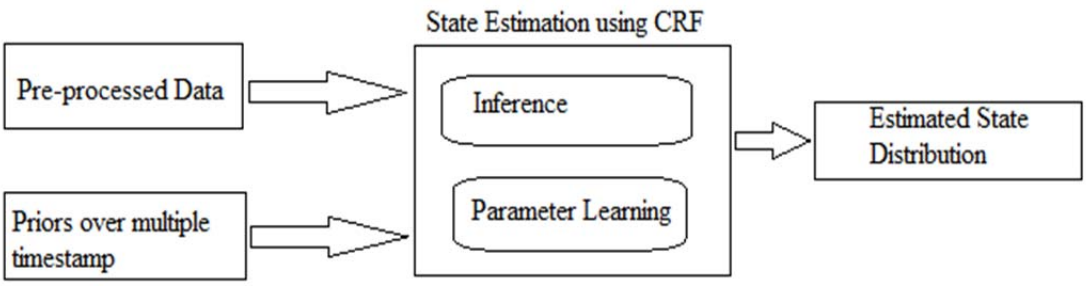


Fig.3.4 State Estimation

3.5.1 Conditional Random Factor (CRF)

CRF has the advantage of both discriminative classification and graphical modeling. By having the ability to leverage a large no of input features X for predictions, it built the multivariate outputs Y [1].

Inference

It computes the marginal distributions of $P(y/x)$ and then computes the maximum probability assignment

$$Y^* = \arg\max_y P(y/x)$$

Learning

Parameter estimation task will determine $P(y/x)$ by choosing parameter θ in order to best fit a set of training samples $\{x^{(i)}, y^{(i)}\}_{i=1}^{N}$[1]

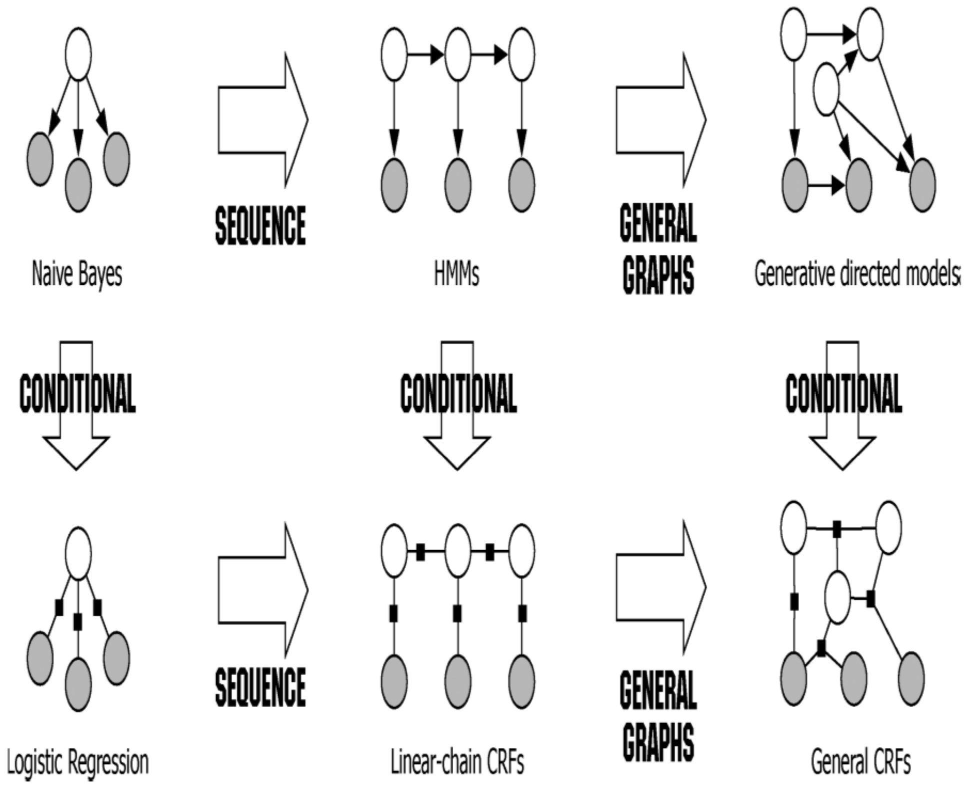


Fig. 3.5 Diagram of the relationship between naive Bayes, logistic regression, HMMs, linearchain CRFs, generative models, and general CRFs.

3.6 Accuracy Estimation

Then the estimated state information together with the pre-processed sensor measurements are forwarded into the accuracy estimation layer, where the accuracy of the measurements are estimated with the metric specified by the sensing application.

Consider two accuracy metrics, a proximity-based and a similarity-based that can be used in sensing applications which require different types of accuracy metric.

The proximity-based accuracy metric defines the accuracy of a probabilistic measurement based on its distance to the estimated state. It uses Eucidean distance.

$f_e^p(Z_t^m; \hat{X}_t)=\int_{\varsigma,x} pz_t^m(\zeta)p\hat{X}_t(\chi)C(\varsigma-\chi)d\varsigma d\chi$

The similarity-based metric defines accuracy as the divergence between the probabilistic measurements and the estimated state distributions. It uses kullback-Leibler (KL) divergence.

$f_e^s(Z_t^m, \hat{X}_t) = D_{KL}(\hat{X}_t || Z_t^m)$

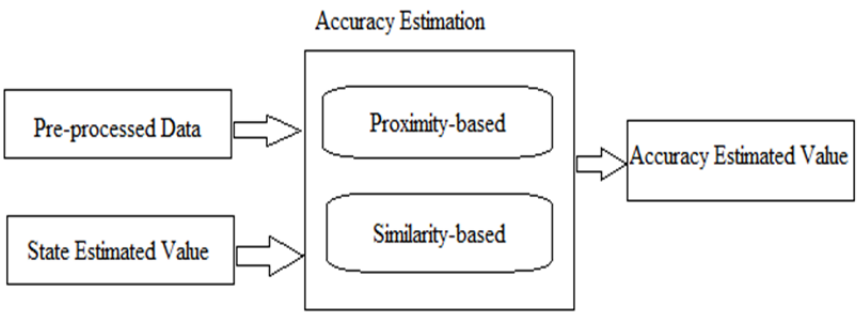


Fig.3.6 Accuracy Estimation

3.7 Accuracy Indexing

The proposed accuracy indexing layer builds the accuracy indices in two steps:

a) The accuracy aggregation step.

The first aggregation step is straightforward; here the estimated accuracy is averaged according to the measurements for the individual sensor.

b) The accuracy interpolation step.

In the accuracy interpolation step the unknown accuracy is filled based on computed accuracy information[10].

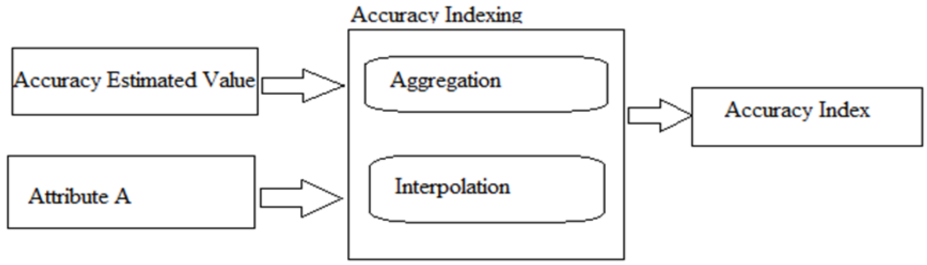
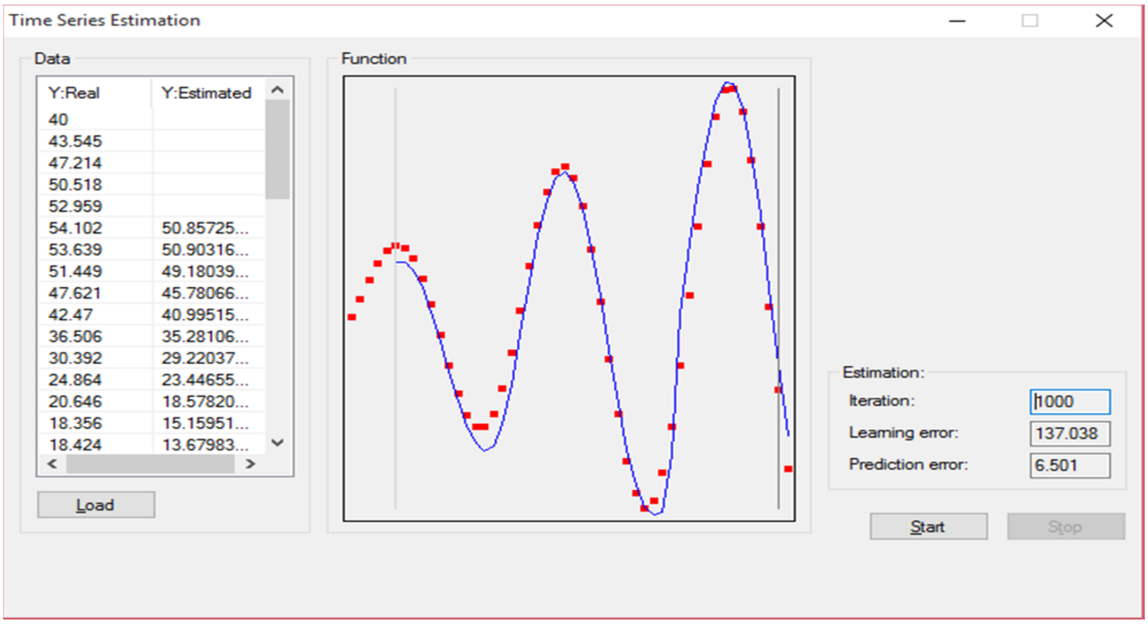


Fig3.7 Accuracy Indexing

4. RESULT AND DISCUSSION

Proposed accuracy estimation framework is implemented. For simplicity pre-processing and accuracy indexing layers are same for all the application scenarios. State estimation and accuracy estimation is varies over time and space.

In the time series estimation the prediction error and the learning error is calculated using the 1000 iterated values. Depending upon the prediction error and learning error, graph is created to show the accurate values that can be used for the specific application.



5. CONCLUSION

The problem of estimating the accuracy of one or more coexisting sensor systems is discussed based on the probabilistic measurements they generate. Now the proposed system has a general accuracy estimation framework, which has many layers to breaks the problem down and addresses it step by step.

This framework can be used in various cases, and implemented in different ways: In the state estimation layer, created a taxonomy of techniques, including simple inference-based and learning-based approaches, and CRF is used to infer and learn the parameter of the probability measurements over multiple timestamp; In the accuracy estimation layer, introduced two accuracy metrics, a proximity based and a similarity-based, and shown that they can be used by applications with different types of accuracy requirements using undirected graph.

REFERENCES

[1] Charles Sutton and Andrew McCallum, An Introduction to Conditional Random Fields
[2] Dempster A.P, Laird N.M, Rubin D.B, "Maximum likelihood from incomplete data via the EM algorithm," Journal of the Royal statistical society series B (methodological), vol. 39, no. 1, pp.1-38, 1997.
[3] Dong Wang, Tarek Abdelzaher, Hossein Ahmadi, Jeff Pasternack, Dan Roth, Manish Gupta, Jiawei Han, Omid Fatemieh, and Hieu Le, "On Bayesian Interpretation of Fact- f nding in InformationNetworks".
[4] Elnahrawy.E and Nath.B, "Cleaning and querying noisy sensors, "in Proc. WSNA, 2003.
[5] Fox D, Hightower J, Liao.L, Schulz.D, Borriello G, "Baysian filtering for location estimation," in IEEE Pervasive computing, 2003.
[6] Hasenfratz.D, Saukh.O, and Thiele.L, "Model-driven accuracy bounds for noisy sensor readings," in Proc. DCOSS, 2013, pp. 165–174.
[7] Huang.J, Millman.D, Quigley.M, Stavens.D, Thrun.S, and Aggarwal.A, "Efficient, generalized indoor wifi graphslam," in Proc.ICRA, 2011, pp. 1038–1043.
[8] James Biagioni, Tomas Gerlich, Timothy Merrif eld, Jakob Eriksson,"EasyTracker: Automatic Transit Tracking, Mapping, and Arrival Time Prediction Using Smartphones".
[9] Moghtadaiee.V, Dempster.A, and Lim.S, "Indoor localization using fm radio signals: A fingerprinting approach," in Proc. IPIN, 2011.
[10] Rasmussen.C and Williams.C, Gaussian processes for machine learning. MIT press Cambridge, MA, 2006, vol. 1.
[11] Reznik.L and Kreinovich.V, "Fuzzy and probabilistic models of association information in sensor networks," in Proc. FUZZ-IEEE, 2004.
[12] Osborne M. A. And Roberts S. J. , "Real-Time Information Processing of Environmental Sensor Network Data using Bayesian Gaussian Processes".
[13] Wang.D, Abdelzaher.T, Kaplan.L, and Aggarwal.C, "Recursive fact-finding: A streaming approach to truth estimation in crowd sourcing applications," in Proc. ICDCS, 2013, pp. 530-539.
[14] Wang.D, Kaplan.L, Le.H, and Abdelzaher.T, "On truth discovery in social sensing: a maximum likelihood estimation approach," in Proc. IPSN, 2012.
[15] Wang.D, Kaplan.L, and Abdelzaher.F.T, "Maximum likelihood analysis of conflicting observations in social sensing," ACM Trans.Sen. Netw., vol. 10, no. 2, pp. 30:1–30:27, 2014.
[16] Wen.H, Xiao.Z, Trigoni.N, and Blunsom.P, "On assessing the accuracy of positioning systems in indoor environments," in Proc. EWSN, 2013, pp. 1–17.