Data Mining for Internet of Things: A Survey

Chun-Wei Tsai, Chin-Feng Lai, Ming-Chao Chiang, and Laurence T. Yang

Abstract—It sounds like mission impossible to connect everything on the earth together via internet, but Internet of Things (IoT) will dramatically change our life in the foreseeable future, by making many "impossibles" possible. To many, the massive data generated or captured by IoT are considered having highly useful and valuable information. Data mining will no doubt play a critical role in making this kind of system smart enough to provide more convenient services and environments. This paper begins with a discussion of the IoT. Then, a brief review of the features of "data from IoT" and "data mining for IoT' is given. Finally, changes, potentials, open issues, and future trends of this field are addressed.

Index Terms—Internet of Things, data mining.

I. INTRODUCTION

SINCE the first day it was conceived, the Internet of Things (IoT) [1], [2], [3] has been considered the technology for seamlessly integrating classical networks and networked objects [4]. The basic idea of IoT is to connect all *things*[1] in the world to the internet. It is expected that things can be identified automatically, can communicate with each other, and can even make decisions by themselves. In fact, it is even anticipated that the emergence of IoT as a service provider will be a trend although not all the technologies used are brand-new. In addition to the great progress on computer communication and relevant technologies that make many applications possible, a good news [5] is that IoT and machine-to-machine were worth $44.0 billion in 2011 and are expected to grow up to $290.0 billion by 2017. Thus, researchers from different industries, research groups, academics, and government departments have paid attention to revolutionizing the internet, by constructing a more convenient environment that is composed of various intelligent systems, such as smart home, intelligent transportation, global supply chain logistics, and health care [3], [4], [6], [7].

To clarify what the IoT refers to, several good surveys were presented recently each of which view the IoT from a different perspective: challenges [4], applications [7], standards [8], and smartness [9], [10]. Among these surveys, a comprehensive overview of the IoT from three different angles, things, internet, and semantics, was presented by Atzori and his colleagues [3]. Another recent study [6], inspired by [3], presented a generic five-layer architecture to describe the overall design of IoT. From the bottom up, the five layers are: edge technology, access gateway, internet, middleware, and application. Besides describing the infrastructures and things of IoT, a number of more recent studies [11], [12], [13], [10], [4] emphasized that most things on the IoT are supposed to have intelligence, thus are called "smart objects" (SO) and are assumed capable of being identified, sensing events, interacting with others, and making decisions by themselves [13], [10].

One of the most important questions that arise now is, how do we convert the data generated or captured by IoT into knowledge to provide a more convenient environment to people? This is where knowledge discovery in databases (KDD) and data mining technologies come into play, for these technologies provide possible solutions to find out the information hidden in the data of IoT, which can be used to enhance the performance of the system or to improve the quality of services this new environment can provide. A numerous researches are therefore focusing on using or developing effective data mining technologies for the IoT. The results described in [14], [15], [16], [17] show that data mining algorithms can be used to make IoT more intelligent, thus providing smarter services.

This paper gives not only a systematic description of data mining but also a detailed discussion about how to connect it to the IoT so as to provide a guideline for the researchers focusing on data mining or the IoT to shift to the next generation internet environment. The remainder of the paper, as the roadmap given in Fig. 1 shows, is organized as follows. Section I begins with a brief introduction to the IoT, followed by the motivation for applying data mining technologies to the IoT. Section II describes first the features of data from IoT and then the challenges in acquisition, management, and analysis of the data from IoT. Also described in this section are the possible solutions of using KDD for large scale data. The main purpose of Section III is to provide a comprehensive survey of data mining algorithms for the IoT, including classification, clustering, and frequent pattern mining. Several simple examples are given to associate the data mining technologies described herein with each other so as to reduce the effort needed in learning these algorithms. The focus in this section is on what the data mining algorithms described herein can do and how they do their job. The changes of the IoT itself, the potentials of applying data mining to the IoT, and the open issues of data mining technologies for IoT are presented in Section IV. Conclusions and future trends are drawn in Section V.

C. W. Tsai is with the Department of Applied Informatics and Multimedia, Chia Nan University of Pharmacy & Science, 71710 Tainan, Taiwan, ROC (e-mail: cwtsai0807@gmail.com).

C. F. Lai is with the Department of Computer Science and Information Engineering, National Chung Cheng University, 62102 Chia-Yi, Taiwan, ROC (e-mail: cinfon@ieee.com).

M. C. Chiang is with the Department of Computer Science and Engineering, National Sun Yat-sen University, 80424 Kaohsiung, Taiwan, ROC (e-mail: mcchiang@cse.nsysu.edu.tw).

L. T. Yang is with the the School of Computer Science and Technology, Huazhong University of Science and Technology, China (e-mail: ltyang@ieee.org). He is also with the Department of Computer Science, St. Francis Xavier University, Canada.

[1]Since no confusion is possible, throughout the rest of the paper, the terms "things," "devices," "objects," and etc. will be used interchangably.
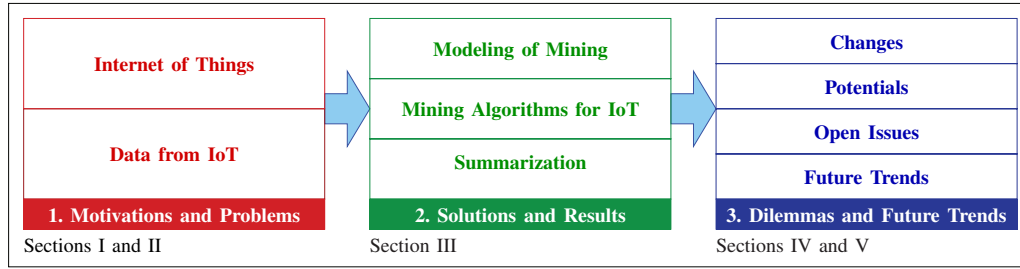
Fig. 1. Roadmap of this survey.

II. DATA FROM IoT

Technically, all *things* on the IoT may create a data deluge that contains different kinds of valuable information. However, technical issues and challenges on how to handle these data and how to dig out the useful information have emerged in recent years. A simple taxonomy [18] to differentiate the types of data from IoT is to use "data about things" to refer to data that describe things themselves (e.g., state, location, identity, and so on) and "data generated by things" to refer to data generated or captured by things. Normally, the former contains data that can be used to optimize the performance of the systems, infrastructures, and things of IoT whereas the latter contains data that are the results of interaction between humans, between human and systems, and between systems that can be used to enhance the services provided by IoT.

Owing to these characteristics, this new type of data (i.e., data captured by sensor or RFID) has been defined as a kind of "big data"[2] [20]. These characteristics let us quickly realize that big data is no longer a business hype, but something that is so real that we have to face seriously today. Although the amounts of data generated worldwide every year estimated by different studies [21], [22], [23], [24] are different, it is thought that the total amount of data generated has exceeded one zettabyte in recent years. It is evident that data analysis tools available today are simply not powerful enough to handle and analyze big data of IoT. There is no doubt that it is still a difficult problem to put more than one zettabyte into a single storage system. The next issue we need to take into account is that the data from IoT are generally too big and too hard to be processed by the tools available today [25], [14], [6], [26]. As Baraniuk observed [21], the bottleneck of data processing will be shifted from sensor to the data processing, communication, and storage capability of sensor. This observation also implies that the design and implementation issues of information system will be changed because of IoT.

Of course, since the problem of handling and analyzing large scale data has been around for years, it is not surprising that several traditional but efficient methods [27] presented in the past may be used to solve or mitigate the issues of handling the big data problem. These methods can be found in some previous data mining studies, such as random sampling [28], data condensation [29], divide and conquer [30], and incremental learning [31]. Among them, a possible way [21] to solve the big data problem of IoT is to have

---

[2]The *volume* of data, the *variety* of data types, and the *velocity* of data generated are the major characteristics of big data [19].

---

sensors acquire only the interesting data instead of all the data. One of the representative research trends has been on reducing the complexity of input data. One intuitive method is to use the principal component analysis (PCA) [32] or other dimension reduction methods to reduce the number of features of the input data. Recently, another promising trend called pattern reduction (PR), which is based on a different consideration, was presented in [33], [34]. Compared with PCA, this technology is aimed at reducing the number of patterns instead of the number of dimensions during the convergence process. Thus, it can also be used to reduce the complexity of input data. Different from these methods, some promising directions for handling the big data issue in recent years have been feature selection, distributed computing, and cloud computing [21], [35], [36], [37].

It is anticipated that organizations will own more and more large scale data (or big data) from the services, applications, and platforms they provide in the near future [38]. In addition to handling these massive data, how to find information hidden in these data has become an important task of these organizations because this task can put them in an unassailable position (i.e., a position to provide unique services) or create for them much more commercial values (i.e., innovation of products). As described in [39], [40], data analysis for sensors and devices may help us develop friendlier systems for smart city or smart home. From these examples, we realize that many potential applications can be developed by analyzing the big data. Since KDD was successfully applied to various domains to find hidden information out of the data in question, it has become the foundation of several information systems for years. It is naturally anticipated that KDD is able to find "something" from IoT, by using the following steps: selection, preprocessing, transformation, data mining, and interpretation/evaluation [41]. Of these steps, the data mining step, as the name suggests, plays the key role in extracting interesting patterns (rules) from the data. The other steps can be broadly divided into two steps: the data processing step (consisting of the selection, preprocessing, and transformation steps), which is to be taken before the data mining step, and the decision making step (consisting of the interpretation/evaluation step), which is to be taken after the data mining step.

As depicted in Fig. 2, IoT collects data from different sources, which may contain data for the IoT itself. KDD, when applied to IoT, will convert the *data* collected by IoT into useful *information* that can then be converted into *knowledge*. The data mining step is responsible for extracting patterns from the output of the data processing step and then feeding
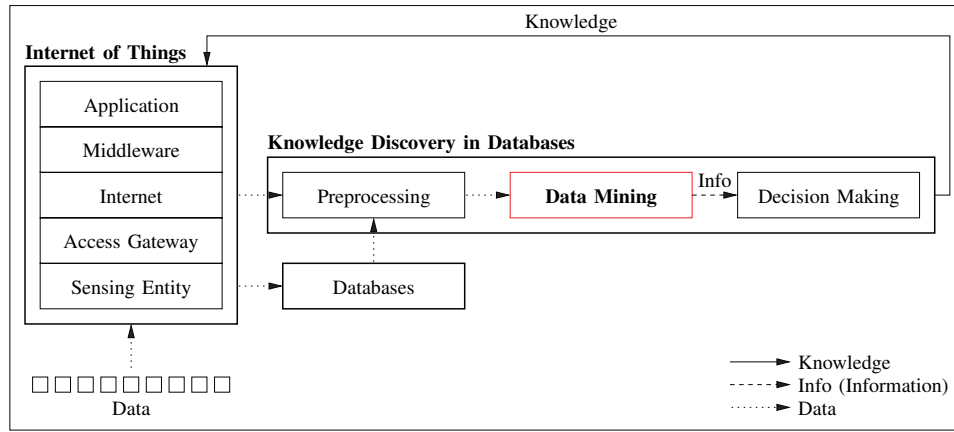
Fig. 2. The architecture of IoT with KDD.

them into the decision making step, which takes care of transforming its input into useful knowledge. It is important to note that all the steps of the KDD process may have a strong impact on the final results of mining. For example, not all the attributes of the data are useful for mining; so, feature selection is usually used to select the key attributes from each record in the database for mining. The consequence is that data mining algorithms may have a hard time to find useful information (e.g., putting patterns into appropriate groups) if the selected attributes cannot fully represent the characteristics of the data. It is also important to note that the data fusion, large scale data, data transmission, and decentralized computing issues may have a stronger impact on the system performance and service quality of IoT than KDD or data mining algorithms alone may have on the traditional applications.

III. DATA MINING FOR IoT

The relationships between big data, KDD, and data mining for IoT will be discussed in this section. A simple model for determining the applicable mining technologies and a brief introduction to the well-known data mining technologies for IoT will also be given in this section, by using a unified data mining framework and a few simple examples. After that, a detailed analysis and summarization of mining technologies for the IoT will be given.

A. Basic Idea of Using Data Mining for IoT

It is much easier to create data than to analyze data. The explosion of data will certainly become a serious problem of IoT. Until now, a numerous studies [14], [15], [16], [17] have attempted to solve the problem of inquiring big data on IoT. Without effective and efficient analysis tools, we, and all the systems, will definitely be submerged by this unprecedented amount of data. When KDD is applied to IoT, from the perspective of hardware, cloud computing and relevant distributed technologies are the possible solutions for big data; nevertheless, from the perspective of software, most mining technologies are designed and developed to run on a single system. In the circumstances of big data, it is almost certain that most KDD systems available today and most traditional mining algorithms cannot be applied directly

to process the large amount of data of IoT. Generally speaking, either the preprocessing operator of KDD or the data mining technologies need to be redesigned for IoT that can produce a large amount of data. Otherwise, the data mining technologies today can only be applied to small scale IoT system that can produce nothing but a small amount of data.

To develop a high-performance data mining module of KDD for IoT, the three key considerations in choosing the applicable mining technologies for the problem to be solved by the KDD technology—the objective, characteristics of data, and mining algorithm—are as given below.

- Objective ($\mathcal{O}$): The assumptions, limitations, and measurements of the problem need to be specified first so as to precisely define the problem to be solved. With this information, the objective of the problem can be made crystal clear.
- Data ($\mathcal{D}$): Another important concern of data mining is the characteristics of data, such as size, distribution, and representation. Different data usually need to be processed differently. Although data coming from different problems, say, D_i and D_j, may be similar to each other, they may have to be analyzed differently if the meanings of the data are different.
- Mining algorithm ($\mathcal{A}$): With needs (objective) and data clearly specified above, data mining algorithm can be easily determined, as to be discussed in Section III-E2.

Whether or not to develop a new mining algorithm can be easily justified by using these factors. For instance, from the characteristics of data, if the amount of data exceeds the capability of a system and if there is no feasible solution to reduce the complexity of the data, then a novel mining algorithm is definitely needed; otherwise, the current mining algorithm suffices. Another consideration is related to the property and objective of the problem itself. If a novel mining algorithm can enhance the performance of a system, then the new mining algorithm is also needed. An example is the clustering algorithm for a wireless sensor network, which needs to take into account the load of computation, but most traditional clustering algorithms simply ignore this issue.

Now that the objective of the problem is decided, the characteristics of the input data are understood, and the particular goals of mining and the mining algorithms are chosen, a

Algorithm 1 Unified Data Mining Framework	
1 Input data D	
2 Initialize candidate solutions r	
3 **While** the termination criterion is not met	
4 $d = Scan(D)$ [Optional]	**S**
5 $v = Construct(d, r, o)$	**C**
6 $r = Update(v)$	**U**
7 **End**	
8 Output rules r	

Algorithm 2 k-means algorithm	
1 Input data D	
2 Randomly create a set of centroids c	
3 **While** the termination criterion is not met	
4 $v = Assign(D, c)$	**SC**
5 $c = Update(v)$	**U**
6 **End**	
7 Output centroids c	

unified framework is presented here and used throughout the rest of the paper to describe all the mining algorithms presented in this paper to provide the audience a systematic way to understand data mining algorithms [42], [43], [44], [45], [46], [45]. As Algorithm 1 shows, the framework employs the following operations: initialization, data input and output, data scan, rules construction, and rules update. Note that in Algorithm 1, D represents the data; d the data read in by the scan operator; r the rules selected by the rules update operator; o the predefined measurement for the objective of the problem; and v the candidate rules created by the rules construction operator. To simplify the descriptions that follow, the core operators—data scan, rules construction, and rules update—will be denoted by **S**, **C**, and **U**, respectively.

- Scan (**S**): This operator plays the role of getting the needed data or information from the input patterns. Since not all the mining technologies need to read or check the input data at each iteration, the scan operator is optional for the proposed framework.
- Construct (**C**): Three parameters, d, r, and o, are usually employed to guide the search directions of this operator. The most important thing of the rules construction operator is to create the candidate rules (solutions) by using operators, such as selection, construction, and perturbation. For deterministic mining algorithms, a general way to create the candidate rules is by selection. For metaheuristic algorithms, it is the perturbation and construction that are usually used.
- Update (**U**): This operator is responsible for evaluating the candidate rules by some predefined criteria[3] first and then determining which rules will be retained for the next iteration.

The input, output, data scan, rules construction, and rules update operators are widely considered having a strong impact on the performance of data mining algorithms. As described above, this framework can be used to describe not only the traditional deterministic mining algorithms (e.g., k-means and apriori algorithm) but also the metaheuristic algorithms (e.g., simulated annealing and genetic algorithm). Several simple examples will be given in the following sections to show how this framework can be used to provide a unified perspective for different data mining algorithms, such as k-means for clustering, decision tree for classification, and apriori algorithm for association rules. The discussions on the approaches of different mining algorithms in the following sections are divided into two classes: infrastructures of IoT and services of IoT so as to make it easy to differentiate the goal of data mining for IoT, which is either for enhancing the performance of the system or for providing better services.

B. Clustering for IoT

1) Basic Idea of Clustering: The clustering problem [42], [43] is generally defined as follows: Given as input a set of unlabeled patterns[4] $X = \{x_1, x_2, \ldots, x_n\}$ in the d-dimensional space, the output of an optimal clustering is a partitioning of X into k clusters $\Pi = \{\pi_1, \pi_2, \ldots, \pi_k\}$ based on a predefined similarity metric, such as minimization of the intra-cluster distance and maximization of the inter-cluster distance. To represent the clustering result, the output of a clustering problem also consists of a set of means or centroids $C = \{c_1, c_2, \ldots, c_k\}$ to depict the characteristics of each group. A simple method for computing the means is as follows:

$$c_i = \frac{1}{|\pi_i|} \sum_{\forall x \in \pi_i} x. \tag{1}$$

How the quality of the clustering result is evaluated depends to a large extent on the application to which a clustering algorithm is applied. For instance, the sum of squared errors (SSE) is widely used for data clustering while the peak-signal-to-noise ratio (PSNR) is widely used for image clustering [27].

The k-means [47] is certainly one of the most well-known algorithms for the hard clustering problems. As shown in Algorithm 2, first, a set of centroids c are created randomly to represent how the input patterns are divided into different groups. Then, the assignment operator will compute the distances between patterns and centroids to find out to which group each pattern belongs. Because this operator needs to scan all the input patterns so as to assign each pattern to the group to which it belongs, it plays the role of scanning the patterns and constructing the candidate rules. The update operator plays the role of updating the centroids after all the patterns are assigned to the groups to which they belong by the assignment operator. The development of clustering algorithms afterwards has been more or less influenced by the k-means and its variants [48]. A simple example which combines genetic algorithm with k-means [49] shows that k-means can be used to fine-tune the clustering results of other metaheuristic algorithms. For the soft clustering problems, k-means can also be combined with fuzzy logic to provide better clustering results.

A simple example is shown in Fig. 3 where the incoming patterns are partitioned into three different groups according to a predefined proximity measure. The set of unlabeled input patterns, of course, can be any kind of data, such as the user behavior captured by the sensors of the IoT.

---

[3]The criteria depend on the kind of problem to be solved.

[4]That is, there is no prior knowledge about which pattern belongs to which group.
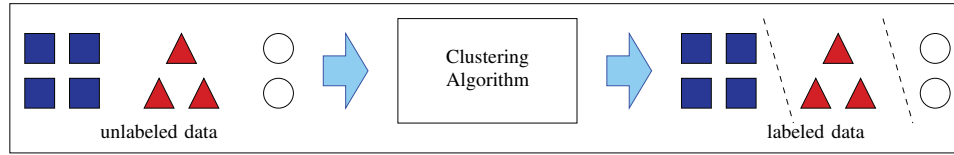
Fig. 3. Outline of the clustering process.

2) Clustering for Infrastructures of IoT: Unlike traditional clustering algorithms that employ deterministic local search methods to find the clustering results (e.g., k-means), in recent years, many metaheuristic algorithms [50] used stochastic methods to *guess* the clustering results. Owing to the characteristics of randomization in searching for the solutions, metaheuristic algorithms are less likely to fall into local optimum at the early iterations, thus having a higher chance to find better results than deterministic local search methods do, especially for complex and large data set.

The accuracy of the clustering results is not the only goal of clustering for the IoT. The other two important goals are, respectively, making clustering appropriate for these new situations and making it satisfy the conditions of the problems. One of the focuses of studies on clustering for the IoT is finding out the user behavior to provide the services the user needs. Another focus is on the distributed clustering [51] that is the basic requirement for a wireless sensor network (WSN) to prolong its lifetime. These observations lead to the development of several new clustering algorithms.

As an important kind of things of the IoT, although sensors nowadays provide small size, cheap, and energy efficient solutions for sensing events or monitoring the environment [52], they suffer from some dilemmas, especially energy conservation to maximize the system lifetime of a WSN [53]. According to the observation of [54], clustering is an efficient way to enhance the performance of IoT on the integration of identification, sensing, and actuation. That is why many new clustering algorithms are developed for the WSNs that are probably the most common devices to be found on the IoT.

One of the most well-known clustering algorithms that take into consideration the energy conservation of WSN is apparently the low-energy adaptive clustering hierarchy (LEACH)[5] [55]. Although it may not be the first study on applying clustering to WSN, it does significantly influence later researches on using clustering to minimize the global energy usage of WSN [56]. The basic idea of LEACH is very simple. Each sensor can be elected as the cluster head (CH) based on a probability model. The major problem [57] is that because the cluster heads are randomly elected, there may be too many CHs or no CHs at all. Another early clustering method was presented in [58]. It works by first constructing a maximum spanning tree and then using a simple greedy algorithm to fine-tune the partition results to balance the number of clusters and to minimize the total distance between the sensor nodes and the CHs of a WSN. A later study [51] used the estimated current residual energy of each node and the maximum energy to compute the probability of selecting the node to be the CH to provide a better solution for the

lifetime of a WSN. To speed up the transmission of data in a WSN, a two-phase clustering (TPC) algorithm for reducing the average transmission distance of WSN by creating an additional relay link for all the sensors was presented in [59]. In the first phase, TPC will first find out the cluster head that has the minimum transmission distance to the other sensors in the same group. Then, a direct link to the cluster head is created for each sensor node in the same group. In the second phase, each sensor will search for the nearest neighbor sensor node (not the CH) to construct an energy-saving data relay link.

In order to take into consideration more of the characteristics of the IoT, a dynamic clustering algorithm for exchanging information between nodes was presented in [60]. It works as follows: Each node chooses a remote node that is known and is interested, the two nodes then gossip (exchange) information with each other. As a result, each node can select an interested sensor node based on the current needs of the application dynamically. Since things may be able to make decisions by themselves, López et al. [61] pointed out that the design of clustering algorithm for the IoT needs to take into account the use of active network to create collaborative, multi-hop, and dynamic interaction or to exchange information between objects. How the information is forwarded to a network and other distant objects is essential in the development of clustering algorithm for the IoT.

3) Clustering for Services of IoT: Another research issue in using clustering technologies for the IoT is having services and applications of the IoT make decisions by themselves or provide better services, such as detecting the falling event of older people by using a smart home system.

An interesting smart home was presented in [62] the focus of which is on mining the frequent patterns as well as recognizing and tracking the user behavior in a smart home. After using discontinuous varied-order sequential miner (DVSM) to mine the useful patterns, the activity discovery method (ADM) is designed to cluster the sequences based on the simple k-means algorithm. However, rather than the Euclidean distance, it is the order and attribute distances that are used to measure the similarity between sequences of frequent patterns in such a system. More precisely, the clustering algorithm takes as input the output of the frequent patterns mining procedure (DVSM in this case). Then, the classification procedure based on the hidden Markov model (HMM, a statistical model for modeling the relationships or probabilities between patterns) takes as input the clustering results to identify activities in such a smart environment.

Although applying a clustering algorithm to a smart home needs to take into account new conditions for the smart home, just like clustering for WSN, the traditional clustering algorithms fortunately also work for smart homes the situa-

---

[5]It is worth mentioning in passing that LEACH is a routing protocol with adaptive clustering as a stage.

tions and conditions of which do not change dramatically. To automatically and more accurately take the location tracking data from different objects (e.g., RFID), the density-based spatial clustering of applications with noise (DBSCAN) is employed to group patterns to find the possible location of services [63].

Since the sensing and other relevant technologies have been applied to agriculture, how to use these technologies to improve the agricultural productivity has become a promising research area [64], [65] in recent years. In [64], Schuster et al. presented a two-step k-means algorithm to identify the management zones; i.e., to classify the zones with different planting strategies based on the slope and electrical conductivity. In the first step, the number of clusters is set to a large number, and the patterns are clustered based on the spatial proximity. Then, in the second step, the number of clusters is decreased, and the clusters obtained in the first step are clustered in terms of the spatial proximity and other similarities, such as the slope and electrical conductivity. Different from the above studies which are aimed to analyze the massive data captured from the sensors, RFID's, or other objects, the focuses of [66], [67], [68] are on the behavior or habits of a social network. The so-called things in these researches are the smart phones, personal digital assistants (PDA), and so on. Also, in this case, the clustering algorithm is used to find out the social relationships between the users and mobile nodes instead of the personal behavior in a smart home. In [67], the clustering results are used to predict the trace of mobile nodes and to provide the awareness information to the users. In other words, clustering can be used to find out latent communities existing in the IoT [68], which will be useful for connecting things to things, people to things, or people to people.

C. Classification for IoT

1) Basic Idea of Classification: Unlike clustering that assumes no prior knowledge to guide the partitioning process, classification [44], [45], [46] assumes some prior knowledge to guide the partitioning process to construct a set of classifiers to represent the possible distribution of patterns. It can be easily understood that classification is a supervised learning process whereas clustering is an unsupervised learning process. Mathematically, the classification algorithm can be defined as follows: Given a set of labeled data and a set of unlabeled data, the set of labeled data is used to train the classifier (i.e., the hyperline or prediction function) while the set of unlabeled data will then be classified by the classifier.

To evaluate the quality of the classification results, an intuitive way is to count the number of test patterns that are assigned to the right groups, which is also referred to as the accuracy rate (AR) defined by

$$AR = \frac{N_c}{N_t}, \tag{2}$$

where N_c denotes the number of test patterns correctly assigned to the groups to which they belong; N_t the number of test patterns. To measure the details of the classification results, the so-called precision (P) and recall (R) are commonly used [45]. Given that the four possible outcomes are

---

Algorithm 3 Decision tree algorithm

1	Input data D	
2	Initialize the tree t	
3	**While** the termination criterion is not met	
4	$h = Scan(d)$	**S**
5	$v = Split(h, t, o)$	**C**
6	$t = Update(v)$	**U**
7	**End**	
8	Output tree t	

---

true positive (TP), false negative (FN), false positive (FP), and true negative (TN), the precision (P) and recall (R) are generally defined as

$$P = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{3}$$

and

$$R = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{4}$$

respectively. Given P and R, a simple way to represent the precision and recall of the overall classification results, called F-score or F-measure, is defined as:

$$F = \frac{2PR}{P + R}. \tag{5}$$

Since the focus of a classification algorithm is on constructing a set of classifiers that can be used to represent the distribution of training patterns (such as how the patterns can be partitioned), a number of variants [46], such as decision tree, k-nearest neighbor, naïve Bayesian classification, adaboost, and support vector machine (SVM), have been developed to construct a set of better classifiers. As shown in Algorithm 3, most of the traditional classification algorithms, ID3, C4.5, and C5.0, are widely used in constructing the decision tree [69], [70] of a classifier; that is, each branch represents a test or check for partitioning the patterns. This implies that the information (e.g., entropy and diversity) required for constructing the decision tree needs to be computed at each iteration. This in turn implies that decision tree-based algorithms need the scan operator of Algorithm 3 to read the information contents of input patterns. In the construction phase (i.e., the split operator on line 5 of Algorithm 3), the decision node of a classifier is created based on the labeled patterns and the information contents. After the scan and construction operators finish their tasks, the decision tree will then be updated.

Once the decision tree or classifier was constructed, in the recognition phase, the unlabeled input patterns are assigned to the groups to which they belong based on the decision tree constructed in the construction phase. By walking from the root of the decision tree down to the leaf (i.e., passing through all the if-then-else checks along the path), each input pattern will be assigned to the group to which it belongs. Different from the decision tree, the naïve Bayesian classification [71], [72] uses the so-called probability model to create the classifiers which are estimated by the training patterns. Recently, the support vector machine (SVM) [73] has become a promising classification algorithm due to one of the important characteristics of SVM. That is, SVM can efficiently separate non-linearly separable patterns by transforming them to a high-dimensional space using the kernel model.
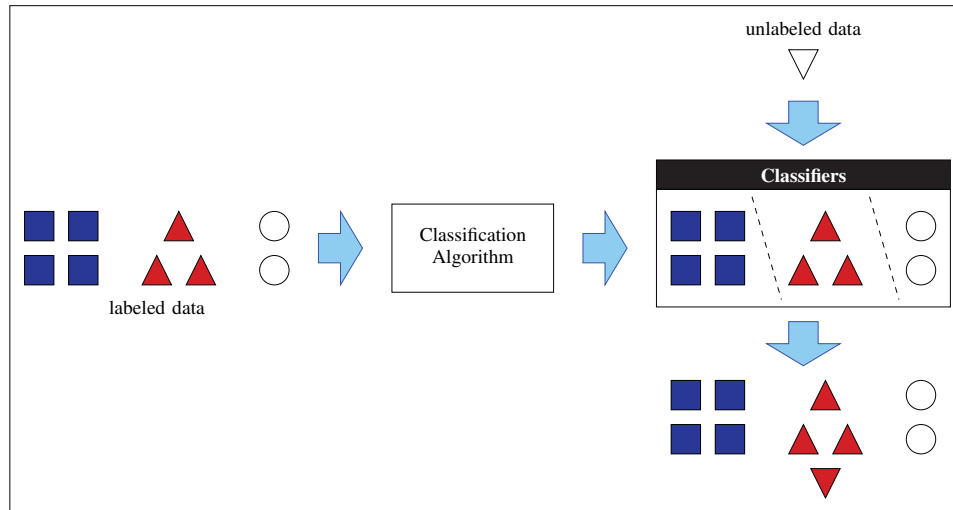
Fig. 4. Outline of the classification process.

The example described in Fig. 4 shows that before classifying the unlabeled input patterns, a classification algorithm will use the labeled input patterns to create a set of classifiers that act like a knowledge base. Then, all the unlabeled input patterns will be classified by the set of classifiers created previously. By using this information (the classifiers or hyperlines), the classification algorithm can easily partition all the input patterns into the groups to which they belong. In the case patterns are far away from any groups, some studies create new groups to describe these patterns.

2) Classification for Infrastructures of IoT: Although researches on classification for the IoT are still at their infancy, they do have the potential for enhancing the performance of the infrastructures and systems of the IoT. One of the situations to be expected is a number of unique item identifier (UII) schemes for the IoT today or in the future because several standards or applications have developed their own schemes in recent years. To avoid or mitigate the problem of UII query being increased dramatically (even more than the number of DNS queries in the current internet environment), Lee et al. [74] developed a binary tree-based classification algorithm to solve this problem. By checking only a certain number of the initial bits of a device, the tree-based classification can quickly recognize the type of the device; that is, there is no need to check the complete header of a device.

3) Classification for Outdoor Services of IoT: Researches on using classification for the IoT can be divided into two categories: outdoor and indoor, depending on the region. One of the nightmares occurred frequently in our daily life is the traffic jam problem, especially when you are living in a big city. More and more researches therefore are focusing on the traffic jam problem, by using mobile devices or smart phones to interchange information to avoid getting stuck on road, such as traffic forecasting services [75] and accident detection [76]. Since getting information about the traffic situation via the internet and social network has become a trend, in [77], a driver guidance tool was developed by integrating the location information provided by the GPS, geographical information tracking from vehicles, and other information that can be collected from the internet to predict the future traffic situation. The decision tree classification algorithm is then used by this kind of system to predict and suggest the routing path for a driver, based on real-time information, historical data, and so on.

Like the traffic jam problem that happens way too often in a big city, another problem is spending a considerable amount of time to look for the parking space. The sensor technologies, such as pressure sensors and video cameras, have successfully applied to parking lot monitoring systems. A different method presented in [78] uses the so-called passive infra-red (PIR) sensor to detect the motion of vehicles based on the information that comes from both the inner and outer nodes to classify the vehicles that are entering or leaving to count the number of cars parked in the parking lot. In addition, the presented method also integrates the short message service (SMS) notification system with the parking lot monitoring system to send the availability information to the user who needs a parking space.

4) Classification for Indoor Services of IoT: For the indoor systems of the IoT, a variety of successful applications have been developed by using different sensor technologies. Among them, the researches on health care are inseparable from the smart home because one of the important goals of a smart home is to classify the activities of the human at home or in an indoor environment.

Noury and his colleagues [79], [80] have shown the ability of the current smart home to recognize the activities of human. This kind of smart home incorporates several sensor technologies and web cameras to monitor the space of a smart home. Infrared presence sensors, microphones, or even wearable kinematic sensors are used to collect information of the monitored space. In addition to defining the activities (e.g., bathing, dressing, and toilet use) and integrating the domain knowledge from different devices (e.g., feature extraction from cameras and sensors), the performance of a classification algorithm will also impact the accuracy of this kind of system. One-against-one support vector machine (OAO-SVM) is used in [79] for multi-class classification to provide an

accurate activity detection and prediction. The SVM-based classification algorithm, of course, has also been used to track the human behavior in the other smart home researches [81]. The k-nearest neighbor [82], the naïve Bayesian and dynamic time warping [83], and the C4.5 classification algorithm [84] are also used to monitor the human activities (standing posture, temporary posture, and lying posture) to detect the falling event of older people.

The technologies relevant to the researches mentioned above can also be used to predict the preference of users in order to provide the services that they need [85], [86]. The video-based classification is one of the important researches which include object identification [87], face recognition [88], and facial expression classification [89]. To increase the accuracy of this kind of application, integration with other devices seems to be a feasible way to go. Yang et al. [89] collected data from multiple devices (including camera, microphone, and other devices) to recognize the painful, sad, and other facial expressions to monitor humans, called NCHMS. Also, a three-dimensional emotion model (pleasure-arousal-dominance, abbreviated PAD) is then used to classify the human emotions.

Owing to the privacy concern of users, the sound signal has become a preferred approach in capturing the data of a smart home. This can be complemented by the image or video signals to improve the accuracy rate of detection, such as location awareness. In [90], the data collected by multiple microphones are employed to recognize the sound events, such as door clapping, glass breaking, phone ringing, step sounds, screaming, dish sounds, and door locking. The training process of such system contains the expectation maximization (EM) and k-means algorithms to create the classifier for the multi-channel sound detection.

Several other studies [91], [92] focus on the health care of a smart home, by using different detection technologies to avoid sudden death events, such as cardiovascular disease. Different from the falling event detection that can use the information captured by video camera or microphone, this kind of event detection needs to take into consideration the signal (data) of human body. Also different from the classical electrocardiogram (ECG) researches that were aimed at analyzing the data stored in the data center of a hospital or in a public database, the focus of smart home and smart environment is on real-time event detection. The wireless ECG sensor (WES) has become one of the most important devices to replace large instruments in a hospital. To address the real-time issue in the detection of events, the adaptive neuro-fuzzy inference method was used in [92] to train the classifier for detecting sudden polarity change events in this kind of system. Another key point to take into account is that feature extraction may significantly impact the accuracy of the classification (detection) results in this kind of system, such as peak and wave data in the ECG.

D. Frequent Pattern Mining for IoT

1) Basic Idea of Frequent Pattern Mining: The main difference between frequent pattern mining [44], [93], [94], [45] and other mining technologies is in that the former is focused on finding out the "interesting patterns" from a large set of data items. One of the well-known frequent pattern mining

---

Algorithm 4 Apriori algorithm	
1 Input data D	
2 Initialize the rule r	
3 **While** the termination criterion is not met	
4 $d = Scan(D)$	**S**
5 $v = FindFrequentPatterns(d, r, o)$	**C**
6 $r = FindAssociationRules(v)$	**U**
7 **End**	
8 Output rules r	

techniques is the association rule [95], [44], [93], which is usually used to find all the co-occurrence relationships from a set of data items called associations. A simple example that is often used to explain the concept of association rule is discovering items that will be purchased together with items already purchased. Mathematically, the problem can be defined as follows: Given a set of items $I = \{i_1, i_2, \ldots, i_m\}$ and a set of transactions $T = (t_1, t_2, \ldots, t_n)$ where $t_i \subseteq I$, find the set of association rules[6] which are greater than or equal to the predefined threshold values of support and confidence. In other words, the two important conditions to evaluate the mining results, support and confidence, are predefined by the user. For example, a transaction rule for buying bread and milk together, denoted {bread} $\Rightarrow$ {milk}, with the value of support set equal to 10% and the value of confidence set equal to 70%, indicates that 10% of the customers buy bread and milk together while each customer has a 70% of chance to buy milk if he or she bought bread. Mathematically, the support is defined as

$$\text{support}(X \Rightarrow Y) = P(X \cup Y) = \frac{\text{TC}(X \cup Y)}{n}, \qquad (6)$$

and the confidence is defined as

$$\text{confidence}(X \Rightarrow Y) = P(Y|X) = \frac{\text{TC}(X \cup Y)}{\text{TC}(X)}, \qquad (7)$$

where $\text{TC}(\psi)$ denotes the number of transactions in T that contain ψ, i.e., $\psi = X \cup Y$ or $\psi = X$.

Another key issue of frequent pattern mining is the "order" of transactions, called sequential pattern [94], [45]. The difference between association rule and sequential pattern is in that the association rule is aimed at finding interesting patterns from a transaction while the sequential pattern is aimed at discovering interesting patterns from a sequence of transactions [94]. The problem of sequential patterns can be defined as follows: Given a set of items I and a set of transactions T, the goal of the sequential patterns problem [45], [94] is to discover all the sequences with a minimum support where the minimum support of a sequence, support(s), is defined as the fraction of all the data sequences that contain the particular sequence s. In addition, the three important parameters that can be used to define the situation we need to observe are the duration of a time sequence, the event folding window, and the time interval for events [44].

Since the frequent pattern mining problem is aimed at finding hidden information from a database of transactions, several researches have attempted to develop more efficient algorithms for this problem, which look like the apriori algorithm [95]. As

---

[6]For example, the set of items Y co-occurrenced with another set of items X is denoted by $X \Rightarrow Y$ where $X \subseteq I$, $Y \subseteq I$, and $X \cup Y \neq \emptyset$.

shown in Algorithm 4, the apriori algorithm for the association rule problem is aimed at (1) scanning the dataset to discover all the frequent item sets f_s and (2) generating all the association rules from f_s with a certain confidence value.

The example given in Fig. 5 illustrates that as the name suggests, the basic idea of frequent pattern mining is to find patterns that occur frequently in a database and that satisfy the predefined support and confidence conditions.

2) Frequent Pattern for Infrastructures of IoT: A couple of comprehensive surveys on the association rules and sequential patterns were given in [93], [94]. Several studies, such as temporal data mining [96], privacy preserving on the IoT [97], web usage mining, bioinformatics, and intrusion detection [98], also provided successful applications and showed the potential. Among these studies, how to handle and analyze a large number of transactions has become more and more important in recent years owing to the fact that 90% of the data created today was born in electronic form [99]. To deal with the situation of having billions of RFID tags or objects in a smart environment or the IoT, how to manage these objects effectively and find information from them efficiently has become an interesting research topic. In [100], [101], Li et al. used the time to live (TTL) information to manage the 870,000 RFID events entering their system per second. By detecting the event sequence and the four types (absolute, relative, periodical, and sequential) of the TTLs, the system can accurately identify the RFID events at real time. Another interesting application is the spatial collocation pattern mining [102], [103], [104] which can be used in the ubiquitous geographic information system (GIS) or as part of the foundation of the IoT to discover the relationships between spatial collocation patterns. For example, the relationships between two collocation spatial patterns X and Y can be represented by the association rule $X \Rightarrow Y(\text{PI}, \text{CP})$ where the participation index PI and conditional probability CP are the two measures for this problem [104]. Although the spatial collocation pattern mining problem is also aimed at finding the frequent patterns, just like the association rule problem, the classical association rule algorithms need to be modified for this problem [103], [104], which includes not only the properties of the spatial data but also the time-consuming issue.

3) Frequent Pattern for Services of IoT: Like most approaches for the association rule, analyzing the purchase behavior still attracts the attention of researchers and companies using the RFID or even IoT approaches. In [123], Schwenke et al. provided a high level method to describe the agent states and the customer behavior in a supermarket, namely, thinking, moving, and action. To solve the problem of customers not being able to find the products they are looking for quickly, in [111], customer specific rules (e.g., age and family state), category-based rules, and association rules are integrated into the same buying suggestion system to help the customers of a supermarket find the products they are looking for. The rule-based and case-based reasonings are used to analyze the data to make the system be able to suggest to the customers what they need more accurately based on (1) the personal favor of a customer, (2) the personal purchase history, and (3) the behavior of other people.

The other approach based on the RFID and sensor technologies is the smart environment which also needs mining technologies to make it more intelligent so as to provide more convenient services. Different from most studies on classification that employed the activities of daily living (ADL) to describe the activities, a set of motions are used to facilitate the classification algorithm to differentiate different events, as discussed in Section IV-C. The focus of [112], [113], [114] is on the relations between temporal activities. Inspired by [124], several studies [112] were focused on describing and defining the relations of temporal activities. The relations are first divided into before, after, meets, met-by, overlaps, overlapped-by, starts, started-by, finishes, finished-by, during, contains, and equals, which can then be used to describe the relations between temporal activities. To predict the activities in a smart environment, a probability method was proposed to compute the possibility of event X appearing with another event Y as follows:

$$
\begin{aligned}
P(X|Y) = (&|\text{After}(Y,X)| + |\text{During}(Y,Z)| \\
&+ |\text{OverlappedBy}(Y,Z)| + |\text{MetBy}(Y,Z)| \\
&+ |\text{Starts}(Y,Z)| + |\text{StartedBy}(Y,Z)| \\
&+ |\text{Finishes}(Y,Z)| + |\text{FinishBy}(Y,Z)| \\
&+ |\text{Equals}(Y,Z)|)/|Y|.
\end{aligned}
\tag{8}
$$

The apriori algorithm is also used in [112] to discover the frequent patterns.

Later studies [119], [120] add the k-means clustering algorithm to classify the activities to create a normal mixture model for each activity before the association rules mining is applied. Another study [121] integrates the association rule mining with the linear support vector machine (LSVM) classification algorithm to improve the accuracy rate of a behavior prediction system for the health care at home. Patterns captured by four different kinds of sensors—the ECG sensor, temperature sensor, network camera for the human location and motion, and facial expression sensor—are used to detect the events. The LSVM is used to recognize the home services while the association rule plays the role of analyzing the home services for the human if any sudden events occurred, such as a human suffering from stress. To take into account the performance issue of these systems, especially when the data is large and the response time is too long, in [115], a fast association rule algorithm for health care, especially for the elderly surveillance system, was presented. The key idea of this research is based on the perspective of statistics to create the association rules, which employ the p-value to check the dependency among a set of events. With the acquisition, activity, and wellness determination modules, the study described in [116] also defined two wellness functions (β_1 and β_2) to analyze the data from the sensor networks to determine the status of human in a smart environment. The wellness functions β_1 and β_2 for inhabitants are based on the inactive and excess usage measures of appliances, respectively.

The activity sequences, of course, can provide abundant information for a user behavior detection system; however, how to define or describe the information is a difficult issue. As mentioned previously in this section, the activities of daily living (ADL) [125], [126] or relations of Allen [124] can be used to describe daily activities. An alternative is to use
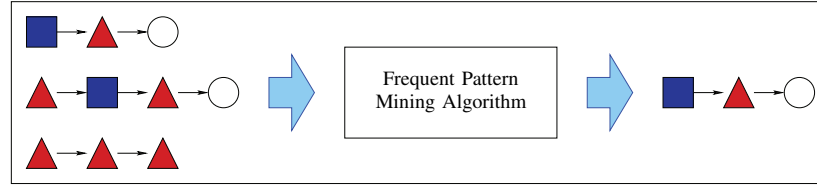
Fig. 5. Outline of the frequent pattern mining process.

TABLE I
COMPARISON OF MINING TECHNOLOGIES.

Mining Algorithm	Goal	Data Source	References
Clustering	Network performance enhancement	wireless sensor	[55], [105], [56], [57], [58], [51], [106], [59], [60], [61]
	Inhabitant action prediction	X10 lamp and home appliances	[107]
	Provisioning of the needed services	Raw location tracking data	[63]
	Housekeeping	Vacuum sensor	[108]
	Managing the plant zones	GPS and sensor for agriculture	[64], [65]
	Relationships in a social network	RFID, smart phone, PDA, and so on	[66], [67], [68]
Classification	Device recognition	RFID	[74]
	Traffic event detection	GPS, smart phone, and vehicle sensor	[75], [76], [77]
	Parking lot management	Passive infrared sensor	[78]
	Inhabitant action prediction	RFID, sensor, video camera, microphone, wearable kinematic sensor, and so on	[79], [109], [110], [80], [81], [82], [89] [83], [84]
	Inhabitant action prediction	Video camera	[87], [88]
	Inhabitant action prediction	microphone	[90]
	physiology signal analysis	wireless ECG sensor	[91], [92]
Frequent Pattern	RFID tag management	RFID	[100], [101]
	Spatial colocation pattern analysis	GPS and sensor	[102], [103], [104]
	Purchase behavior analysis	RFID and sensor	[111]
	Inhabitant action prediction	RFID and sensor	[112], [113], [114], [115], [116], [117]
Hybrid	Inhabitant action prediction	RFID and sensor	[62], [118], [119], [120], [121], [122]

the ontology technologies to describe the relations between objects, activities, and services [122] to provide more services to the users of such a system. More precisely, the system described in [122] first models the relations between activities and objects. The user behavior or patterns are then used to find out the services that the user may need in the future.

E. Summary

1) Overall Comparison: A simple comparison of the data mining technologies from different perspectives is given in Table I to outline the data mining technologies for the IoT. The "Goal" column of this table gives the problem that needs to be solved; the "Mining Algorithm" column the method that can be used; and the "Data Source" the input patterns that will be used. This table shows that all the mining algorithms described in the table can be used to increase the intelligence of the IoT. The clustering, classification, and frequent pattern mining technologies can be used to predict the action of inhabitants in a smart environment [107], [79], [109], [110], [80], [81], [82], [89], [83], [84], [87], [88], [90], [112], [113], [114], [115], [116], [117]. Of course, the application of these mining technologies is not limited to a smart environment. They can also be used in other domains, such as enhancing the performance of WSNs [55], [105], [56], [57], [58], [51], [106], [59], [60], [61] and detecting and managing traffic events [75], [76], [77].

A promising trend that can be easily found from these successful examples is that data mining technologies are able to make the IoT or relevant infrastructures smarter or even more intelligent. A brief historical retrospect of the researches

```
Algorithm 5 Metaheuristic algorithms
1   Input data D
2   Randomly create a set of initial solutions s
3   While the termination criterion is not met
4       r = Transit(s)                          C
5       f = Evaluate(d, r, o)                    S
6       s = Determine(r, f)                      U
7   End
```

on the IoT shows that some studies [96], [62], [117] have implied that mining technologies can be used to increase the smartness of the IoT or relevant infrastructures. The focus of data mining technologies for the IoT is not only on the smart home but also on the environment of our daily life, such as the selling strategy of a supermarket and the event detection on a transportation management system. Also, data mining technologies are used in some studies to enhance the performance of the IoT, such as minimizing the global energy usage of WSN [56] and accelerating the recognition time of RFID [100], [101]. From the perspective of the "Data Source" column of Table I, RFID and a variety of sensors are employed in different researches, some of which use a single device while some of which use multiple devices for mining. These examples show that data mining technologies can be employed to analyze the data if the system or device is able to collect the needed data. The deeper meaning is that data mining technologies can be applied to different types of data.

Another promising trend is to use metaheuristic algorithms [50] to solve a variety of mining and optimization problems [9] of IoT. A simple example for illustrating how metaheuristic
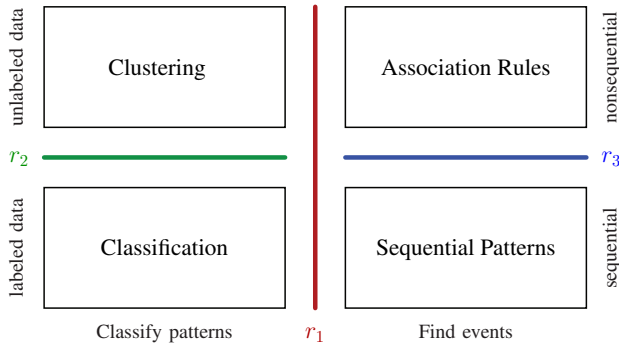
Fig. 6. The mining technology classification matrix.

algorithms can be used to solve mining problems under the framework of Algorithm 1 is given in Algorithm 5, which consists of the initialization, transition, evaluation, and determination operators [127]. In Algorithm 5, s denotes the current solutions; r the candidate solutions; f the evaluated values of the candidate solutions in r. This example shows that to apply metaheuristic algorithms to mining problems, the move, recombination, or perturbation of the transition operator of metaheuristic algorithms corresponds to the construction operator of Algorithm 1 while the solution evaluation and search direction determination of metaheuristic algorithms can be regarded as the scan and update operators.

2) Which Mining Technologies Are Applicable: Now that a unified framework has been presented to give a systematic description of data mining algorithms and how they are applied to IoT, the question that arises is, what mining algorithm is suitable for the development of a high-performance system or for the provisioning of a better service in different IoT environments that we may encounter. Fig. 6 gives a classification matrix to help us differentiate clustering, classification, association rules, and sequential patterns algorithms we have by three simple rules as given below.

- The first rule, denoted r_1, is employed to differentiate the four mining technologies into two classes based on the characteristics of the problem. The first class contains technologies to divide or classify patterns into different groups, namely, clustering and classification while the second class consists of technologies to find out the most frequently occurring patterns, namely, association rules and sequential patterns.
- The second rule, denoted r_2, is presented to differentiate clustering from classification. That is, clustering may be a more suitable solution for problems for which all the input patterns are unlabeled whereas classification may be a better choice for problems for which some of the input patterns are labeled.
- The third rule, denoted r_3, is used to differentiate association rules from sequential patterns mining based on whether the events are sequential or not.

Given the classification matrix and the three rules, we can now easily determine which mining technology will be applicable to the problem we have at hand, as follows:

1) Clustering: if the goal is to *classify patterns* all of which are unlabeled.

2) Classification: if the goal is to *classify patterns* some of which are labeled and some of which are unlabeled.
3) Association Rules: if the goal is to *find events* from the input patterns that occur in no particular order.
4) Sequential Patterns: if the goal is to *find events* from the input patterns that occur in some particular order.

Although most mining technologies are designed for finding useful information that is hidden in the data, the original objectives are different. For example, clustering is used to classify a set of patterns whereas frequent pattern mining is often employed to find patterns that occur frequently in a set of patterns. More and more studies attempted to combine different mining technologies to provide better services in recent years because all the mining technologies have pros and cons. If we consider only the main data mining technologies on clustering, classification, and frequent pattern mining, as shown in Fig. 7, the total number of possible combinations is 12.[7] Several researches showed that some of these combinations can potentially be applied to the IoT to find useful knowledge.

The problems that arise now are, which kind of mining technology is better and how does this mining technology cope with the other mining technologies or modules in a system? A comprehensive consideration and design of the system is definitively needed, especially for an intelligent system, because a single algorithm or module will not work out for extracting information, thinking, and making decisions, which need cooperation between algorithms and modules. For this reason, the classification of mining technologies, possible combinations, and simple examples will be given in the following sections to show one of the many possible ways to combine mining technologies into a single system.

The first combination is "clustering $\rightarrow$ classification," i.e., clustering first and then classification, that can be considered as the unsupervised classification system. In this case, clustering is usually used to automatically create a set of classifiers without a priori knowledge on the input patterns while classification then uses these classifiers to classify the incoming patterns. Different from the first combination, the second combination is "classification $\rightarrow$ clustering," i.e., classification first and then clustering, that can be considered as an incremental classification system or the so-called semi-supervised learning system. Although they use the same mining technolgies, the goals are very different. For the second combination, it is classification that is responsible for creating a set of classifers from a set of labeled patterns. Then, clustering is used to incrementally (dynamically) add new classifers to the set of classifiers created by using classification, when necessary. In other words, it is this kind of combination that enables the possibilities of handling patterns entering the IoT dynamically, i.e., to recognize the human faces or behavior not previously in the knowledge database (the set of classifiers). The combinations, clustering $\rightarrow$ classification $\rightarrow$ frequent pattern

---

[7]The total number of possible combinations is 12 in terms of how the three different mining technologies are ordered. There are total six combinations when we use only two mining technologies, i.e., we can first employ clustering and then use classification, we can first employ classification and then use clustering, and so on. Another six combinations come up for the situations when we employ three mining technologies.
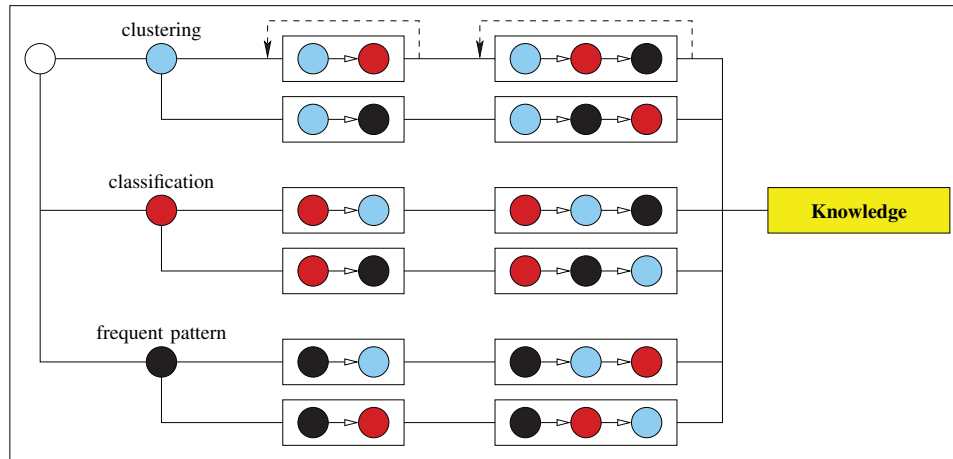
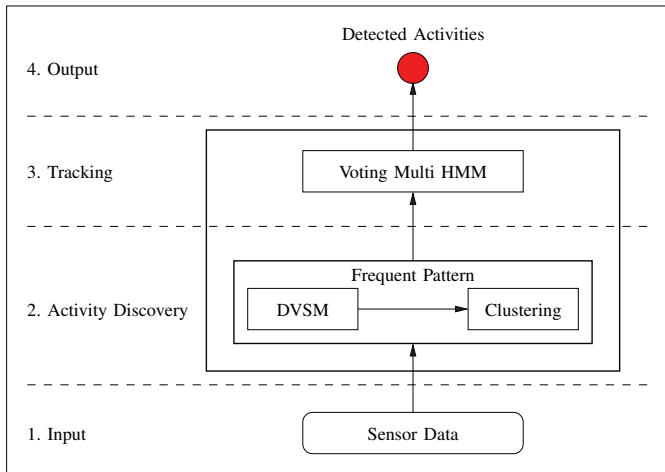Fig. 7.   Different combinations of mining technologies for the IoT.

Fig. 8.   Main components of a smart home system [62].

and classification → clustering → frequent pattern, integrate three mining technologies mentioned above. An interesting example is the activities detection system described in [62], which integrates clustering, classification, and association rule mining technologies into a single system. Another interesting combination is one that performs a combination repeatedly for several times to enable the iterative learning or to find better solutions. As shown in Fig. 7, the combination can be clustering first, then classification, then back to clustering, then classification, and so on, or it can be the loop of clustering → classification → frequent pattern. This is conceptually very similar to that of iterative soft computing algorithms.

3) Simple Examples: An integrated system for a smart home to recognize and track human health was presented in [62]. It can be easily found that mining technologies (frequent pattern mining, clustering, and classification) are indispensable as far as smart home and other systems of the IoT are concerned. Of course, such a system does not appear suddenly. Rather, each module, which uses different mining technologies, can be considered as a building block presented in a series of studies by Cook and her colleagues [117], [120], [62]. In the early works of Cook, it is the basic idea of MavHome [117] that becomes the foundation of the following

studies on the smart home, especially for the prediction of human activities in their daily life.

Fig. 8 shows the main components of a smart home system, from the collection of sensor data to the discovery of the frequent and repeatable patterns to the detection of events. Clustering technologies can use the pattern sequences extracted by the frequent pattern mining to classify the activities into different groups. Finally, by integrating classification, clustering, and frequent pattern mining, an intelligent model can be constructed to recognize human activities in such a smart environment. An interesting study was also presented in [108] for the wireless vacuum system. Like the WSN, the smart home vacuum (SHV) [108] environment has the cluster head and sensor nodes and attempts to maximize the network lifetime. Unlike the WSN, the SHV also concerns the cleaning efficiency. It is explained that the relevant experience and mining technologies after some modifications may be useful for other environments or problems if the needs or concerns for the IoT are similar because it is often an interdisciplinary research, meaning the need of knowledge from different domains.

As we all know, once the relevant technologies and approaches for smart home are mature, the high level integration of smart city will become obvious automatically. However, smart city is far more than a composition of a set of smart homes, which, among others, needs to take into account power system, public transportation system, security, emergency, and other public facilities. To realize an ideal smart city (i.e., to make it as *Utopia* as possible), cloud computing [128], [129], smart grid [130], [131], [132], and data mining [14], [15], [16], [17] are definitely the essential components for connecting things together and for enhancing the performance of the systems so as to provide better services to the residents. A simple example for the integration of different technologies is plug-in hybrid electric vehicle (PHEV) and gridable vehicle (GV) [133], [134] which use smart grid[8] as the infrastructure for charging the vehicle, for accessing the information of

---

[8]Smart grid is a promising power grid which can provide two-way communication instead of one-way communication as the traditional power grid does. Thus, many novel approaches can be easily realized, such as monitoring the usage of electrical devices and predicting the power consumption.

neighboring buildings available on a cloud, and for sensing the events of things. Then, data mining can be used to find out the information to make the plan for driving a vehicle.

A simple framework for the smart city is given on the left of Fig. 9 while an example is given on the right to explain how to apply the mining technologies to IoT and how to integrate them with other necessary technologies. As shown in Fig. 9, the IoT for a smart city is described as a hypercube that contains a set of cubes (sensing entities, applications, and integrated services). More precisely, the framework is used to explain: (1) all the things (sensing entities) can be connected to the internet; (2) some applications need data collected by different things; (3) applications interact with each other via the internet; (4) a smart city contains a set of integrated services, a bunch of applications, and a large number of things; and (5) each application cube may be composed of many sensing entities while each integrated service may be composed of many applications.

Two applications (parking lot and smart home) are given in Fig. 9 to show that data collected by things can be shared by different applications. Each application is composed of a number of particular sensing entities. For instance, a parking lot uses data collected by cameras and pressure sensors to monitor the current state of the parking lot. The smart home example gives the basic idea of how to integrate different kinds of renewable energy appliances, such as wind power generator, solar panel, and PHEV. In other words, the energy can be used more efficiently when these appliances (things) are connected to the internet. On the level of a city, the main issues are how to coordinate and integrate things, applications, and systems. Not only can the data mining technologies be applied to adjust things in the level of sensing entities, but they can also be applied to things in the levels of applications (e.g., store) and integrated services (e.g., city government). In summary, the major differences between the mining technologies for different levels of applications and for different applications in the same level of IoT are the goal (i.e., the objective and limitations) and the input data types. Thus, to provide better services of IoT for a smart city, we need to take into account not only data sensing and preprocessing but also the goal of the problem so as to determine the applicable mining technologies.

IV. DISCUSSIONS

One possible way to make a system smart, or "think" like a human, is using data mining technologies, but how to make it work as expected is still a difficult problem today, just like we have gone a long way trying to make computers think by themselves. To develop a convenient, high-performance, and intelligent IoT system using the data mining technologies, we need to know the key issues they face. The focus in this section is thus on three different research aspects: the changes, the potentials, and the open issues of the IoT in using data mining technologies in the design, integration, development, and use of a smart IoT system.

A. Changes Caused by IoT

In this section, changes for the IoT inspired by [3] are discussed from three different perspectives—thing-oriented, internet-oriented, and semantics-oriented.

The very first kind of change is on the "thing." A very simple way to differentiate a device is if it can automatically become part of the IoT; that is, if it can be automatically connected to the internet, either directly or indirectly through other devices disregarding its capability. Another change is that devices are made smaller and smaller, implying that the lifetime of battery can be prolonged and thus more functions can be integrated into each device, such as integrating RFID and wireless sensor into a single device. These changes make it possible for every thing to upload the data it collected to the server or application. We can then imagine that all the things on the planet being connected together [135] will come true in the foreseeable future. Owing to these changes on things, data mining technologies are now able to enhance the performance of devices, to give them basic autonomous or cognitive ability to make decisions by themselves (e.g., to give air-conditioner the autonomous ability to automatically adjust the temperature of a room) and to interact with other things (i.e., machine to machine).

The second kind of change is on the "internet" of the IoT in that many more IP addresses are required for the devices, and an enormous amount of data will be created by the devices that may flood the internet. A good news is that IPv6 provides a possible solution to the IP address issue. However, the large amount of data will certainly be a critical issue. Many researchers have started paying their attention to the big data issue, for it is self-evident that data generated by IoT will flood the internet. The data we are talking here are not only the data for the interaction between devices, servers, and applications but also the data collected by all the devices (e.g., the temperature data and the pressure data). The rapid speed of devices joining and leaving the internet environment is another significant change to be expected because devices may not connect to the IoT all the time for several reasons. Considering the battery usage, a device may turn itself on or off, depending on whether it has data to transfer or not. Another example is that devices of a vehicle, when turned on to collect data, may connect to the internet most of the time, but not necessarily all the time. The next generation of internet will then face the problem of large amount of data that it may not be able to handle (i.e., process or store them), especially when taking into account the rapid speed devices join or leave the IoT.

One of the important changes to the "semantics" is definitively the so-called smart object. According to [13], the foundation of smart objects is the ability to identify themselves, to sense data, to make decision, to reach other devices and services, and to exchange information with others. Different from the perspective of intelligent things, for intelligent management of the information, Atzori et al. [3] pointed out that how to represent, store, interconnect, search, and organize the information is still a very difficult problem. The changes of things and internet generally are due to the advance of computer technologies; however, the semantics for making the IoT more intelligent is due to the fact that we want to improve the level of services for this new environment. In summary, for the intelligence of the IoT, the changes depend on what we need and what we will do about it, which is no longer limited to integrating existing technologies into the system.
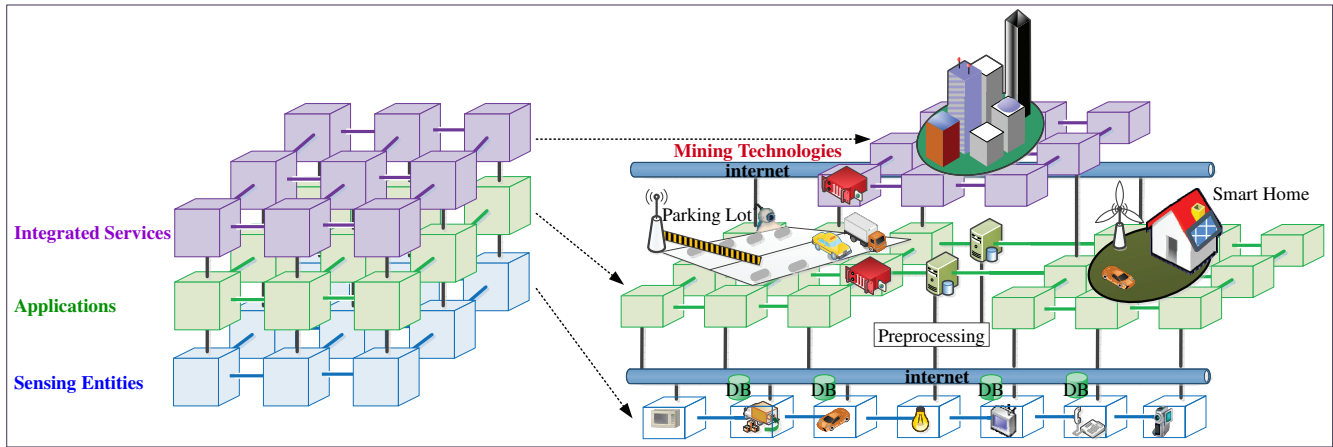
Fig. 9. Framework of a smart city.

B. Potentials of Using IoT

Several possible applications of the IoT have been presented or are about to be presented in the near future as are mentioned in several technical reports [39], [40], research papers [6], [7], and books [136], [137], [138] and by international companies [135], [139]. Because it is expected that the IoT will create a considerable sum of profits to many domains, from the high-technology researches to our daily life, several nationwise research organizations [140], [141] have focused on it. In [3], Atzori et al. classified the possible applications into transportation, health care, smart environment, and personal and social domains. Another study [142] divided the application domain into society, industry, and environment based on the activities of the IoT on the kind of events, such as where it will be used and who will use it. A recent study [39] showed 50 interesting applications for building a smart world which are classified into thirteen categories: (1) smart cities, (2) smart environment, (3) smart water, (4) smart metering, (5) security & emergencies, (6) retail, (7) logistics, (8) industrial control, (10) smart agriculture, (11) smart animal farming, (12) domotic & home automation, and (13) eHealth.

Now, we can imagine that every thing in our daily life, whatever size it is, is possible to become part of the IoT. From the perspective of using data mining technologies to enhance the performance, applications of the IoT can be divided into two groups: systems and services. To enhance the performance of the IoT, Section III-B2, Section III-C2, and Section III-D2 give examples to show the focus of current researches. Examples are prolonging the lifetime of battery or recognizing the device type quickly. To provide a better service of the IoT, Section III-B3, Section III-C3, Section III-C4, and Section III-D3 give some possible solutions, such as increasing the accuracy rate of the falling event detection and predicting the services users need. With these examples and looking back at [26], the gap between the current work on the IoT and the ultimate goal is how to move from knowledge to wisdom even though some works are arguably a smart system. Of course, it somehow depends on the definition of smartness and intelligence for a system.

Most researches on applying mining technologies to the IoT-oriented systems today are applying simple mining al-

gorithm to create rules to split the data or the sequence of interesting behavior. These results are not the end, but the start of further analysis. Therefore, our observations show that the potentials of applying data mining technologies to the IoT can be summarized as follows:

1) To people: One of the potentials is to provide a more accurate suggestion; that is, the IoT acts like a semi-automatic intelligent system. Different from a fully automatic system, this kind of system may require a high accuracy rate or a high intelligence, such as medical diagnosis. For systems that demand a high accuracy rate, mining technologies play the role of supporting decision making or providing suggestions (data or hidden information) to the users rather than participating in the final decision. Combining mining technologies with iterative learning methods will bring the IoT up to another level in the sense that the system now may be able to learn incrementally and even interact with people.

2) To themselves: One of the promising researches on the smart object is for things to think by themselves. In practice, mining technologies have the potential to filter out the redundant data and to decide what kind of data or information needs to be uploaded to the system that will be useful for applications on a broad region with limited resources, such as precision agriculture and natural resources monitoring. In addition, reasoning the current situation for providing a possible solution is another potential of data mining technologies, such as dynamically fine-tuning the air-conditioning for a comfortable room temperature.

3) To other things: Another key research issue on the IoT or M2M is communication and collaboration with other things. However, the focus is not on transmitting data to the others and finishing the work together. If the objects or systems are similar to each other, thus having similar goals or requirements, clustering technologies will be able to be used to group them into the same group so that the objects can easily know which objects need the information they hold (those on the same group) and which objects do not (those on different groups).

C. Open Issues of IoT

Although the focus of mining problems for the IoT differs from that of the traditional mining problems, they still inherit

many of the open issues of the traditional mining algorithms and problems in the development of the mining algorithms for IoT, such as scalability for large-scale data set. There is no need to say that many other open issues need to be addressed in the design of the mining algorithms for IoT. With the technologies available today, here are some of the open issues:

1) Infrastructure Perspective: Two characteristics, *decentralization* and *heterogeneity*, of IoT [142] strongly impact the development of data mining algorithms. Although most researches on IoT emphasize that computing and data storage of IoT need to be decentralized, most data mining technologies are designed in such a way that they need to have an overall view of the data in order to find useful information. That is why several of the current researches on the smart environment using data mining technologies still employ computing that is centralized. This means that we need to reconsider how to appropriately decentralize mining technologies for the IoT.

A typical example is clustering for WSN, as mentioned in Section III-B2. Since most traditional clustering algorithms are designed to perform on a single system (centralized) with all the data, they apparently are inappropriate for saving the energy of sensors of a WSN. Another example is the classification algorithm for IoT which also suffers from the issue of decentralized computing. According to our observations, however, not all the computations need to be distributed to many things (i.e., sensors or devices), for doing this will not bring you all the benefits of using the decentralized strategy. For instance, the accuracy rate of classification may degrade because the classification algorithm does not have access to all the data. We do believe that "decentralized computing" is just one way to make the algorithm fit to the external environment of the situations it faces, but not the only one. Hence, whether to decentralize or not depends to a large extent on the application, the system, or the application domain. But mining algorithms for the IoT nowadays are often centralized for a smart home for which the scale is small because its focus is on the accuracy. For the scale of, say, a smart city, decentralization is essential because the data needed to be processed at the same time are way too large. Our observations also show that the centralized computing on a local site (i.e, a group of devices) and the distributed computing on the global site (i.e., the application or system) may ease the shift of traditional mining technologies to the IoT environment with a minimum number of changes.

The issue of low computation and high throughput of IoT comes from the fact that mining algorithms are designed to be performed on devices of small size and low power consumption. That is why the mining algorithms cannot be designed to use too much memory and CPU nowadays even though the computation ability of sensors is expected to become more and more powerful in the near future. A possible solution is to let cloud-based systems play the role of computing and storage. The reality is that cloud computing is absolutely not a catholicon because the cost of communication, the demand for a large amount of data storage, and the demand for computing power still exist but cloud computing may not be able to satisfy all these demands. The limitation of network bandwidth will bring up several dilemmas and open issues.

A trade-off solution is thus to have things cache temporary data "locally" and cloud keep refined data "globally." For this reason, when integrating IoT with cloud computing systems, the critical issue is how to use cloud systems to manage (e.g., data replication) and analyze data collected by things.

2) Data Perspective: In order to deal with the open issues of *large-scale* data entering the system from *different data resources quickly* and *dynamically*, several relevant technologies, such as data preprocessing, information extraction, and information retrieval [143], are usually employed. Simply speaking, determining what data need to be kept on the sensor and what data need not is a difficult issue for a sensor. The limited memory size is not the only reason for this issue. The other reason is in that we need to filter out redundant data so as to avoid their influence on the performance of the system. That is why several recent studies attempted to provide solutions to handle the big data issue from things, such as dimension reduction, data compression, and data sampling. For the big data issue of IoT, it can be easily found that acquisition, deposition, analysis, and integration involve several vital open issues that may strongly impact the performance of an IoT system [144], [145].

The other important issue is that most data are captured by different sensors, RFID's, and devices; thus, the data are either heterogeneous or dispersed to different regions, systems, or applications. Without an effective way to connect the systems on IoT, each will do things in its own way. The question that arises now is, how is the knowledge shared with other applications, systems, or services, which is much more important than in the traditional environment? To define the devices of the IoT, several standards, protocols, and interfaces have been presented, such as ISO18000-1 to ISO18000-7 for RFID and IEEE 802.15.4 for WSN. However, the "meaning" of the input data from different sensors, devices, applications, and systems, may not be exactly the same. It may significantly impact the quality of the data mining results, for the similarity measure is not as precise as it used to be. A simple example was given in [146] to explain the situation, which says that "tall" people in two different systems may represent different degrees; therefore, a mining algorithm may classify them into the same group even though the degrees are very different. The reason is very simple. Data captured from different resources do not use the same representation method, which may be caused by different representative ranking or different relational data [48].

Ontology, semantic web, extensible markup language (XML), and other related technologies seem to be able to provide a solution to this issue. But they are technologies, not the final solutions. Our observations show that interoperability and interactivity between things are key to solving this issue. Numerous researches used classification algorithms to recognize the activities of users in a smart home or smart environment in recent years. An intuitive solution to this issue is to use the activities of daily living (ADL) [125], [126] or instrumental activities of daily living (IADL) [147], [148] to define the activities of human. How to detect the posture of human motion from a video will take us back to some of the open issues of computer vision or image processing. Fortunately, sensor technologies (e.g., microphone and cam-

era) provide these detection systems (health care systems in a smart home) additional data or information to improve the accuracy rate of detection and classification. According to our observations, in addition to ontology, semantic web, XML, ADL, IADL, and other related technologies, the collective intelligence (CI) [149] for the internet can enrich the ontology and semantic web technologies to automatically or semi-automatically construct the knowledge base to describe and define the particular patterns (i.e., association rules) because some of the human knowledge and wisdom can be found on the internet.

3) Algorithm Perspective: Since sensors or devices may join or leave the system at any time or the network topology may be changed, the characteristics of data that need to be analyzed are not always the same or static. In some cases, the stream data will enter the system not at the same time. Dynamically adjusting the mining results and rerunning the mining algorithm are two intuitive ways to solve the problem faced in a dynamic environment, but both approaches increase the computation time. Therefore, what we need to have so as to solve this issue is a flexible and precise mining technique.

Another representative example is that classification for the IoT needs to either dynamically add classifiers into the system or statically adjust the classifiers for the situations the system faces. It is interesting to note that not all the systems need to be fully dynamic. The smart home example gives us a chance to look at the issue from a different angle. That is, do we need to dynamically add classifiers to the system when the incoming and outgoing people are the members of a family? The example also shows that a static system may have a lower accuracy rate because it cannot handle the events which are not predefined for the system, such as the invasion of strangers. In summary, we suggest that a dynamic mining algorithm be used except that the thresholds or conditions for adding classifiers or for adjusting the classifiers of such a system can be made more restrictive to retain the accuracy rate while providing the flexibility.

The open issue of how to combine mining technologies with other analysis methods or mining technologies is due to the fact that more and more intelligent systems for a smart environment or application of IoT emphasize this kind of integration, such as integration with clustering [119], [120] and with classification [121]. One possible solution is to use frequent pattern mining technologies as either the front end or the back end of the other mining modules. The input and output will therefore become a more critical issue than ever before. In the worst case, the algorithm may need to be redesigned to cope with the other modules or algorithms of such a system. The possible ways to integrate mining technologies are given in Section III-E2 and Fig. 7. The open issues on the integration of different mining technologies are: how the data from a mining algorithm (or module) are transferred to another and how the computation load and services are balanced.

Another critical issue, i.e., the so-called overfitting, is not caused by the mining problems but by the mining algorithms. For example, although the labeled patterns act like the knowledge base (experience or knowledge from experts) of the system that can be used to construct the classifier (model)

to classify the unlabeled patterns, the answer to how many patterns are needed to train the classifier is not unique. The more the training patterns, the higher the accuracy rate. This can be easily justified by observing that a larger number of training patterns implies a more precise training. Labeling patterns is very expensive in terms of the computation time or cost. Thus, the number of training patterns used depends to a large extent on the desire to achieve a balance between cost and accuracy. This is due to the fact that some mining algorithms employ iterative process to find the solutions; thus, after too many iterations of training, the candidate solutions created by the mining algorithms may fit too well to the training patterns. This kind of situation happens not only on the traditional mining algorithms but also on the metaheuristic-based mining algorithms. For this reason, to avoid falling into local optimum at the early iterations, the timings to restart the mining algorithms, to adjust (i.e., to delete, add, or modify) the candidate solutions, and to select the features are the main issues on over training, which may have a strong impact on the mining technologies of IoT.

4) Open Issues on Privacy and Security: Even though some of the camera-based detection and recognition technologies are mature enough, the privacy concern of users will make most people uncomfortable. Similar concerns when using data mining technologies to analyze the data from IoT appear in recent years. For example, companies today can easily collect various kinds of data of consumers from different sources or devices and then use data mining technologies to find the information that can be used to make marketing tactics to increase the volume and revenue of sales, but the fact is that not many consumers would like to have their privacy, such as shopping behavior, collected. Another example is applications related to the health care in a smart home or hospital. In this case, the sensitive information, such as the behavior of patients, needs to be protected. According to these observations, an efficient and effective method which preserves the performance of mining in terms of the computation time also has to protect the personal information. To mitigate these privacy issues, some technologies have been proposed in recent years, such as anonymization [150], randomly delay [151], temporary identification [152], and encryption [153].

An alternative solution is thus the sound sensor, RFID, or other wearing sensors. But the problem that arises now is, although the new sensor or RFID technologies can benefit the mining module of the IoT, they will also bring up new challenges, such as how to handle data captured by multiple sensors and how to deal with the security problems [154]. Data mining technologies can still recognize the identity and behavior of a particular user if enough data are collected by the IoT even though they are not collected from camera-based devices. Also because most researches adopt classical ways to analyze the data of IoT, the new environment will inherit the open issues of these algorithms and hardware. Thus, the security problems become an extremely important issue after the data analysis systems find the hidden or sensitive information from the IoT. In summary, data mining provides the potential to add additional values to the raw data collected by IoT. Some of the additional values are so incredible that they are simply beyond our imagination in the past century.

The problems we have to deal with are how to protect this information and how to avoid undesired influences while we are enjoying these benefits.

V. Conclusions

In this paper, we review studies on applying data mining technologies to the IoT, which consist of clustering, classification, and frequent patterns mining technologies, from the perspective of infrastructures and from the perspective of services. The analysis and discussions on the scale of each mining technology and of the overall integrated system are also given. To make it easier for the audience of the paper to fully understand the changes brought about by the IoT, a discussion, which goes from the changes caused by the IoT in using mining technologies to the potentials to the open issues we are facing nowadays, is presented.

Since the development of IoT is still at the early stage of Nolan's stages of growth model,[9] it is often the case that the focus is first on the development of *efficient* preprocessing mechanisms to make the IoT system capable of handling big data and then on the development of *effective* mining technologies to find out the rules to describe the data of IoT. From the perspective of Nolan's stages of growth model, the issue to surface in the development of data mining for IoT is how to summarize and represent the mining results. To help the audience of the paper find *something* to proceed, the possible high impact research trends are given below:

- There is no doubt at all that big data is the future trend because most, if not all, of the devices will upload data to the internet once they are connected to the internet. IoT also inherits several signal processing problems from classical data mining researches, namely, data fusion, data abstraction, and data summarization. To handle the flood of big data, sampling technologies, compression technologies, incremental learning technologies, and filtering technologies all will become more and more important, especially when using data mining technologies to analyze the data of IoT.
- Ontology and semantic web technologies provide some possible solutions to these problems, but they cannot fully solve these problems because these two technologies are not mature enough. The fuzzy logic and multi-objective methods have the potential to enable a sensor to make decisions by itself. As a consequence, we do believe that these two technologies will be the trend of data mining technologies for the IoT.
- Needless to say, many app's available today make it possible to use handheld devices to control our household appliances. Two representative applications are app to control the multimedia devices and app to control the energy of a home. If the applications of app's to control devices are everywhere, then how to mine the data from these app's will become another interesting and potential research trend.

[9]Nolan's stages of growth model, which consists of the initiation, contagion, control, integration, data administration, and maturity stages, is a well-known theoretical model developed by Nolan in the 1970s [155] to describe the growth of information technology in a business or the like.

- When we argue whether the anthropogenic impact is the main reason for this change of environment and ecology or not, the data of climate and power consumption are collected by a large number of sensors located almost everywhere at the same time. The importance of analyzing these data can be easily justified, for data mining technologies can definitely be used to better achieve the ecological balance while at the same time enhancing the energy efficiency so as to reduce the carbon emissions.
- Many technical reports, research papers, and even speeches put the IoT, big data, smart grid, and cloud computing together partially because these issues and their products may significantly impact our life and partially because they are very closely connected and in fact inseparable. For the big data that comes from the IoT and smart grid, cloud computing provides a powerful resource that is able to achieve both the computing work and the incogitable data storage space needed. As a result, how to integrate them with IoT will also become a further research trend.
- Feedback mechanisms are usually used in the intelligent system or traditional data mining researches; however, they are not usually used in the household appliances or other IoT applications. To improve the accuracy, the interactive learning and feedback procedures will be needed, and they may become the mainstream research trends in using data mining technologies for the IoT.
- With respect to the computing and data flow characteristics of the IoT, the swarm intelligence algorithm fits perfectly to this environment because particles of particle swarm optimization [156] or ants of ant colony optimization [157] can be put on the devices to enable decentralized computing and cooperative learning with others when the work is complex and large.

References

[1] K. Ashton, "That 'Internet of Things' Thing," 2009, RFID Journal, Available at http://www.rfidjournal.com/article/print/4986.

[2] Auto-ID Labs, Massachusetts Institute of Technology, 2012, available at http://www.autoidlabs.org/.

[3] L. Atzori, A. Iera, and G. Morabito, "The internet of things: A survey," *Computer Networks*, vol. 54, no. 15, pp. 2787–2805, 2010.

[4] D. Miorandi, S. Sicari, F. De Pellegrini, and I. Chlamtac, "Internet of things: Vision, applications and research challenges," *Ad Hoc Networks*, vol. 10, no. 7, pp. 1497–1516, 2012.

[5] M&M Research Group, "Internet of Things (IoT) & M2M communication market - advanced technologies, future cities & adoption trends, roadmaps & worldwide forecasts 2012-2017," Electronics.ca Publications, Tech. Rep., 2012.

[6] D. Bandyopadhyay and J. Sen, "Internet of things: Applications and challenges in technology and standardization," *Wireless Personal Communications*, vol. 58, no. 1, pp. 49–69, 2011.

[7] M. C. Domingo, "An overview of the internet of things for people with disabilities," *Journal of Network and Computer Applications*, vol. 35, no. 2, pp. 584–596, 2012.

[8] M. Palattella, N. Accettura, X. Vilajosana, T. Watteyne, L. Grieco, G. Boggia, and M. Dohler, "Standardized protocol stack for the internet of (important) things," *IEEE Commun. Surveys Tutorials, In Press*, 2013.

[9] R. Kulkarni, A. Forster, and G. Venayagamoorthy, "Computational intelligence in wireless sensor networks: A survey," *IEEE Commun. Surveys Tutorials*, vol. 13, no. 1, pp. 68–96, 2011.

[10] T. Sánchez López, D. C. Ranasinghe, M. Harrison, and D. Mcfarlane, "Adding sense to the internet of things," *Personal and Ubiquitous Computing*, vol. 16, no. 3, pp. 291–308, 2012.

[11] F. Siegemund, "A Context-Aware communication platform for smart objects," in *Proc. International Conference on Pervasive Computing*, 2004, pp. 69–86.

[12] G. Kortuem, F. Kawsar, V. Sundramoorthy, and D. Fitton, "Smart objects as building blocks for the internet of things," *IEEE Internet Computing*, vol. 14, no. 1, pp. 44–51, 2010.

[13] T. S. López, D. C. Ranasinghe, B. Patkai, and D. C. McFarlane, "Taxonomy, technology and applications of smart objects," *Information Systems Frontiers*, vol. 13, no. 2, pp. 281–300, 2011.

[14] V. Cantoni, L. Lombardi, and P. Lombardi, "Challenges for data mining in distributed sensor networks," in *Proc. International Conference on Pattern Recognition*, vol. 1, 2006, pp. 1000–1007.

[15] T. Keller, "Mining the internet of things: Detection of false-positive RFID tag reads using low-level reader data," Ph.D. dissertation, The University of St. Gallen, Germany, 2011.

[16] E. Masciari, "A framework for outlier mining in RFID data," in *Proc. International Database Engineering and Applications Symposium*, 2007, pp. 263–267.

[17] S. Bin, L. Yuan, and W. Xiaoyi, "Research on data mining models for the internet of things," in *Proc. International Conference on Image Analysis and Signal Processing*, 2010, pp. 127–132.

[18] N. Ali and M. Abu-Elkheir, "Data management for the internet of things: Green directions," in *Proc. IEEE Globecom Workshops*, 2012, pp. 386–390.

[19] P. Russom, "Big data analytics," TDWI, Tech. Rep., 2011.

[20] S. Berkovich and D. Liao, "On clusterization of "big data" streams," in *Proc. International Conference on Computing for Geospatial Research and Applications*, 2012, pp. 3:1–3:1.

[21] R. G. Baraniuk, "More is less: Signal processing and the data deluge," *Science*, vol. 331, no. 6018, pp. 717–719, 2011.

[22] D. Reed, D. Gannon, and J. Larus, "Imagining the future: Thoughts on computing," *Computer*, vol. 45, no. 1, pp. 25–30, 2012.

[23] M. Hilbert and P. López, "The world's technological capacity to store, communicate, and compute information," *Science*, vol. 332, no. 6025, pp. 60–65, 2011.

[24] J. Gantz and D. Reinsel, "Extracting value from chaos," 2011, IDC IVEW, Available at http://www.itu.dk/people/rkva/2011-Fall-SMA/readings/ExtractingValuefromChaos.pdf.

[25] S. Madden, "From databases to big data," *IEEE Internet Computing*, vol. 16, no. 3, pp. 4–6, 2012.

[26] Y. K. Chen, "Challenges and opportunities of internet of things," in *Proc. Asia and South Pacific Design Automation Conference*, 2012, pp. 383–388.

[27] R. Xu and D. C. Wunsch-II, *Clustering*. Wiley, John & Sons, Inc., 2008.

[28] R. T. Ng and J. Han, "Clarans: A method for clustering objects for spatial data mining," *IEEE Trans. Knowl. Data Eng.*, vol. 14, no. 5, pp. 1003–1016, 2002.

[29] T. Zhang, R. Ramakrishnan, and M. Livny, "BIRCH: an efficient data clustering method for very large databases," in *Proc. ACM SIGMOD international conference on Management of data*, 1996, pp. 103–114.

[30] S. Guha, A. Meyerson, N. Mishra, and R. Motwani, "Clustering data streams: Theory and practice," *IEEE Trans. Knowl. Data Eng.*, vol. 15, no. 3, pp. 515–528, 2003.

[31] K. M. Hammouda and M. S. Kamel, "Efficient phrase-based document indexing for web document clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 16, no. 10, pp. 1279–1296, 2004.

[32] C. Ding and X. He, "K-means clustering via principal component analysis," in *Proc. International Conference on Machine Learning*, vol. 69, 2004, pp. 225–232.

[33] M. C. Chiang, C. W. Tsai, and C. S. Yang, "A time-efficient pattern reduction algorithm for k-means clustering," *Information Sciences*, vol. 181, no. 4, pp. 716–731, 2011.

[34] C. W. Tsai, S. P. Tseng, C. S. Yang, and M. C. Chiang, "Preaco: A fast ant colony optimization for codebook generation," *Applied Soft Computing*, vol. 13, no. 6, pp. 3008–3020, 2013.

[35] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of things (IoT): A vision, architectural elements, and future directions," *Future Generation Computer Systems*, vol. 29, no. 7, pp. 1645–1660, 2013.

[36] R. Roman, J. Zhou, and J. Lopez, "On the features and challenges of security and privacy in distributed internet of things," *Computer Networks, In Press*, 2013.

[37] C. C. Aggarwal, N. Ashish, and A. Sheth, "The internet of things: A survey from the data-centric perspective," in *Managing and Mining Sensor Data*, C. C. Aggarwal, Ed. Springer US, 2013, pp. 383–428.

[38] V. Mayer-Schonberger and K. Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Boston: Houghton Mifflin Harcourt, 2013.

[39] Libelium Comunicaciones Distribuidas S.L., "50 sensor applications for a smarter world," 2012, available at http://www.libelium.com/top_50_iot_sensor_applications_ranking/.

[40] R. Macmanus, "Top 10 internet of things products of 2009," 2009, available at http://www.readwriteweb.com/archives/top_10_internet_of_things_products_of_2009.php.

[41] U. M. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery in databases," *AI Magazine*, vol. 17, no. 3, pp. 37–54, 1996.

[42] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: A review," *ACM Computing Surveys*, vol. 31, no. 3, pp. 264–323, 1999.

[43] R. Xu and D. C. Wunsch-II, "Survey of clustering algorithms," *IEEE Trans. Neural Netw.*, vol. 16, no. 3, pp. 645–678, 2005.

[44] J. Han, *Data Mining: Concepts and Techniques*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2005.

[45] B. Liu, *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data*. Berlin, Heidelberg: Springer-Verlag, 2007.

[46] X. Wu, V. Kumar, J. Ross Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, P. S. Yu, Z. H. Zhou, M. Steinbach, D. J. Hand, and D. Steinberg, "Top 10 algorithms in data mining," *Knowledge and Information Systems*, vol. 14, no. 1, pp. 1–37, 2007.

[47] J. B. McQueen, "Some methods of classification and analysis of multivariate observations," in *Proc. Berkeley Symposium on Mathematical Statistics and Probability*, 1967, pp. 281–297.

[48] A. K. Jain, "Data clustering: 50 years beyond k-means," *Pattern Recognition Letters*, vol. 31, no. 8, pp. 651–666, 2010.

[49] K. Krishna and M. N. Murty, "Genetic k-means algorithm," *IEEE Trans. Syst. Man Cybern. B Cybern.*, vol. 29, no. 3, pp. 433–439, 1999.

[50] C. Blum and A. Roli, "Metaheuristics in combinatorial optimization: Overview and conceptual comparison," *ACM Computing Surveys*, vol. 35, no. 3, pp. 268–308, 2003.

[51] O. Younis and S. Fahmy, "Distributed clustering in ad-hoc sensor networks: A hybrid, energy-efficient approach," in *Proc. IEEE INFOCOM*, 2004.

[52] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey," *Computer Networks*, vol. 38, no. 4, pp. 393–422, 2002.

[53] A. Chamam and S. Pierre, "A distributed energy-efficient clustering protocol for wireless sensor networks," *Computers & Electrical Engineering*, vol. 36, no. 2, pp. 303 – 312, 2010.

[54] D. Uckelmann, M. Harrison, and F. Michahelles, "An architectural approach towards the future internet of things," in *Architecting the Internet of Things*, 2011, pp. 1–24.

[55] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in *Proc. Hawaii International Conference on System Sciences*, 2000.

[56] A. A. Abbasi and M. Younis, "A survey on clustering algorithms for wireless sensor networks," *Computer Communications*, vol. 30, no. 14-15, pp. 2826–2841, 2007.

[57] J. Zheng and A. Jamalipour, *Wireless Sensor Networks: A Networking Perspective*. John Wiley & Sons, 2009.

[58] S. Ghiasi, A. Srivastava, X. Yang, and Sarrafzadeh, "Optimal energy aware clustering in sensor networks," *Sensors*, vol. 2, no. 7, pp. 258–269, 2002.

[59] W. Choi, P. Shah, and S. K. Das, "A framework for energy-saving data gathering using two-phase clustering in wireless sensor networks," in *Proc. International Conference on Mobile and Ubiquitous Systems*, 2004, pp. 203–212.

[60] V. Kardeby, "Automatic sensor clustering : connectivity for the internet of things," Licentiate thesis, *Mid Sweden University, Department of Information Technology and Media*, 2011.

[61] T. S. López, A. Brintrup, M.-A. Isenberg, and J. Mansfeld, "Resource management in the Internet of Things: Clustering, synchronisation and software agents," in *Architecting the Internet of Things*, 2011, pp. 159–193.

[62] P. Rashidi, D. J. Cook, L. B. Holder, and M. Schmitter-Edgecombe, "Discovering activities to recognize and track in a smart environment," *IEEE Trans. Knowl. Data Eng.*, vol. 23, no. 4, pp. 527–539, 2011.

[63] N. Anastasiou, T.-C. Horng, and W. Knottenbelt, "Deriving generalised stochastic petri net performance models from high-precision location tracking data," in *Proc. International ICST Conference on Performance Evaluation Methodologies and Tools*, 2011, pp. 91–100.

[64] E. Schuster, S. Kumar, S. Sarma, J. Willers, and G. Milliken, "Infrastructure for data-driven agriculture: identifying management zones for cotton using statistical modeling and machine learning techniques," in *Proc. International Conference Expo on Emerging Technologies for a Smarter World*, 2011, pp. 1–6.

[65] Z. Min, W. Bei, G. Chunyuan, and S. Zhao qian, "Application study of precision agriculture based on ontology in the internet of things environment," in *Applied Informatics and Communication*, J. Zhang, Ed. Berlin, Heidelberg: Springer-Verlag, 2011, vol. 227, pp. 374–380.

[66] S. White and P. Smyth, "A spectral clustering approach to finding communities in graph," in *Proc. SIAM International Conference on Data Mining*, 2005.

[67] L. Wang, "Using the relationship of shared neighbors to find hierarchical overlapping communities for effective connectivity in IoT," in *Proc. International Conference on Pervasive Computing and Applications*, 2011, pp. 400–406.

[68] J. An, X. Gui, W. Zhang, and J. Jiang, "Nodes social relations cognition for mobility-aware in the internet of things," in *Proc. International Conference on Internet of Things and Cyber, Physical and Social Computing*, 2011, pp. 687–691.

[69] S. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE Trans. Syst. Man Cybern.*, vol. 21, no. 3, pp. 660–674, 1991.

[70] M. Friedl and C. Brodley, "Decision tree classification of land cover from remotely sensed data," *Remote Sensing of Environment*, vol. 61, no. 3, pp. 399 – 409, 1997.

[71] A. McCallum and K. Nigam, "A comparison of event models for naive bayes text classification," in *Proc. National Conference on Artificial Intelligence*, 1998, pp. 41–48.

[72] P. Langley, W. Iba, and K. Thompson, "An analysis of bayesian classifiers," in *Proc. National Conference on Artificial Intelligence*, 1992, pp. 223–228.

[73] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proc. annual workshop on Computational learning theory*, 1992, pp. 144–152.

[74] Y. Lee, H. Kim, B.-h. Roh, S. Yoo, and Y. Oh, "Tree-based classification algorithm for heterogeneous unique item id schemes," in *Proc. Workshops on Embedded and Ubiquitous Computing*, vol. 3823, 2005, pp. 1078–1087.

[75] E. Horvitz, J. Apacible, R. Sarin, and L. Liao, "Prediction, expectation, and surprise: Methods, designs, and study of a deployed traffic forecasting service," in *Proc. Conference on Uncertainty in Artificial Intelligence*, 2005, pp. 244–257.

[76] C. Thompson, J. White, B. Dougherty, A. Albright, and D. C. Schmidt, "Using smartphones to detect car accidents and provide situational awareness to emergency responders," in *Proc. International Conference on Mobile Wireless Middleware, Operating Systems, and Applications*, vol. 48, 2010, pp. 29–42.

[77] K. Perera and D. Dias, "An intelligent driver guidance tool using location based services," in *Proc. International Conference on Spatial Data Mining and Geographical Knowledge Services*, 2011, pp. 246–251.

[78] R. S. Sudhaakar, A. Sanzgiri, M. Demirbas, and C. Qiao, "A plant-and-play wireless sensor network system for gate monitoring," in *Proc. International Symposium on a World of Wireless, Mobile and Multimedia Networks*, 2009, pp. 1–9.

[79] A. Fleury, M. Vacher, and N. Noury, "SVM-based multimodal classification of activities of daily living in health smart homes: sensors, algorithms, and first experimental results," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 2, pp. 274–283, 2010.

[80] P. Barralon, N. Vuillerme, and N. Noury, "Walk detection with a kinematic sensor: frequency and wavelet comparison," in *Proc. Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 1, 2006, pp. 1711–1714.

[81] Q. C. Nguyen, D. Shin, D. Shin, and J. Kim, "Real-time human tracker based on location and motion recognition of user for smart home," in *Proc. International Conference on Multimedia and Ubiquitous Engineering*, 2009, pp. 243–250.

[82] C. L. Liu, C. H. Lee, and P. M. Lin, "A fall detection system using k-nearest neighbor classifier," *Expert Systems with Applications*, vol. 37, no. 10, pp. 7174–7181, 2010.

[83] A. Parnandi, K. Le, P. Vaghela, A. Kolli, K. Dantu, S. Poduri, and G. S. Sukhatme, "Coarse in-building localization with smartphones," in *Proc. International ICST Conference Mobile Computing, Applications, and Services*, 2009, pp. 343–354.

[84] M. Viswanathan, T. K. Whangbo, K. J. Lee, and Y. K. Yang, "A distributed data mining system for a novel ubiquitous healthcare framework," in *Proc. International Conference on Computational Science*, 2007, pp. 701–708.

[85] J. Hong, E.-H. Suh, J. Kim, and S. Kim, "Context-aware system for proactive personalized service based on context history," *Expert Systems with Applications*, vol. 36, no. 4, pp. 7448–7457, 2009.

[86] E. Leon, G. Clarke, V. Callaghan, and F. Doctor, "Affect-aware behaviour modelling and control inside an intelligent environment," *Pervasive and Mobile Computing*, vol. 6, no. 5, pp. 559–574, 2010.

[87] S. Chi and C. Caldas, "Automated object identification using optical video cameras on construction sites," *Computer-Aided Civil and Infrastructure Engineering*, vol. 26, no. 5, pp. 368–380, 2011.

[88] H. K. Ekenel, J. Stallkamp, and R. Stiefelhagen, "A video-based door monitoring system using local appearance-based face models," *Computer Vision and Image Understanding*, vol. 114, no. 5, pp. 596–608, 2010.

[89] N. Yang, X. Zhao, and H. Zhang, "A non-contact health monitoring model based on the internet of things," in *Proc. International Conference on Natural Computation*, 2012, pp. 506–510.

[90] D. Istrate, M. Vacher, E. Castelli, and C.-P. Nguyen, "Sound processing for health smart home," in *Proc. International Conference on Smart homes and health Informatics*, 2004, pp. 41–48.

[91] H. Zhou, K. M. Hou, and D.-C. Zuo, "Real-time automatic ecg diagnosis method dedicated to pervasive cardiac care," *Wireless Sensor Network*, vol. 1, no. 4, pp. 276–283, 2009.

[92] L. Pang, I. Tchoudovski, M. Braecklein, K. Egorouchkina, K. W., and A. Bolz, "Real time heart ischemia detection in the smart home care system," in *Proc. Annual Conference on Engineering in Medicine and Biology*, 2005, pp. 3703–3706.

[93] J. Han, H. Cheng, D. Xin, and X. Yan, "Frequent pattern mining: current status and future directions," *Data Mining and Knowledge Discovery*, vol. 15, no. 1, pp. 55–86, 2007.

[94] Q. Zhao and S. S. Bhowmick, "Sequential pattern mining: A survey," Technical Report CAIS Nayang Technological University Singapore, Tech. Rep., 2003.

[95] R. Agrawal, T. Imieliński, and A. Swami, "Mining association rules between sets of items in large databases," in *Proc. ACM SIGMOD International Conference on Management of Data*, vol. 22, no. 2, 1993, pp. 207–216.

[96] M. Galushka, D. Patterson, and N. Rooney, "Temporal data mining for smart homes," in *Designing Smart Homes*, J. C. Augusto and C. D. Nugent, Eds. Berlin, Heidelberg: Springer-Verlag, 2006, pp. 85–108.

[97] L. Chen and G. Ren, "The research of data mining technology of privacy preserving in sharing platform of internet of things," in *Internet of Things*, Y. Wang and X. Zhang, Eds. Springer Berlin Heidelberg, 2012, vol. 312, pp. 481–485.

[98] F. E. Bekri and A. Govardhan, "Association of data mining and healthcare domain: Issues and current state of the art," *Global Journals Inc.*, vol. 11, no. 21, pp. 1–8, 2011.

[99] F. J. Romano, "An investigation into printing industry trends," Rochester, NY: Rochester Institute of Technology, Printing Industry Center, Tech. Rep. PICRM-2004-05, 2004.

[100] X. Li, J. Liu, Q. Z. Sheng, and W. Zhong, "Time to live: Temporal management of large-scale RFID applications," The University of Queensland, Tech. Rep. SSE-2008-02, 2008.

[101] X. Li, J. Liu, Q. Sheng, S. Zeadally, and W. Zhong, "TMS-RFID: Temporal management of large-scale RFID applications," *Information Systems Frontiers*, vol. 13, pp. 481–500, 2011.

[102] K. Koperski and J. Han, "Discovery of spatial association rules in geographic information databases," in *Proc. International Symposium Advances in Spatial Databases*, vol. 951, 1995, pp. 47–66.

[103] J. S. Yoo and S. Shekhar, "A joinless approach for mining spatial colocation patterns," *IEEE Trans. Knowl. Data Eng.*, vol. 18, no. 10, pp. 1323–1337, 2006.

[104] S. K. Kim, J. H. Lee, K. H. Ryu, and U. Kim, "A framework of spatial co-location pattern mining for ubiquitous GIS," *Multimedia Tools and Applications*, pp. 1–20, 2012.

[105] W. R. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Trans. Wireless Commun.*, vol. 1, no. 4, pp. 660–670, 2002.

[106] O. Younis and S. Fahmy, "HEED: A hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks," *IEEE Trans. Mobile Computing*, vol. 3, no. 4, pp. 366–379, 2004.

[107] S. P. Rao and D. J. Cook, "Predicting inhabitant action using action and task models with application to smart homes," *International Journal on Artificial Intelligence Tools*, vol. 13, no. 1, pp. 81–99, 2004.

[108] H. Chen, B. C. Cheng, C. C. Cheng, and L. K. Tsai, "Smart home sensor networks pose goal-driven solutions to wireless vacuum systems," in *Proc. International Conference on Hybrid Information Technology*, vol. 2, 2006, pp. 364–373.

[109] A. Fleury, N. Noury, and M. Vacher, "Supervised classification of activities of daily living in health smart homes using SVM," in *Proc. Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2009, pp. 6099–6102.

[110] ——, "A wavelet-based pattern recognition algorithm to classify postural transitions in humans," in *Proc. European Signal Processing Conference*, 2009, pp. 2047–2051.

[111] S. Somasundaram, P. Khandavilli, and S. Sampalli, "An intelligent RFID system for consumer businesses," in *Proc. International Conference on Green Computing and Communications and International Conference on Cyber, Physical and Social Computing*, 2010, pp. 539–545.

[112] V. Jakkula, D. Cook, and A. Crandall, "Temporal pattern discovery for anomaly detection in a smart home," in *Proc. IET International Conference on Intelligent Environments*, 2007, pp. 339–345.

[113] G. Papamatthaiakis, G. C. Polyzos, and G. Xylomenos, "Monitoring and modeling simple everyday activities of the elderly at home," in *Proc. IEEE Conference on Consumer Communications and Networking Conference*, 2010, pp. 617–621.

[114] S. Lühr, G. West, and S. Venkatesh, "Recognition of emergent human behaviour in a smart home: A data mining approach," *Pervasive and Mobile Computing*, vol. 3, no. 2, pp. 95–116, 2007.

[115] K. J. Kang, B. Ka, and S. J. Kim, "A service scenario generation scheme based on association rule mining for elderly surveillance system in a smart home environment," *Engineering Applications of Artificial Intelligence*, vol. 25, no. 7, pp. 1355–1364, 2012.

[116] N. Suryadevara, A. Gaddam, S. Mukhopadhyay, and R. Rayudu, "Wellness determination of inhabitant based on daily activity behaviour in real-time monitoring using sensor networks," in *Proc. International Conference on Sensing Technology*, 2011, pp. 474–481.

[117] D. J. Cook, G. M. Youngblood, E. O. Heierman III, K. Gopalratnam, S. Rao, A. Litvin, and F. Khawaja, "MavHome: An agent-based smart home," in *Proc. IEEE International Conference on Pervasive Computing and Communications*, 2003, pp. 521–524.

[118] P. Rashidi and D. J. Cook, "An adaptive sensor mining framework for pervasive computing applications," in *Proc. KDD Workshop on Knowledge Discovery from Sensor Data*, 2008, pp. 154–174.

[119] E. Nazerfard, P. Rashidi, and D. Cook, "Discovering temporal features and relations of activity patterns," in *Proc. IEEE International Conference on Data Mining Workshops*, 2010, pp. 1069–1075.

[120] E. Nazerfard, P. Rashidi, and D. J. Cook, "Using association rule mining to discover temporal relations of daily activities," in *Proc. International Conference on Toward Useful Services for Elderly and People with Disabilities: Smart Homes and Health Telematics*, 2011, pp. 49–56.

[121] J. Choi, D. Shin, and D. Shin, "Ubiquitous intelligent sensing system for a smart home," in *Proc. International Conference on Structural, Syntactic, and Statistical Pattern Recognition*, 2006, pp. 322–330.

[122] H. S. Kim and J. H. Son, "User-pattern analysis framework to predict future service in intelligent home network," in *Proc. IEEE International Conference on Intelligent Computing and Intelligent Systems*, 2009, pp. 818–822.

[123] C. Schwenke, V. Vasyutynskyy, and K. Kabitzsch, "Simulation and analysis of buying behavior in supermarkets," in *Proc. IEEE International Conference on Emerging Technologies and Factory Automation*, 2010, pp. 1–4.

[124] J. F. Allen and G. Ferguson, "Actions and events in interval temporal logic," University of Rochester, Tech. Rep., 1994.

[125] S. Katz, "Assessing self-maintenance: activities of daily living, mobility, and instrumental activities of daily living," *Journal of the American Geriatrics Society*, vol. 31, no. 12, pp. 721–727, 1983.

[126] Activities of daily living (ADL), 2012, available at http://son.uth.tmc.edu/coa/FDGN_1/RESOURCES/ADLandIADL.pdf.

[127] C. W. Tsai and J. Rodrigues, "Metaheuristic scheduling for cloud: A survey," *IEEE Syst. J.*, 2013.

[128] I. Foster, Y. Zhao, I. Raicu, and S. Lu, "Cloud computing and grid computing 360-Degree compared," in *IEEE Grid Computing Environments Workshop*, no. 5, 2008, pp. 1–10.

[129] R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg, and I. Brandic, "Cloud computing and emerging it platforms: Vision, hype, and reality for delivering computing as the 5th utility," *Future Generation Computer Systems*, vol. 25, no. 6, pp. 599–616, 2009.

[130] R. Hassan and G. Radman, "Survey on smart grid," in *Proc. IEEE SoutheastCon*, march 2010, pp. 210–213.

[131] J. Gao, Y. Xiao, J. Liu, W. Liang, and C. P. Chen, "A survey of communication/networking in smart grids," *Future Generation Computer Systems*, vol. 28, no. 2, pp. 391–404, 2012.

[132] National Institute of Standards and Technology, "NIST Framework and Roadmap for Smart Grid Interoperability Standards, Release 2.0," 2012, available at http://www.nist.gov/smartgrid/framework-022812.cfm.

[133] T. Sousa, H. Morais, Z. Vale, P. Faria, and J. Soares, "Intelligent energy resource management considering vehicle-to-grid: A simulated annealing approach," *IEEE Trans. Smart Grid*, vol. 3, no. 1, pp. 535–542, 2012.

[134] W. Su and M.-Y. Chow, "Performance evaluation of an EDA-based large-scale plug-in hybrid electric vehicle charging algorithm," *IEEE Tran. Smart Grid*, vol. 3, no. 1, pp. 308–315, 2012.

[135] Smarter Planet, 2012, available at http://www.ibm.com/smarterplanet/th/en/overview/ideas/index.html.

[136] H. Sundmaeker, P. Guillemin, P. Friess, and S. Woelfflé, *Vision and Challenges for Realising the Internet of Things*. Cluster of European Research Projects on the Internet of Things, European Commision, 2010.

[137] INFSO D.4 Networked Enterprise & RFID INFSO G.2 Micro & Nanosystems, in: Co-operation with the Working Group RFID of the ETP EPOSS, *Internet of Things in 2020: Roadmap for the future*. European Technology Platform on Smart Systems Integration (EPoSS), 2008.

[138] D. Uckelmann, M. Harrison, and F. Michahelles, Eds., *Architecting the Internet of Things*. Berlin, Heidelberg: Springer-Verlag, 2011.

[139] Eye on Earth, 2012, available at http://watch.eyeonearth.org.

[140] European Technology Platform on Smart Systems Integration, 2012, available at http://www.smart-systems-integration.org/public/documents/publications.

[141] Internet of Things System Technology Laboratory, Shanghai Institute of Microsystem And Information Technology, 2012, available at http://www.sim.cas.cn/zzjg/kybm/wlwjssys/.

[142] A. de Saint-Exupery, "Internet of things strategic research roadmap," European Research Cluster on the Internet of Things, Tech. Rep., 2009.

[143] R. A. Baeza-Yates and B. A. Ribeiro-Neto, *Modern Information Retrieval*. ACM Press / Addison-Wesley, 1999.

[144] O. J. Reichman, M. B. Jones, and M. P. Schildhauer, "Challenges and opportunities of open data in ecology," *Science*, vol. 331, pp. 703–705, 2011.

[145] Special Online Collection: Dealing with Data, 2011, available at http://www.sciencemag.org/site/special/data/.

[146] E. Portmann, A. Andrushevich, R. Kistler, and A. Klapproth, "Prometheus – fuzzy information retrieval for semantic homes and environments," in *Proc. International Conference on Human System Interaction*, 2010, pp. 757–762.

[147] R. Kane and R. Kane, *Assessment of older people: Self maintaining and instrumental activities of dailyliving*. Toronto, USA: Lexington Books, 1981.

[148] Instrumental activities of dailyliving (IADL), 2012, available at http://www.abramsoncenter.org/PRI/documents/IADL.pdf.

[149] T. Segaran, *Programming Collective Intelligence: Building Smart Web 2.0 Applications*. Sebastopol, CA: O'Reilly, 2007.

[150] C. Efthymiou and G. Kalogridis, "Smart grid privacy via anonymization of smart metering data," in *Proc. IEEE International Conference on Smart Grid Communications*, 2010, pp. 238–243.

[151] J. Deng, R. Han, and S. Mishra, "Intrusion tolerance and anti-traffic analysis strategies for wireless sensor networks," in *Proc. International Conference on Dependable Systems and Networks*, 2004, pp. 637–646.

[152] P. Kamat, W. Xu, W. Trappe, and Y. Zhang, "Temporal privacy in wireless sensor networks: Theory and practice," *ACM Trans. Sensor Networks*, vol. 5, no. 4, pp. 28:1–28:24, 2009.

[153] A. Perrig, R. Szewczyk, J. D. Tygar, V. Wen, and D. E. Culler, "Spins: security protocols for sensor networks," *Wireless Networks*, vol. 8, no. 5, pp. 521–534, 2002.

[154] R. H. Weber, "Internet of things new security and privacy challenges," *Computer Law & Security Review*, vol. 26, no. 1, pp. 23–30, 2010.

[155] R. L. Nolan, "Managing the crises in data processing," *Harvard Business Review*, vol. 57, no. 1, pp. 115–126, 1979.

[156] A. P. Engelbrecht, *Fundamentals of Computational Swarm Intelligence*. John Wiley & Sons, 2006.

[157] M. Dorigo and T. Stützle, *Ant Colony Optimization (Bradford Books)*. The MIT Press, 2004.

Ming-Chao Chiang received the B.S. degree in Management Science from National Chiao Tung University, Hsinchu, Taiwan, ROC in 1978 and the M.S., M.Phil., and Ph.D. degrees in Computer Science from Columbia University, New York, NY, USA in 1991, 1998, and 1998, respectively. He has over 12 years of experience in the software industry encompassing a wide variety of roles and responsibilities in both large and start-up companies before joining the faculty of the Department of Computer Science and Engineering, National Sun Yat-sen University, Kaohsiung, Taiwan, ROC in 2003, where he is currently an Associate Professor. His current research interests include image processing, evolutionary computation, and system software.

Chun-Wei Tsai received the M.S. degree in Management Information System from National Pingtung University of Science and Technology, Pingtung, Taiwan in 2002, and the Ph.D. degree in Computer Science from National Sun Yat-sen University, Kaohsiung, Taiwan in 2009. He was also a postdoctoral fellow with the Department of Electrical Engineering, National Cheng Kung University, Tainan, Taiwan before joining the faculty of Applied Geoinformatics and then the faculty of Applied Informatics and Multimedia, Chia Nan University of Pharmacy & Science, Tainan, Taiwan in 2010 and 2012, respectively, where he is currently an Assistant Professor. His research interests include data mining, combinatorial optimization, internet technology, and metaheuristic algorithms.

Laurence T. Yang is a professor at School of Computer Science and Technology, Huazhong University of Science and Technology, China and at Department of Computer Science, St. Francis Xavier University, Canada. He received the B.E. degree in computer science and technology from Tsinghua University, Beijing, China, and the Ph.D. degree in computer science from the University of Victoria, BC, Canada. His current research interests include parallel and distributed computing, embedded and ubiquitous/pervasive computing. His research is supported by the National Sciences and Engineering Research Council, Canada and Canada Foundation for Innovation.

Chin-Feng Lai is an assistant professor at Department of Computer Science and Information Engineering, National Chung Cheng University since 2013. He received the Ph.D. degree in department of engineering science from the National Cheng Kung University, Taiwan, in 2008. His research interests include multimedia communications, sensor-based healthcare and embedded systems. After receiving the Ph.D degree, he has authored/co-authored over 100 refereed papers in journals, conference, and workshop proceedings about his research areas within four years. Now he is making efforts to publish his latest research in the IEEE Trans. multimedia and the IEEE Trans. circuit and system on video Technology. He is also a member of the IEEE as well as IEEE Circuits and Systems Society and IEEE Communication Society.