IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received October 5, 2019, accepted October 21, 2019, date of publication October 25, 2019, date of current version November 5, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2949612

A Link Quality Prediction Method for Wireless Sensor Networks Based on XGBoost

YI FENG[1], LINLAN LIU[1], AND JIAN SHU[2]

[1]School of Information Engineering, Nanchang Hangkong University, Nanchang 330063, China
[2]School of Software, Nanchang Hangkong University, Nanchang 330063, China

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

This work was supported by the National Natural Science Foundation of China under Grant 61962037 and Grant 61762065, in part by the Natural Science Foundation of Jiangxi Province under Grant 20171ACB20018, Grant 20171BBH80022, and Grant 20181BAB202015, and in part by the Innovation Foundation for Postgraduate Student of Jiangxi Province under Grant YC2018023.

ABSTRACT Link quality is an important factor for nodes selecting communication links in wireless sensor networks. Effective link quality prediction helps to select high quality links for communication, so as to improve stability of communication. We propose the improved fuzzy C-means clustering algorithm (SUBXBFCM) and use it to adaptively divide the link quality grades according to the packet reception rate. The Pearson correlation coefficient is employed to analyse the correlation between the hardware parameters and packet reception rate. The averages of the received signal strength indicator, link quality indicator and the signal to noise ratio are selected as the inputs of the link quality estimation model based on the XGBoost (XGB_LQE). The XGB_LQE is constructed to estimate the current link quality grade, which takes the classification advantages of XGBoost. Based on the estimated results of the XGB_LQE, the link quality prediction model (XGB_LQP) is constructed by using the XGBoost regression algorithm, which can predict the link quality grade at the next moment with historical link quality information. Experiment results in single-hop scenarios of square, laboratory, and grove show that the SUBXBFCM algorithm is effective at dividing the link quality grades compared with the normal division methods. Compared with link quality prediction methods based on the Support Vector Regression and 4C, XGB_LQP makes better predictions in single-hop wireless sensor networks.

INDEX TERMS Wireless sensor networks, link quality prediction, XGBoost, improved fuzzy C-means algorithm.

I. INTRODUCTION

Wireless sensor networks (WSNs) are multi-hop self-organizing networks that are formed by wireless communication, and they consist of large numbers of inexpensive micro sensor nodes that are deployed in a monitoring area [1]. Their purposes are to collaboratively perceive, collect and process the information of the area and then send it to the observer. Sensors, perceptual objects, and observers form the three elements of a wireless sensor network. WSNs currently have broad application prospects in the fields of smart home, urban transportation and so on.

In WSNs, nodes communicate by radio frequencies, and links have characteristics such as asymmetry, irregularity, and directionality [2], [3]. The WSN links are susceptible to multipath effect, loss and adjacent channel interference, which result in unreliable links, low channel quality and frequent topological changes [4]. The quality of the communication links has a significant impact on the performance of the wireless sensor network, such as the life cycle of the network, the throughput of the network and the reliability of transmissions [5]. Selecting a link with poor link quality for communications is likely to cause a large amount of packet loss. The retransmission due to packet loss recovery is the main cause of the increase of network energy consumption [6].

Effective link quality prediction helps to select good links for data transmissions, so as to improve network throughput and to maintain normal network operations and topological stability of networks, thereby improving the performance of wireless sensor networks [7]. It can be seen that the link quality prediction for WSNs is of great significance. To provide support for the upper layer routing protocol

The associate editor coordinating the review of this manuscript and approving it for publication was Kai Yang.

(the MAC protocol of the data link layer, the routing protocol of the network layer, the topological control of the network management layer, etc.), a novel link quality prediction method for WSNs based on XGBoost is proposed in this paper.

The key contributions of this paper are as follows:

1) An adaptive method is proposed for dividing link quality grades based on the improved fuzzy C-means algorithm (Subtraction Xie-Beni Fuzzy C-means Algorithm, SUBXBFCM). Due to the different distribution of link quality parameters in different scenarios, the current methods for dividing the link quality grades do not have wide applicability. We propose the SUBXBFCM and use it to adaptively divide the link quality grades in different scenarios, which does not need to determine the initial cluster centres and the optimal number of clusters artificially.

2) An approach reducing the influence of the unbalanced link quality samples on estimation models is proposed. In the construction of link quality estimation model XGB_LQE, the maximum step of each tree's weight change is controlled to reduce the influence of the unbalanced link quality samples on the XGB_LQE, and further to improve the accuracy of link quality estimation.

3) A link quality prediction method based on XGBoost regression algorithm is proposed. According to the results of the XGB_LQE, the XGB_LQP based on the XGBoost regression algorithm is constructed to predict the link quality grade at the next moment. The experimental results show that the proposed link quality prediction method has better prediction accuracy compared with the 4C prediction model and the SVR prediction model in single-hop WSNs.

The remainder of this paper is structured as follows. Section II introduces the latest related research on link quality prediction and XGBoost. Section III addresses the method of division of link quality grades. The link quality estimation approach for WSNs is presented in Section IV. Then, we introduce the link quality prediction for WSNs in Section V. Section VI verifies the performance of our method. We conclude in Section VII.

II. RELATED WORK

A. LINK QUALITY PREDICTION METHODS

Link quality prediction methods for WSNs fall into the following categories– link quality predictions based on link characteristics, statistics, and machine learning.

1) LINK CHARACTERISTIC-BASED PREDICTION METHODS

Link quality prediction methods based on link characteristics predict the link quality by analysing the relationships between the information in physical layer and information in link layer. The information in physical layer mainly includes the received signal strength indicator (RSSI), the link quality indicator (LQI) and the signal to noise ratio (SNR). The link layer information mainly includes the packet reception rate (PRR).

Baccour *et al.* [8] propose a link quality evaluation method F-LQE based on fuzzy logic, which combines four link quality parameters, including packet delivery, asymmetry, stability, and channel quality. The membership degree of each parameter is calculated by the membership function. The experimental results show that the method has good stability but poor agility. Ansar and Dargie [9] propose a protocol that estimates the duration of good or bad links from received ACK packets and shares the link states with neighbouring nodes to improve the utilization of the shared channel. A packet is sent when the link is in good link cycle and is not sent when the link is in bad link cycle. Experiments show that the proposed prediction method can effectively improve the throughput of the network and reduce the energy consumption during transmissions. By analyzing the influence of the asynchronous nature of widely adopted duty-cycled radio control on link quality estimation, Liu *et al.* [10] design a distributed frame counter meter to count the successfully decoded wake-up frames and the corrupted ones, separately. The meter is applied to link quality estimation models Four-Bit. Experiments in indoor and outdoor scenarios show that link quality estimation model has been improved. Aswale and Ghorpade [11] propose a triangle link quality metric and minimum inter-path interference-based geographic multipath routing (TIGMR). By using the averages of the SNR and LQI as the right angle, the value of the hypotenuse represents the triangle metric. The simulation results indicate that the TIGMR protocol optimizes the overall performance and improves the network lifetime compared with the state-of-the-art two-phase geographic forwarding (TPGF) and link quality and energy-aware routing (LQEAR) protocols. Sun *et al.* [12] propose a reliable link estimation model for end-to-end data transmission in WSNs applications for industrial monitoring and control. The model uses an alpha-stable distribution to accurately represent the background noise, and uses a modified log-normal path loss model to more realistically describe the RSS. The link quality is obtained by mapping relationship between the physical channel signal state and the PRR. Experimental results show that it has good estimation effect.

2) STATISTICS-BASED PREDICTION METHODS

Link quality prediction methods based on statistics predict the link quality by combining statistical methods to analyse the large amount of collected data. Fonseca *et al.* [13] propose an estimator 4-Bit, which combines physical layer, link layer and network layer information. The physical layer provides immediate information on the quality of the decoding of a packet. The link layer provides the ACK bit that is written to the transfer buffer when the packet is successfully received. The network layer provides the compare bit and the pin bit for the routing table. The experimental results show that significant improvements on cost and delivery ratio over the state of the art. From a multidimensional view of characterizing the wireless link quality, Guo *et al.* [14] present the fuzzy logic based link quality indicator (FLI), and implement a FLI-based wireless link quality estimator in Collection Tree Protocol (CTP). By comparing the performance between 4Bit-CTP and FLI-CTP, FLI-CTP reduces the average path length and topological changes in different network configurations. Sharma *et al.* [15] select a high-quality path with a minimum number of hops in a mobile ad hoc network based on node mobility and variable channel conditions during data transmissions. This approach improves the communication efficiency of mobile ad hoc networks by modifying the expected number of transmissions and adding the link loss variance to the prediction. Experiments show that the method acquires significant improvements in the throughput, packet transmission score, normalized routing and end-to-end delay. Qin *et al.* [16] propose a three-layer impulse response framework to characterize the fading effect of wireless links in industrial environments. Among them, an augmented Kalman-filter-based link quality estimator is designed to track both the specular and scattered power in the distribution parameter space with constant noise covariance matrices. Experiments from industrial sites show that it significantly increases the accuracy. Bildea *et al.* [17] divide the links quality into three categories which are good, middle and bad according to the PRR. The Gilbert-Elliot two-state Markov model is used to analyse the PRR. Based on this, the LQI is converted into two states according to the threshold. The Gibert-Elliot model is constructed to estimate the PRR according to LQI.

3) MACHINE LEARNING-BASED PREDICTION METHODS

Link quality prediction methods based on machine learning predict the link quality by utilizing pattern matching, supervised learning and other modelling technologies. Ancillotti *et al.* [18] propose a novel strategy for link quality monitoring called the RL-probe in the RPL, which accurately measures the link quality with minimal overhead and energy waste. The RL-probe leverages both synchronous and asynchronous monitoring schemes to maintain up-to-date information on the link quality and to promptly react to sudden topological changes. A reinforcement learning model is used to drive the monitoring procedures and minimize the overhead that is caused by active probing operations. The results demonstrate that the RL-probe helps to effectively improve packet loss rates, thus allowing nodes to promptly react to link quality variations and to link failures due to nodes' mobility. Shu *et al.* [19] propose a multi-class link quality estimation mechanism based on support vector machine. The RSSI and LQI are selected as the estimation parameters, and the link quality is divided into five grades according to the PRR. The model can accurately estimate the current link quality using a small number of probe packets. Lowrance *et al.* [20] propose a unique approach using fuzzy logic to infer the LQ based on the collective output from a series of offset c1assifiers and their posterior probabilities. The proposed model leverages machine learning to extract the underlying functional relationships between the input and output variables. Compared with the traditional linear regression approach, the results show that there are statistically significant improvements in three out of the six real-world indoor and outdoor environments where the robot is operated. Liu and Cerpa [21] propose 4C, which is a novel link estimator that applies the link quality prediction along with the link estimation. The authors utilize logistic regression to construct the prediction model, which combines the PRR and the information in physical layer, i.e., the RSSI, SNR and LQI, as inputs, and outputs the success probability of delivering the next packet. The experimental results show that the improvements in the order of 20% to 30% compared with 4Bit and STLE estimators in single and multiple sender experiments, with some cases improving performance by more than 45%.

Some of these methods, such as [17], did not consider that the distribution characteristics of the link quality parameters are different in different scenarios. The specific thresholds for dividing the link quality grades do not have wide applicability. Based on this, we propose a method for adaptively dividing link quality grades based on SUBXBFCM algorithm in different scenarios. Others, such as those presented in [9], [10] and [11] only use hardware parameters in physical layer which ignore the calibration errors of the hardware are easy to overestimate the link quality due to ignore the packet loss. Other works, such as [13] and [15], only use the software parameters in link layer which need to send many probe packets and are lack of real-time. Based on this, we combine hardware parameters and software parameters, and establish a mapping relationship between hardware parameters and link quality grades. Generally, the existing link quality estimation models, such as those in [19] and [21], ignore the unbalanced characteristics of link quality samples caused by external factors. In the construction of the link quality estimation model (XGB_LQE), we control the maximum step of each tree's weight change to reduce the influence of the unbalanced link quality samples on the link quality estimation model, so as to further improve the accuracy of link quality estimation. The link quality prediction method proposed in this paper belongs to the machine learning-based method.

B. ANALYSIS AND APPLICATION OF XGBOOST

Extreme gradient boosting (XGBoost) is an ensemble algorithm based on trees or linear classifiers. It integrates several weak classifiers to form a strong classifier that has a better classification effect or regression effect [22]. Different from the conventional ensemble algorithms, the regularization term of XGBoost's objective function contains leaf nodes' weights and the tree depth, which can control the complexity of the model and prevent over-fitting. In addition, the objective function of XGBoost is approximated by a second-order Taylor expansion, which optimizes the objective function and guarantees the prediction accuracy [23]. XGBoost supports parallelism in the process of splitting, which greatly improves the efficiency of building trees. Chen *et al.* [24] use extreme gradient boosting (XGBoost) as the detection method in a software defined network (SDN)-based cloud to prevent the SDN controller from being attacked by distributed-denial-of-service (DDoS) that may paralyze the entire network. The XGBoost classifier uses the flow packet data set collected by TcpDump for the DDoS detection, and it is compared with other classifiers. The results validate that the method has higher accuracy, a lower false positive rate, fast speed and scalability. Due to the development of technologies, such as the Internet of Things, bio-sensing and data mining, Guo *et al.* [25] propose a machine learning-based physical fitness estimation model that is designed for wearable running monitoring for teenagers. This model includes four modules: PPG signal preprocessing, physiological data estimation, feature engineering and classification. The classification module utilizes the XGBoost algorithm for the classification of each teenager's physical fitness level. Compared with several existing ones, the proposed model is feasible and effective and can be expected to be a promising candidate solution for physical fitness evaluations by using smart wearable technologies in the future. Tian *et al.* [26] change the traditional combustion method, and transmit the combustion data that were collected by the smart sensors to a remote server in real time using GPRS data transmission technology and set up a database for data storage. They use the XGBoost algorithm to build the data analysis model, which is a multi-input and multi-output simulation model of cremation equipment. Then, they simulate the actual working conditions in cremation equipment and establish an intelligent monitoring system for cremation equipment. The experiments show that the oxygen content can be bound to a certain range by controlling certain parameters in advance based on the prediction results, which is of great significance for the efficient operation of equipment, energy savings and emission reductions. In addition, recent studies show that XGBoost has better analytical and prediction effects on the UEA and UCR time series datasets [27].

The link quality prediction for WSNs predicts the link quality at the next moment by using the historical link quality information. The essence of link quality prediction is time series prediction. Firstly, we use the SUBXBFCM algorithm to adaptively divide the link quality grades according to the PRR. The Pearson correlation coefficient is used to analyse the correlation between the hardware parameters and the PRR. Hardware parameters with high correlations are selected as the inputs of the link quality estimation model (XGB_LQE). Then, we construct the XGB_LQE based on the XGBoost classification algorithm to estimate the link quality grades at the current time. In the construct of the XGB_LQE, we control the maximum step of each tree's weight change to reduce the influence of the unbalanced link quality samples on the XGB_LQE. Based on the temporal characteristic of the link quality information, the link quality prediction model (XGB_LQP) is constructed based on the XGBoost regression algorithm to predict the link quality grade at the next moment. We use the mean square error to verify the performance of the XGB_LQP.

III. DIVISION OF LINK QUALITY GRADE

The transmission range is traditionally defined by using three regions that are measured by the PRR: the connected region, transitional region, and disconnected region. In some literatures, there are some different standards for dividing link quality grades according to the PRR. Bildea *et al.* [17] divide the link quality into three grades according to the PRR: good links, where $PRR \geq 80\%$; intermediate links, where $20\% \leq PRR < 80\%$; and bad links, where $0 < PRR < 20\%$. Luo *et al.* [28] divide the link quality into five grades according to the PRR: very good links, where $PRR \geq 90\%$; good links, where $75\% \leq PRR < 90\%$; common links, where $60\% \leq PRR < 75\%$; bad links, where $20\% \leq PRR < 60\%$, and very bad links, where $0 < PRR < 20\%$.The distribution characteristics of the link quality parameters are different in different scenarios. Hence, the specific thresholds for dividing the link quality grades do not have wide applicability. Based on this, we propose a method to adaptively divide the link quality grades based on the improved fuzzy C-means algorithm (SUBXBFCM).

The fuzzy C-means (FCM) algorithm is a fuzzy flexible partitioning method. It treats each cluster that is generated by the cluster as a fuzzy set. The membership degree of each sample is calculated to determine which cluster they belong to. The clustering algorithm is more accurate and flexible than the hard partition, such as those in literatures [17] and [28]. However, it has the following key issues that need to be solved.

1) The selection of the p value which plays an important role for noisy data and the smoothness of the membership degree has an impact on the clustering results.

2) The FCM is very sensitive to the selection of the initial cluster centres because its objective function is a nonlinear convex function. If the cluster centres are not properly selected, the FCM easily falls into the local optimum, which affects the global judgement of the algorithm.

3) The number of clusters in an actual situation is often unknown. The optimal number of clusters for different problems is different, which affects the clustering performance. Hence, it is necessary to select appropriate number of clusters.

To further improve the performance of the FCM algorithm, we use the fast and independent subtractive clustering algorithm [29] to find the initial cluster centres of the FCM to avoid falling into the local optimum and improve the quality and efficiency of the FCM. We select the fuzzy measure as $p = 2$ [30]. Considering the fuzzy membership and geometric structure of the data set, the validity index XB [31] that is based on the compactness and separation of clusters is adopted to adaptively determine the number of clusters. Hence, we can adaptively determine how many grades to divide the link quality into.

We use the PRR as the input of the SUBXBFCM algorithm. The purpose of clustering is to cluster the PRR to divide the link quality grades, and make the cluster objective function reach its minimum value. The objective function is shown in equation (1).

$$J(U, V) = \sum_{j=1}^{N} \sum_{i=1}^{c} (u_{ij})^2 (d_{ij})^2 \tag{1}$$

Here, u_{ij} denotes the membership degree of the *j-th* PRR belonging to the *i-th* link quality grade. Where the degree of membership indicates the degree to which the *j-th* PRR belongs to the *i-th* link quality grade. Among them, $0 \leq u_{ij} \leq 1$, $u_{ij} = 1$ indicates that the *j-th* PRR is completely subordinate to *i-th* grade, and $u_{ij} = 0$ indicates that the *j-th* PRR is not belong to the *i-th* grade. The *j-th* PRR belongs to the link quality grade corresponding to the maximum membership degree. d_{ij} denotes the distance from the *j-th* PRR to the *i-th* link quality grade, and it is calculated by using the Euclidean distance. p is the fuzzy measure. The process of dividing the link quality grades based on SUBXBFCM is shown as Algorithm 1.

Algorithm 1 link Quality Grades Division Algorithm

Input: Data set $D_{PRR} = \{PRR_1, PRR_2, \ldots, PRR_N\}$

Output: The optimal number of clusters c^*, link quality grades $D_{level} = \{Level_i\}, i = 1, 2, \ldots, N$

1: specify fuzzy measure $p = 2$, initial number of cluster $c = 2$, number of iteration $t = 0$, the convergence precision ε, set the optimal number of clusters to c^*, set the maximum number of clusters to $c_{\max}$;

2: **for** $c < c_{\max}$

3: increment $c(c = c + 1)$;

4: initialize cluster centres $V^{(0)} = \{v_1, v_2, \ldots v_c\}$ by subtractive clustering based on D_{PRR};

5: **for** $||U^{(t+1)} - U^{(t)}|| > \varepsilon$

6: increment $t(t = t + 1)$;

7: calculate the membership degree u_{ij} of PRR, and update the membership matrix $U^{(t+1)} = \{u_{ij}\}$; $u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{||PRR_i - v_j||}{||PRR_i - v_k||} \right)^2}$

8: calculate the cluster centre v_j of PRR, and update the cluster centre matrix $V^{(t+1)} = \{v_j\}$; $v_j = \frac{\sum_{i=1}^{N} (u_{ij})^2 \cdot PRR_i}{\sum_{i=1}^{N} (u_{ij})^2}$

9: calculate the value of the validity indicator XB, and output output XB and c;

10: **if** $XB <$ the lasted XB

11: set $c^* = c$, $U^* = U$, $V^* = V$;

12: **return** c^* and D_{level} according to U^*, V^*

By dividing link quality grades based on Algorithm 1, the link quality parameters in the same grade are consistent, and the link quality parameters are different between different grades. Hence, the link quality grades are determined, which reflect whether the links are good or bad.

IV. LINK QUALITY ESTIMATION

We adopt the Pearson correlation coefficient to analyse the correlation between the hardware parameters and the PRR. The hardware parameters, RSSI, LQI and SNR, are selected as the inputs of the link quality estimation model (XGB_LQE). Sample preprocessing is performed on the collected link quality parameters. The XGB_LQE is constructed based on the XGBoost classification algorithm to estimate the current link quality grade.

A. SELECTION OF LINK QUALITY PARAMETERS

Although estimating link quality using hardware parameters has the characteristic of being in real-time, calibration errors in the hardware and the problem of ignoring packet loss are likely to overestimate the link quality [7]. The link quality estimation method based on software parameters needs to send a large number of probe packets, which is not agile enough to reflect the link quality in real time [1]. By comprehensively considering the hardware parameters and software parameters, we study the correlation between the hardware parameters and the PRR and select the hardware parameters which have highest correlation with the PRR as the inputs of the link quality estimation model (XGB_LQE).

We use the self-developed wireless sensor networks link quality testbed (WSNs-LQT) [19] to collect the link quality parameters in square, laboratory, and grove scenarios. We use Pearson correlation coefficient to analyse the correlation between the averages of hardware parameters and PRR. The greater the absolute value of Pearson correlation coefficient, the stronger the correlation, and vice versa [25]. The correlation between hardware parameters and PRR in three scenarios is shown in Table 1.

TABLE 1. Correlation betweet the averages of hardware parameters and PRR.

	Square	Laboratory	Grove
Average of RSSI	0.396	0.656	0.356
Average of LQI	0.444	0.873	0.491
Average of SNR	0.398	0.781	0.687

It can be seen from Table 1 that the hardware parameters have significant correlations with the PRR. In the laboratory scenario, the hardware parameters have a strong correlation with the PRR, and the values of the Pearson correlation coefficients are all greater than 0.5. In the square scenario, the average of the LQI has the strongest correlation with the PRR compared with other parameters. In the grove scenario, the average of the SNR has the strongest correlation with the PRR compared with other parameters, and the Pearson correlation coefficient is 0.687.

In summary, in the three scenarios, the averages of the RSSI, LQI and SNR have correlation with the PRR. Hence, we select the averages of the RSSI, LQI and SNR as the inputs of the XGB_LQE.

B. SAMPLE PREPROCESSING

Substantial noise and numerous outliers in the sample space may affect the prediction accuracy. Excessive training sample sets will increase the complexity of solving the problem. Therefore, it is necessary to study the sample preprocessing method to distinguish the noisy samples from the samples. Extracting high-information parameters from valid samples will increase the prediction speed and ensure the prediction accuracy.

1) REMOVING ABNORMAL SAMPLES

The complete packet loss phenomenon occurs in the samples that are collected by the WSNs-LQT experimental platform, and the obtained link quality parameters are zero at this time. However, the corresponding PRR is not zero definitely when the hardware parameters are all zero in actual situations, especially when the RSSI is negative. In this paper, all the abnormal samples with zeros are excluded during sample processing.

2) DENOISING OF SAMPLES

Noisy samples exist among the raw samples due to external and internal interference. Noisy samples reduce the accuracy of the prediction model and do not fully reflect the real link quality. Kalman filtering is an ideal method for continuously changing systems, and it has the advantages of less memory consumption and fast speed [32]. Based on the above analysis, we use Kalman filtering to denoise the link quality parameters of the RSSI, LQI and SNR.

C. LINK QUALITY ESTIMATION MODEL

The link quality estimation is a process of determining the current link quality grade according to the selected link quality parameters. Different link quality parameters reflect different link characteristics. Link quality grades can intuitively reflect link quality. Hence, the link quality estimation problem is transformed into a classification problem. The link quality estimation model (XGB_LQE) is constructed based on XGBoost classification algorithm, which takes the classification advantages of XGBoost. The link quality estimation is implemented by the mapping relationship between hardware parameters and link quality grades. The link quality grades at the current time are obtained by using the XGB_LQE.

The averages of the RSSI, LQI and SNR after preprocessing are combined into a triple $\{\mu(RSSI_i), \mu(LQI_i), \mu(SNR_i)\}$, $(i = 1, 2, \ldots, N)$, which is the input of the XGB_LQE. The output is the link quality grade $\{Level_i\}$. The data set of the XGB_LQE is $D_{\text{eva}} = \{(X_i, Y_i)\}, (i = 1, 2, \ldots, N)$, where $X_i = \{\mu(RSSI_i), \mu(LQI_i), \mu(SNR_i)\}$ which is the input of the XGB_LQE and $Y_i = \{Level_i\}$ which is the output of the XGB_LQE.

We construct decision trees using the hardware parameters X_i and the link quality grades Y_i. The correlation between the hardware parameters X_i and the link quality grades Y_i is mined. Hence, the link quality estimation model XGB_LQE is constructed. The basic idea is to add multiple decision trees to the XGB_LQE step by step. When one decision tree is added to the model, the objective function is relatively reduced. By minimizing the objective function, we construct multiple decision trees, and combine them into a strong classifier. The leaf nodes of each decision tree are assigned weights that correspond to the estimated link quality grade. When the link quality parameters are input into the classifier, corresponding link quality grades are determined according to the attributes.

The objective function of the XGB_LQE includes the loss function and regularization term. Because we transform the link quality estimation problem into a classification problem, the Softmax function is used as the loss function.

The splitting process of the decision trees in the XGB_LQE is different from that of general decision trees. General decision trees do not consider the complexity when splitting, and rely on subsequent pruning operations to control the complexity. The XGB_LQE adds a regularization term to the objective function to control the complexity of the tree and to avoid over-fitting when splitting. The construction of each decision tree is iterated based on the previous decision tree. Every time a decision tree is constructed, it calculates the gain value of each leaf node and selects the highest valued leaf node to split.

When the gain of the split leaf node is less than zero or the depth of the tree reaches the specified maximum depth, the decision tree stops splitting. Hence, the optimization of the leaf node value and the decision tree structure is completed, and final link quality estimation model XGB_LQE is obtained.

The classifier sequence of M decision trees $\{I_1(X), I_2(X), \ldots, I_M(X)\}$ is obtained through training. The link quality estimation model XGB_LQE is formed by using additive combinations, as shown in equation (2).

$$G(X) = I_1(X) + I_2(X) + \ldots + I_M(X) = \sum_{m=1}^{M} I_m(X) \tag{2}$$

Among them, $G(X)$ is the final strong classifier, namely, the link quality estimation model XGB_LQE, and $I_m(X)$ is the *m-th* decision tree.

The interference sources in different scenarios are different, and the collected link quality parameters have different distribution characteristics. For example, due to the large numbers of trees and stones in the forest scenario, the reflectors have great impacts on the link quality for WSNs link, which results in poor link quality as a whole. That is, there are many bad links and a few good links, which result in imbalanced link quality data. Thus, the classification performance of the XGB_LQE for the link data of the minority class is poor.

Aiming at solving this problem, we control the maximum step of each tree's weight change in the iterative process of the link quality estimation model XGB_LQE, and adjust the weight of each tree. This process is done to avoid the influence of the good link data which is the minority category on the classifier and to reduce the errors that are caused by the imbalanced training data.

V. LINK QUALITY PREDICTION

Time series of link quality grades are obtained by using the link quality estimation model (XGB_LQE), and the essence of the link quality prediction problem is a regression process. Hence, the link quality prediction is transformed into a regression problem. We construct multiple regression trees with the sample set of time series which is built by using sliding time window, and combine the regression trees via addition to construct the XGB_LQP. The predicted link quality grade reflects the link quality at the next moment.

A. CONSTRUCTION OF TIME SERIES SAMPLE SETS

The link quality for WSNs has strong regularity in the short term [33]. Thus we use the estimated results of the XGB_LQE $\{Level_i\}$, $(i = 1, 2, \ldots, N)$ to construct the link quality sample set of time series using time intervals. As shown in figure 1, the sample set of time series $\mathrm{D_{pre}} = \{(X_i, Y_i)\}$, $(i = 1, 2, \ldots, N)$ is constructed by using a sliding time window of size w (there are w link quality grades in the sliding time window), where $X_i = \{Level_i, Level_{i+1}, \ldots, Level_{i+w-1}\}$ is the input of the link quality prediction model XGB_LQP, $Y_i = \{Level_{i+w}\}$ is the output, and i is the sequence number of the sample set.

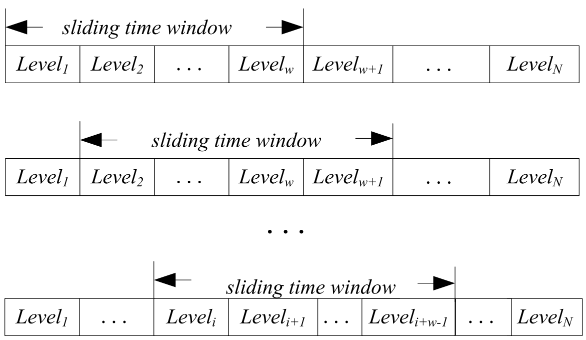

FIGURE 1. Construction process of a time series sample set.

The size of the sliding time window w has a certain influence on the prediction accuracy of the link quality prediction model XGB_LQP. An excessively large sliding time window will contain too many time features, which may result in increasing training complexity and longer training time for the XGB_LQP. In contrast, the XGB_LQP will not be able to learn the features in the time domain, which will result in unsatisfactory prediction results. In this paper, the size of w is determined by using a comparison experiment.

B. LINK QUALITY PREDICTION MODEL

We use the XGBoost regression algorithm to construct the link quality prediction model XGB_LQP for each experimental scenario. The XGB_LQP is shown in figure 2.

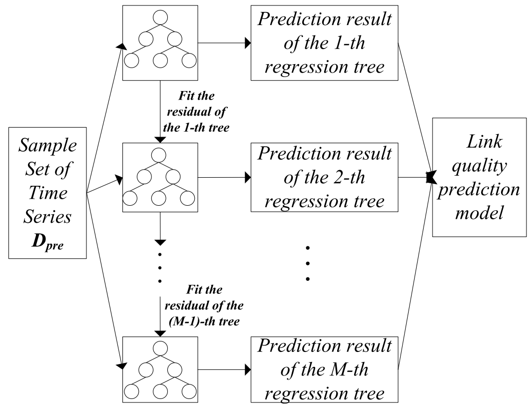

FIGURE 2. Link quality prediction model based on XGBoost regression algorithm.

The link quality prediction model XGB_LQP based on the XGBoost regression algorithm that is shown in figure 2 can be expressed as equation (3).

$$\hat{Y}_m = \sum_{i=1}^{M} \hat{Y}_i = \sum_{m=1}^{M} f_m(X_i) \tag{3}$$

Among them, f_m is the m-th regression tree, M is the total number of regression trees, $\hat{Y}_i$ is the i-th predicted link quality grade, and $\hat{Y}_m$ is the final link quality prediction model. We build M regression trees by optimizing the objective function step by step, and finally add the M regression trees together to obtain the link quality prediction model XGB_LQP.

First, the definition of the link quality prediction model's objective function is shown in equation (4).

$$O = \sum_{i=1}^{N} l\left(\hat{Y}_i, Y_i\right) + \sum_{m=1}^{M} \Omega(f_m) \tag{4}$$

where $\Omega(f_m) = \gamma T + 1/2\lambda \|\omega\|^2$ is the regularization term that measures the complexity of the regression tree and includes the number of leaf nodes of the regression trees and $L2$ regularization, T is the number of leaf nodes, and ω represents the score on a leaf. γ and λ are weight parameters, and l is the loss function of the XGB_LQP, which is the square loss function, as shown in equation (5).

$$l(\hat{Y}_i, Y_i) = \sum_{i=1}^{N} (\hat{Y}_i - Y_i)^2 \tag{5}$$

The link quality prediction model XGB_LQP optimizes the objective function step by step, and the main process is as follows.

$$\hat{Y}_i^{(0)} = 0 \tag{6}$$

$$\hat{Y}_i^{(1)} = f_1(X_i) = \hat{Y}_i^{(0)} + f_1(X_i) \tag{7}$$

$$\ldots$$

$$\hat{Y}_i^{(t)} = \sum_{m=1}^{t} f_m(X_i) = \hat{Y}_i^{(t-1)} + f_t(X_i) \tag{8}$$

The objective function of the XGB_LQP at the t-th step is as shown in equation (9).

$$O^{(t)} = \sum_{i=1}^{N} l\left[Y_i, \hat{Y}_i^{(t-1)} + f_t(X_i)\right] + \Omega(f_m) \tag{9}$$

According to the second-order Taylor expansion, the first-order derivative g_i and the second-order derivative h_i of the loss function are substituted into the objective function. The objective function at step t is shown in equation (10).

$$O^{(t)} = \sum_{i=1}^{N}\left(g_i f_t(X_i) + \frac{1}{2} h_i f_t^2(X_i)\right) + \Omega(f_m) + C \tag{10}$$

where C is a constant, and $I_j = \{i | q(Y_i) = j\}$ is the instance set of leaf j, which is the corresponding link quality values. We can rewrite equation (10), which can be used as a scoring function to measure the quality of a tree structure, as shown in equation (11).

$$O^{(t)} = \sum_{j=1}^{T}\left[(\sum_{i\in I_j} g_i)\omega_j + \frac{1}{2}(\sum_{i\in I_j} h_i + \lambda)\omega_j^2\right] + \gamma T \tag{11}$$

We find the minimum of equation (11) and compute the optimal weight ω_j^* of leaf j, as shown in equation (12).

$$\omega_j^* = -\frac{\sum_{i\in I_j} g_i}{\sum_{i\in I_j} h_i + \lambda} \tag{12}$$

The corresponding objective function is shown in equation (13).

$$O^* = -\frac{1}{2}\sum_{j=1}^{T}\frac{\left(\sum_{i\in I_j} g_i\right)^2}{\sum_{i\in I_j} h_i + \lambda} + \gamma T \tag{13}$$

Next, we use the greedy algorithm that starts from a single leaf and iteratively adds the branches to the tree. Then, the loss reduction after the split is given by equation (14).

$$Gain = \frac{1}{2}\left[\frac{(\sum_{i\in I_L} g_i)^2}{\sum_{i\in I_L} h_i + \lambda} + \frac{(\sum_{i\in I_R} g_i)^2}{\sum_{i\in I_L} h_i + \lambda} - \frac{(\sum_{i\in I} g_i)^2}{\sum_{i\in I_L} h_i + \lambda}\right] - \gamma \tag{14}$$

where I_L and I_R are the instance sets of the left and right nodes after the split, respectively, and $I = I_L + I_R$. We iterate this process until $Gain$ is less than zero or the number of iterations reaches the maximum depth.

After constructing M regression trees, the XGB_LQP is obtained by adding the M regression trees together. By inputting $\{Level_i, Level_{i+1}, \ldots, Level_{i+w-1}\}$ into the XGB_LQP, we get the link quality at the next moment by adding the predicted results of the M regression trees together.

VI. EXPERIMENTS AND ANALYSIS

To verify the validity of the model, the link quality data are acquired from multiple application scenarios. We use the TelosB node that is created by CrossBow to send and receive packets, and use the wireless sensor network link quality testbed (WSNs-LQT) that is developed by the lab to collect the link quality parameters, including the RSSI, SNR and LQI. The WSNs-LQT platform, which is shown in figure 3, dynamically displays the link quality information from nodes.

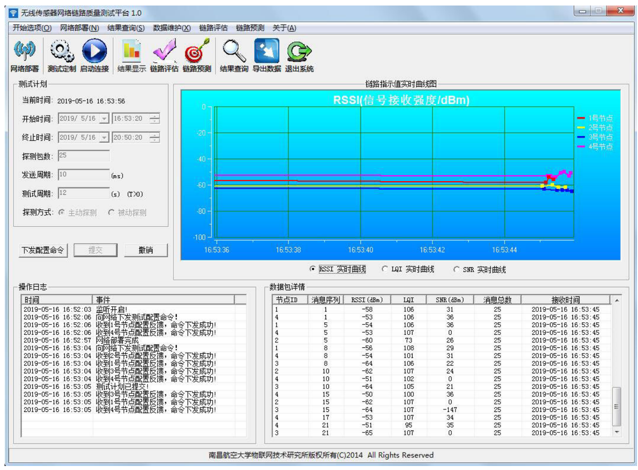

FIGURE 3. WSN Link Quality Testbed (WSNs-LQT).

A. EXPERIMENTAL DESIGN

We carry out our experiments on a server consisting of four GPUs. The experiment steps are mainly organized as follow: (1) In order to ensure the validity of the data set, the pre-processed data set is divided into training set and testing set according to the ratio of 7:3. (2)The GridSearchCV method in Scikit-learn is used to train multiple sets of different hyperparameters. The mean squared error which is shown in equation (15) of the training model is sorted and parameters of the highest score are taken as the optimal parameters of the link quality prediction model.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \hat{Y}_i\right)^2 \tag{15}$$

where Y_i is the real value of the link quality grade at the next moment, $\hat{Y}_i$ is the predicted value of the link quality grade at the next moment and N is the total number of samples. The mean squared error (MSE) is a measure of the degree of difference between the predicted value and the real value. We use the mean squared error to verify the link quality prediction model XGB_LQP since it reflects the credibility of the predicted value of the XGB_LQP. The smaller the MSE of the XGB_LQP' results is, the higher the accuracy of the XGB_LQP for predicting. (3) The validity of dividing link quality grades by using SUBXBFCM algorithm is verified by comparison experiments with other methods of dividing link quality grades. (4)The validity of XGB_LQP is verified by the MSE of model in different experimental scenarios by taking comparison experiments with SVR_LQP and 4C_LQP methods.

B. EXPERIMENTAL SCENARIOS AND DATA ANALYSIS

The link quality of WSNs experiences highly dynamic changes due to interferences from external environmental factors. We design experiments in three different scenarios with different interferences. Aiming at the interference caused by moving objects including moving vehicles and pedestrians, we design an experiment in square scenario. Aiming at the electromagnetic interference caused by Wi-Fi, Bluetooth, electronic devices, etc., we design an experiment in laboratory scenario. Aiming at the interference caused by the reflection and shelter of stationary objects, including trees and leaves, we design an experiment in grove scenario. Three experimental scenarios are shown in figure 4.

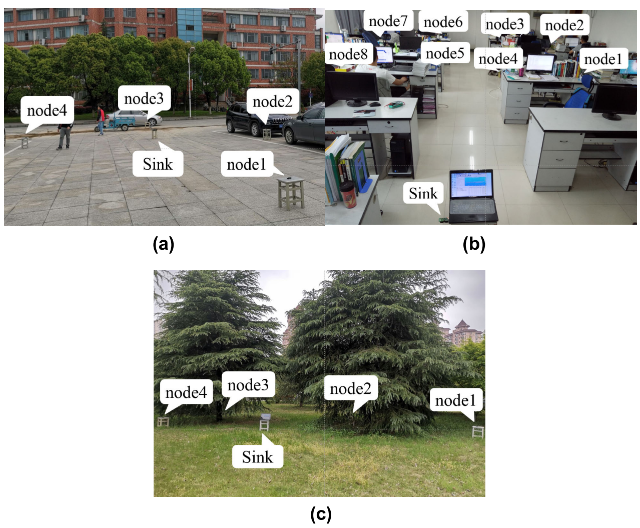

FIGURE 4. **Experimental scenarios: (a) square scenario, (b) laboratory scenario, (c) grove scenario.**

As shown in figure 4, different wireless sensor networks are deployed in each experimental scenario. In these three experimental scenarios, sensor nodes send packets to the sink node through one hop, and the sink node sends the collected packet to the personal computer.

The experimental parameter settings are shown in Table 2. To ensure that the diversity and reliability of the data, the time series of PRR are obtained by taking measurements for several consecutive days in three scenarios. The link quality data of PRR, which are collected by the experimental testbed in the three scenarios, are plotted into three time series diagrams, as shown in figures 5∼7.

It can be seen from figure 5 that the link quality is abnormally unstable and poor in the square scenario. The values of the PRR are mostly below 0.4. The interference sources in the square scenario mainly include moving vehicles and pedestrians, which will result in electromagnetic interference

TABLE 2. **Testbed parameters setting.**

Parameter	Values
Transmission power	0 dBm
Channel	26
Number of probe packets	30
Detection method	Active detection
Packet rate	10 pcs/s
Test cycle	10 s

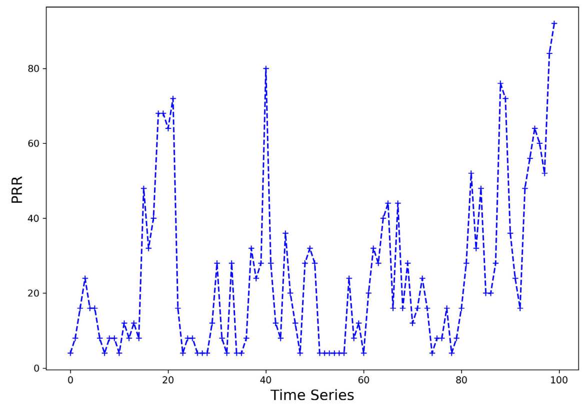

FIGURE 5. **Time series diagram of the PRR in the square scenario.**

and high attenuation to the WSN link quality and make the wireless channel unstable.

Figure 6 shows that the fluctuation of the link quality is relatively small in the laboratory scenario. The link quality is very good and stable, with an average PRR of 0.9 after number 25 in the series, as shown in figure 6. However, the link quality fluctuates greatly, and the values of PRR are very low before number 25. The interferences in the laboratory scenario are mainly for electromagnetic interference and radio frequency interference caused by Wi-Fi, Bluetooth, electronic devices, etc., which result in fading, including path

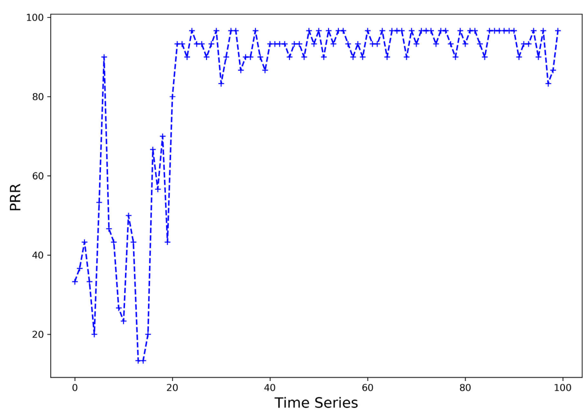

FIGURE 6. **Time series diagram of the PRR in the laboratory scenario.**

loss, channel fading, and multipath. These interferences make the link quality unstable.

In the grove scenario of figure 7, the overall link quality fluctuates greatly, and the PRR is unstable and fluctuates at approximately 0.4. Large numbers of reflection and shelter caused by trees and leaves in the grove scenario have great impacts on the WSN link quality. The link quality becomes very bad and unstable.

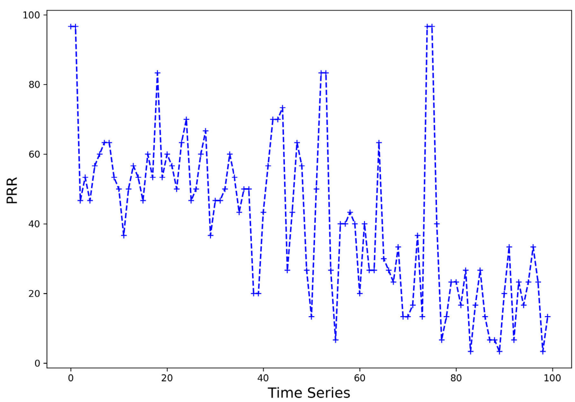

FIGURE 7. **Time series diagram of the PRR in the grove scenario.**

C. DETERMINATION OF CLUSTERING NUMBERS

We use SUBXBFCM algorithm to adaptively divide the link quality grades in different scenarios which can determine the optimal number of clusters by using the validity index *XB*. That is, SUBXBFCM algorithm does not need to run multiple times, which greatly reduces the calculation load and improves the efficiency of the algorithm. After the division of the link quality grades by the SUBXBFCM algorithm, the values of the *XB* corresponding to different clustering numbers are shown in Table 3.

TABLE 3. **The values of *XB* corresponding to different clustering numbers.**

The number of clusters	Value of XB
3	0.1521
4	0.1750
5	0.1735
6	0.1788
7	0.1613
8	0.1687
9	0.1822
10	0.1700

Table 3 shows that as the clustering number changes, the values of *XB* fluctuate accordingly. When the samples are clustered into different clusters, the compactness within the cluster and the separation between clusters are different. The smaller the value of *XB* is, the better the compactness within the grade and the separation between grades. When the number of clusters is 3, the value of *XB* is the smallest. Hence, we divide link quality into 3 grades. To verify the effectiveness of using the SUBXBFCM algorithm to divide the link quality grades, we compare the proposed method with two other methods, which are *Divide*_5 [28] and *Divide*_3 [17]. The experimental results are shown in figure 8.

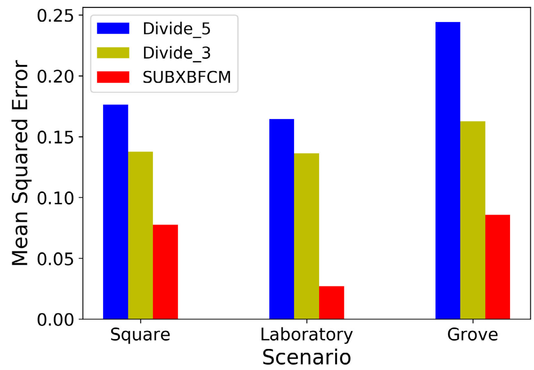

FIGURE 8. **MSEs of XGB_LQP using different link quality grade division methods.**

Figure 8 shows the prediction effect of the link quality prediction model XGB_LQP with different link quality grade division methods.

It shows that the prediction effect of the XGB_LQP using the proposed link quality grade division method is better than that of the others in the three scenarios. The MSE values of the prediction model using the SUBXBFCM algorithm to divide the link quality grades are respectively decreased by 15% and 11% in the laboratory scenario compared with the other two methods; these results verify the effectiveness of using the SUBXBFCM algorithm to divide the link quality grades.

D. DETERMINATION OF SLIDING TIME WINDOW SIZE

The number of samples in the sliding time window indicates how many historical link quality grades are used to predict the link quality grade at the next moment. Figure 9 shows the average MSE of the XGB_LQP in the three scenarios. In this experiment, we use a sliding time window with a size ranging from 2 to 15 [34].

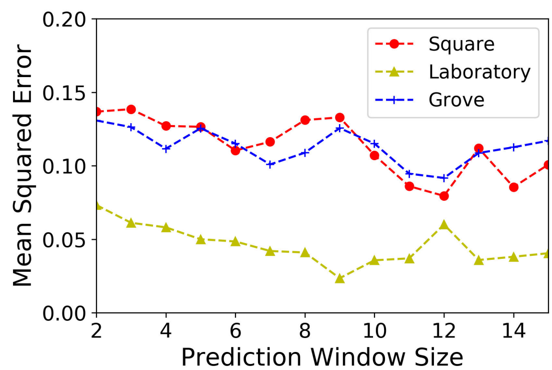

FIGURE 9. **MSEs of the XGB_LQP using different sliding time window sizes.**

It can be seen from figure 9 that the MSE of the link quality prediction model XGB_LQP fluctuates with the change of the sliding time window size. Hence, the size of the sliding time window has a certain impact on the prediction of the next link quality. The MSE of the XGB_LQP fluctuates the most in the square scenario, and the best prediction result is obtained for a window size of 12. In the laboratory and grove scenarios, the MSE of the XGB_LQP fluctuates slightly, but there are still minimum values. In the square, laboratory, and grove scenarios, the optimal sliding time window sizes are 12, 9 and 12, respectively.

E. VERIFICATION AND COMPARISON OF XGB_LQP MODEL

To verify the validity of the link quality prediction model XGB_LQP, we carry out comparison experiments with the support vector regression (SVR)-based model (SVR_LQP) and the 4C-based model (4C-LQP) using the same datasets in three scenarios. Among them, the SVR is the current popular machine learning algorithm, and the 4C is the classical link quality prediction method. The experimental results are shown in figures 10∼12. Real_Value is the measured value of the PRR.

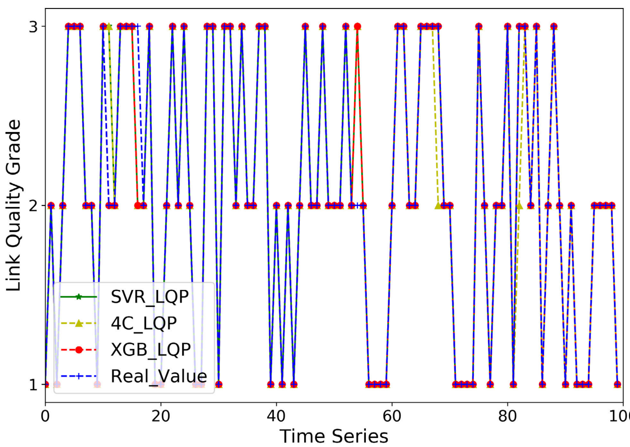

FIGURE 10. **Comparison of the prediction results in the square scenario.**

Figures 10∼12 show that the link quality mainly fluctuates in the first and second grades in the square scenario. The proposed link quality prediction model XGB_LQP can still make accurate predictions, while the prediction effects of the SVR_LQP and 4C_LQP are relatively poor. In the laboratory and grove scenarios, the link quality frequently fluctuates among three grades, which indicates that the link quality is very unstable. The SVR_LQP and 4C-LQP make more mistakes when the grade suddenly changes. However, the XGB_LQP can effectively predict the link quality grade at the next moment even when the link quality grade suddenly changes. Based on the experimental analysis of the three scenarios, the predicted results of the XGB_LQP are more consistent with the real values. Even in the case of unstable links, the link quality grades at the next moment can still be effectively predicted by the proposed prediction model.

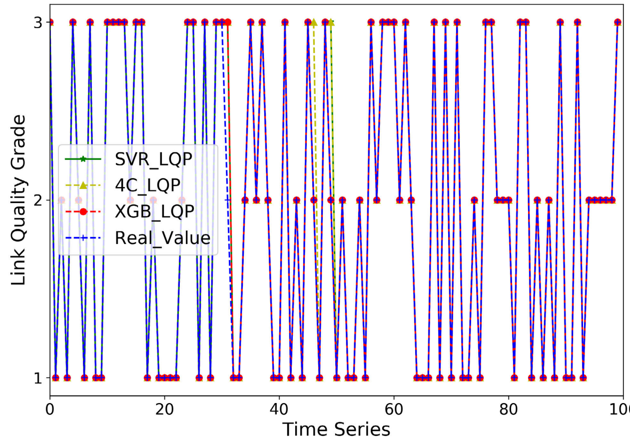

FIGURE 11. **Comparison of the prediction results in the laboratory scenario.**

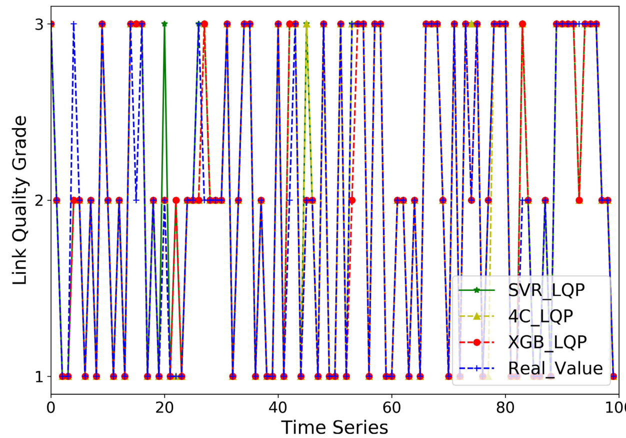

FIGURE 12. **Comparison of the prediction results in the grove scenario.**

To further accurately verify the prediction effect of the link quality prediction model XGB_LQP, the MSEs of three prediction models in the square, laboratory and grove scenarios are shown in Table 4.

TABLE 4. **Comparison between the XGB_LQP, SVR_LQP and 4C_LQP models.**

Scenario	MSE		
	SVR_LQP	XGB_LQP	4C-LQP
Square	0.0922	0.0774	0.1263
Laboratory	0.0303	0.0269	0.0419
Grove	0.1018	0.0857	0.1007

The results in Table 4 demonstrate that the MSE of the 4C_LQP is the biggest among the three scenarios. The link quality prediction model XGB_LQP has the smallest MSEs in the three scenarios compared with the other link quality prediction models. The three prediction models have the best prediction effect in the laboratory scenario, which is a relatively stable scenario compared with other scenarios. Among them, the XGB_LQP has the best prediction accuracy, with an MSE of 0.0269. Due to the interferences caused by moving vehicles and pedestrians in the square scenario, the link quality for WSNs fluctuates greatly and exhibits frequent sudden changes. The 4C_LQP and SVR_LQP have poor predictive performances compared with the XGB_LQP. Hence, the XGB_LQP can predict fluctuations of link quality and has a better anti-interference ability. In the grove scenario, all link quality prediction models have the worst prediction accuracy compared with the other scenarios due to the reflection and shelter caused by trees and leaves which will result in multipath. The MSE of the XGB_LQP is 0.0857, which is still smaller than 4C_LQP and SVR_LQP.

The comprehensive analysis of the above experimental results shows that the XGB_LQP link quality prediction model has better prediction accuracy compared with the 4C_LQP and SVR_LQP in the three experimental scenarios with different interference sources for single-hop WSNs. It can be seen that the proposed link quality prediction model has a better prediction effect in link quality prediction for single-hop WSNs and has better generalization abilities in different scenarios.

VII. CONCLUSION

In this paper, we propose a link quality prediction method based on XGBoost for WSNs. Firstly, link quality grades are divided adaptively based on SUBXBFCM according to the PRR. Pearson correlation coefficient is employed to analyse the correlation between hardware parameters and PRR. Based on the averages of RSSI, LQI, SNR and link quality grades, the link quality estimation model (XGB_LQE) is constructed by using the XGBoost classification algorithm to estimate the current link quality grade. In the construction of the XGB_LQE, the maximum step of each tree's weight change is controlled to reduce the influence of the unbalanced link quality samples on the XGB_LQE. Based on results of the XGB_LQE, the link quality prediction model (XGB_LQP) is constructed based on the XGBoost regression algorithm. Experiments are carried out in square, laboratory and grove scenarios which have different interferences. In these three experimental scenarios, we deployed three single-hop WSNs. Experimental results show that the prediction accuracy of the XGB_LQP using the SUBXBFCM algorithm is higher than those of the other two division methods, which verifies the rationality of using SUBXBFCM algorithm to divide the link quality grades. Compared with the SVR_LQP and 4C_LQP models in the three scenarios with different interferences, XGB_LQP has better prediction accuracy and good environmental adaptability for single-hop WSNs.

In the future, we will consider applying our proposed method for multi-hop WSNs to predict the link quality at the next moment. We should not only consider the link quality but also the path cost when the next hop node is selected. Selecting links with good quality may need more hops to delivery a packet, which will result in high path cost. Therefore, it is necessary to combine the predicted results of the proposed model with the path cost of each node when selecting the next hop node.

REFERENCES

[1] S. Rekik, N. Baccour, M. Jmaiel, and K. Drira, "Experiencing low power wireless links in distribution smart grid environments," in *Proc. IEEE/ACS 15th Int. Conf. Comput. Syst. Appl. (AICCSA)*, Aqaba, Jordan, Oct./Nov. 2018, pp. 1–8.

[2] N. Baccour, A. Koubâa, L. Mottola, M. A. Zúñiga, H. Youssef, C. A. Boano, and M. Alves, "Radio link quality estimation in wireless sensor networks: A survey," *ACM Trans. Sensor Netw.*, vol. 8, no. 4, Sep. 2012, Art. no. 34.

[3] J. An, K. Yang, J. Wu, N. Ye, S. Guo, and Z. Liao, "Achieving sustainable ultra-dense heterogeneous networks for 5G," *IEEE Commun. Mag.*, vol. 55, no. 12, pp. 84–90, Dec. 2017.

[4] X. Gao, P. Wang, D. Niyato, K. Yang, and J. An, "Auction-based time scheduling for backscatter-aided RF-powered cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 18, no. 3, pp. 1684–1697, Mar. 2019.

[5] N. Baccour, A. Koubâa, M. Ben Jamâa, H. Youssef, M. Zuniga, and M. Alves, "A comparative simulation study of link quality estimators in wireless sensor networks," in *Proc. IEEE Int. Symp. Modeling, Anal. Simulation Comput. Telecommun. Syst.*, London, U.K., Sep. 2009, pp. 1–10.

[6] K. Yang, N. Yang, N. Ye, M. Jia, Z. Gao, and R. Fan, "Non-orthogonal multiple access: Achieving sustainable future radio access," *IEEE Commun. Mag.*, vol. 57, no. 2, pp. 116–121, Feb. 2019.

[7] N. Baccour, A. Koubâa, H. Youssef, and M. Alves, "Reliable link quality estimation in low-power wireless networks and its impact on tree-routing," *Ad Hoc Netw.*, vol. 27, pp. 1–25, Apr. 2015.

[8] N. Baccour, A. Koubâa, H. Youssef, M. Ben Jamâa, D. do Rosário, M. Alves, and L. B. Becker Becker, "F-LQE: A fuzzy link quality estimator for wireless sensor networks," in *Proc. Eur. Conf. Wireless Sensor Netw. (EWSN)*, Coimbra, Portugal, Feb. 2010, pp. 240–255.

[9] Z. Ansar and W. Dargie, "Adaptive burst transmission scheme for WSNs," in *Proc. 26th Int. Conf. Comput. Commun. Netw.*, Vancouver, BC, Canada, Jul./Aug. 2017, pp. 1–7.

[10] D. Liu, Z. Cao, Y. Zhang, and M. Hou, "Achieving accurate and real-time link estimation for low power wireless sensor networks," *IEEE/ACM Trans. Netw.*, vol. 25, no. 4, pp. 2096–2109, Apr. 2017.

[11] S. Aswale and V. R. Ghorpade, "Geographic multipath routing based on triangle link quality metric with minimum Inter-path interference for wireless multimedia sensor networks," *J. King Saud Univ.-Comput. Inf. Sci.*, to be published. doi: 10.1016/j.jksuci.2018.02.001.

[12] 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.

[13] R. Fonseca, O. Gnawali, P. Levis, and K. Ben Jamâa, "Four-bit wireless link estimation," in *Proc. 6th Workshop Hot Topics Netw.*, Atlanta, GA, USA, Nov. 2007, pp. 1–7.

[14] Z.-Q. Guo, Q. Wang, M.-H. Li, and J. He, "Fuzzy logic based multidimensional link quality estimation for multi-hop wireless sensor networks," *IEEE Sensors J.*, vol. 13, no. 10, pp. 3605–3615, Oct. 2013.

[15] A. Sharma, A. Bansal, and V. Rishiwal, "Selection of high quality path through MXAODV in mobile ad-hoc network," *Int. J. Syst., Control Commun.*, vol. 10, no. 1, pp. 1–17, Jan. 2019.

[16] F. Qin, Q. Zhang, W. Zhang, Y. Yang, J. Ding, and X. Dai, "Link quality estimation in industrial temporal fading channel with augmented Kalman filter," *IEEE Trans. Ind. Informat.*, vol. 15, no. 4, pp. 1936–1946, Apr. 2019.

[17] A. Bildea, O. Alphand, F. Rousseau, and A. Duda, "Link quality estimation with the Gilbert–Elliot model for wireless sensor networks," in *Proc. IEEE 26th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Hong Kong, Aug./Sep. 2015, pp. 2049–2054.

[18] E. Ancillotti, C. Vallati, R. Bruno, and E. Mingozzi, "A reinforcement learning-based link quality estimation strategy for RPL and its impact on topology management," *Comput. Commun.*, vol. 112, no. 1, pp. 1–13, Nov. 2017.

[19] 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.

[20] C. J. Lowrance, A. P. Lauf, and M. Kantardzic, "A fuzzy-based machine learning model for robot prediction of link quality," in *Proc. IEEE Symp. Comput. Intell. (SSCI)*, Athens, Greece, Dec. 2016, pp. 1–8.

[21] T. Liu and A. E. Cerpa, "Data-driven link quality prediction using link features," *ACM Trans. Sensor Netw.*, vol. 10, no. 2, Jan. 2014, Art. no. 37.

[22] J. Fan, W. Yue, L. Wu, F. Zhang, H. Cai, X. Wang, X. Lu, and Y. Xiang, "Evaluation of SVM, ELM and four tree-based ensemble models for predicting daily reference evapotranspiration using limited meteorological data in different climates of China," *Agricult. Forest Meteorol.*, vol. 263, no. 1, pp. 225–241, Dec. 2018.

[23] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (SIGKDD)*, San Francisco, CA, USA, Jun. 2016, pp. 785–794.

[24] Z. Chen, F. Jiang, Y. Cheng, X. Gu, W. Liu, and J. Peng, "XGBoost classifier for DDoS attack detection and analysis in SDN-based cloud," in *Proc. IEEE Int. Conf. Big Data Smart Comput. (BigComp)*, Shanghai, China, Jan. 2018, pp. 251–256.

[25] J. Guo, L. Yang, R. Bie, J. Yu, Y. Gao, Y. Shen, and A. Kos, "An XGBoost-based physical fitness evaluation model using advanced feature selection and Bayesian hyper-parameter optimization for wearable running monitoring," *Comput. Netw.*, vol. 151, no. 1, pp. 166–180, Mar. 2019.

[26] L. Tian, F. Huang, L. Fang, and Y. Bai, "Intelligent monitoring system of cremation equipment based on Internet of Things," in *Proc. Chin. Intell. Syst. Conf.*, Wenzhou, China, Oct. 2018, pp. 249–258.

[27] C. Ji, X. Zou, Y. Hu, L. Lyu, X. Zheng, and S. Liu, "XG-SF: An XGBoost classifier based on shapelet features for time series classification," *Procedia Comput. Sci.*, vol. 147, no. 1, pp. 24–28, Feb. 2019.

[28] X. Luo, L. Liu, J. Shu, and M. Al-Kali, "Link quality estimation method for wireless sensor networks based on stacked autoencoder," *IEEE Access*, vol. 7, pp. 21572–21583, 2019.

[29] H. Sabit, A. Al-Anbuky, and H. Gholam-Hosseini, "Distributed WSN data stream mining based on fuzzy clustering," in *Proc. Symp. Workshops Ubiquitous, Auton. Trusted Comput.*, Brisbane, QLD, Australia, Jul. 2009, pp. 395–400.

[30] Z. Lu, Y. Qiu, Y. Fu, and B. Lu, "An improved FCM method for image segmentation based on wavelet transform and particle swarm," in *Proc. 33rd Youth Acad. Annu. Conf. Chin. Assoc. Automat. (YAC)*, Nanjing, China, May 2018, pp. 112–117.

[31] M. Singh, R. Bhattacharjee, N. Sharma, and A. Verma, "An improved Xie-Beni index for cluster validity measure," in *Proc. 4th Int. Conf. Image Inf. Process. (ICIIP)*, Shimla, India, Dec. 2017, pp. 1–5.

[32] R. Singh, R. Mehra, and L. Sharma, "Design of Kalman filter for wireless sensor network," in *Proc. IEEE Int. Conf. Internet Things Appl. (IOTA)*, Pune, India, Jan. 2016, pp. 63–67.

[33] A. Cerpa, J. L. Wong, M. Potkonjak, and D. Estrin, "Temporal properties of low power wireless links: Modeling and implications on multi-hop routing," in *Proc. 6th ACM Int. Symp. Mobile Ad Hoc Netw. Comput. (MobiHoc)*, Champaign, IL, USA, May 2005, pp. 414–425.

[34] M. L. Bote-Lorenzo, E. Gómez-Sánchez, C. Mediavilla-Pastor, and J. I. Asensio-Pérez, "Online machine learning algorithms to predict link quality in community wireless mesh networks," *Comput. Netw.*, vol. 132, no. 1, pp. 68–80, Feb. 2018.

YI FENG was born in Changzhou, Jiangsu, China, in 1995. She is currently pursuing the master's degree with Nanchang Hangkong University. Her research interest includes wireless sensor networks.

LINLAN LIU was 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. Her research interests include wireless sensor networks, software engineering, and distributed systems. She is also a member of CCF.

JIAN SHU was 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 also a Senior Member of CCF.