Research Article

Missing Value Imputation Based on Gaussian Mixture Model for the Internet of Things

Xiaobo Yan,¹ Weiqing Xiong,² Liang Hu,¹ Feng Wang,¹ and Kuo Zhao¹

¹College of Computer Science and Technology, Jilin University, Changchun, Jilin 130012, China

²School of Computer Science and Technology, Harbin Institute of Technology, Harbin, Heilongjiang 150001, China

Correspondence should be addressed to Liang Hu; hul@jlu.edu.cn

Received 6 June 2014; Revised 25 December 2014; Accepted 13 January 2015

Academic Editor: Thomas Hanne

Copyright © 2015 Xiaobo Yan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper addresses missing value imputation for the Internet of Things (IoT). Nowadays, the IoT has been used widely and commonly by a variety of domains, such as transportation and logistics domain and healthcare domain. However, missing values are very common in the IoT for a variety of reasons, which results in the fact that the experimental data are incomplete. As a result of this, some work, which is related to the data of the IoT, can't be carried out normally. And it leads to the reduction in the accuracy and reliability of the data analysis results. This paper, for the characteristics of the data itself and the features of missing data in IoT, divides the missing data into three types and defines three corresponding missing value imputation problems. Then, we propose three new models to solve the corresponding problems, and they are model of missing value imputation based on context and linear mean (MCL), model of missing value imputation based on binary search (MBS), and model of missing value imputation based on Gaussian mixture model (MGI). Experimental results showed that the three models can improve the accuracy, reliability, and stability of missing value imputation greatly and effectively.

1. Introduction

Recently, with the rapid development of key technology of the Internet of Things, most platforms and systems, based on the Internet of Things, are widely applied to various fields and industries, such as intelligent transportation, smart buildings, healthcare field, positioning and navigation field, and logistics field [1–3]. Meanwhile, it generates a lot of urgent problems to be solved, and one of these problems is missing value imputation for the Internet of Things [4]. Most learning algorithms, which are used to analyze data collected in the Internet of Things, generally assume that data is complete so each attribute of all instances is filled with a valid value. However, missing values are very common in the IoT for a variety of reasons, which results in people not being able to use the experimental data normally. For this reason, the accuracy and reliability of the experimental results will be greatly reduced. Therefore, it is very necessary for the IoT to estimate missing values. Missing value imputation exists in various domains, such as computer and network domain [5],

economy domain [6], medical domain [7], and psychological domain [8]. This paper addresses missing value imputation for the Internet of Things. In order to solve the problem, firstly, it is necessary to find the reasons for missing data [9]. There are many reasons for this phenomenon, such as unstable network communication, synchronization issues, unreliable sensors, and other equipment failure. Secondly, we need to study the mechanism of missing data [10]. Missing data mechanism can be divided into three categories: missing completely at random (MCAR), missing at random (MAR), not missing at random (NMAR). Thirdly, we have to study the pattern of missing data. Currently, there are two kinds of missing data patterns, and they are monotone missing pattern (MMP) and arbitrary missing pattern (AMP). Finally, we will have to establish missing value imputation model for the IoT and then give the estimated values of the missing data with the model. Missing values imputation algorithms, which are used commonly, are single imputation algorithm, multivariate imputation algorithm, and MCMC algorithm, and so forth.

Data collected in the IoT has many features, such as temporal and spatial correlation, correlation between individual properties, and correlation between individual properties. However, most studies of missing value imputation did not take into account the characteristics of the IoT data itself. And missing value imputation in the IoT field is rarely seen. This paper, for the characteristics of the data itself and the features of missing data in IoT, divides the missing data into three types and defines three corresponding missing value imputation problems, and then we propose three new models and solutions to solve the corresponding problems. Experimental results showed that the three models can improve the accuracy, reliability, and stability of missing value imputation greatly and effectively.

The rest of the paper is organized as follows: we first review the existing literature in Section 2. In Section 3, we illustrate Gaussian mixture model based on expectation-maximization algorithm. In Section 4, we define the problems of missing value imputation for the IoT. In Section 5, we establish three models to solve the missing value imputation problems. We present our experiments in Section 6. Finally, we conclude our study in Section 7.

2. Related Work

Inspired by ROUSTIDA, which is an incomplete data analysis method without statistical analysis and probability analysis, literature [11] proposes a new method for missing data imputation based on incomplete data clustering (MIBOI). This method requires a prespecified key parameter, which is the upper limit of constraint tolerance set dissimilarity, but it is very difficult in practical application.

Literature [12] proposes three data imputation algorithms based on the probabilistic path-event model to solve the problem of missing data imputation in RFID field. However, the matching algorithm, which is used by the data imputation algorithm, is inefficient.

Researchers propose an imputation technique for missing data based on spatial-temporal and association rule mining (STARM) to solve the context data missing problem in [13]. This technique can lead to data overflow, but this paper did not deal with the issue of data overflow.

To solve the problem of coverage holes in wireless sensor network (WSN), [14] proposes a new moving-neighborhood interpolation algorithm based on Delaunay triangulation technique. But this algorithm did not take into account the characteristics of the data itself.

Literature [15] proposes a multiple regression model-based missing values imputation algorithm to estimate the missing data as accurately as possible in wireless sensor networks. The estimation performance of the algorithm is more stable and reliable, but this algorithm has many default preconditions.

Literature [16] proposes a new single imputation algorithm based on locally linear reconstruction (LLR) to solve the problem of missing data imputation. This method can improve the prediction performance of missing values imputation, but it has to be used in a smaller range.

In [17], researchers propose a novel nearest neighbor (NN) imputation algorithm to estimate missing values in wireless sensor network by learning spatial-temporal correlation between wireless sensor nodes. This method has no assumptions for the WSN applications, but researchers do not extend the approach to other fields, such as the Internet of Things. And this algorithm lacks analysis of correlation between the properties.

3. Background

3.1. Gaussian Mixture Model. Gaussian mixture model (GMM) is commonly used for clustering. Each GMM consists of several Gaussian distributions, and each of them is called a component, which represents a different cluster. All the components are linearly added together to form a probability density function (PDF) of GMM.

There is a set of observed data, which is $X = \{x_1, \dots, x_i, \dots, x_n\}$. Each vector x_i is a p -dimensional vector. We assume that X is generated by a GMM with K components. The function $f_k(x_i)$ represents the probability density function of the k component, expressing the probability of x_i generated by the k component. So, the PDF of the GMM is as follows:

$$P(x_i) = \sum_{k=1}^K \pi_k f_k(x_i | \mu_k, \epsilon_k). \quad (1)$$

In the above formula, π_k represents the weight of the k component in the GMM; μ_k and ϵ_k represent the mean vector and covariance matrix of the k component; $P(x_i)$ represents the probability of x_i generated by the GMM. Moreover, π_k satisfies the following conditions:

$$\pi_k \geq 0, \quad \sum_{k=1}^G \pi_k = 1. \quad (2)$$

The PDF of the k component $f_k(x_i | \mu_k, \epsilon_k)$ is expressed as follows:

$$f_k(x_i | \mu_k, \epsilon_k) = \frac{\exp\left\{-(1/2)(x_i - \mu_k)^T \Sigma_k^{-1}(x_i - \mu_k)\right\}}{(2\pi)^{p/2} |\epsilon_k|^{1/2}}. \quad (3)$$

3.2. Expectation-Maximization Algorithm. Expectation-maximization algorithm (EM algorithm), formally proposed by Dempster, Laird, and Rubin in 1977, is an effective method to deal with incomplete data. EM algorithm is also a maximum likelihood estimation algorithm to solve model parameters from incomplete data. EM algorithm, with good local convergence, is simple and stable.

The basic principle of EM algorithm can be expressed as follows: X is the data that we can observe, but it is incomplete.

The complete data is $Y = (X, Z)$, and Z is the missing data. If θ is used to represent model parameter, $P(\theta | X)$ is the posterior distribution of θ with X . However, $P(\theta | X)$ is too complex to perform various statistical calculations. Assuming that the missing data Z is known, we can get a posterior

distribution of θ , $P(\theta | X)$, which is relatively simple so that we can perform a variety of statistical calculations. Looking back, the assumption about the value of Z can be checked and improved. Following this cycle, we will convert a complex maximization problem to a series of simple maximization problems.

There is a set of observed data $X = \{x_1, \dots, x_i, \dots, x_n\}$, and data is independent from each other. The implicit category of the data, Z , is introduced, and $Z = \{z_1, \dots, z_n\}$, $z_i \in \{1, 2, \dots, M\}$. M is a finite integer that is needed to be specified in advance. X and Z are combined together to form a complete data set, $Y = (X, Z) = \{(x_1, z_1), \dots, (x_n, z_n)\}$. The likelihood function of the complete data is as follows:

$$L(\theta | X, Z) = P(X, Z | \theta) = \prod_{i=1}^n P(x_i, z_i | \theta). \quad (4)$$

The expected value of the likelihood function is as follows:

$$E(L(\theta | X, Z)) = \int p(X, Z | \theta) f(Z) dz. \quad (5)$$

Each round iteration of EM algorithm consists of two steps: E-step and M-step. Each operation will produce a new value of θ parameter, but we have to give its initial value. Assuming that the initial value of θ parameter is θ^0 , the two steps are as follows.

E-Step. Define an auxiliary function, $Q(\theta, \theta^{(i-1)})$:

$$\begin{aligned} Q(\theta, \theta^{(i-1)}) &= E(\log L(\theta | X, Z)) \\ &= \int \log L(\theta | X, Z) f(Z | \theta^{(i-1)}) dz. \end{aligned} \quad (6)$$

In the above function, $Q(\theta, \theta^{(i-1)})$ is the function of θ parameter; $\theta^{(i-1)}$ is the value of θ from the last iteration. The value of $Q(\theta, \theta^{(i-1)})$ is the expected value of $\log L(\theta | X, Z)$.

M-Step. Calculate the value of θ^* , and $Q(\theta^*, \theta^{(i-1)})$ is the great value:

$$\theta^* = \theta^i = \arg \max Q(\theta, \theta^{(i-1)}). \quad (7)$$

It can be seen that the distribution of random vector Z is determined by X and $\theta^{(i-1)}$. If θ_i^* represents the maximum likelihood function value of the i iteration and θ_{i-1}^* represents the maximum likelihood function value of the $i-1$ iteration, it can be proved that EM algorithm can guarantee that $\theta_i^* \geq \theta_{i-1}^*$, and EM algorithm is convergent.

3.3. Gaussian Mixture Model Based on EM Algorithm. In a GMM, we assume that X is generated by a GMM with K components, and $X = \{x_1, \dots, x_i, \dots, x_n\}$. The complete data is $Y = (X, Z) = \{(x_1, z_1), \dots, (x_n, z_n)\}$. Z is the implicit category of the data. z_i represents the cluster of x_i , and $z_i = (z_{i1}, \dots, z_{iK})$:

$$z_{ik} = \begin{cases} 1 & \text{if } x_i \text{ belongs to group } k \\ 0 & \text{otherwise.} \end{cases} \quad (8)$$

Assuming that there are K clusters and the weight of the k component in the GMM is π_k , θ_k is the corresponding parameter. The density of z_i based on x_i is as follows:

$$\prod_{k=1}^K f_k(x_i | \theta_k)^{z_{ik}} \quad (9)$$

The logarithmic likelihood function of the complete data is

$$L(\theta_k, \pi_k, z_{ik} | x) = \sum_{i=1}^n \sum_{k=1}^K z_{ik} [\log \pi_k f_k(x_i | \theta_k)]. \quad (10)$$

Here, it is difficult to calculate the maximum value with the method of solving equations directly, so we will use EM algorithm to solve problem. In E-step of EM algorithm, the conditional expectation of the logarithmic likelihood function of complete data can be calculated with observable variables X and the estimated value of current parameter; in M-step, the weight, mean, and covariance matrix, which meet the log-likelihood function value which is the greatest, will be calculated with the results from E-step.

The parameter estimation of GMM based on EM algorithm is as follows.

Initialize weight, mean, and covariance matrix.

Repeat.

E-Step. z_{ik} is calculated:

$$z_{ik} = \frac{(\pi_k f_k(x_i | \mu_k, \epsilon_k))}{\left(\sum_{j=1}^G \pi_j f_j(x_i | \mu_j, \epsilon_j)\right)}. \quad (11)$$

M-Step. Calculate the estimated value of the maximum likelihood parameter with z_{ik} from E-step:

$$\begin{aligned} n_k &= \sum_{i=1}^n z_{ik}, \\ \pi_k &= \frac{n_k}{n}, \\ \mu_k &= \frac{\left(\sum_{i=1}^n z_{ik} x_i\right)}{n_k}, \\ \epsilon_k &= \frac{\left(\sum_{i=1}^n z_{ik} (x_i - \mu_k)(x_i - \mu_k)'\right)}{n_k} \end{aligned} \quad (12)$$

until reaching the satisfactory convergence criteria.

4. Problem Definition

Nowadays, the Internet of Things is widely used in a variety of fields, such as intelligent transportation, smart buildings, healthcare field, positioning and navigation field, and logistics field. The IoT is usually divided into three layers, from bottom to top, which are perception layer, transport layer, and application layer. Among them, the perception layer is responsible for data collection. There

are wireless sensor nodes, weird sensor nodes, and other hardware at the perception layer, which are integrated with the sensor module so that they can collect environmental data. Among these data collecting equipment, some nodes only integrate a kind of sensor module, such as temperature sensor module, and this type of sensor nodes can only collect a kind of environmental data, such as temperature. However, some nodes integrate multiple sensor modules, such as temperature module, humidity module, light module, and acceleration module. These sensor nodes can acquire diverse environmental data, such as temperature, humidity, light, and acceleration.

Data collected in the IoT are usually time-series data. A data collected at a time is called an instance, which generally contains multiple attributes: time, node number, sequence number, temperature, humidity, light, voltage, and acceleration. This phenomenon of data missing is very common in the IoT for various reasons, such as unstable network communication, synchronization issues, unreliable sensors, and other types of equipment failure. It results in people not being able to use the experimental data normally. As a result, the accuracy and reliability of the experimental results will be greatly reduced. So, we will have to estimate missing values. First of all, some definitions will be given as follows.

Definition 1. Data collected at a time in the Internet of Things is called an instance.

Definition 2. Data set consists of multiple instances.

Definition 3. An instance consists of basic attributes and observation attributes. Time, node number, and sequence number belong to basic attributes; temperature, humidity, light, voltage, and acceleration belong to observation attributes.

In the next section, D is used to represent a data set; X represents an instance, while x represents an attribute. The relationship between these variables is as follows:

$$\begin{aligned} D &= [X_1, \dots, X_i, \dots, X_m]^T, \\ X_i &= [x_{i1}, \dots, x_{ij}, \dots, x_{in}], \\ 1 &\leq i \leq m, \quad 1 \leq j \leq n. \end{aligned} \quad (13)$$

The basic attributes uniquely identify an instance. And we determine whether the data is missing through its basic attributes. We determine the basic attributes of an instance with the following methods: time can be obtained with historical time and data collecting frequency; node number and sequence number can be acquired with historical sequence number of nodes. Therefore, this paper addresses missing value imputation about observation attributes. In the next section, x_{ij} represents observation attributes; $\widehat{x}_{ij}$ represents the estimated value of missing data.

Definition 4. $\forall X_i \in D, \forall x_{ij} \in X_i$, if x_{ij} is missing, the problem of solving the estimated value of x_{ij} and making $|\widehat{x}_{ij} - x_{ij}|$ the minimum value is called missing data imputation problem.

Definition 5. $\forall X_i \in D, \forall x_{ij} \in X_i$, X_i has only one observation attribute. If x_{ij} is missing and $\exists(x_{(i-1)j}, x_{(i+1)j})$, $x_{(i-1)j}$ and $x_{(i+1)j}$ are not missing, the problem of solving the estimated value of x_{ij} and making $|\widehat{x}_{ij} - x_{ij}|$ the minimum value is called the first class of problem of missing data imputation.

Definition 6. $\forall X_i \in D, \forall x_{ij} \in X_i$, X_i has only one observation attribute. If x_{ij} is missing and $\exists x_{(i-1)j}$, $x_{(i-1)j}$ is missing (or $\exists x_{(i+1)j}$, $x_{(i+1)j}$ is missing), the problem of solving the estimated value of x_{ij} and making $|\widehat{x}_{ij} - x_{ij}|$ the minimum value is called the second class of problem of missing data imputation.

Definition 7. $\forall X_i \in D, \forall x_{ij} \in X_i$, X_i has two or more observation attributes. If x_{ij} is missing, the problem of solving the estimated value of x_{ij} and making $|\widehat{x}_{ij} - x_{ij}|$ the minimum value is called the third class of problem of missing data imputation.

5. Missing Value Imputation Model Based on GMM for the IoT

5.1. Model of Data Missing. In Section 4, the problems of missing value imputation for the IoT are divided into three categories. Next, the three corresponding models of data missing will be given.

Definition 8. $\forall X_i \in D, \forall x_{ij} \in X_i$,

$$x_{ij} = \begin{cases} 1 & \text{if } x_{ij} \text{ is not missing} \\ 0 & \text{if } x_{ij} \text{ is missing.} \end{cases} \quad (14)$$

The model of data missing, corresponding to the first class of problem of missing data imputation, is shown below:

$$D = [1, 0, 11, 0, 1, 1, 1, \dots, 1, 1, 0, 1]^T. \quad (15)$$

In the first class of data missing model, each instance has only one observation attribute. The values of observation attribute are missing at random, but the situation of data missing continuously does not exist.

The model of data missing, corresponding to the second class of problem of missing data imputation, is shown as follows:

$$D = [1, 0, 0, 0, 0, 0, 1, 1, \dots, 1, 0, 0, 1]^T. \quad (16)$$

In the second class of data missing model, each instance has only one observation attribute. The values of observation attribute are missing at random. If data miss, data will miss continuously in a certain period of time.

The model of data missing, corresponding to the third class of problem of missing data imputation, is shown below:

$$D = \left\{ \begin{array}{cccc} 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & \dots & 0 \\ 1 & 0 & 0 & & 1 \\ \vdots & & \ddots & & \vdots \\ 1 & 1 & 1 & & 1 \\ 1 & 0 & 1 & \dots & 0 \\ 1 & 1 & 0 & & 1 \end{array} \right\}. \quad (17)$$

In the third class of data missing model, each instance has two or more observation attributes. The values of an instance may all be missing or partially missing; the values of an attribute may be missing continuously or intermittently.

5.2. Missing Value Imputation Model Based on GMM for the IoT. Data collected in the IoT are usually time-series data, which are generated by kinds of nodes at different locations and different regions. These data has strong spatial-temporal correlation, attribute correlation, instance correlation, and nodes correlation, though the correlation is strong or weak.

According to the strong spatial-temporal correlation and other types of potential relevance, we build three models of missing value imputation, for three models of data missing, to solve the above three types of problems.

The model of missing value imputation based on context and linear mean (MCL) is as follows.

Model 1. $\forall X_i \in D, \exists x_{ij} \in X_i$, if x_{ij} is missing, $\exists x_{(i-1)j} \in X_{(i-1)}$, $\exists x_{(i+1)j} \in X_{(i+1)}$, but $x_{(i-1)j}$ and $x_{(i+1)j}$ are not missing; then,

$$x_{ij} = \frac{1}{2} (x_{(i-1)j} + x_{(i+1)j}). \quad (18)$$

In the first model of data missing, each instance has only one observation attribute, so $x_{ij} = x_i$.

The model of missing value imputation based on binary search (MBS) is as follows.

Model 2. $\forall X_i \in D, \exists x_{hj} \in X_h, \exists x_{tj} \in X_t$, if x_{hj} and x_{tj} are not missing, $\exists x_{ij} \in X_i$, x_{ij} is missing, and $t < i < h, i = \lfloor (1/2)(h + t) \rfloor$, then,

$$x_{ij} = \frac{1}{2} (x_{tj} + x_{hj}). \quad (19)$$

Model 3. The detailed process of missing value imputation model based on GMM for the IoT (MGI) is as follows.

Step 1. Data set D is divided into two data sets: D_y and D_n . D_y is a complete data set, which contains all instances without missing values; But D_n is an incomplete data set, which contains all instances with missing values.

Data partitioning rule is as follows. $\forall X_i \in D, \exists x_{ij} \in X_i$, if x_{ij} is missing, then $X_i \in D_n$; else $X_i \in D_y$.

Step 2. Cluster the complete data D_y with the GMM model based on EM algorithm, and find cluster center of each cluster. Then, determine the cluster of each instance from D_y .

Step 3. The incomplete data D_n are classified according to the result of clustering. Then, determine the cluster of each instance from D_n .

Classification rules are as follows. $\forall X_i \in D_n$, X_i belongs to a cluster, whose cluster center is the closest to X_i by the Euclidean distance.

Step 4. $\forall X_i \in D_n$, find one or more complete instances, which is the closest to X_i , in the cluster of X_i by the Euclidean distance.

Step 5. $\forall X_i \in D_n$, determine the value of X_i with the mean value of complete instances from Step 4.

The flow diagram of the MGI model is shown in Figure 1.

6. Experiment

6.1. Experimental Data and Settings. The programming software and data statistical analysis tool of this paper are Matlab (Release 2012a) and SPSS 20. Experimental data are collected from 54 sensors deployed in the Intel Berkeley Research lab between February 28 and April 5, 2004, [18]. The attributes of the data from the IoT includes time, node number, sequence number, temperature, humidity, light, and voltage. The sampling frequency of the sensor nodes is 2 times per minute. This paper, for three different data missing models, set three corresponding experiments.

6.2. Experiment of MCL Model. This experimental sample data includes 220 instances, which only contain temperature attribute. In the 220 instances, the temperature values of 20 instances miss at random, and this data missing situation is consistent with the first type of data missing model. This experiment adopts the MCL model to solve the first type of missing value imputation problems.

The results of this experiment are shown in Figures 2 and 3.

6.3. Experiment of MBS Model. In this experiment, sample data contain 220 instances, only including temperature attribute. In the 220 instances, the temperature values of 21 instances miss at random, and this data missing situation is consistent with the second type of data missing model. This experiment adopts the MBS model to solve the second type of missing value imputation problems.

The experimental results are shown in Figures 4 and 5.

6.4. Experiment of MGI Model. This experimental sample data of MGI includes 220 instances, which contain temperature attribute and humidity attribute. In the 220 instances, the values of 20 instances miss at random, and this data missing

FIGURE 1: Flow diagram of the MGI model.

FIGURE 2: Experiment results of MCL model.

FIGURE 3: Error analysis of MCL model.

situation is consistent with the third type of data missing model. This experiment adopts the MGI model to solve the third type of missing value imputation problems.

The experimental results are shown in Figures 6, 7, and 8.

6.5. Analysis and Evaluation of Experimental Results. Through the above experiments for the tree types of data missing, it can be seen that the three models of missing values imputation can estimate the missing values effectively. According to the experimental results and error analysis, we can see that MCL model, MBS model, and MGI model are

stable and reliable. And the estimate values from the three models have high accuracy and small error.

7. Conclusions

This paper, for the characteristics of the data itself and the features of missing data in the Internet of Things, divides the missing value into three types and defines three corresponding missing value imputation problems. Then, we propose three new models, which are MCL model, MBS model, and MGI model, to solve the corresponding problems. Experimental results showed that the three models can

FIGURE 4: Experiment results of MBS model.

FIGURE 5: Error analysis of MBS model.

improve the accuracy, reliability, and stability of missing value imputation greatly and effectively.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work is funded by European Framework Program (FP7) under Grant no. FP7-PEOPLE-2011-IRSES, by National Natural Science Foundation of China under Grants nos. 61073009 and 60873235, by National High Technology R&D Program 863 of China under Grant no. 2011AA010101, by National Key Technology R&D Program of China under Grant no.

FIGURE 6: Clustering results.

FIGURE 7: Experiment results of MGI model.

FIGURE 8: Error analysis of MGI model.

SQ2013GX11E00316-F03, and by National Science and Technology Major Projects of China under Grant no. SinoProbe-09-01-03&2012ZX01039-004-04-3. In addition, one of the authors wants to express their gratitude to one of the corresponding authors, Zhao Kuo. The work is also funded by Key Science Technology Program of Jilin Province of China under Grant no. 2011ZDGG007.

References

- [1] L. Atzori, A. Iera, and G. Morabito, "The internet of things: a survey," *Computer Networks*, vol. 54, no. 15, pp. 2787–2805, 2010.
- [2] Q.-B. Sun, J. Liu, S. Li, C.-X. Fan, and J.-J. Sun, "Internet of things: summarize on concepts, architecture and key technology problem," *Journal of Beijing University of Posts and Telecommunications*, vol. 33, no. 3, pp. 1–9, 2010.
- [3] Y. Mao, V. Bhuse, Z. Zhou, P. Pichappan, M. Abdel-Aty, and Y. Hayafuji, "Applied mathematics and algorithms for cloud computing and iot," *Mathematical Problems in Engineering*, vol. 2014, Article ID 946860, 2 pages, 2014.
- [4] Y. Liu, Y. Yang, X. Lv, and L. Wang, "A self-learning sensor fault detection framework for industry monitoring IoT," *Mathematical Problems in Engineering*, vol. 2013, Article ID 712028, 8 pages, 2013.
- [5] J. Yick, B. Mukherjee, and D. Ghosal, "Wireless sensor network survey," *Computer Networks*, vol. 52, no. 12, pp. 2292–2330, 2008.
- [6] Y. Huang, J.-F. Martínez, V. H. Díaz, and J. Sendra, "Localized and energy-efficient topology control in wireless sensor networks using fuzzy-logic control approaches," *Mathematical Problems in Engineering*, vol. 2014, Article ID 973163, 11 pages, 2014.
- [7] J. Wang, H. Zhao, Y. Bi, X. Chen, R. Zeng, and Y. Wang, "An improved task scheduling algorithm for intelligent control in tiny mechanical system," *Mathematical Problems in Engineering*, vol. 2014, Article ID 307869, 8 pages, 2014.
- [8] X. Zhang, J. Fang, F. Meng, and X. Wei, "A novel self-powered wireless sensor node based on energy harvesting for mechanical vibration monitoring," *Mathematical Problems in Engineering*, vol. 2014, Article ID 642365, 5 pages, 2014.
- [9] K. Zhao, K. Pan, and B. Zhang, "Cooperative transmission in mobile wireless sensor networks with multiple carrier frequency offsets: a double-differential approach," *Mathematical Problems in Engineering*, vol. 2014, Article ID 853821, 13 pages, 2014.
- [10] Y. Tian, Y. Ou, H. Reza Karimi, Y. T. Liu, and J. Q. Han, "Distributed multitarget probabilistic coverage control algorithm for wireless sensor networks," *Mathematical Problems in Engineering*, vol. 2014, Article ID 421753, 8 pages, 2014.
- [11] S. Wu, X.-D. Feng, and Z.-G. Shan, "Missing data imputation approach based on incomplete data clustering," *Chinese Journal of Computers*, vol. 35, no. 8, pp. 1726–1738, 2012.
- [12] Y. Gu, G. Yu, X.-J. Li, and Y. Wang, "RFID data interpolation algorithm based on dynamic probabilistic path-event model," *Journal of Software*, vol. 21, no. 3, pp. 438–451, 2010.
- [13] Y.-X. Wang, X.-Q. Qiao, X.-F. Li, and L.-M. Meng, "An imputation technique for missing context data based on spatial-temporal and association rule mining," *Journal of Electronics and Information Technology*, vol. 32, no. 12, pp. 2913–2918, 2010.
- [14] L. Bin, L. Yaping, Z. Siwang, L. Qing, and Y. Bo, "An interpolation algorithm based on sliding neighborhood in wireless sensor networks," *Journal of Computer Research and Development*, vol. 49, no. 6, pp. 1196–1203, 2012.
- [15] L. Pan and J. Li, "A multiple-regression-model-based missing values imputation algorithm in wireless sensor network," *Journal of Computer Research and Development*, vol. 46, no. 12, pp. 2101–2110, 2009.
- [16] P. Kang, "Locally linear reconstruction based missing value imputation for supervised learning," *Neurocomputing*, vol. 118, pp. 65–78, 2013.
- [17] Y.-Y. Li and L. E. Parker, "Nearest neighbor imputation using spatial-temporal correlations in wireless sensor networks," *Information Fusion*, vol. 15, no. 1, pp. 64–79, 2014.
- [18] S. Madden, *Intel Berkeley Research Lab Data [OL]*, 2006, <http://db.csail.mit.edu/labdata/data.txt.gz>.