IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received January 17, 2019, accepted January 30, 2019, date of publication February 12, 2019, date of current version February 27, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2898690

Link Quality Estimation Method for Wireless Sensor Networks Based on Stacked Autoencoder

XIONGHUI LUO[1], LINLAN LIU[1], JIAN SHU[2], AND MANAR AL-KALI[2]
[1]School of Information Engineering, Nanchang Hangkong University, Nanchang 330063, China
[2]School of Software, Nanchang Hangkong University, Nanchang 330063, China

Corresponding author: Linlan Liu (765693987@qq.com) and Jian Shu (shujian@nchu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61762065 and Grant 61363015, and in part by the Natural Science Foundation of Jiangxi Province under Grant 20171ACB20018, Grant 20071BBH80022, and Grant GJJ150702.

ABSTRACT In wireless sensor networks, effective link quality estimation is the basis of topology management and routing control. Effective link quality estimation can guarantee the transmission of data, as well as improve the throughput rate, and hence, extend the life of the entire network. For this reason, a stacked autoencoder-based link quality estimator (LQE-SAE) is proposed. Specifically, the zero-filling method is developed to process the original missing link information. Then, the SAE model is used to extract the asymmetric characteristics of the uplink and downlink from the received signal strength indicator, link quality indicator, and signal-to-noise ratio, respectively. These characteristics are fused by SAE to construct the link features' vectors, which are given as inputs to the support vector classifier (SVC), for which the link quality grade is taken as its label. The experimental results in different scenarios show that the LQE-SAE has better accuracy than link quality estimators based on the SVC, ELM, and WNN.

INDEX TERMS Wireless sensor networks, link quality estimation, stacked autoencoder, deep learning, asymmetry of link.

I. INTRODUCTION

Wireless Sensor Networks (WSNs) [1] are made up of a large number of nodes deployed in a monitoring area with self-organized manner. With data functions, such as acquisition, dispatch, and forwarding, sensor nodes are used to perceive the physical world information. WSNs, as an important part of the Internet, are widely applied in the military, health care, environmental monitoring, smart homes and so on.

Links are the basis of WSNs to realize the interconnection of nodes and multihop communication. The limited resources of sensing nodes, the complexity of the monitoring environment, and the diversity of noise lead to the directionality, asymmetry [2], [3], instability, and "gray zone" of communication and other space-time characteristics [4], [5], which cause volatility of communication and affect data transmission. Research shows that most of the energy consumption of the node is in the data transmission of the wireless communication module [6] and the routing process [7]. Effective link quality estimation model can assist the routing protocol [8] to select high-quality links to transmit data. Then it can avoid unnecessary energy consumption of re-transmission, and can improve the throughput so as to extend the life of the entire network. Most of them take the mean of physical layer parameters to establish the relationship with PRR, however, the mean will lose part of the link quality information. Some link quality estimation methods only consider forward link information, ignoring the asymmetry of the link, others take simple way, adding or subtracting, to handle the asymmetry of links.

This paper aims to provide a novel link quality estimator based on the SAE; thus, we provide an overview of the existing link quality estimation model in Section 2. Section 3 introduces the Zero-filling method for preprocessing parameters. The LQE-SAE is proposed in Section 4. Section 5 verifies the performance of our method. Finally, Section 6 is the conclusion.

II. RELATED WORK

In recent years, much research has been conducted on the link quality estimation of WSNs, including estimation methods based on link characteristics, statistics, and machine learning.

The associate editor coordinating the review of this manuscript and approving it for publication was Jafar A. Alzubi.

Estimation methods based on link characteristics evaluate link quality by link characteristics such as volatility, instability and so on. Jayasri and Hemalatha [9] filtered the RSSI; then, comprehensively evaluated link quality by applying the fuzzy rules to dealt with the filtered RSSI, the LQI, the packet reception rate (PRR) as well as the stability coefficient. Ansar and Dargie [10] analyzed the volatility of link quality under static nodes, dynamic nodes and multinode networking; estimated the duration of the good and bad link quality from the statistics of received ACK packets; and thereby proposed the adaptive burst transmission protocol, which improves the utilization of shared channels by sharing the link state with neighboring nodes so that packets are sent during a good link quality cycle and go to sleep during bad cycles. Lu *et al.* [11] selected the sink node for data upload according to the communication quality of the upper layer network and the link quality of the bidirectional link to achieve higher reliability by the WSNs in the transmission line monitoring system; selected the next node according to the reliability and residual energy of the node path and the number of hops of the neighbor node; evaluated the link quality by the packet loss rate, residual energy, and hop count of the neighbor node; and then, adjusted energy by energy weight. Deb *et al.* [12] proposed the link verification metric (LVM) model, which is based on the existing comprehensive link quality metric. This model evaluates link quality by integrating the parameters PRR, RSSI and SNR standardized; this integration enhances the impact of PRR and SNR on link quality and weakens the impact of RSSI on it.

Statistics-based methods of evaluating link quality are based on a large number of data packets. Rekik *et al.* [13] combined the three link attributes of packet retransmissions, packet transmissions, and channel quality by fuzzy logic; utilized the overall link quality estimation method; and applied it in the routing protocol. Sun *et al.* [14] proposed a link quality estimation algorithm based on a wavelet neural network. The algorithm utilizes a moving average algorithm to decompose SNR to a time-varying nonlinear and nonstationary random two divisors. The SNR and its variance at the next moment are obtained by a wavelet neural network so that the confidence interval of the SNR is calculated. The tail-integral function of the unit Gaussian probability density function is used to obtain the mapping relationship between the SNR and the packet acceptance rate to calculate confidence intervals of the packet acceptance rate. Sun *et al.* [15] proposed a reliable link estimation model for end-to-end data transmission in WSNs applications for industrial monitoring and control. The model not only considers the distance between the receiving node and the sending node, but it also proposes a submetric model according to significant factors affecting channel signal propagation. The statistical characteristics of background noise were accurately represented by the alpha-stable distribution, and the path loss model based on the RSSI was proposed through the improved log normal distribution, achieving the mapping relation between the physical channel and the PRR.

Methods based on machine learning convert the estimation of link quality into classification, mainly utilizing pattern matching, supervised learning and other technologies for modeling. Gomes *et al.* [2] pointed out that energy consumption is caused by sending probe packets used in the existing link quality estimation. By contrast, in analyzing the relationship between the RSSI, LQI and PRR, the link quality estimation node is designed to indicate that the mean value of the RSSI which has a strong correlation with the PRR. The three link quality indicators (p_f, c_a, and p_b) defined by the RSSI, and the number of received packet copies are used to capture the link-path multipath effect, channel interference, and link quality variance caused by the asymmetry link. Lowrance [16] proposed a new link quality prediction model mixing machine learning and fuzzy logic to meet the requirements of dramatic changes in the environment of wireless robot networks by utilizing fuzzy logic to conclude and model wireless LQ as a fluidized and dynamic process. The model adopts an active incremental learning framework to minimize the sampling overhead in the network and achieve the purpose of quickly building and maintaining a predictive model. Fu *et al.* [17] proposed an estimation model for detecting the degradation of abnormal link quality in low-power wireless links. This model, based on Bayesian theory, designs a threshold-based decision method that takes the sampled RSSI mean, variance, and appropriate prior probability as input and returns the minimum error rate. Kulin *et al.* [18] divided the prediction method into three steps: data acquisition, offline modeling and performance estimation. The neural network training prediction model takes the number of detected nodes, the interval between packets, the number of accepted packets, and the number of error frames as input; and returns the packet loss rate, thereby predicting link quality.

Some of these methods, such as those presented in [9] and [10], use an insufficient criterion (e.g., ''good or bad'') for evaluating link quality. Others, such as those presented in [11], [15], [16], and [18] use statistical values, so they cannot quickly reflect link quality. Other works, such as [12], [13], [2], and [17] use the mean of link quality parameter; thus, the link quality information may be lost. In general, the existing link quality estimation models, such as [14], mostly consider forward link information and ignore the role of reverse information in link quality estimation. Therefore, to fully utilize link asymmetry information, this paper proposes the LQE-SAE, a method based on a Stacked Autoencoder for evaluating the link quality of WSNs. The specific flow is shown in Figure 1. In summary, the LQE-SAE determines the link quality grade hierarchy according to the PRR; extracts the asymmetry characteristics of each physical layer parameters between the uplink and downlink; and obtains asymmetry features of the link, which are taken as inputs to the support vector classifier to classify the link quality, thereby improving the accuracy of the link quality estimation.

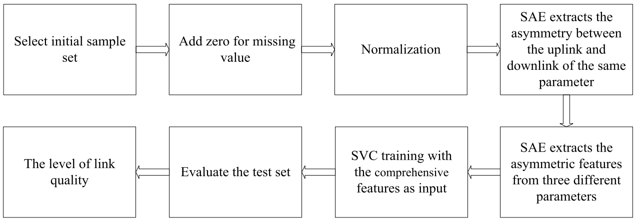


FIGURE 1. LQE-SAE model.

III. LINK QUALITY PARAMETER PREPROCESSING

Packet loss in the detection process leads to the loss of link quality information. Thus, the missing information is filled in with 0's; then, the normalized data, which were processed through the missing values, are later used as inputs to the LQE-SAE algorithm.

A. SELECTION OF LINK QUALITY PARAMETERS

In WSNs applications, due to objective factors, such as different transmit powers of the sensor nodes and the temporal and spatial characteristics of the background noise, the links are asymmetric, i.e., the asymmetry of uplink and downlink quality. Compared with the WSNs link layer estimation parameters, the physical layer parameters have the advantages of rapid response to link quality, easily reading parameters, and low overhead. In this paper, the uplink and downlink of the RSSI, SNR and LQI are selected as the link quality estimation parameters.

B. PROCESSING THE MISSING VALUES OF THE PARAMETER

Considering that the averaging of parameters inevitably loses part of the link information, the original value of each cycle is used to evaluate link quality. In the real environment, due to actual factors (such as interference), random packet loss occurs, and some of the original link quality information values are missing. To adapt to the SAE model, this paper proposes the zero value method of filling the missing link information with 0's; this method is consistent with the packet loss of WSNs. The specific Zero-filling process is shown in Algorithm 1.

After the Zero-filling operation, the simultaneous uplink and downlink quality parameters within one detection period are spliced and recombined into one input; for example, the input of LQI $\{\mathrm{LQI}_{\mathrm{up}}, \mathrm{LQI}_{\mathrm{down}}\}$. $\mathrm{LQI}_{\mathrm{up}}$ denotes the uplink LQI vector, and $\mathrm{LQI}_{\mathrm{down}}$ is the downlink LQI vector. Figure 2 shows an example of the input of a detection period after 0's padding in which the third packet is lost during a detection period; that is, the third message sequence value is missing. $\mathrm{LQI}^1_{\mathrm{up}}$ is the first uplink LQI value and $\mathrm{LQI}^1_{\mathrm{down}}$ is the first downlink LQI value in the detection period. If a probe period includes 30 probe packets, then n is 30, and it forms a 60-element input vector.

C. NORMALIZATION

The range of each link quality parameters in the wireless sensor network is different. To eliminate the impact of range

---

Algorithm 1 Corresponding Zero-Filling Operation

Input: Raw data set
Output: The model input dataset
Step 1: Get the set D of message sequences detected in each cycle in the data from the data sample
Step 2: Create a collection A, where the elements are the total number of probe packets that are needed during the period
Step 3: Compare message sequence set D with set A
Step 4: if the elements in set D are in set A
    Keep the original value of the parameter at this position
  else
    Use 0 to assign the link quality information of the corresponding position of the message sequence
Step 5: If the elements in set D are all compared to the elements in set A
    End the algorithm
  else
    Return to step 4

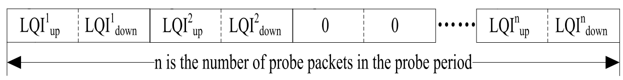


FIGURE 2. One cycle of input after filling.

and reduce model error, the data are normalized so that the data are between 0 and 1. The normalization process is shown in (1) as follows:

$$X^* = \frac{X - \min(X)}{\max(X) - \min(X)} \tag{1}$$

where X denotes the raw physical layer link quality parameter, X^* refers to the normalized physical layer link quality parameter, $\max(X)$ refers to the maximum value of the original physical layer link quality parameter, and $\min(X)$ is the minimum value of the raw physical layer link quality parameter.

D. LINK QUALITY ESTIMATION

The asymmetry information of each physical layer parameter is extracted by the SAE, then the asymmetry feature of the link is obtained so that the relationship between the link quality and its parameters is constructed. The link quality estimation problem is transformed into a classification problem by dividing the quality grade of the link.

E. LINK ASYMMETRY INFORMATION

Single physical layer parameter could characterize the link quality of WSNs to some extent, however the comprehensive link quality parameters can characterize link quality better. The SAE4 is used to fuse the asymmetric information of three physical layer parameters which are extracted from SAE1, SAE2 and SAE3, hence it can obtain the asymmetric characteristics of the link, shown in Figure 3.The relationship

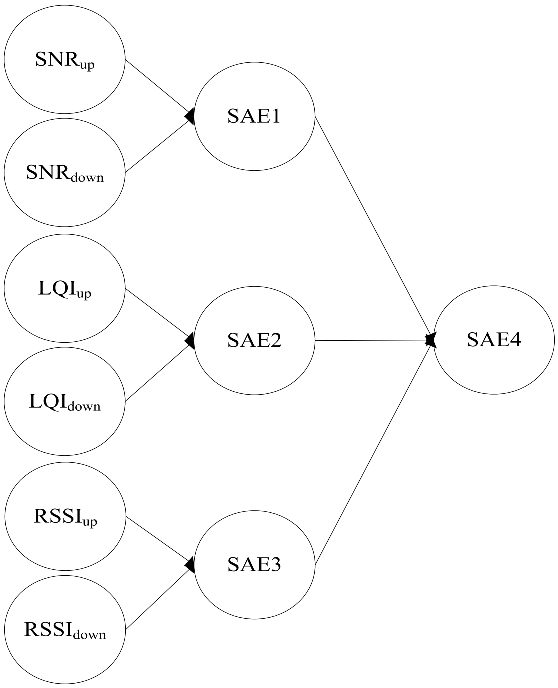


FIGURE 3. The process of deep learning.

between the parameters and the link grade is achieved through their features.

In this work, this paper deals with multiple features extraction and fusion processes. As shown in Figure 3, the SAE1 is used to extract the asymmetric feature of SNR_{up} and SNR_{down}. SNR_{up} denotes the uplink SNR, and SNR_{down} is the downlink SNR. Then, SAE2 is used to extract the asymmetric feature of LQI_{up} and LQI_{down}. LQI_{up} denotes the uplink LQI, and LQI_{down} is the downlink LQI. Afterwards, SAE3 is used to extract the asymmetric feature of $RSSI_{up}$ and $RSSI_{down}$. $RSSI_{up}$ denotes the uplink RSSI, and $RSSI_{down}$ is the downlink RSSI. Finally, the SAE4 fuses all the three distinguished asymmetric features extracted previously.

F. ASYMMETRIC INFORMATION EXTRACTION OF PHYSICAL LAYER PARAMETERS

The SAE is an autoencoder-stacking algorithm, which is a multihidden layer deep learning algorithm. It adopts greedy layer wise pretraining, the hidden layer of the trained autoencoder is used as the input of the next autoencoder, and layer-by-layer training is performed.

In the data collection process, this paper adopts the method of active detection and simultaneous acquisition of uplink and downlink quality parameter information. The SAE is used to obtain the deep asymmetry information between the uplink and downlink of each estimation parameter to instead of simply adding or subtracting. That is, the SAE is used for feature extraction of the uplink and downlink for the SNR, RSSI and LQI, respectively.

The inputs are $\{SNR_{up}, SNR_{down}\}$, $\{LQI_{up}, LQI_{down}\}$, and $\{RSSI_{up}, RSSI_{down}\}$. In this method, SNR_{up} is the uplink SNR vector, SNR_{down} is the downlink SNR vector, LQI_{up} is the uplink LQI vector, and LQI_{down} is the downlink LQI vector. $RSSI_{up}$ denotes the uplink RSSI vector, and $RSSI_{down}$ denotes the downlink RSSI vector.

The SAE maps the uplink and downlink for the physical layer parameters to the hidden layer and replaces it with a new feature representation. Xavier initialization method is used to automatically adjust the input node and output node to the most appropriate distribution. The learning rate is considered to be 0.001. Epoch is 20. And batch size is 128. In this paper, the squared error is used as a loss function to determine if the SAE still needs to continue training. When the value of the loss function no longer decrease, it is considered that SAE has completed the training. The loss function is shown in (2) as follows:

$$f(X', X) = \frac{1}{2}\sum_{\mathrm{i}=1}^{n}\left(X' - X\right)^2 \tag{2}$$

where X' refers to the physical layer link quality parameter reconstructed by the model, and X is the input physical layer link quality parameter, n is the number of data.

This paper uses Softplus as the "activation function" The Softplus expression is shown in (3) as follows:

$$f\ (x) = \log\left(1 + e^x\right) \tag{3}$$

The final estimation accuracy in this paper is used to determine the structure of the SAE model. The SAE1, the SAE2 and SAE3 adopt a 7-layer structure determined by experiments. Furthermore, the input layer has 60 nodes and the number of hidden nodes are 40, 25, 16, 14, 5, 1 respectively. The structure is shown in Figure 4.

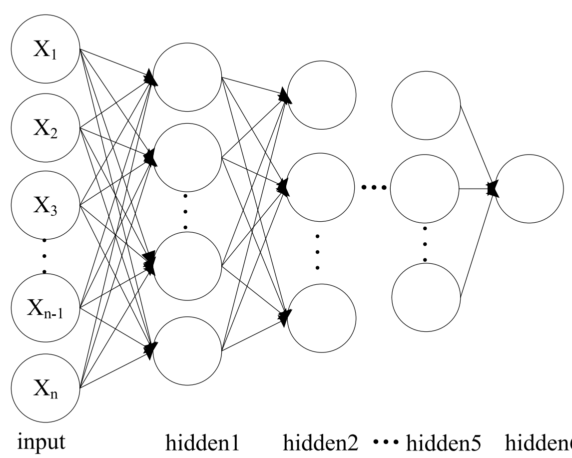


FIGURE 4. The structure of the SAE1, SAE2 and SAE3.

Experiments show that the SAE4 has a good performance when the number of hidden layer is 1 or 2. When the number of hidden nodes is 5 in the indoor corridor, the accuracy is the best. When the number of hidden nodes is 1 in the grove and parking lot scene, the accuracy is the best. When the number of hidden nodes is 2 in the road scene, the accuracy is the best. So, considering the complexity of the model, SAE4 adopts a

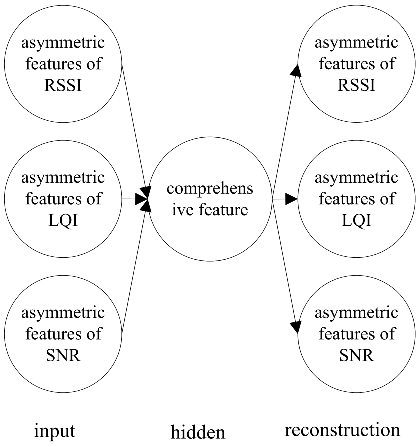


FIGURE 5. The structure of the SAE4.

structure which only has one hidden layer with one node. The structure of SAE4 is shown in Figure 5.

G. SVC CLASSIFICATION

The SVC is a process of nonlinear mapping of low-dimensional input space to high-dimensional attribute space [19], [20]. The SVC is often applied to multiple classification problems, such as text recognition, material classification, and spectral analysis. The SVC belongs to the supervised learning process. The link quality grade according to the PRR is used as the class label, and the asymmetric feature extracted by the SAE is used as the input. Seventy percent of the data set is used for training, and 30% is used for model testing. The link quality estimation process based on the SAE is shown in Algorithm 2. Furthermore, in this paper, the link quality is subdivided into five grades according to the PRR, as shown in Table 1:

Algorithm 2 LQE-SAE Algorithm

Input: data set D, number of input nodes n_{input}, number of nodes in each hidden layer n_{hidden}, activation function f, learning rate r

Output: The grade of link quality

Step 1: Load the dataset D that have been processed with missing values, and create four sets of $\{\mathrm{SNR_{up}}, \mathrm{SNR_{down}}\}$, $\{\mathrm{LQI_{up}}, \mathrm{LQI_{down}}\}$, $\{\mathrm{RSSI_{up}}, \mathrm{RSSI_{down}}\}$ and $\{\mathrm{PRR}\}$

Step 2: Convert the $\{\mathrm{SNR_{up}}, \mathrm{SNR_{down}}\}$, $\{\mathrm{LQI_{up}}, \mathrm{LQI_{down}}\}$ and $\{\mathrm{RSSI_{up}}, \mathrm{RSSI_{down}}\}$ into 60-dimension data by pre-processing, splicing, and converting $\{\mathrm{PRR}\}$ into a label set representing the link quality grade $\{\mathrm{label}\}$ according to the rules in Table 1

Step 3: Randomly sample data $\{\mathrm{SNR_{up}}, \mathrm{SNR_{down}}\}$, $\{\mathrm{LQI_{up}}, \mathrm{LQI_{down}}\}$, $\{\mathrm{RSSI_{up}}, \mathrm{RSSI_{down}}\}$ and $\{\mathrm{label}\}$. Among them, 70% of the data consists of $\{\mathrm{train}\}$ for training, and 30% of the data consists of $\{\mathrm{test}\}$ for model testing

Step 4: Normalize $\{\mathrm{train}\}$ and $\{\mathrm{test}\}$

Step 5: Define the structure of the SAE model. n_{input}, n_{hidden}, f, r were included in hyperparameter settings

Step 6: Extract features for three data sets $\{\mathrm{train}\}$, which are used to evaluate parameters using a well-defined SAE, and determine the optimal model according to the f

Step 7: Redefine the SAE4 model; use the three higher-order features extracted by the optimal model in Step 6 as its input; and, finally, output a comprehensive feature

Step 8: Take the comprehensive feature as the input of the SVC, and train the training set of $\{\mathrm{label}\}$ as the label in the last layer of the coding phase of the SAE4

Step 9: Test in $\{\mathrm{test}\}$ and compare the test results to the results tested in $\{\mathrm{label}\}$

end

TABLE 1. The classification of link quality based on PRR.

The grade of link quality	The range of PRR	Description
1	0.9<PRR<1	Very good
2	0.75<PRR<0.9	Good
3	0.6<PRR<0.75	Common
4	0.2<PRR<0.6	Bad
5	0<PRR<0.2	Very bad

IV. EXPERIMENTAL DESIGN AND ANALYSIS

To verify the validity of the model, the link quality data are acquired from multiple application scenarios. This paper uses the TelosB node that is created by CrossBow to send and receive data, and uses the wireless sensor network link quality testbed (WSNs-LQT) developed by the lab to collect link quality information, including the RSSI, SNR and LQI for the uplink and downlink. For example, RSSI can be obtained by formula (4).

$$P = RSSI\_VAL + RSSI\_OFFSET(dBm) \tag{4}$$

where P is the RSSI in this paper, $RSSI_VAL$ is the value obtained through a register. $RSSI_OFFSET$ is an empirical value, generally -45 dBm. LQI is the chip error rate of the$DSSS - OQPSK$ modulation of the CC2420 chip. And SNR can be obtained by formula (5).

$$SNR = \frac{P_{signal}}{P_{noise}}(dB) \tag{5}$$

where P_{signal} and P_{noise} represent useful signal power and noise power, respectively. Used to dynamically display link quality information among nodes, the test platform WSNs-LQT is shown in Figure 6. The collected link quality information is analyzed and processed using TensorFlow and Jupyter Notebook.

A. EXPERIMENTAL SCENARIOS

Four experimental scenarios are designed: indoor corridors, grove, parking lots and road. Four experimental scenarios are shown in Figure 7(a), Figure 7(b), Figure 7(c), and Figure 7(d). As shown in each of these figures, small star-shaped WSNs networks are deployed in each selected

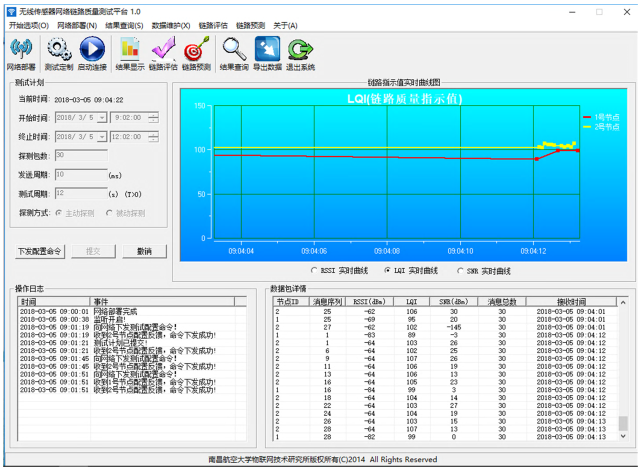


FIGURE 6. **The test platform of wireless sensor network link quality.**

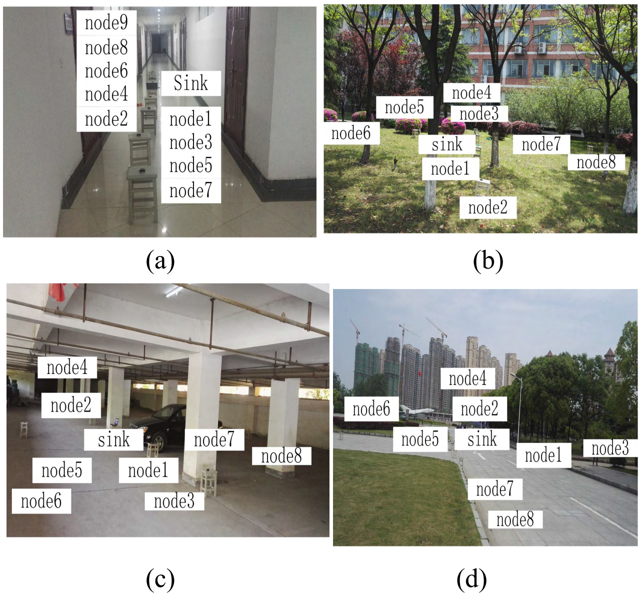


FIGURE 7. **Various of experimental scenarios. (a) Interior corridors. (b) Grove. (c) Parking lots. (d) Road.**

TABLE 2. **The settings of experimental parameter.**

Parameter attribute	Parameter value
Transmit power	0 dBm
Channel	26
Number of probes	30
Detection method	Active detection
Packet rate	10 pcs/s
Test period	10 s

experimental scenario. The experimental data collection process parameter settings are shown in Table 2:

B. EXPERIMENTAL ANALYSIS

The experiment is mainly organized as follow: firstly, the validity of LQE-SAE is verified by the accuracy of the model in different experimental scenarios, and compared with LQE-SVC, LQE-ELM and LQE-WNN method. In LQE-SVC, the kernel function uses RBF and the penalty term is 1. In LQE-ELM, the activation function uses the sigmoid function, and the number of hidden layer nodes is 320. And in LQE-WNN, the learning rates are 0.001 and 0.0001, the number of iterations is 10000, and the Morlet wavelet function is defined by the formula (6).

$$\varphi(x) = e^{-x^2/2}\cos(5x) \tag{6}$$

Then, the stability of LQE-SAE is analyzed. Lastly, the rationality of the missing value processing method is verified by experiments.

1) VERIFICATION OF THE ACCURACY OF THE ESTIMATION MODEL

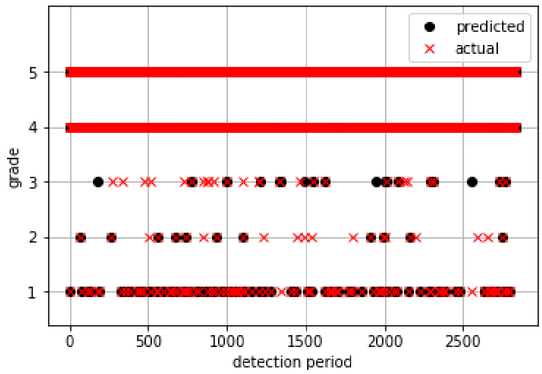
The accuracy of the estimation model refers to the classification of the correct sample divided by the total number of samples. Among them, the correct classification sample refers to the same link quality grade obtained by the proposed model and the true link quality grade.

Figure 8(a) shows the experimental results in the interior corridors scene. The accuracy of the LQE-SAE is 95.76%. Figure 8(b) shows the experimental results under the grove scenario. The accuracy is 94.85%. Figure 8(c) shows the experimental results in the parking lots scenario. The accuracy of the LQE-SAE is 99.49%. Figure 8(d) shows the experimental results in a road scenario. The accuracy of the model reaches 99.85%. It is seen from the figure that the estimation result based on the LQE-SAE is basically the same as the real link quality grade. It is also observed that there are very few deviations between the estimation results of the 1st grade and the real link quality grade. The 4th and 5th grades are completely coincident, and some deviations occur during the estimation of the 2nd and 3rd grades.

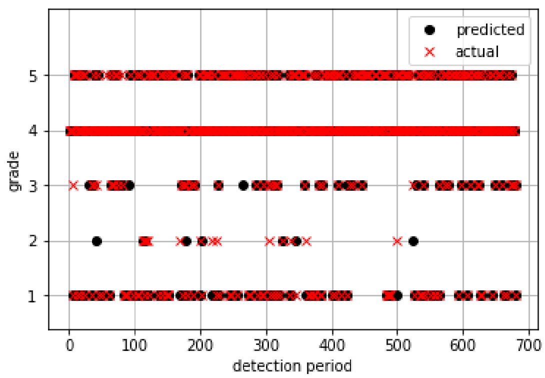
Figure 9 shows the top 60 estimation results in the interior corridors experimental scenario. As shown in Figure 9, the real link quality grade is the 4th grade in the seventh estimation result of the first 60 results. However, the result of LQE-SAE is grade 5. It has one grade of error. It is shown from the figure that same grade of error also appears in the 37th estimation result. The LQE-SAE mistakes grade 5 for grade 4.

Figure 10 also selects 60 estimation results for comparison. The parking lots is selected from the 60th estimation result to the 119th estimation result. From the figure, it is shown that the 60 estimations are basically completely coincident; that is, the estimation results of the LQE-SAE are exactly the same as the real link quality grade. However, in the 24th estimation result in the figure, there is one grade deviation, i.e., the LQE-SAE mistakes grade 4 for grade 5.

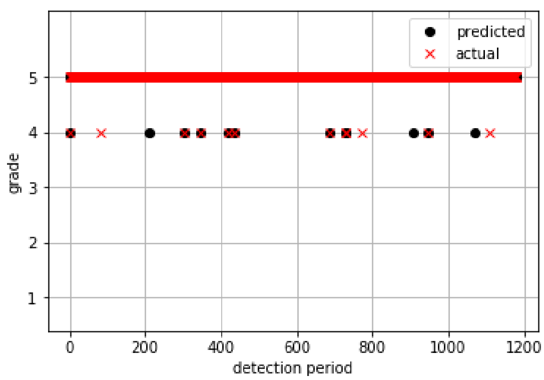
In the grove experiment scene, 60 estimation results were also selected for analysis. From Figure 11, it is seen that the estimation results are basically coincident with the real link quality grade, but there is a very small part of one grade of deviation. As shown in the sixth estimation result



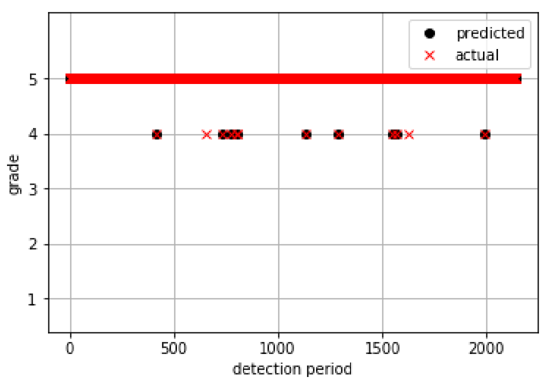
(a)



(b)



(c)



(d)

FIGURE 8. The results of link quality for estimation in the four scenarios. (a) Interior corridors. (b) Grove. (c) Parking lots. (d) Road.

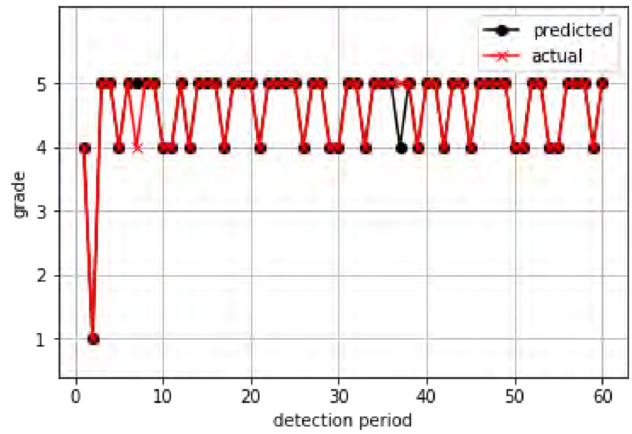


FIGURE 9. Sixty results of estimation in interior corridors.

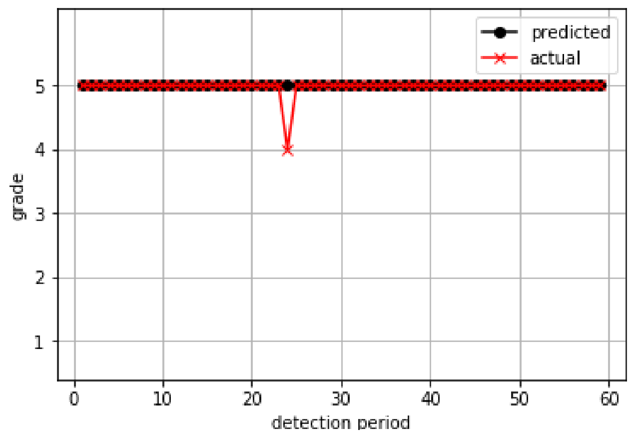


FIGURE 10. Sixty results of estimation in parking lots.

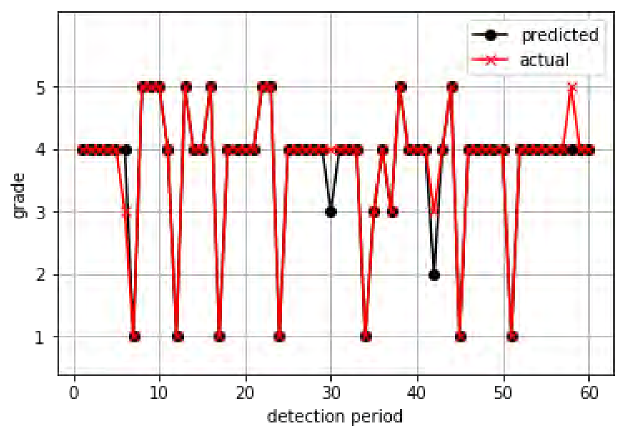


FIGURE 11. 60 results of estimation in grove.

in Figure 11, the true 3rd link quality grade is evaluated as the 4th grade, and one grade error occurs. Similarly, in the 30th evaluation, the true link quality grade of the 4th grade is evaluated as the 3rd grade. In the 42nd estimation, the real link quality of the 3rd grade was evaluated as the 2nd grade. In the 58th evaluation, the real link quality of the 5th grade was evaluated as the 4th grade. In general, the maximum

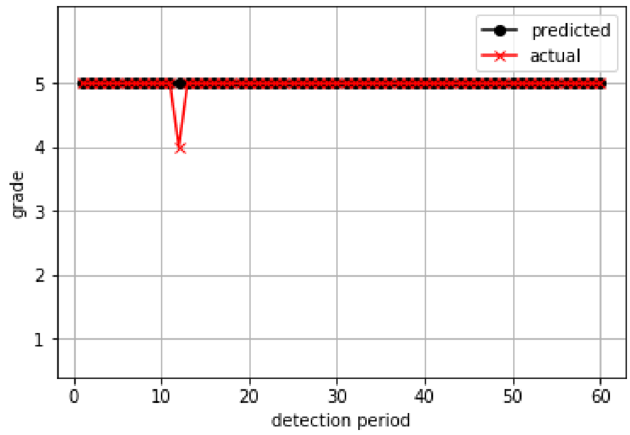


FIGURE 12. Sixty results of estimation in road.

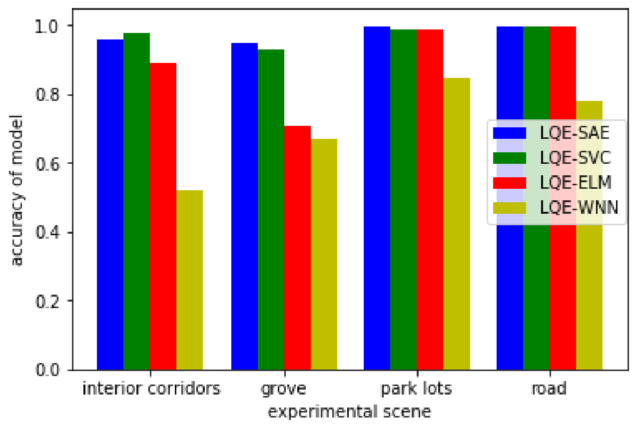


FIGURE 13. Comparison with LQE-SVC, LQE-ELM and LQE-WNN models.

grade deviation in the grove experiment scenario is within one grade of levels.

In Figure 12, the 60 results selected in the experimental scene of the road are seen to be completely coincident. Only in the 11th estimation result, the real link quality is 4th grade, but evaluated as the 5th grade.

Through the analysis of four scenarios, the LQE-SAE has higher accuracy in the link quality estimation. If there is an error, between most of the real link quality grade and the evaluated link quality grade, there is only one grade difference, and there is no large deviation.

Figure 13 is a comparison of the accuracy among LQE-SAE, the LQE-SVC, the LQE-ELM [21] and the LQE-WNN [14] in four scenarios. As shown in figure, our model is superior to the LQE-ELM and LQE-WNN in the four scenarios. It is also seen from the figure that our model is superior to the LQE-SVC in the grove, parking lots, and road scenarios. Due to the interference of scenes, such as grove and road, the sample is in the 5th grade; that is, the PRR is less than 0.2. It is also seen that the LQE-SAE has better performance for the very poor link environment.

Generally, deep learning may increase the complexity of the algorithm while ensuring accuracy. However, LQE-SAE is pre-trained offline with a large number of link samples.

TABLE 3. The analysis of time complexity.

The estimation method of link quality	Running time(s)
LQE-SAE	0.005
LQE-SVC	0.003
LQE-ELM	0.003
LQE-WNN	0.0175

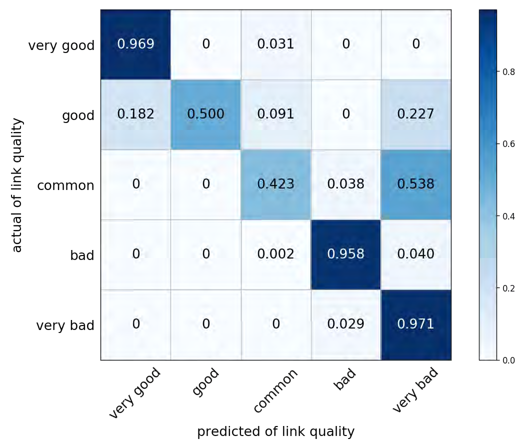


FIGURE 14. Confusion matrix in interior corridors scenarios.

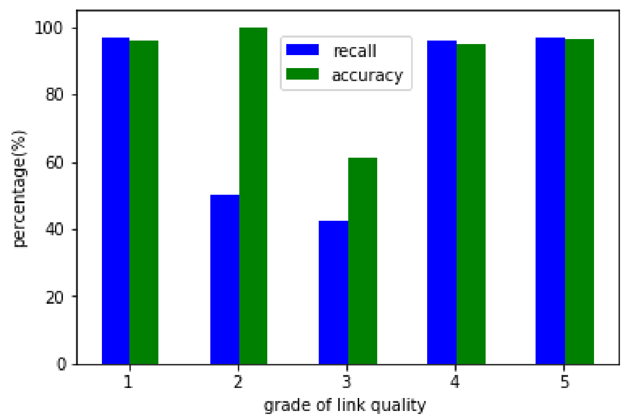


FIGURE 15. Recall rates and accuracy in interior corridors scenarios.

In practical applications, it will use the already trained model. Hence using test data to analyze the running time. Under the same conditions, as shown in Table 3, the running time of LQE-SAE is 0.005s. The running time of the LQE-SVC and the LQE-ELM is 0.003s. Finally, the running time of LQE-WNN is 0.0175s. Therefore, the method proposed in this paper has the same order of magnitude compared to the traditional machine learning method.

The confusion matrix, as shown in Figure 14, is used to calculate the accuracy and recall rate of the model for each link quality grade. The results are shown in Figure 15.

As shown in Figure 15, the LQE-SAE has excellent performance in both the accuracy and recall rate, especially in the 1st, 4th and 5th grades. That is, the LQE-SAE is outstanding in the case of very good link quality, bad link quality, and very

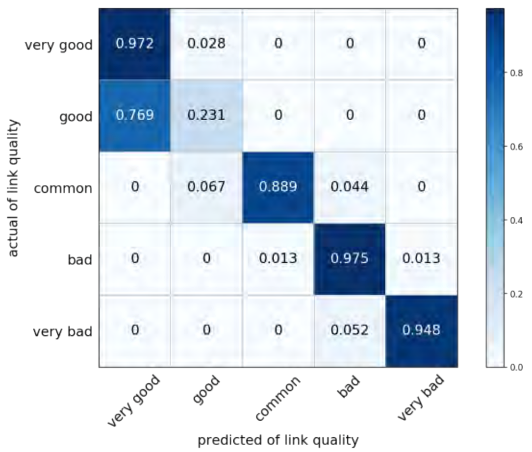


FIGURE 16. **Confusion matrix in grove scenarios.**

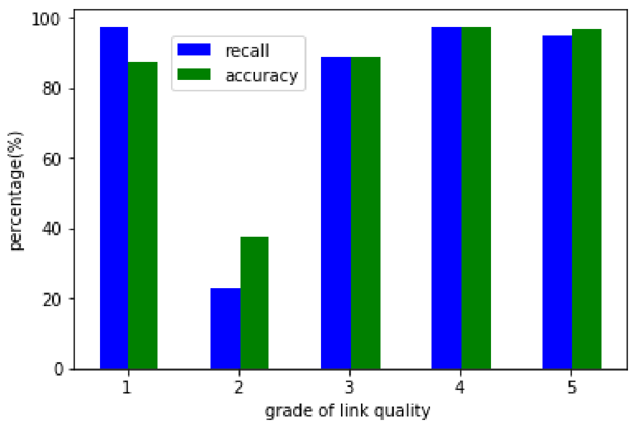


FIGURE 17. **Recall rates and accuracy in grove scenarios.**

bad link quality respectively. In the case of good link quality, the accuracy is up to 100%.

As shown in Figure 16 and Figure 17, the LQE-SAE still has excellent performance in terms of accuracy and recalls in the grove scene, especially in the 1st, 3rd, 4th and 5th grades. That is, the model is outstanding in four cases: the link quality is very good, common, bad, and very bad.

As shown in Figure 18, the recall rate of the 4th grade in the parking lots scene is 0.75, and the accuracy rate is 0.75. The recall rate of the 5th grade is 0.998, and the accuracy rate is 0.75. The LQE-SAE still performs well in the 4th and 5th grades. Due to the large interference of the parking lots, the overall link quality is poor. The test sets used in the experimental results of this paper randomly take 70% of the original samples as training samples and the remaining samples as test samples. Therefore, only the 4th and 5th grade samples are included in the test results.

As shown in Figure 19, there are only samples of the 4th and 5th grades in the road scene. The 4th grade has a recall rate of 0.75 and an accuracy rate of 1. The 5th grade has a recall rate of 1, with an accuracy of 0.998. The 4th and 5th grades in the road scene perform well. The 4th grade of accuracy is as high as 100%, and the 5th grade of recall rate is as high as 100% as well.

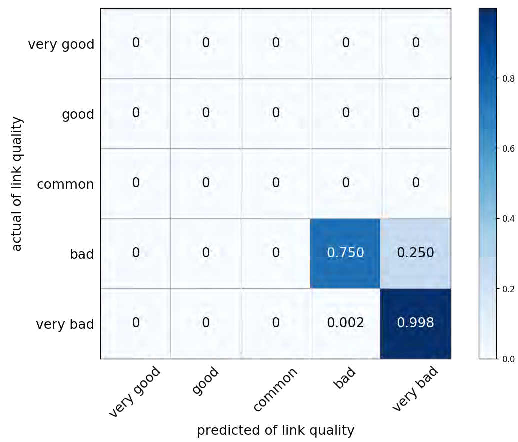


FIGURE 18. **Confusion matrix in parking lots scenarios.**

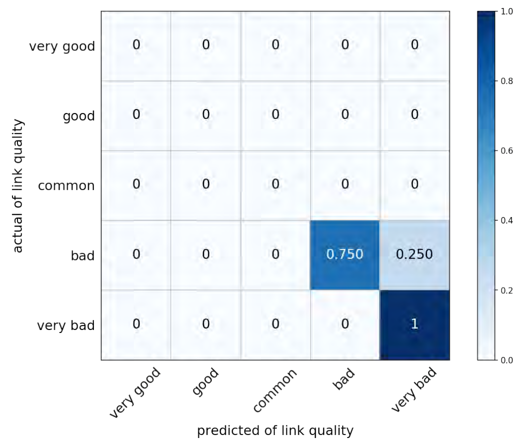


FIGURE 19. **Confusion matrix in road scenarios.**

Through the LQE-SAE, the recall rate and accuracy analysis in the four scenarios show that the whole model has good performance in all grades of estimation, and in the case of misclassification, most of the misclassifications appear in an adjacent grade, that is, one grade deviation.

C. THE STABILITY OF ESTIMATION MODEL

The variable a_i is the link quality at the current time, and a_{i+1} is the link quality at the next moment. The stability of the estimation model is defined as estimation keeping unchanged when giving a certain minor interference. If the link quality undergoes mutations of n grades as $|a_{i+1} - a_i| = n$, the following consecutive moments of a_{i+1} are equal to the link quality of a_i; that is, only the current moment is changed. While the estimation model does not change with the current moment and maintains the prior estimation result, the model does not perform real-time reaction to this kind of mutation, which is called n-grade stability. Among the n grades, the n is 1, 2, 3 or 4. For the stability grade, the model is considered the most stable if it is stable at the 1st grade.

To verify the stability of the LQE-SAE, this paper uses other nodes to interfere with the target node. For a single node to interfere with the target node, the interference size is

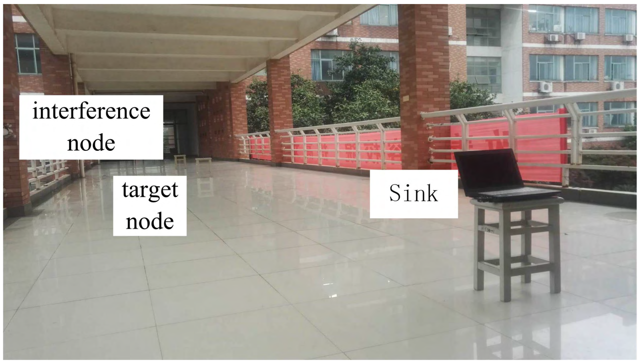


FIGURE 20. The interference experiment in corridors.

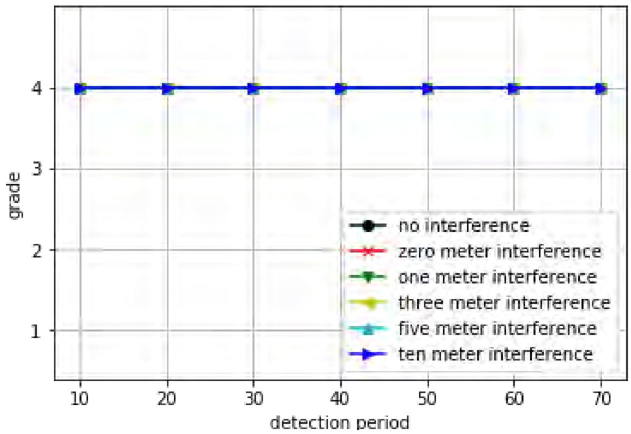


FIGURE 21. The estimation of Interference under different distances.

adjusted by the distance between the interference node and the target node; the interference node and the target node are, respectively, set in the same position, one meter apart, three meters apart, five meters apart and ten meters apart. The experimental scenario is shown in Figure 20.

For convenience of observation, display multiples of ten times. As shown in Figure 21, the results of the estimation of the link nodes at the same location and the target nodes of one meter, three meters, five meters, and ten meters are exactly the same as the results of the model estimation without interference. That is, under these five conditions, the interference is not enough to destroy the stability of the model.

This paper analyzes the estimation of interference separately. Figure 22 shows the link quality of the target node when the interference node is one meter away from the target node. In the selected 68 sample observations, it is seen that in the 50th estimation result of the real link, the link quality is changed from the fourth-grade PRR of 0.32 to the fifth-grade PRR of 0.16. In addition, the last few moments remain the same as the 50th real link quality, but the estimation model does not change, and the 4th grade estimation is maintained; that is, if the link PRR differs by 0.16, the model is embodied in the first-grade stability range in the definition, the same situation occurs in the 57th evaluation, and the model is also in the 1st-grade stable range. Figure 23 shows the link quality estimation of the target node when the interference

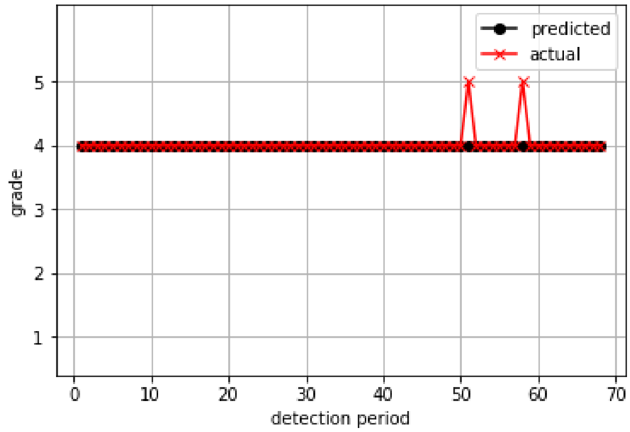


FIGURE 22. The estimation of Interference under different distances.

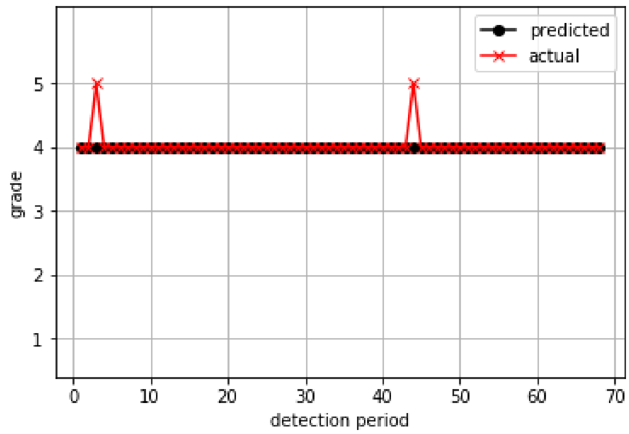


FIGURE 23. The estimation of interference with five-meter distances.

node is five meters away from the target node. In the selected 68 sample observations, it is seen that in the third estimation result of the real link, the link quality is changed from the fourth-grade PRR of 0.24 to the fifth-grade PRR of 0.16, but the estimation model does not change and maintains the 4th grade of evaluation; that is, if the link PRR differs by 0.08, the model reflects the first-grade stability range in the definition. In the 44th estimation result, the link PRR is abruptly changed from 0.32 to 0.16, and the model is also in the first-grade stable range when the link PRR differs by 0.16. Experiments show that the model has strong stability in the interference experiment.

D. ANALYSIS OF MISSING VALUE PROCESSING

The common missing value processing method used by previous works is the mean filling method. In this section, this paper compares the performance of the mean filling method with the proposed Zero-filling method.

The same experimental parameters are set to perform missing value processing experiments on the four scenarios. As shown in Figure 24, the Zero-filling and the mean filling method are compared in the interior corridors scene, the grove, the park and the road, respectively. In the interior

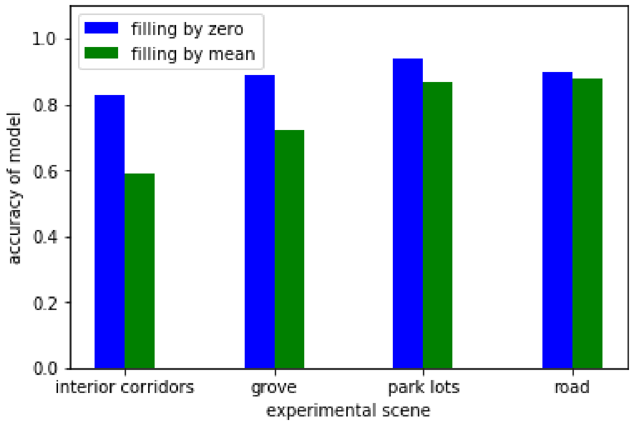


FIGURE 24. Comparison of filling methods.

corridors scene, it can be seen that the Zero-filling method achieves an accuracy of 0.83, while the mean filling method achieves 0.59. In the grove scene, it is seen that the proposed method has 0.89, and the mean filling has 0.72. Then, the comparison of these methods in the park scenario show that the Zero-filling method achieves 0.94, and the mean filling method achieves 0.87. Finally, the comparison made in the road scene shows that the accuracy of Zero-filling is slightly higher with accuracy of 0.9.

In summary, as it is seen from the results, the proposed Zero-filling achieves a significantly higher accuracy performance for filling the missing values than the mean filling method. After padding with a value of 0, there are many 0's in the input matrix which increase the sparsity. And the zero value is more appropriate in the physical sense.

V. CONCLUSION

In this paper, the asymmetry of the link is taken as the entry point, and the physical layer parameters of uplink and downlink are selected. Zero-filling method is used to fill the missing link information. The link quality grade is divided according to the PRR, and the LQE-SAE is used to evaluate link quality. The validity and stability of the model are verified by experiments in four scenarios. Experimental results in different scenarios have shown that the LQE-SAE has better accuracy than link estimation based on the SVC, ELM and WNN. And the experimental results have been proved that the LQE-SAE has better performance in the case of a poor link among the defined link quality grades. The main contributions of this paper are as follows: (1) The Zero-filling method is proposed to process the missing values in the detection period, thus this paper uses the original value for link quality assessment, instead of the mean of value and (2) deep learning is used to mine the asymmetry features between uplink and downlink and the deep features among parameters, and it improves the accuracy and stability of the estimation model, lastly (3) this paper defines stability. However, in the actual application, the model needs to be pre-trained through a large number of link samples.

REFERENCES

[1] K. Li, Z. Wen, and Z. Wang, "Optimised placement of wireless sensor networks by evolutionary algorithm," *Int. J. Comput. Sci. Eng.*, vol. 15, nos. 1–2, pp. 74–86, Jan. 2017.

[2] R. D. Gomes, D. V. Queiroz, A. C. L. Filho, I. E. Fonseca, and M. S. Alencar, "Real-time link quality estimation for industrial wireless sensor networks using dedicated nodes," *Ad Hoc Netw.*, vol. 59, pp. 116–133, May 2017.

[3] J. Shu, S. Liu, L. Liu, L. Zhan, and G. Hu, "Research on link quality estimation mechanism for wireless sensor networks based on support vector machine," *Chin. J. Electron.*, vol. 26, no. 2, pp. 377–384, Mar. 2017.

[4] N. Baccour, A. Koubâa, L. Mottola, M. A. Zúñiga, H. Youssef, C. A. Boano, M. Alves, "Radio link quality estimation in wireless sensor networks: A survey," *ACM Trans. Sensor Netw.*, vol. 8, no. 4, pp. 1–33, Sep. 2012.

[5] C. U. Bas and S. C. Ergen, "Spatio-temporal characteristics of link quality in wireless sensor networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Paris, France, Apr. 2012, pp. 1152–1157.

[6] L. Xu, J. Wang, and H. Zhang, T. A. Gulliver, "Performance analysis of IAF relaying mobile D2D cooperative networks," *J. Franklin Inst.*, vol. 354, no. 2, pp. 902–916, Jan. 2017.

[7] D. Estrin, "Tutorial 'wireless sensor network' part IV: Sensor network protocols," in *Proc. 8th Annu. Int. Conf. Mobile Comput. Netw.*, Atlanta, GA, USA, 2002, pp. 140–147.

[8] J. Zhao, D.-A. Li, H. Wen, and Q. Tang, "Load balanced and efficient data collection protocol for wireless sensor networks," *Int. J. High Perform. Comput. Netw.*, vol. 10, no. 6, pp. 463–473, 2017.

[9] T. Jayasri and M. Hemalatha, "Link quality estimation for adaptive data streaming in WSN," *Wireless Pers. Commun.*, vol. 94, no. 3, pp. 1543–1562, Jun. 2017.

[10] Z. Ansar and W. Dargie, "Adaptive burst transmission scheme for WSNs," in *Proc. Int. Conf. Comput. Commun. Netw.*, Vancouver, BC, Canada, Jul./Aug. 2017, pp. 1–7.

[11] J. Lu, Y. P. Zhu, and Z. Q. Xu, "A reliable wireless sensor network routing method for power transmission line monitoring," *Power Syst. Technol.*, vol. 41, no. 2, pp. 644–650, Feb. 2017.

[12] M. Deb, S. Roy, B. Saha, P. Das, and M. Das, "Designing a new link quality estimator for sensor nodes by combining available estimators," in *Proc. IEEE 7th Int. Advance Comput. Conf. (IACC)*, Hyderabad, India, pp. 179–183, Jan. 2017.

[13] S. Rekik, N. Baccour, M. Jmaiel, and K. Drira, "Holistic link quality estimation-based routing metric for RPL networks in smart grids," in *Proc. IEEE 27th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Valencia, Spain, Sep. 2016, pp. 1–6.

[14] W. Sun, W. Lu, Q. Li, L. Chen, D. Mu, and X. Yuan, "WNN-LQE: Wavelet-neural-network-based link quality estimation for smart grid WSNs," *IEEE Access*, vol. 5, pp. 12788–12797, 2017.

[15] W. Sun, X. Yuan, J. Wang, Q. Li, L. Chen, and D. Mu, "End-to-end data delivery reliability model for estimating and optimizing the link quality of industrial WSNs," *IEEE Trans. Autom. Sci. Eng.*, vol. 15, no. 3, pp. 1127–1137, Jul. 2017.

[16] C. J. Lowrance, "An adaptable fuzzy-based model for predicting link quality in robot networks," Univ. Louisville, Louisville, KY, USA, Tech. Rep., 2016.

[17] S. W. Fu, C. Y. Shih, and Y. M. Jiang, "RADIUS: A system for detecting anomalous link quality degradation in wireless sensor networks," Cornell Univ., New York, NY, USA, pp. 1–13,Jan. 2017.

[18] M. Kulin, E. De Poorter, T. Kazaz, and I. Moerman, "Towards a cognitive MAC layer: Predicting the MAC-level performance in dynamic WSN using Machine learning," in *Proc. 14th Int. Conf. Embedded Wireless Syst. Netw. (EWSN)*, Uppsala, Sweden, Feb. 2017, pp. 214–215.

[19] H. Y. Li and B. Zhong, "Modifying kernel function for support vector machines classifier," *Comput. Eng. Appl.*, vol. 45, no. 24, pp. 53–55, 2009.

[20] W. Fang, H. Jiang, J. Li, W. Miao, K. Ma, and W. Xiao, "An intelligent production fluctuation monitoring system for giant oilfield development," *Int. J. Embedded Syst.*, vol. 9, no. 1, pp. 36–44, 2017.

[21] L. L. Liu and J. B. Xu, "A link quality estimation method for WSNs based on extreme learning machine," *J. Beijing, China Univ. Posts Telecommun.*, vol. 41, no. 1, pp. 134–138, Feb. 2018.

XIONGHUI LUO born in Nanchang, Jiangxi, China, in 1993. He is currently pursuing the master's degree with Nanchang Hangkong University. His research interest includes wireless sensor networks.

LINLAN LIU born in 1968. She received the B.Sc. degree in computer application from the National University of Defense Technology. She is currently a Professor with Nanchang Hangkong University. She is a member of CCF. Her research interests include wireless sensor networks, software engineering, and distributed systems.

JIAN SHU born in 1964. He received the M.Sc. degree in computer networks from Northwestern Polytechnical University. He is currently a Professor with Nanchang Hangkong University. His research interests include wireless sensor networks, embedded systems, and software engineering. He is a Senior Member of CCF.

MANAR AL-KALI received the M.S. and Ph.D. degrees majoring in communication and information systems from the Department of Telecommunication and Information Engineering, Huazhong University of Science and Technology, China, in 2011 and 2015, respectively. He is currently a Researcher and a Lecturer with the School of Software, Nanchang Hangkong University. His research interests include wireless sensor networks, cognitive radio, cooperative communications, MIMO OFDM, and information theory.