IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received June 10, 2019, accepted July 3, 2019, date of publication July 17, 2019, date of current version August 2, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2929550

ANN-Based Outlier Detection for Wireless Sensor Networks in Smart Buildings

KAI ZHANG[1], KE YANG[2], SHAOYI LI[1], DISHAN JING[2], AND HAI-BAO CHEN[2]

[1]Institute of Aeronautics and Astronautics, Northwestern Polytechnic University, Xi'an 710129, China
[2]Department of Micro/Nano-electronics, Shanghai Jiao Tong University, Shanghai 200240, China

Corresponding author: Hai-Bao Chen (haibaochen@sjtu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61703337, and in part by the Shanghai Academy of Spaceflight Technology Innovation Foundation under Grant SAST2017-082.

ABSTRACT Wireless sensor network (WSN) is an emerging technology with a wide range of potential applications in smart buildings. The measuring process by using WSNs in the actual environment always introduces noise, errors, accidents, and other potential outliers to the data collected by the sensors. It is crucial to establish an effective approach for outlier detection and recovery in the real applications of WSNs. In this paper, we propose an outlier detection and recovery approach using artificial neural network (ANN), which can be used to determine whether the temperature values measured by the sensors in WSNs are outliers. The experimental results in real building show that the proposed ANN-based models can provide a reasonably good prediction of the temperature and high accuracy in buildings compared with the hidden Markov model (HMM)-based approach, which can potentially be used for outlier detecting and thermal controlling in the Internet of Things (IoT) applications.

INDEX TERMS Outlier detection, wireless sensor networks, thermal controlling, artificial neural network.

I. INTRODUCTION

The Internet of Things (IoT) is defined as a system which consists of connected application-specific embedded devices including actuators and sensors, software and network connectivity that enable the new service for meeting people in the community. Sensors as a rapidly growing component of IoT have gained much attention recently due to their incredible potentials to enable emerging applications of high societal and economic impacts [1]–[4]. Smart building has been wildly utilized for detecting buildings' physical conditions (physiological attributes, physical activities, etc.) to improve automatic control for their users. Despite promising future for smart buildings, the data connection related utility in the system is currently limited in part because of the challenges associated with the battery lifetime leading to poor integration and user adherence, and the unreliability of the system.

Smart building is the systematic study on the applications in new technology in buildings, in order to improve occupant comfort and convenience in the building, and reduce energy consumption [5]–[7]. As one of the important research areas in smart buildings, the study on wireless sensor networks (WSN) has attracted more and more attention. Each sensor node in the WSN is usually constituted by a wireless radio transceiver, a processor, a battery and multiple sensors [8]–[10]. The sensor node should be able to sample the temperature value at a fixed period and transmit the data to the main processor for advanced processing through network communication [11]. The function of the WSN system varies widely, including environmental monitoring, package tracking, industrial safety and control, etc. For example, the real-time forest fire detection with wireless sensor networks has been studied in [12]. To satisfy the requirement of these applications, the whole system should be capable to work for a long lifetime with only some small coin cell batteries, since the users do not want to recharge the detached nodes frequently [13].

As the automation and sensing technology advances, the data collected by the WSN system is an important source for further analysis, and high precision and reliability is required for data processing [14], [15]. Because of the complexity of the environment, the quality of the data measured by sensors are always affected by surrounding noise, system error and environment incidents. During the work process of wireless sensor network, the measured data detected by sensor may not be accurate enough. There are several factors

The associate editor coordinating the review of this manuscript and approving it for publication was Malik Jahan Khan.

that affect the sensor measurement accuracy: 1) change of temperature detected by sensor is induced by some man-made factors; 2) noises caused by various accidental factors in sensor circuits, data storage and processing modules and data transmission circuits results in detection errors during data transmission or storage; 3) the sensor itself or the related circuit is damaged, resulting in the inability to measure the correct data. To improve credibility of sensor read values, reliable detection methods of unusual behavior or outlier/offset are required. Many existing works such as classification and clustering techniques have been proposed to detect outliers from observed data.

Outlier detection techniques provide a way to search for values that do not follow the normal pattern of temperature data in the network, guarantee the quality of measured data, and improve the robustness of the data analysis under the presence of noise and faulty sensors. Outlier detection methods can be divided into three categories [16]: Type 1 detectors determine outliers without prior knowledge of data; type 2 detectors work with knowledge of both normality and abnormality; type 3 detectors work with knowledge of normality and very little knowledge of abnormality. In this paper, we target on developing type 1 outlier detection technique based on artificial neural network (ANN). For the purpose to decrease the influence of the outliers, the topic of outlier detection in WSN has attracted more and more attention in recent years.

Outlier detection techniques for WSN are classified into nearest neighbor-based approach, clustering-based approach and classification-based approach [17]. Classification-based approach, which provides an exact set of outliers by building a classification model to classify, are the most widely used in the literature, such as support vector machine-based approach, Bayesian network-based approach, Hidden Markov Model (HMM) based approach [18] and neural network-based approach [19]. Neural network approach is an efficient approach for doing data analysis, which derives approximate functions or models based on a large amount of training data. In this paper, we propose an artificial neural network (ANN) based outlier detection method for detecting the outliers in the temperature values by the WSN system in smart buildings. For the proposed ANN-based outlier detection technique, we need to assume that the sensors, which constitute the WSN system are placed in a limited space randomly. This assumption indicates that there is a relationship between the measured temperature value of each sensor [20]. In this work, an ANN model is built to reveal this relationship and thus predict the accurate temperature value. By this technique the outliers can be identified by computing the posterior probability of the measured values in the WSN system. Also, we propose an optimized ANN model to improve the detection accuracy in the case of a big change of the environment temperature. Experiment and simulation results show that both the two ANN-based models are effective in predicting the temperature value in the WSN system in smart buildings, and the optimized ANN model can achieve better performance when the environment temperature changes significantly [21].

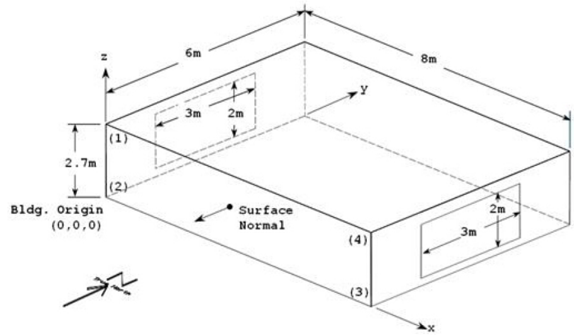


FIGURE 1. The building model.

The rest of the paper is organized as follows. In Section II, we describe the information-theoretical framework for the heat exchange problem. Section III presents the ANN-based forecast models. The efficacy of the proposed approach is demonstrated by experiment results in Section IV. The work is finally concluded in Section V.

II. THERMAL BEHAVIOR IN BUILDINGS

In this section, we will demonstrate the thermal behavior in buildings which will show the feasibility of the ANN-based outlier detection method for predicting the temperature profile in the WSN system in smart buildings.

A. BUILDING MODEL

The approach we proposed is able to predict the anomalous behaviors of specific circumstances of an room or office. Outdoor temperatures and indoor temperatures are taken as explanatory variables to predict results. In this model, we do not require sensors to be high-end or extremely accurate which is highly demanded in a number of previous models, because the ANN-based model we propose in this paper does not predict anomalous behaviors through precise detection of high-end sensors. For example, the occupancy of inhabitants can be predicted under statistic approach which is not rapidly changed due to subtle changes of indoor temperature variance or outdoor temperature variance.

The building behaviors are highly related to the infrastructure including internal materials, internal loads, space conditioning, location, simulation period, ground temperatures, walls, floors, roofs, etc. For the purpose of simulation, the size and shape of an office is demonstrated in Fig. 1 generated by EnergyPlus. Before we establish this model, we need to confirm some parameters of the office of a building. This model has a high demand for knowing the detail of a certain room or office, even the infrastructure of a building. We need the specific parameters of a room to establish a model and get access to a vast amount of data generated via EnergyPlus. We set up the parameters of an office in a building. In this model, the details of the building construction and operation are shown in Table.1. These details are listed in a fashion to make for easy entry into EnergyPlus.

TABLE 1. Surface constructions.

Material (listed from outside to inside)	Conductivity (W/m-K)	Thickness (m)
Wall		
WOOD SIDING-1	0.140	0.009
FIBERGLASS QUILT-1	0.040	0.066
PLASTERBOARD-1	0.160	0.012
Roof		
ROOF DECK	0.140	0.019
FIBERGLASS QUILT-2	0.040	0.066
PLASTERBOARD-2	0.160	0.010
Floor		
C5 CONCRETE	1.73	0.1015

U (W/m^2-K)	R (m^2-K/W)	Density (kg/m^3)	C_p (J/kg-K)
Wall			
15.556	0.064	530	900
0.606	1.650	12	840
13.333	0.075	950	840
Roof			
7.368	0.136	530	900
0.606	1.650	12	840
1.60	0.625	950	840
Floor			
17.04	0.059	2243	837

EnergyPlus is a whole building energy simulation program that utilized by researchers, architects and engineers to model both energy consumption, e.g., ventilation, lighting, heating, cooling, and plug and process loads, and water use in buildings. U.S. Department of Energy Building Technologies Office fund its development. EnergyPlus plays in the role as a modular and structured code which contains most popular features of DOE-2.1E and BLAST. It mainly is a functional engine with texts written in format of inputs and outputs. Via a heat balance engine, loads calculated are able to be modified by users in specific time step to simulate the building system. The EnergyPlus based building system simulation module which is with the variable time step is capable of calculating cooling and heating system and the response of electrical system. This system works more like an integrated simulation and it is playing an important role in calculations for plant sizing, system, occupant health calculations and occupant comfort while providing precise temperature prediction. Moreover, integrated simulation also offers user authority to evaluate moisture adsorption and desorption, controls of realistic system in building elements, air flow in an interzone, and radiant cooling and heating systems. The energy consumed in buildings is directly related to space heating, cooling and ventilation. It is important to control the heating, ventilation and air conditioning (HVAC) system in a more energy-efficient way. Neural network based accurate detection of unusual behavior in temperature sensors (outliers) can help reduce or prevent waste of energy consumption in a Heating, Ventilation and Air Conditioning (HVAC) system. Reliable outlier detection methods are required for improving the credibility of sensor values. Reliable detection means the chance of reporting correct outliers is maximized and the chance for reporting false alarms is minimized. It should be noted that an existing fault detector provided by building management system can report more than 10000 alarms per day asserting that the system is not running under correct status. Due to that these faults could be caused by temperature sensor outliers, it is important to improve the precision of automatic outlier detection in smart buildings.

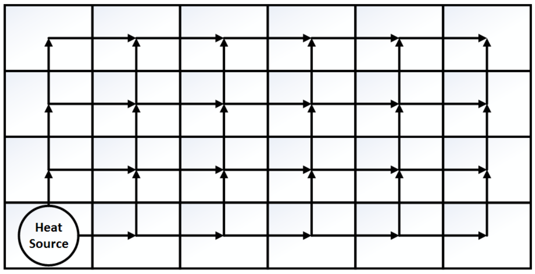


FIGURE 2. The heat exchange process in the room where there is a heat source in the lower left corner.

B. HEAT EXCHANGE PROCESS

In the applications of the WSN system including those relating to personal and industrial such as environmental and habitat monitoring, sensors and communication module can be equipped in the working space in the preparatory step. Temperature sensors can monitor the surrounding environment, collect the temperature data and transmit it through wireless communication. In order to improve the accuracy of the data collected by WSN, we need to develop effective approaches to detect the outliers and replace them by an accessible estimation. Before proposing our ANN-based outlier detection method, we need to introduce the theoretical framework of the heat exchange problem in buildings.

In order to better describe the heat exchange process, the object we study is limited in a room and only the horizontal heat transfer is taken into consideration. In the model, we separate the space into some rectangle zones and we assume that the heat exchange can only occur between the neighboring zones. If the room contains one or multiple heat sources, the thermal state will be influenced because of the heat production. Figure 2 shows the heat transfer process in the room where there is a heat source in the lower left corner.

In our analysis, we divide the heat exchange problem into the two cases: the one and the multiple heat source problems. If many heat sources are placed in the room but one of them produces much more heat than the others, it will eliminate the other heat sources' effects on the environment in the room. For this case, it can be simplified to the case of the one heat source problem.

C. ONE HEAT SOURCE PROBLEM

It can be seen from Figure 2 that the zones that have the same distance to the heat source share the identical heat exchange quantity when only one heat source is placed in the room. The influence of the walls is ignored due to their high heat capacity. Based on the symmetry of space, the thermal state

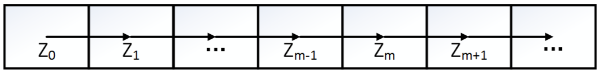


FIGURE 3. The heat exchange process between zones in one direction.

in each zone is determined by its distance to the heat source. In this case, the heat exchange problem in the room can be simplified to an one-dimensional heat transfer problem [22]. Figure 3 shows the heat exchange process between zones in one direction.

To quantify the heat exchange process between two neighboring zones z_1 and z_2 during the heating process, we define a function $h(z_1, z_2)$ to describe the heat transfer process from the zone z_1 to the zone z_2, where the temperature values are T_1 and T_2 for the two zones, respectively. Since heat transfers from one zone with high temperature to another zone with relatively low-temperature region naturally, the value of the function $h(z_1, z_2)$ will be positive if $T_1 > T_2$, else negative. Also, for any positive integers i and j, if $i < j$, we have $h(z_i, z_{i+1}) > h(z_j, z_{j+1})$. The function $h(z_1, z_2)$ is determined by the temperatures T_1, T_2 and the heating time. Without loss of generality, we assume that all the zones in each direction of the heat source get the same thermal energy H from the heat source during the heating process, all the zones in the room have the same temperature value at the beginning, and the zone containing the heat source is z_0. During the heating process, temperature change for each zone in the room is caused by the heat source. For the zone $z_m(m = 1, 2, \cdots)$, its corresponding heat transfer can be described as follows

$$\Delta h_m = h(z_{m-1}, z_m) - h(z_m, z_{m+1})$$

In the heating process, the temperature changes in every zone and the total thermal energy H contributes to all heat exchanges. After stopping heating, the temperature in each zone is determined only by the total exchange energy and the heating time. Let the initial temperature of the zone z_m be T_{m0}, and the air weight and air heat capacity are m and c, respectively, then the temperature value in the zone after heating can be given as

$$T_{m1} = T_{m0} + \frac{\Delta h_m}{mc}$$

Correspondingly, the total exchange heat H from the heat source zone z_0 along one given direction can be written as

$$\begin{aligned} H &= mc(\Delta T_1 + \Delta T_2 + \Delta T_3 + \cdots) \\ &= \Delta h_1 + \Delta h_2 + \Delta h_3 + \cdots \end{aligned} \tag{1}$$

Equation (1) indicates that the zones in the direction have a unique heat distribution pattern after stopping heat exchanging. On the other hand, if we regard the zones z_0 and z_1 as a new zone $\tilde{z}$, the total heat exchange $\tilde{H}$ from the zone $\tilde{z}$ along this direction is

$$\begin{aligned} \tilde{H} &= H - mc\Delta T_1 = mc(\Delta T_2 + \Delta T_3 + \Delta T_4 + \cdots) \\ &= \Delta h_2 + \Delta h_3 + \Delta h_4 + \cdots \end{aligned} \tag{2}$$

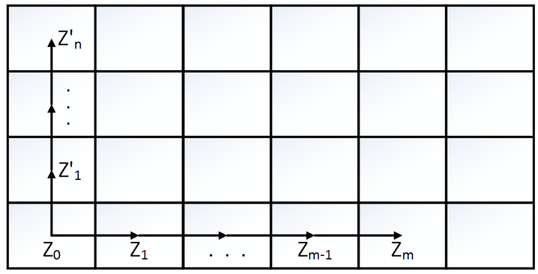


FIGURE 4. A heat exchange route between z_m and z'_n in the room where there is a heat source in the lower left corner.

Through similar treatments for the remaining zones $z_m(m = 2, 3, \cdots)$, we can obtain the total heat exchange of the zone z_m and the corresponding temperature value.

For any two random zones in the room, there is a heat exchange route which goes across the heat source point linking the two zones. Figure 4 shows a route going through the heat source in this building. Each temperature value for the two zones is a function value of the exchange energy and the heating time. The temperature values in the other zones can be obtained once we get the temperature values T_m and T'_n. In such condition, we can deduce the temperature distribution pattern for any zone in the room by the artificial neural network (ANN) method which will be later shown in Section III.

D. MULTIPLE HEAT SOURCE PROBLEM

For the multiple heat source problem, we assume that there are n heat sources distributed randomly in the room. For example, Figure 5 shows the heat exchange process in a room with three heat sources. For each heat source, there is a corresponding one heat source problem when ignoring the effects of the other heat sources. For the one heat source problem, the thermal condition is determined by the produced heat and the heat exchange time. The approach to solve the one heat source problem can be available in [23] and the temperature distribution can be generated in the whole room [24]. The method of treating the one heat source problem can be extended to the case where there are multiple heat sources in buildings. There is also a relationship between the temperature values of different zones for the multiple heat source problem. We can simulate this relationship by using the ANN-based method, and thus deduce the temperature distribution in the room with multiple heat sources. In the next section, we will present the ANN-based methods which will be used to detect the outliers of the temperature values in the zones. The temperature data used in the proposed artificial neural network model is obtained in real time by the temperature sensors. The ANN model can be powerful in automatically extracting the characterization of slow or fast signals and can be capable of learning long-term time dependencies in vibration signals. The temperature outlier can be detected rapidly with pre-trained ANN-based models. For guaranteeing the rapid response to fast signals, a multi-layer

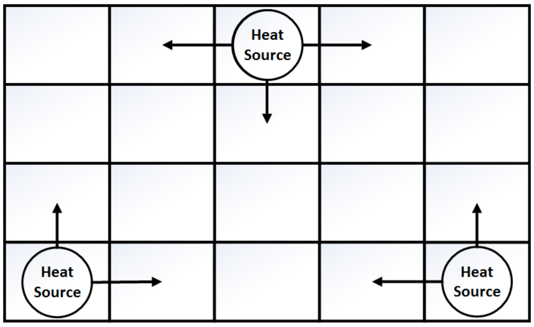


FIGURE 5. The heat exchange in a room with three heat sources.

neural network can be trained and tuned for more complex signal analysis and testing.

III. ANN BASED TEMPERATURE FORECAST MODEL

In this section, we will present practicable ANN-based forecast models which will be used to simulate the relationship between the temperature values of different zones in buildings. By using the proposed models, the outliers of the temperatures can be detected in the WSN system.

A. ARTIFICIAL NEURAL NETWORK METHOD

In cognitive science and machine learning, artificial neural networks are a set of methods inspired by biological neural networks such as central nervous systems of creatures or brains, which are applied to anticipate outcomes that heavily rely on many of inputs. The connections possess numeric weights and biases that are able to be triggered on training and make detections on neural nets adapted to inputs and learning. Artificial neural networks are methods for optimization and learning, and they generally comprise five elements: a directed graph known as network topology whose arcs are named as links, a state variable linked to each node, a weight linked to each link, a bias linked to each node, and a transfer function for each node that is able to decide the state of a node. In many cases, the transfer function often takes the form as a sigmoid function or a step function.

In this work, we use an artificial neural network (ANN) topology to establish the model for temperature prediction in building. The ANN model used in this work has no closed paths, its output nodes have no arcs away from them while its input nodes have no arcs to them, and the other nodes are named hidden nodes. Through the network, all the nodes in the neural network can be set by propagation when the states of all the input nodes are assigned. Given a set of inputs, the ANN model is able to calculate outputs through their implemented mechanisms.

For outlier detection in smart buildings we need to have access to label training examples $\left(x^{(i)}, y^{(i)}\right)$. Neural networks make an approach into defining a complicated and non-linear form of hypotheses $h_{W,b}(x)$ with weights and biases parameters W, b that are established to fit the sampling dataset. The neurons are virtually computational units that take inputs $[x_1, x_2, \ldots, x_n]$, and outputs $h_{W,b}(x) = g\left(W^T x\right) = g\left(\sum_{i=1}^{n} W_i x_i + b\right)$, where g is called the activation function. In this neural network, we choose $g(.)$ to be the sigmoid function

$$g(z) = \frac{1}{1 + \exp(-z)}.$$

Artificial neural networks with layers become prevalent for the following reasons: Most importantly, they are found in practice to fit well in detection and estimation both on sparse set of data points and sufficient data points, and they are capable of providing the outcome for an input not existing in a training set. Through a training method, the neural network can swiftly seek out a relatively good set of weights and biases to fit a current problem, and the topology is able to adjust the network automatically after every time the epoch finishes, thus providing flexibility in applications.

In this article, we use the conventional ANN model that aims at outputting accurate detected value for temperature in a certain office. The focus on the model is to acquire a good set of sampling data which is to feed the input layer of the neural network, and to avoid training out an obsessively complex neural network, which may lead to over-fitting. After the model is established through a training process, figures will be drawn to manifest its performance.

B. ARTIFICIAL NEURAL NETWORK FORECAST MODEL

Artificial neural network is a machine learning methodology inspired by biological neural network. An artificial neural network consists of several layers which are connected with respective weights. ANN can be used to approximate a function, and solve classification and regression problems. In the model generation process for the ANN method, a large amount of input data and the corresponding output data are required. After the self-modification procedures such as back-propagation adjustment, the output of the network will be the approximation of the value corresponding to the given input data [25], [26].

In the WSN system in buildings, ANN is utilized to simulate the relationship between the temperatures of different zones. We assume that the temperature sensors are distributed in the room randomly. The temperature value the sensor measures represents the average temperature of the corresponding zone in the room. After long-time measurement of temperature, time series of temperature can be collected by sensors. We construct an artificial neural network model by using the temperature values gathered by many sensors. This model can be used to predict the temperature values of one sensor by using the synchronous temperature values collected by the other sensors in the system. If the gap between the prediction and the real value exceeds a threshold value, this temperature value will be classified to an outlier and require recovery [27].

We assume that n temperature sensors are employed in the WSN system and our target is to predict one measured value based on the other $n-1$ values. For this purpose, the

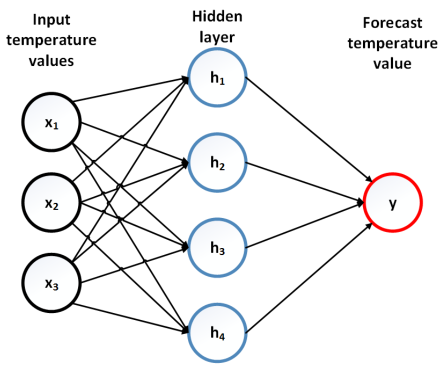


FIGURE 6. The architecture of an artificial neural network for temperature prediction application.

ANN should have $n-1$ input units and one output unit. The training method used in the proposed ANN model is back-propagation with an optimization algorithm such as gradient descent. For simplicity, in the following we assume that there are 4 sensors in the WSN system. For our ANN model, the target error, the learning rate and the epoch number will be initialized and adjusted in the learning process. The number of units in the hidden layer is set to be 4. Figure 6 shows the architecture of the artificial neural network we apply in this system for temperature prediction.

In our ANN model, the measured temperature values of the WSN system, x_1, x_2 and x_3 corresponding to three sensors, are the input values, and y is the output value of the model. In the training process, the target value is x_4 which is measured by the fourth sensor. For $k = 1, 2,$ and 3, we have

$$\begin{aligned} h_k &= f(W_{k1}x_1 + W_{k2}x_2 + W_{k3}x_3) \\ y &= g(W_1h_1 + W_2h_2 + W_3h_3 + W_4h_4) \end{aligned}$$

in which W_i is the connection weight between each units, f is a sigmoid function and g is a linear function. In training process, the weights W_i are set randomly at the beginning. All the measured values are divided into the training and test sets. The cost function of the neural network is defined to be the difference between output y and target value x_4.

$$f_c = |g(W_1h_1 + W_2h_2 + W_3h_3 + W_4h_4) - x_4|$$

Through the self-modulation process, all weights will be adjusted to minimize the cost function. In each epoch, the output and the error can be calculated at the output unit, and thus we can obtain the weights $W_k(k = 1, 2, 3, 4)$ and $W_{ij}(i, j = 1, 2, 3)$ based on the error. By using the back-propagation adjustment process, the weights of the forecast model will converge to optimal values. In the test process, the ANN forecast model will make prediction with the input test values. However, this forecast model has some limitations in some practical applications. When the temperature distribution pattern changes, the accuracy of the prediction of this model will significantly decrease. We need to propose an optimized ANN outlier detection model for treating this case.

C. OPTIMIZED ANN FORECAST MODEL

The temperature data, collected by the WSN for constructing the forecast model, are always confined in a limited scope. Thus the output value of the model is also limited in this range because of the limitation of ANN. When this model is used to predict the temperature when the average environment temperature value exceeds this scope, such as the season change, big error will be introduced to the predicted value. In the following, we propose an optimized ANN forecast model to improve the prediction accuracy in different situations.

To reduce the deviation caused by the change of average environment temperature, our optimized model predicts the total deviation of all values instead of the absolute value of temperature. In the optimized ANN model, the output y in the neural network is replaced by y' and the target value is assumed to be $e_1+e_2+e_3$ where $e_i = x_i - x_4$. Thus, the range of the forecast value will not vary much in different situations, which leads to better accuracy when average environment temperature changes. It should be noted that the structure of the optimized ANN model is same to the ANN model proposed in Section III-B. For this optimized ANN model, we have

$$\begin{aligned} h'_k &= f(W'_{k1}x_1 + W'_{k2}x_2 + W'_{k3}x_3) \\ y' &= g(W'_1h'_1 + W'_2h'_2 + W'_3h'_3 + W'_4h'_4) \end{aligned}$$

The cost function of the optimized model is set to be $f'_c = |y' - (e_1 + e_2 + e_3)|$. The target value y will be calculated with the other temperature values x_1, x_2 and x_3, and the prediction y' of this optimized ANN model, given by

$$y = \frac{x_1 + x_2 + x_3 - y'}{3}$$

The process to optimize the weights in this optimized ANN forecast model is the same as in the ANN model in Section III-B. In theory, the error of the value predicted by the optimized model should not exceed the range of the training error. When the error is much larger than the expectation, such as twice of the average training error, the provided value can be regarded as an outlier for the measurement and requires recovery. To compensate the measurement error, the outlier will be substituted by the prediction of the forecast model.

Another method for improving the accuracy is to prolong the measuring time. A complete measured data set can cover most cases which the system will encounter. The performance of the ANN-based forecast model can be optimized when the test data is included in the training set. In the following section, we will show experiment results to demonstrate the effectiveness of the ANN-based forecast model and the corresponding optimized model for temperature prediction in the WSN system in real buildings.

IV. EXPERIMENT AND DISCUSSION

In order to examine the performance of the proposed ANN-based models, we carried out the experiments in a room in Shanghai. For the temperature measurement, we employ

FIGURE 7. **The TI CC2650 development Kit.**

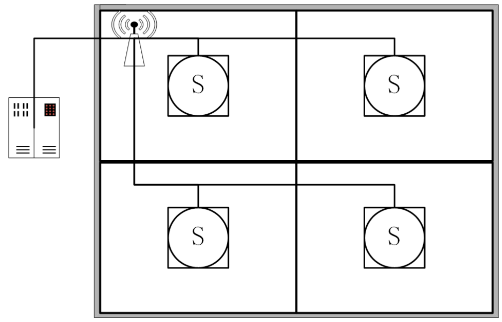


FIGURE 8. **The overview of the wireless sensor network.**

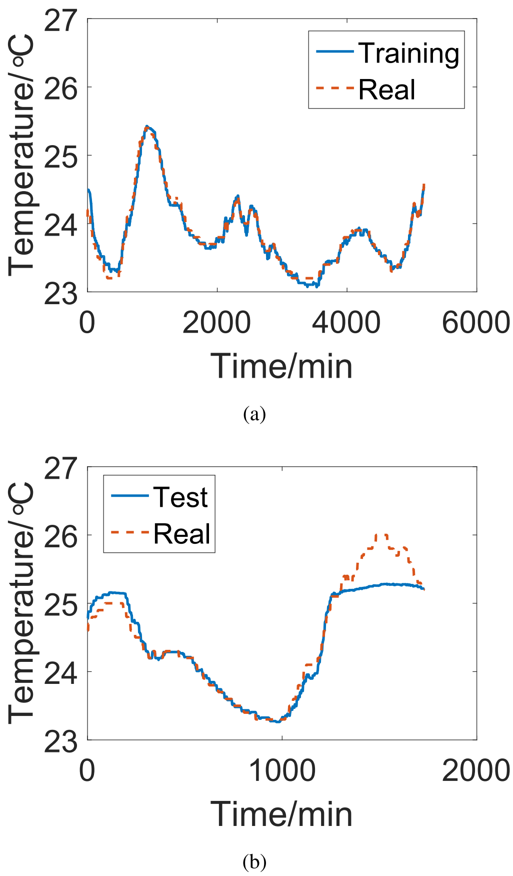


FIGURE 9. **Experiment results obtained by the ANN model: (a) comparison between the training data and the real data; (b) comparison between the test data and the real data.**

four TI CC2650 SensorTags placed in different zones in the room to gather temperature values. In the experiments, each CC2650 SensorTag contains ten low-power sensors such as the temperature sensor, light sensor, humidity sensor, pressure sensor, and magnetic field sensor. The proposed ANN-based method can be used for outlier detection in complex environment for wireless sensor networks which are constituted by many sensors, several controllers, the manager and gateway. In this established network, the temperature data is collected by sensors, and transmitted to controllers for processing. It can be sent to the processor for further analysis and application, or to the internet through the gateway.

The CC2650 device developed by Texas Instrument is a wireless MCU targeting Bluetooth Smart, 6LoWPAN, and ZigBee RF4CE remote control applications, as shown in Figure 7. It is a member of the CC26xx family of cost-effective, low power, 2.4-GHz RF devices. It contains a 32-bit ARM Cortex-M3 processor that runs at 48 MHz as the main processor and a rich peripheral feature set that includes a unique ultralow power sensor controller, which is ideal for interfacing external sensors and for collecting analog and digital data autonomously while the rest of the system is in sleep mode. The CC2650 device has an evaluation board on which array a LCD, 4 LEDs, 5 buttons, a light sensor and many GPIO pins. With these components, experiments about wireless communication and sensors can be carried out and tested. It should be noted that the 32-bit ARM Cortex-M3 processor running at 48 MHz is the main processor is included in the TI CC2650 Development Kit. The micro-controller interfaces external sensors for collecting analog and digital data autonomously and it does not need to run at 48MHz. When reading the temperature data detected by sensors, the ADC module can read the output voltages of sensors through the external pins of the development board and convert them to the corresponding values. It should be also noted that the CC26xx family hardware can be advisable to be used to establish wireless sensor networks for gathering temperature values in smart building.

Figure 8 shows the architecture of the wireless sensor network system and the communication system in our experiment. The main influencing factor of the air temperature fluctuation in the room is the air exchange with the outside environment. Comparing with this factor, the others such as the computer power consumption are trivial for the room temperature change. The collected temperature values have obvious periodicity and regularity, and slowly rise because of the seasonal warming. By extracting and preprocessing the data collected in 5 days, we have about 7000 temperature values for each sensor, each of which represents the average temperature in one minute. We used the earlier 75% of the data as the training data to construct the artificial neural network model, and the others as the test data to examine the accuracy of the proposed ANN-based model. The architecture of this ANN model has been described in Section III.

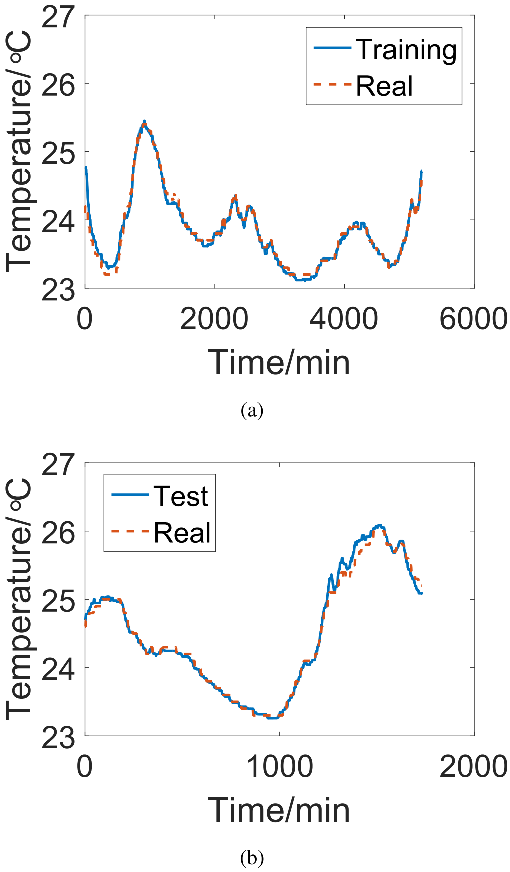


FIGURE 10. Experiment results obtained by the optimized ANN model: (a) comparison between the training data and the real data; (b) comparison between the test data and the real data.

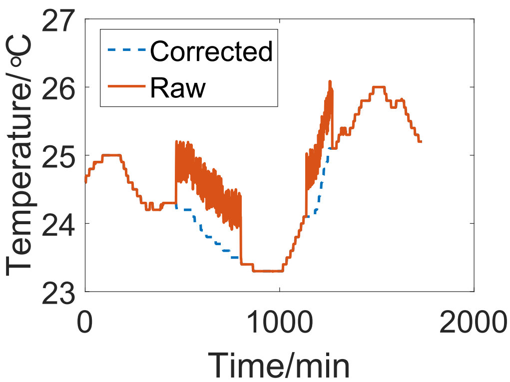


FIGURE 11. The outlier correction process by the optimized ANN model.

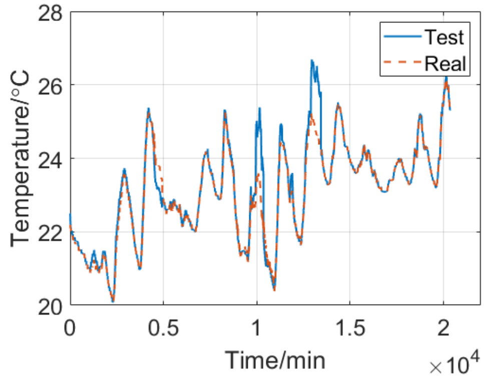


FIGURE 12. Temperature prediction in 14 days using the artificial neural network model.

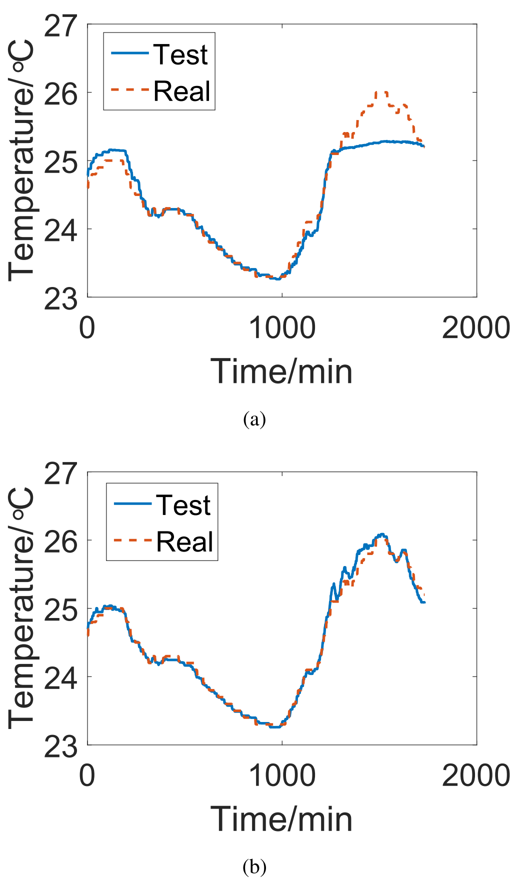


FIGURE 13. The test results of (a) the ANN model and (b) the optimized ANN model.

The comparisons between the training data and the real data, and between the test data and the real data are shown in Figure 9. Experiment results show that the average error for the training process is 0.0545°C, and is 0.1440°C for the test process, and the largest absolute error is 0.2939°C.

It can be seen from Figure 9, the predicted values agree with most of the real values except the data near the second peak (about 26°C). To improve the forecast accuracy, we built an optimized ANN forecast model by the training data as mentioned in Section III-C. The experiment results are shown in Figure 10. For the optimized ANN model, the average errors for the training and test processes are 0.0596°C and 0.0534°C, respectively, and the corresponding largest absolute error is 0.1862°C.

It should be noted that both the ANN and optimized ANN methods are implemented on MATLAB software on a PC with 8-GB RAM and 3.6-GHz Intel i7 Dual-Core CPU. In the training process, the ANN and optimized ANN methods respectively take about 1 second and 6 seconds with

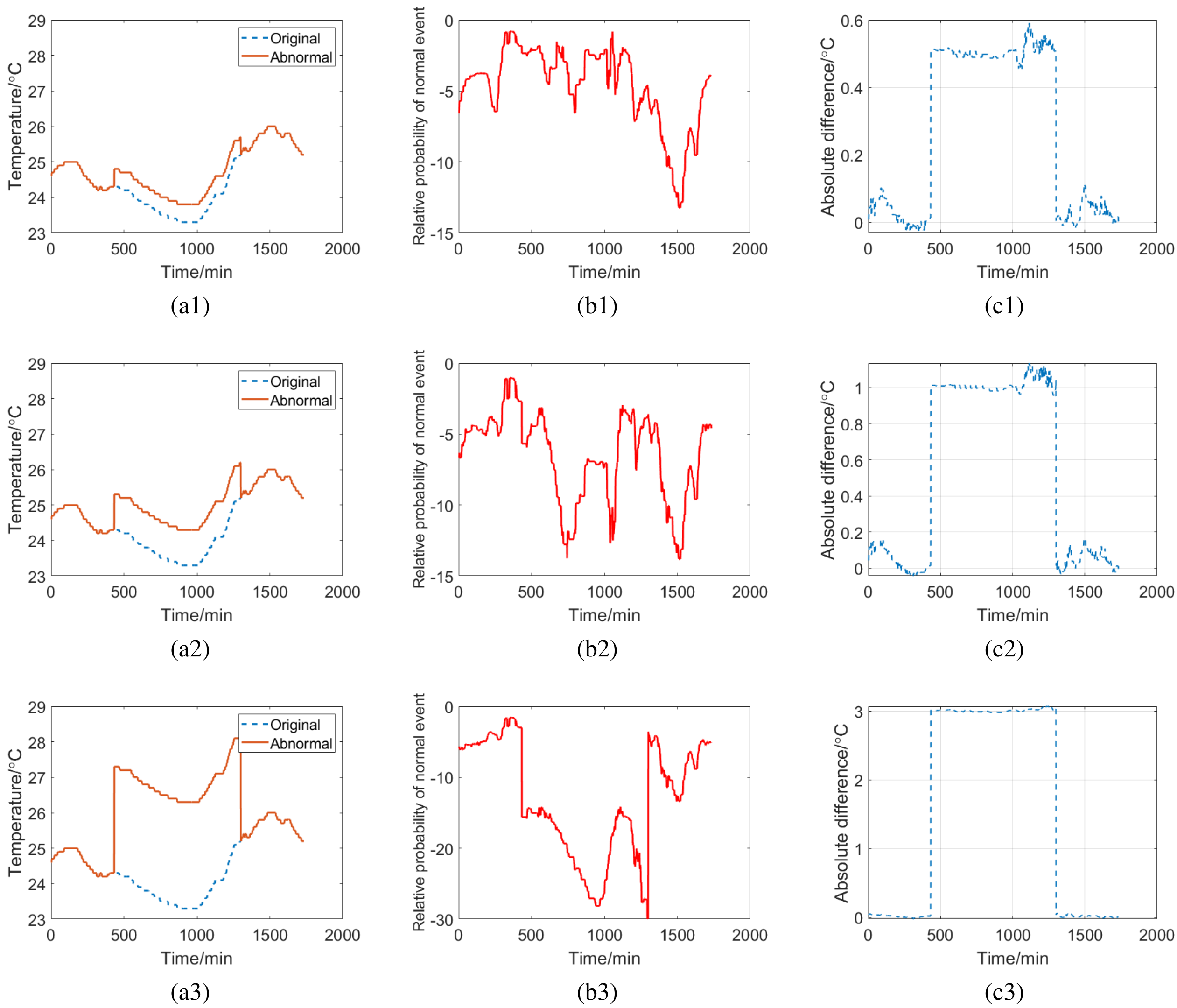


FIGURE 14. **Experimental results using the proposed optimized ANN-based approach and the HMM-based approach for anomaly temperature data detection: (a) the original temperatures and the abnormal values; (b) detection results using the HMM-based approach; (c) detection results using the proposed ANN-based approach.**

about 5000 temperature records. While in the test process, the ANN and optimized ANN models respectively take about 0.007787 second and 0.009211 second with about 2000 temperature records. Experimental results show that we can achieve a real-time outlier detection with high accuracy by a pre-trained optimized ANN model.

To examine the performance of the outlier detection approach, a series of random deviation are introduced to the measured temperature values. If the difference between the measured value and the predicted value is more than 0.4°C, this measured value is regarded as an outlier and need to be replaced by the predicted value. Figure 11 shows the outlier correction process for the optimized ANN model. Figure 12 shows the temperature prediction in 14 days using the artificial neural network model which has a success prediction rate with small average error about 0.09°C. It can be seen from Figures 9, 10, 11 and 12 that both the ANN and optimized ANN models perform well in the training process. However, compared with the ANN model, the optimized ANN forecast model provides a distinctly better result in the test process, which can be seen in Figure 13. The accuracy of the optimized ANN forecast method is 62.92% better than the ANN forecast model. Therefore, the ANN model can be utilized in temperature prediction in the WSN system in buildings, and the optimized ANN forecast method can achieve better performance in outlier detection.

It should be noted that the running times of both the ANN and optimized ANN models are acceptable when the number of the sensors in the WSN system is small. Experimental results show that it takes less than one second to construct both forecast models. If it takes a long time to build the forecast model because of a large number of sensors and a long-time measurement in the WSN system, the structure of the neural network can be substituted by radial basis function neural network for speed optimizing.

Hidden Markov Model which is a typical statistical analysis model different from neural network methods, plays an important role in the field of signal processing. It has been successfully used in speech recognition, behavior recognition, character recognition and outlier detection. In order to further show the high performances of our proposed outlier detection model, the HMM-based outlier detection approach is implemented with the same training and testing data. HMM is a statistical model which is used to describe a Markov process with hidden unknown parameters. An HMM variable λ can be determined by five parameters as the following format

$$\lambda = (Q, V, A, B, \pi) \tag{3}$$

where $Q = \{q_1, \ldots, q_N\}$ represents the set of all possible states, $V = \{v_1, \ldots, v_M\}$ represents the set of all possible observations, A and B are respectively the state transition probability matrix and the observation probability matrix, π is the initial probability distribution vector. As stated in the ANN and the optimized ANN models, we use the earlier 75% as the training data of HMM and the others as the test data for examining the detection accuracy. It should be noted that the constructed HMM is a semi-supervised model. We assume the training data is normal and then obtain a trained HMM representing normal events. Each test record will be examined by a relative probability for determining whether it is normal in the test process. For showing the detection performances by our proposed ANN-based approach and the HMM approach, half of the test data ($433 \leq$ time ≤ 1299) is increased by 0.5, 1 and 3 degrees celsius respectively which is shown in Figures 14 (a1)-(a3). The abnormal degrees obtained by the HMM approach and the ANN based model are shown in Figures 14 (b) and (c). Logarithmic relative probability of normal event as shown in Figure 14 (b) means whether the corresponding temperature record is a normal event. Absolute temperature difference shown in Figure 14 (c) means that the difference between the measured temperature value and the predicted value obtained by the ANN model. The lower the relative probability of normal event, the higher the absolute difference of the record is, the more likely it is not normal. It can be seen from Figures 14 (b) and (c) that the proposed ANN approach has an obvious advantage over the HMM-based detection approach.

V. CONCLUSION

In this paper, we propose an ANN-based model to simulate the relationship between the temperature values of different zones in buildings. By this new approach, the accurate temperature data in different zones can be predicted, and the outliers can be identified by computing the posterior probability of the temperature values measured by the sensors in the WSN system. Also, we propose an optimized ANN model to improve the prediction accuracy in the case of a big change of the environment temperature. Experiment and simulation results show that both the ANN-based approaches are effective in predicting the temperature value in the WSN system in smart buildings, and the optimized ANN model can achieve better performance when the environment temperature changes significantly. The comparison results against the HMM-based approach further show that the proposed ANN detection model is effective for detecting and thermal controlling in the Internet of Things (IoT) applications. Further work will address larger scale WSNs aiming to develop reliable and precise outlier detection approaches for smart building system.

REFERENCES

[1] J. Yick, B. Mukherjee, and D. Ghosal, "Wireless sensor network survey," *Comput. Netw.*, vol. 52, no. 12, pp. 2292–2330, Aug. 2008.

[2] C.-F. Huang and Y.-C. Tseng, "The coverage problem in a wireless sensor network," *Mobile Netw. Appl.*, vol. 10, no. 4, pp. 519–528, 2005.

[3] M. Cardei and D.-Z. Du, "Improving wireless sensor network lifetime through power aware organization," *Wireless Netw.*, vol. 11, no. 3, pp. 333–340, 2005.

[4] A. Perrig, J. Stankovic, and D. Wagner, "Security in wireless sensor networks," *Commun. ACM*, vol. 47, no. 6, pp. 53–57, Jun. 2004.

[5] J. Polastre, J. Hill, and D. Culler, "Versatile low power media access for wireless sensor networks," in *Proc. ACM 2nd Int. Conf. Embedded Netw. Sensor Syst.*, 2004, pp. 95–107.

[6] G. Anastasi, M. Conti, M. Di Francesco, and A. Passarella, "Energy conservation in wireless sensor networks: A survey," *Ad Hoc Netw.*, vol. 7, no. 3, pp. 537–568, May 2009.

[7] S. Percy, C. Knight, F. Cooray, and K. Smart, "Supplying the power requirements to a sensor network using radio frequency power transfer," *Sensors*, vol. 12, no. 7, pp. 8571–8585, 2012.

[8] M. C. Mozer, "The neural network house: An environment hat adapts to its inhabitants," in *Proc. AAAI Spring Symp. Intell. Environ.*, 1998, pp. 110–114.

[9] Z. Zhen-Ya, C. Hong-Mei, and Z. Shu-Guang , "An approach to the identification of temperature in intelligent building based on feed forward neural network and genetic algorithm," in *Proc. 3rd IEEE Int. Conf. Comput. Sci. Inf. Technol. (ICCSIT)*, vol. 2, Jul. 2010, pp. 286–290.

[10] L. Chen, Q.-S. Fang, and Z.-Y. Zhang, "Research on the identification of temperature in intelligent building based on feed forward neural network and particle swarm optimization algorithm," in *Proc. IEEE 6th Int. Conf. Natural Comput. (ICNC)*, vol. 4, Aug. 2010, pp. 1816–1820.

[11] H. Grindvoll, O. Vermesan, T. Crosbie, R. Bahr, N. Dawood, and G. M. Revel, "A wireless sensor network for intelligent building energy management based on multi communication standards—A case study," *J. Inf. Technol. Construct.*, vol. 17, pp. 43–62, May 2012.

[12] L. Yu, N. Wang, and X. Meng, "Real-time forest fire detection with wireless sensor networks," in *Proc. Int. Conf. Wireless Commun., Netw. Mobile Comput.*, vol. 2, Sep. 2005, pp. 1214–1217.

[13] O. H. Uribe, J. P. S. Martin, M. C. Garcia-Alegre, M. Santos, and D. Guinea, "Smart building: Decision making architecture for thermal energy management," *Sensors*, vol. 15, no. 11, pp. 27543–27568, 2015.

[14] P. A. González and J. M. Zamarreño, "Prediction of hourly energy consumption in buildings based on a feedback artificial neural network," *Energy Buildings*, vol. 37, no. 6, pp. 595–601, 2005.

[15] K. Römer and F. Mattern, "The design space of wireless sensor networks," *IEEE Wireless Commun.*, vol. 11, no. 6, pp. 54–61, Dec. 2004.

[16] V. J. Hodge and J. Austin, "A survey of outlier detection methodologies," *Artif. Intell. Rev.*, vol. 22, no. 2, pp. 85–126, 2004.

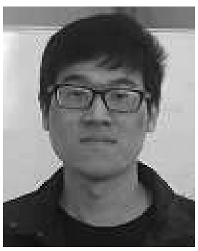
[17] M. Gupta, J. Gao, C. C. Aggarwal, and J. Han, "Outlier detection for temporal data: A survey," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 9, pp. 2250–2267, Sep. 2014.

[18] C. Wang, H. Lin, and H. Jiang, "Trajectory-based multi-dimensional outlier detection in wireless sensor networks using hidden Markov models," *Wireless Netw.*, vol. 20, no. 8, pp. 2409–2418, 2014.

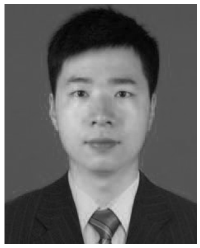
[19] Y. Zhang, N. A. S. Hamm, N. Meratnia, A. Stein, M. van de Voort, and P. J. M. Havinga, "Statistics-based outlier detection for wireless sensor networks," *Int. J. Geograph. Inf. Sci.*, vol. 26, no. 8, pp. 1373–1392, 2012.

[20] O. Sabokpa, A. Zarei-Hanzaki, H. R. Abedi, and N. Haghdadi, "Artificial neural network modeling to predict the high temperature flow behavior of an AZ81 magnesium alloy," *Mater. Des.*, vol. 39, pp. 390–396, Aug. 2012.
[21] H. R. Maier and G. C. Dandy, "Neural networks for the prediction and forecasting of water resources variables: A review of modelling issues and applications," *Environ. Model. Softw.*, vol. 15, no. 1, pp. 101–124, 2000. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1364815299000079
[22] X. Bao, J. Dhliwayo, N. Heron, D. J. Webb, and D. A. Jackson, "Experimental and theoretical studies on a distributed temperature sensor based on Brillouin scattering," *J. Lightw. Technol.*, vol. 13, no. 7, pp. 1340–1348, Jul. 1995.
[23] B. K. Tan, X. Y. Huang, T. N. Wong, and K. T. Ooi, "A study of multiple heat sources on a flat plate heat pipe using a point source approach," *Int. J. Heat Mass Transf.*, vol. 43, no. 20, pp. 3755–3764, 2000.
[24] F. Lefèvre and C. Le Niliot, "Multiple transient point heat sources identification in heat diffusion: Application to experimental 2D problems," *Int. J. Heat Mass Transf.*, vol. 45, no. 9, pp. 1951–1964, 2002.
[25] S. Karatasou, M. Santamouris, and V. Geros, "Modeling and predicting building's energy use with artificial neural networks: Methods and results," *Energy Buildings*, vol. 38, no. 8, pp. 949–958, Aug. 2006.
[26] M. T. Hagan, H. B. Demuth, M. H. Beale, and O. De Jesus, *Neural Network Design*, vol. 20. Boston, MA, USA: PWS Publishing Company, 1996.
[27] A. E. Ruano, E. M. Crispim, E. Z. Conceição, and M. M. J. Lúcio, "Prediction of building's temperature using neural networks models," *Energy Buildings*, vol. 38, no. 6, pp. 682–694, Jun. 2006.

KAI ZHANG received the bachelor's, master's, and Ph.D. degrees in navigation, guidance, and control from Northwestern Polytechnical University, Xi'an, China, in 2001, 2004, and 2009, respectively, where he is currently an Associate Professor with the School of Aerospace. He has authored or coauthored more than 20 papers in scientific journals and conference proceedings. His current research interests include infrared imaging, artificial intelligence, and infrared scene modeling and simulation.



KE YANG received the B.S. degree in electronic science and technology from Xi'an Jiaotong University, Xi'an, China, in 2017. He is currently pursuing the M.S. degree with the Department of Micro/Nano-Electronics, Shanghai Jiao Tong University, Shanghai, China. His current research interests include integrated circuit design automation and infrared small target detection.



SHAOYI LI received the bachelor's degree in measurement and control technology from Southwest Jiaotong University, Chengdu, China, in 2008, and the master's and Ph.D. degrees in navigation, guidance and control from Northwestern Polytechnical University, Xi'an, China, in 2015 and 2011, respectively, where he is currently an Assistant Researcher with the School of Aerospace. He has authored or coauthored more than ten papers in scientific journals and conference proceedings. His current research interests include ATR based on infrared image, artificial intelligence, and image processing.

DISHAN JING received the B.S. degree in electronic science and technology from the Dalian University of Technology, Dalian, China, in 2017. She is currently pursuing the M.S. degree with the Department of Micro/Nano-Electronics, Shanghai Jiao Tong University, Shanghai, China. Her current research interests include the Internet of Things and security intrusion detection.

HAI-BAO CHEN received the B.S. degree in information and computing sciences, and the M.S. and Ph.D. degrees in applied mathematics from Xi'an Jiaotong University, Xi'an, China, in 2006, 2008, and 2012, respectively. He, then, joined Huawei Technologies, where he focused on cloud computing and big data. He was a Postdoctoral Research Fellow with the Electrical Engineering Department, University of California, Riverside, CA, USA, from 2013 to 2014. He is currently an Associate Professor with the Department of Micro/Nano-electronics, Shanghai Jiao Tong University, Shanghai, China. He has authored or coauthored about 50 papers in scientific journals and conference proceedings. His current research interests include model order reduction, system and control theory, circuit simulation, cloud computing and big data, and electromigration reliability. He received one Best Paper Award nomination from Asia and South Pacific Design Automation Conference (ASP-DAC), in 2015. He currently serves as an Associate Editor for *Integration—the VLSI Journal*.