Chinese Journal of Electronics
Vol.26, No.2, Mar. 2017

Research on Link Quality Estimation Mechanism for Wireless Sensor Networks Based on Support Vector Machine*

SHU Jian[1,2], LIU Song[1,2], LIU Linlan[1,3], ZHAN Liqin[2] and HU Gang[1,2]

(1. *Internet of Things Technology Institute, Nanchang Hangkong University, Nanchang 330063, China*)

(2. *School of Software, Nanchang Hangkong University, Nanchang 330063, China*)

(3. *School of Information Engineering, Nanchang Hangkong University, Nanchang 330063, China*)

Abstract — In the application of Wireless sensor networks (WSNs), effective estimation for link quality is a basic issue in guarantying reliable data transmission and upper network protocol performance. A link quality estimation mechanism is proposed, which is based on Support vector machine (SVM) with multi-class classification. Under the analysis of the wireless link characteristics, two physical parameters of communication, Receive signal strength indicator (RSSI) and Link quality indicator (LQI), are chosen as estimation parameters. The link quality is divided into five levels according to Packet reception rate (PRR). A link quality estimation model based on SVM with decision tree is established. The model is built on kernel functions of radial basis and polynomial respectively, in which RSSI, LQI are the input parameters. The experimental results show that the model is reasonable. Compared with the recent published link quality estimation models, our model can estimate the current link quality accurately with a relative small number of probe packets, so that it costs less energy consumption than the one caused by sending a large number of probe packets. So this model which is high efficiency and energy saving can prolong the network life.

Key words — Wireless sensor networks (WSNs), Link quality estimation, Support vector machine (SVM).

I. Introduction

Wireless sensor networks (WSNs) are composed by numerous low-cost and micro sensor nodes[1], which are deployed in the monitored area randomly and constitute to a multi-hop ad hoc network in Wireless communication manner. However, as the general application environment is harsh, the nodes with energy constraints and low-power transceivers are susceptible to background noise and multipath effect that it can result in signal distortion, link quality instability and volatility. In WSNs, although retransmission can maintain data integrity when packets are transmitted over low quality links, it leads to low transmission efficiency, more energy consumption and poor real-time. So, it is necessary to employ link quality estimation method as a support mechanism to select the better stable routes for data delivery. It will improve the network throughput and maximize its lifetime.

II. Related Work

Recent years, there are large amount of literatures on link quality estimation for WSNs at home and abroad. And the existing link quality estimation methods fall in the following categories: data link layer parameter based link estimation, physical layer parameter based link estimation, and comprehensive link quality estimation, *etc.*

In early researches of WSNs, researchers tended to assume that the wireless link was ideal and symmetrical since there was no effective method of link quality measurement. Subsequently, it has been found that the assumption is incorrect, and the wireless link characteristics in WSNs have been discovered, such as time volatility[2], directivity[3], asymmetry, irregularity[4], "grey area"[5] and environment sensitivity[6]. All these characteristics are known to cause the instability of wireless link. Therefore, large numbers of link quality estimation methods are proposed on the analysis of these characteristics.

The metrics of physical layer mainly are RSSI, LQI and SNR[7]. In Ref.[8], a RSSI based link quality estima-

*Manuscript Received July 22, 2014; Accepted July 28, 2016. This work is supported by the National Natural Science Foundation of China (No.61363015, No.61262020, No.61501218, No.61501217), and the Key Research Foundation of Education Bureau of Jiangxi Province of China (No.GJJ150702).

 DOI:10.1049/cje.2017.01.013

tor is proposed. The experiments demonstrate that there exists a good correlation between PRR and RSSI. But the RSSI is vulnerable to outside interference and cannot respond timely enough to environmental changes. With the introduction of Chipcon's CC2420 chip, there is one more measurement metric of link quality named LQI which is provided by the chip itself. Ref.[9] points out that statistical mean LQI can reflect the link quality accurately with a small number of packets. A relationship mapping model between SNR and PRR is proposed in Ref.[10]. These two properties both can be easily obtained from the hardware layer of node and reflect the change of link quality rapidly. Nevertheless, they do not take the lost packets into consideration so that they are incapable of providing holistic information on the link quality.

The PRR is widely adopted as a standard metric in Link quality estimation method based on the data link layer parameters. Ref.[11] proposes a PRR based link quality estimation method applied to the event monitoring. Ref.[12] expounds LQI-PRR model, a new link estimation model between LQI and PRR. Despite the fact that it can be easily obtained, PRR needs to send a large number of probe packets for its computation, so that it results in a dramatic increase in energy consuming and communication overhead. Furthermore, it is not responsive enough to link changes.

Peng *et al.*[13] propose a Multiple time scales link estimation (MTSLE) method by considering long and short time scales link. Guo *et al.*[14] present a Fuzzy logic based link quality indicator (FLI), which reflects single hop reliability of packet delivery, link volatility, and packet loss burst. But in this method the obtaining of input parameters costs lots of energy in sending probe packets and calculation.

Support vector machine (SVM) is a machine learning method based on VC theory and statistical learning theory. It is one of the most important theories in the field of machine learning. Moreover, by using the kernel function, this theory can turn a non-linearity case into a linearity case in order to reduce the algorithm complexity. To a great extent, it overcomes the "dimension disaster" and "over learning" due to its many particular advantages on resolving such problems as small sample, nonlinearity and high dimension. In addition, as its solid theoretical foundation and straightforward mathematical model, it has been widely used in text recognition, handwriting recognition, face image recognition, genetic classification and time series prediction, *etc.*

Ref.[15] presents that the effectiveness of SVM depends on the selection of kernel function and kernel function parameters. However, there is no special method for parameters selection of SVM.

In this paper, we propose a link quality estimation model based on SVM with decision tree, and the link quality estimation problem is converted to the quality level classification. The rest of this paper is organized as follows: In Section III, SVM theory is briefly introduced. In Section IV, the necessary metrics of wireless link quality estimation is selected and analyzed. In Section V, a link quality estimation model based on SVM with decision tree is proposed. In Section VI, the experiment environment is introduced in detail and many groups of comparisons have been conducted on the SVM based estimation model and recent published link quality estimation methods. The conclusion is shown in Section VII. SVM based link quality estimation mechanism can achieve more accuracy and efficiency than the ones of existing approaches.

III. Classification Method of SVM

The SVM theory[16] is designed for solving binary classification problems. The main idea of SVM is to construct linear hyperplanes and to find out one having the largest separation between two classes. Actually, finding out the best hyperplane can be treated as solving the optimization (so-called convex programming). In addition to perform linear classification, SVMs can efficiently perform a non-linear classification and regression using Mercer kernel expansion theorem, implicitly mapping their inputs into high- or infinite-dimensional feature spaces. In short, the SVM theory maps the training data for classification into a higher-dimensional space and makes them linearly separable[15]. Because in reality, many classification problems are not linearly separable, before using the SVM as a classifier these problems need to be transformed into other forms, such as a non-linear classification to a linear one, a multi-class classification to a binary one and a linear non-separable classification to a linear separable one.

1. Linearly separable case

Let $\{x_i, y_i\}_{i=1}^N$ be a set of training samples, each sample $x_i \in R^n$, n being the dimension of the input space, belongs to a class labeled by $y_i \in \{-1, 1\}$ or $y_i \in \{1, 2, \cdots, k\}$, where k is the number of classification category. Suppose the hyperplane as $\boldsymbol{\omega}x + b = 0$ where $\boldsymbol{\omega}$ is the hyperplane normal vector and b is a constant offset. The goal is to define a hyperplane which divides the set of samples such that all the points with the same label are on the same side of the hyperplane. This equals to find $\boldsymbol{\omega}$ and b which would maximize the margin in optimal separating hyperplane.

Fig.1 shows separating hyperplane $L : \boldsymbol{\omega}x + b = 0$; Equation of two boundaries L_1, L_2 is $\boldsymbol{\omega}x + b = \pm 1$ (after identical deformation); and samples on the margin are support vectors, namely the optimal separating sample points. Let X_1 on L_1, X_2 on L_2, *i.e.* $\begin{cases} \boldsymbol{\omega}x_1 + b = +1 \\ \boldsymbol{\omega}x_2 + b = -1 \end{cases}$,

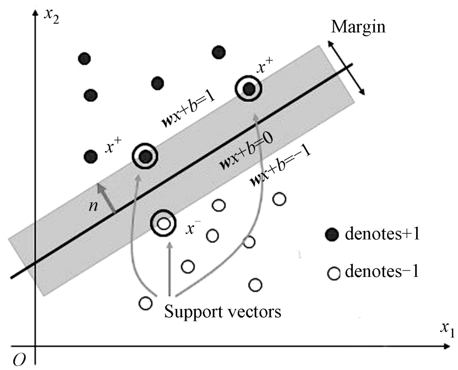


Fig. 1. Solution of the optimal hyperplane

it gets $\boldsymbol{\omega}(x_1-x_2)=2$ after subtraction. The distance, the closest points to hyperplane, can be deduced as:

$$dis=\frac{\boldsymbol{\omega}}{\|\boldsymbol{\omega}\|}(x_1-x_2)=\frac{2}{\|\boldsymbol{\omega}\|} \tag{1}$$

The left of Eq.(1) is the projection that vector (x_1-x_2) projects onto the normal vector $\boldsymbol{\omega}$ of separating hyperplane. So the distance equals to $\frac{2}{\|\boldsymbol{\omega}\|}$, which is called margin. Finding the maximum margin is amounts to minimize $\|\boldsymbol{\omega}\|$. In order to simplify calculation it is equivalent to minimize $\frac{\|\boldsymbol{\omega}\|^2}{2}$. To split the training samples correctly and satisfy maximum margin, the minimizing is done under the constraint $y_i\cdot(\boldsymbol{\omega}x_i+b)\geq 1$ which combines $\boldsymbol{\omega}x_i+b\geq +1, y_i=1$, and $\boldsymbol{\omega}x_i+b\leq -1, y_i=1$. Thus the problem building linear SVM becomes to optimize the following formula:

$$\begin{cases}\min\{\frac{1}{2}\|\boldsymbol{\omega}\|^2\}\\ y_i\cdot(\boldsymbol{\omega}x_i+b)\geq 1,\quad i=1,\cdots,N\end{cases} \tag{2}$$

Note that Eq.(2) is a convex set consisting of a convex function and constraints, called convex programming. Here the Kuhn-Tucker theorem in optimization theory is used for the optimizing. Since the objective function and constraints satisfy Kuhn-Tucker theorem, the Eq.(2) has a unique global minimum. The optimization should be supplemented with a KKT constraint:

$$a_i(y_i(\boldsymbol{\omega}x_i+b)-1)=0,\qquad i=1,\cdots,N \tag{3}$$

As $\|\boldsymbol{\omega}\|^2$ is convex, minimizing it under linear constraints can be achieved with Lagrange multipliers $L(\boldsymbol{\omega},b,a)=\frac{1}{2}\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\omega}-\sum_{i=1}^{N}a_i[y_i\,(\boldsymbol{\omega}x_i+b)-1]$, where a is the Lagrange coefficient. Partial derivatives of $L(\boldsymbol{\omega},b,a)$ with respect to b and $\boldsymbol{\omega}$ are taken respectively and set to zero, then $\sum_{i=1}^{N}a_iy_i=0$ and $\boldsymbol{\omega}=\sum_{i=1}^{N}a_iy_ix_i$ can be achieved. Substituting these two results into Eq.(3), it becomes a quadratic programming optimization problem:

$$\begin{cases}\max J(a)=\max\{\sum_{i=1}^{N}a_i-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}a_ia_jy_iy_jx_i^{\mathrm{T}}x_j\}\\ \sum_{i=1}^{N}a_iy_i=0,\quad 0\leq a_i\leq C,\quad i=1,\cdots,N\end{cases} \tag{4}$$

where C is a penalty parameter.

For Eq.(4), many numerical calculation methods are available, and the corresponding optimal solution $a^*=(a_1^*,a_2^*,\cdots,a_N^*)$ can be worked out easily. In fact, a^* of non support vector is set to zero. a^* is substituted into partial derivatives associated with KKT constraints. $\boldsymbol{\omega}_0$ and b_0 can be deduced as:

$$\boldsymbol{\omega}_0=\sum_{i=1}^{N}a_iy_ix_i,\qquad b_0=-\frac{1}{2}[\boldsymbol{\omega}_0(x_++x_-)] \tag{5}$$

Considering the Eq.(5) of $\boldsymbol{\omega}_0$ in separating hyperplane $\boldsymbol{\omega}x+b=0$, the optimal separating hyperplane decision function can thus be written as:

$$M(x)=\mathrm{sgn}(\boldsymbol{\omega}_0x+b_0)=\mathrm{sgn}(\sum_{i=1}^{N}a_i^*y_i(x_ix)+b_0) \tag{6}$$

2. Linearly non-separable case

When the data is not linearly separable, a slack variable ξ_i and a penalty parameter c are introduced to build the linear SVM model. The separating hyperplane $\boldsymbol{\omega}x+b=0$ satisfies constraints $y_i\cdot(\boldsymbol{\omega}x_i+b)\geq 1-\xi_i,(i=1,\cdots,N)$. ξ_i is the distance from a sample point to the margin.

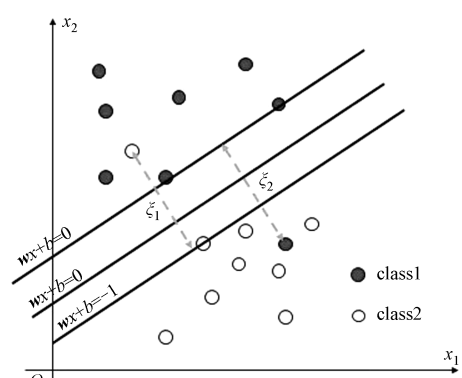


Fig. 2. Linearly non-separable case

Sample points can be classified correctly when $0<\xi_i<1$ but misclassified if $\xi_i\geq 1$. The purpose of the variable ξ_i is to allow the existence of some misclassified points, which have their corresponding $\xi_i\geq 1$. The penalty term $C\sum_{i=1}^{N}\xi_i$ is joined in minimizing $\frac{\|\boldsymbol{\omega}\|^2}{2}$ to control the error rate. The generalized Optimal Separating Hyperplane is then regarded as the solution of the following problem

$$\begin{cases} \min\{\frac{1}{2}\|\boldsymbol{\omega}\|^2 + C\sum_{i=1}^{N}\xi_i\}, \\ y_i \cdot (\boldsymbol{\omega} x_i + b) \geq 1 - \xi_i, \quad i = 1, \cdots, N \end{cases} \tag{7}$$

Consequently, the optimal separating hyperplane decision function can also be written as:

$$M(x) = \mathrm{sgn}(\sum_{i=1}^{N} a_i^* y_i (X_i X) + b_0) \tag{8}$$

For nonlinear classification, the input data is mapped into a high-dimensional feature space through some nonlinear mapping chosen a priori $\varphi(x)$. In this feature space, the optimal separating hyperplane is constructed. One of the key properties of SVM is the use of the kernel $K(x, y)$ to compute the dot product without having to explicitly compute the mapping $\varphi(x)$ according to Ref.[17]. Then the optimal separating hyperplane decision function becomes:

$$M(x) = \mathrm{sgn}(\sum_{i=1}^{N} a_i^* y_i K(X_i X) + b_0) \tag{9}$$

where the kernel K subjects to Mercer's condition $K(X, y) = (\varphi(X) \cdot \varphi(y))$. In this paper, the radial basis function RBF which has good performance on classification has been chosen as the Mercer kernel, while the Polynomial kernel has been chosen for comparison. The kernel radial basis function prototype can be expressed as (g is a parameter):

$$K(x, x_i) = \exp(-g|x - x_i|) \tag{10}$$

And the Polynomial kernel prototype can be expressed as (s, c, d are parameters):

$$K(x, x_i) = (s(x \cdot x_i) + c)^d \tag{11}$$

Note, parameters of the kernel functions and penalty parameter c in this paper are optimized using K-fold cross-validation algorithms. The parameters with best cross-validation accuracy are picked up.

SVM has advantages in performing nonlinear classification and solving small sample problem. So it is well suited for link quality estimation that tries to estimate the link by sending a small number of packets. Hence, the SVM method is introduced to build the estimation model for link quality.

IV. The Metrics of Wireless Link Quality Estimation

The performance of estimation model depends on the training sample set directly. So in the process of modeling, not only the typicality of estimation metrics but also the correlationship among these metrics need to be taken into consideration. In this section, we will select the metrics of link quality estimation, and analyze their linear correlation. Furthermore, the sample set of link quality estimation is obtained, which will be used to train and to test the model based on SVM in the next section.

The RSSI and LQI from physical layer are selected as estimation metrics. A great deal of data is obtained from experiments. The experimental environment is discussed in some detail in Section VI.

121 groups of data obtained in experiments are analyzed using the SPSS statistical analysis tool. The correlations among mean RSSI, mean LQI and PRR are shown in Table 1. Pearson correlation is used to measure the linear relationship between two variables. The closer the absolute value of correlation coefficient to 1, the stronger the correlation is. Statistical significance is the probability value in statistical hypothesis testing in a small probability event. N is the number of samples with 121 groups of experimental data. As shown in Table 1 the Pearson correlation between mean LQI and PRR is 0.829, illustrating a strong correlation, while mean RSSI and PRR is weakly correlated 0.313. This result is consistent with that in Ref.[11], and thus shows the universal relationship among mean RSSI, mean LQI and PRR.

Table 1. Correlation among mean RSSI, mean LQI and PRR

		Mean RSSI	Mean LQI	PRR
Mean RSSI	Pearson correlation	1	-0.12	.313**
	Statistical significance		0.19	0
	N	121	121	121
Mean LQI	Pearson correlation	-0.12	1	.829**
	Statistical significance	0.19		0
	N	121	121	121
PRR	Pearson correlation	.313**	.829**	1
	Statistical significance	0	0	
	N	121	121	121

**Notable correlation (bilateral) on the .01 level.

In order to observe the relationship of RSSI-PRR and LQI-PRR in more details, scatter distribution, shown in Fig.3, is used to depict RSSI-PRR and LQI-PRR. For RSSI greater than −80dBm, inflection point arises in the scatter distribution, and a large number of sample points are distributed in the range of PRR value [90%,100%];

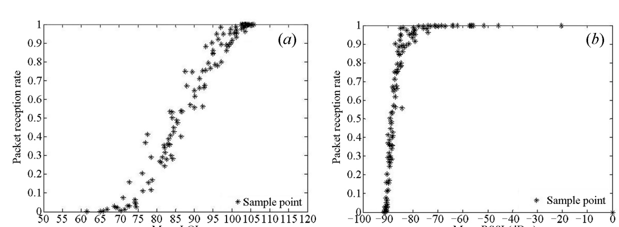


Fig. 3. Scatter distribution of mean RSSI-PRR and mean LQI-PRR

But for mean RSSI lees than −90dBm, distribution of most sample points is within PRR value range [0%,10%]. Similarly, if LQI is less than 75, PRR values of sample points are distributed in [0%,10%]. When the LQI is greater than 100, PRR value distribution of the sample points is in [90%,100%]. Consequently, this fact provides reference for classifying link quality levels. However, the RSSI is not sensitive to variations in packet reception rate, and it's vulnerable to the interference of environmental noise. Furthermore, RSSI or LQI merely considers the successfully received data packet and takes no account of the excessive packet loss in communication, so that one of them alone can not provide estimation accurately. Therefore, the link quality estimators based on a single link property have limitations and they are incapable of providing a fine grain estimation of link quality.

V. Link Quality Estimation Model

The essence of link quality estimation is the classification of the link quality levels. So, SVM is applied to link quality estimation for its excellent performance of classification. In this paper, the link quality will be categorized into 5 levels. Because SVM can only handle binary classification problems, a new link quality estimation model is proposed through combining the SVMs and decision tree.

1. The levels of link quality

To obtain the levels of link quality, the relationship between PRR and distance is analyzed as shown in Fig.4. The PRR is above 90% when distance is within 30 meters, and the link quality is very good. While, PRR decreases to below 10% when distance is over 100 meters, and the link quality is the exact opposite. But for distance between 30 and 100 meters, PRR decreases from 90% to 10%, and the link quality also ranges from good to bad. So, the link quality is divided into five levels.

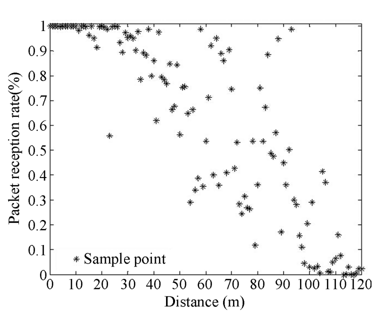


Fig. 4. Scatter distribution of PRR and distance

Table 2 shows the link quality levels in the experiment scenarios of vehicles or pedestrians. If there are different scenarios, or the special requirements of the communication quality, the thresholds of link quality levels should be adjusted.

Very good link corresponds to the effective region, while Good link, Intermediate link and Bad link correspond to the transitional region. Very bad link corresponds to the blank region. Next, SVM method is used to establish the link quality estimation model based on SVM with decision tree, and the link quality estimation is transformed into link quality classification.

Table 2. Link quality estimation classification range

Link quality level	PRR
Very good link	$90\% \le PRR \le 100\%$
Good link	$75\% \le PRR < 90\%$
Intermediate link	$45\% \le PRR < 75\%$
Bad link	$10\% \le PRR < 45\%$
Very bad link	$0\% \le PRR < 10\%$

2. Link quality estimation model base on SVM

The proposed model combines SVM and decision tree to convert the multi-class classification into a binary classification. Fig.5 shows the corresponding link quality levels of each leaf node in the decision tree.

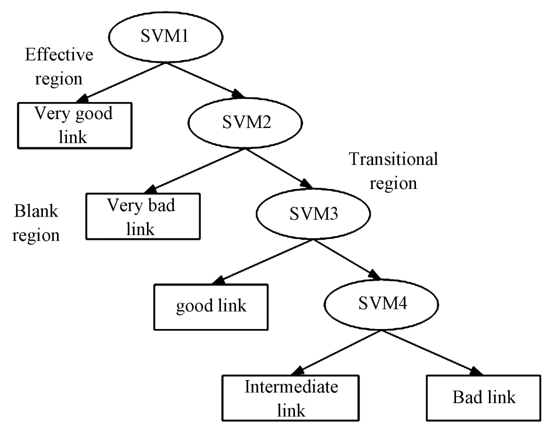


Fig. 5. Link estimation model based on SVM

The levels of link quality have been shown in Table 2. The SVM1 is used to differentiate the effective region of communication link at first, and then the rest data is processed by SVM2 to separate out the link quality of blank region and transitional region. In the transitional region, SVM3 tells the Good link apart. Similarly, the Intermediate link and Bad link are separated by SVM4. In other words, link quality is divided into effective and blank region in a fast way, followed by dividing transitional region into Good link and Bad link. In fact, the process is nested somehow.

The primary task of the model is to make a speedy decision on link for good quality and poor quality through fewer probe packets. But for the division of relative fuzzy transitional region, a more extensive way is still adopted. Essentially, establishing the SVM classification model is a process to select and optimize parameters of SVM. Thus, the link quality estimation model based on SVM with decision tree can be described as follow:

a) Train the estimation model on training sample set from test platform, and get the optimal parameters for SVM.

b) Estimate the current link quality by estimation model obtained in the first step with testing sample set.

c) Verify and analyze the accuracy of the model with experiments.

3. The selection of kernel function

In Ref.[15] the performance of the kernel function for SVM has been analyzed in details by comparing the Linear (Linear), Polynomial (poly real), Radial basis function (RBF), and sigmoid kernel function, *etc.* Wang *et al.* have also put forward several suggestions for the selection of SVM kernel function through studying and calculating of kernel matrix. Their experiments have shown that the RBF kernel has a high generalization performance.

To verify the performance of RBF kernel function on SVM model, Poly kernel function is chosen to make a comparison. Finally the kernel function which performs better is used to build the SVM link quality estimation model.

VI. Experiments and Analysis

1. Experiments

The testbed consists of a single-hop network with a pair of TelosB series nodes TX and RX, positioned in an outdoor environment (a road at the university). These nodes are equipped with a 16-bit low-power MSP430 microprocessor whose work frequency is 8M and a CC2420 wireless transceiver chip which is designed according to IEEE802.15.4. TX is the sender, while RX is the receiver which is connected to the PC through USB. Besides, we have developed a link quality test software platform that runs on the PC to monitor and to analyze the experimental results. The test platform, developed in C++, has the functions of packet parsing, experimental data display, and data logging into a SQL server 2000 database. The data analysis is performed with LIBSVM[18] and SPSS. The nodes are programmed with network embedded system C (nesC) in TinyOS2.1 environment.

In the experiments, a test-packet is consisted of sequence number, RSSI and LQI. The RSSI and LQI value can be read directly from the registers of nodes. To calculate PRR, a counter has been set in the test platform program to record the number of received packets. Therefore, the PRR can be obtained directly from our test platform.

All experiments are done on road, as shown in Fig.6. The sender and the receiver are placed in a distance, which is increased by 1 meter each time. 1000 packets are sent to the receiver node in each experiment.

Fig.7 depicts the experimental results displayed on the PC test platform. It records the packet sequence number and the receiving time, as well as RSSI and LQI.

2. Training and validation for estimation model

The SVM estimation model is trained by the sample set from 121 groups of experimental data in Section IV. The sample set includes RSSI, LQI and the link levels divided by PRR. Moreover, K-fold cross-validation algorithm is selected to solve the optimization of the parameters. The training process of SVM model based on RBF and the classification results are as follows.

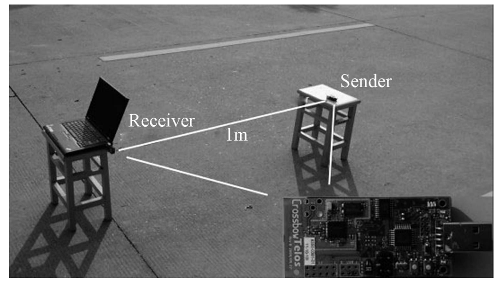


Fig. 6. The experimental scene

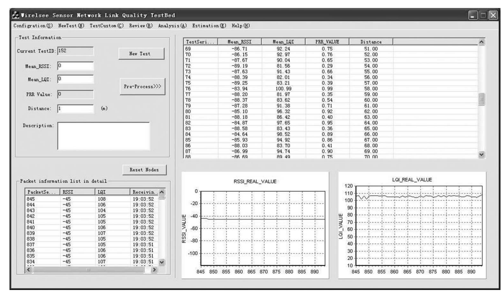


Fig. 7. The Link quality test platform on PC

Firstly, SVM1 is used to tell apart the effective region and other regions (*i.e.* transitional region and blank region). Effective region means Very good link and it is shown as a positive sample in Fig.8, and similarly for the others.

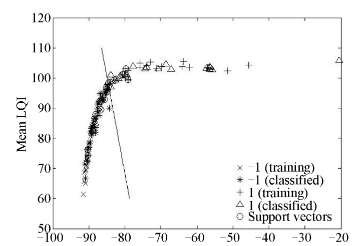


Fig. 8. The efficiency of classification for SVM1

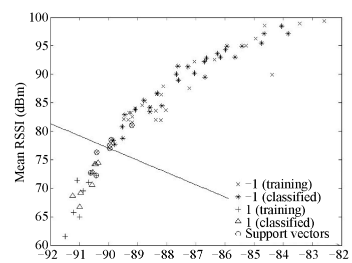


Fig. 9. The efficiency of classification for SVM2

Then the rest of samples are used as input for SVM2 to sort out the link quality of transitional region, namely the corresponding Very bad link, as shown in Fig.9. The division of Link quality at the transitional region is sequentially processed by SVM3 and SVM4 (shown in Fig.10, Fig.11). As a result, the levels of link quality, including Good link, Intermediate link and Bad link, are gained.

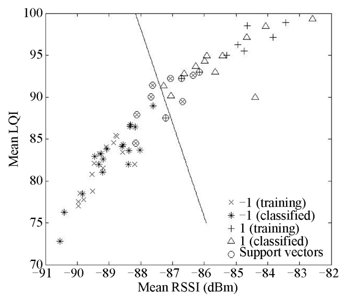


Fig. 10. The efficiency of classification for SVM3

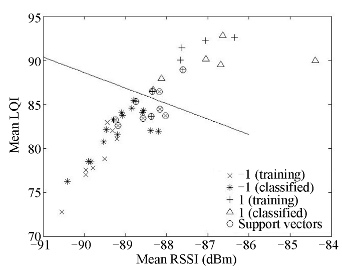


Fig. 11. The efficiency of classification for SVM4

Optimal SVM estimation model is obtained by using K-fold cross-validation. To test SVM's generalization ability, *i.e.* the performance of SVM estimation model, the rest part sample set which has not been used to train the model is used. Table 3 indicates the performance and accuracy of the proposed model.

Table 3. The performance of link quality estimation model based on SVM (RBF)

SVM	C	g	Accuracy
SVM1	100	1.5	100%
SVM2	100	1	87.42%
SVM3	100	1	100%
SVM4	100	1.9	86.11%

To verify the superiority of RBF kernel, Polynomial kernel function is chosen as the contrast of the RBF kernel to test the effect of different kernel functions on SVM model in the experiments. The results are shown in Table 4.

Table 4. The performance of link quality estimation model based on SVM (Poly)

SVM	C	d	Accuracy
SVM1	100	3	77.82%
SVM2	100	4	75.95%
SVM3	100	3	43.48%
SVM4	100	3	93.10%

Comparing Table 3 with Table 4, it can be obviously found that the radial basis kernel function is superior and more suitable for the small sample of the SVM model.

Table 5. The accuracy of link quality estimation model

Model	Accuracy	Input parameters	Output levels
LQI-PRR	88.57%	LQI	\
FLI	92.66%	PRR,CV,DX	2
SVM	93.38%	RSSI,LQI	5

Finally, the accuracy of link quality estimation model proposed in this paper is compared with the ones of LQI-PRR[12] and FLI[14] models. Experiments have been conducted with the same experimental data in the same experimental scenario. It can be seen from Table 5 that the accuracy of estimation model based on SVM is 93.38%, and it is better than other two methods. Compared with the FLI model which has two link quality levels, our model can reflect of link quality more detailed. The FLI method costs lots of energy in sending probe packets, and it needs large number of calculations to obtain CV(Coefficient of variation) and D_α (the correlation distance). Therefore, the proposed model is more efficient and energy saving.

VII. Conclusions

In this paper, we have presented a link quality estimation model based on SVM with decision tree, which applies the SVM theory to the link quality estimation after extensive analysis of the link characteristics. In the model, RSSI and LQI are selected as estimation parameters, and link quality estimation problem is transformed into classification problem. Finally, by using SVM classification method, a link quality estimation model is established. Experimental results demonstrate the rationality of the proposed model. In contrast to the recent published link quality estimation models, the proposed model has better performance on estimation accuracy. Furthermore, it costs less the energy consumption than the one caused by sending a large number of probe packet, and it extends the network lifetime.

References

[1] X.Y. Kui, H.K. Du and J.B. Liang, "An energy-balanced connected dominating sets for data gathering in wireless sensor networks", *Acta Electronica Sinica*, Vol.41, No.8, pp.1521–1528, 2013. (in Chinese)

[2] A. Woo, T. Tong, D. Culler, "Taming the underlying challenges of reliable multihop routing in sensor networks", *Proc. of International Conference on Embedded Networked Sensor Systems*, New York, USA, pp.14–27, 2003.

[3] Y. Xu and W.C. Lee, "Exploring spatial correlation for link quality estimation in wireless sensor networks", *Proc. of IEEE International Conference on Pervasive Computing & Communications (PerCom)*, Pisa, Italy, pp.200–211, 2006.

[4] G. Zhou, T. He and S. Krishnamurthy, "Impact of radio irregularity on wireless sensor networks", *Proc. of International*

Conference on Mobile Systems, Boston, Massachusets, USA, pp.125–138, 2004.

[5] Y.J. Li, Z. Wang and Y.X. Sun, "Analyzing and modeling of the wireless link for sensor networks", *Chinese Journal of Sensors and Actuators*, Vol.20, No.8, pp.1846–1851, 2007. (in Chinese).

[6] N. Baccour, A. Koubâa, M.B. Jamâa, *et al.*, "RadiaLE: A framework for designing and assessing link quality estimators in wireless sensor networks", *Ad Hoc Networks*, Vol.9, No.7, pp.1165–1185, 2011.

[7] P.G. Sun, H. Zhao, D.D. Luo, *et al.*, "Study on measurement of link communication quality in wireless sensor networks", *Communications Journal*, Vol.28, No.10, pp.14–22, 2007. (in Chinese)

[8] N. Reijers, G. Halkes and K. Langendoen, "Link layer measurements in sensor networks", *Proc. of IEEE International Conference on Mobile Ad-hoc & Sensor Systems*, pp.224–234, 2004.

[9] J. Zhu, H. Zhao, X.Y. Zhang, *et al.*, "LQI-based evaluation model of wireless link", *Journal of Northeastern University* (*Natural Science*), Vol.29, No.9, pp.1262–1265, 2008. (in Chinese)

[10] M. Senel, K. Chintalapudi, D. Lal, *et al.*, "A Kalman filter based link quality estimation scheme for wireless sensor networks", *Proc. of IEEE Global Telecommunications Conference* (*GLOBECOM*), Washington, DC., USA, pp.875–880, 2007.

[11] L.L. Liu, Y.L. Fan, J. Shu, *et al.*, "A link quality prediction mechanism for wsns based on time series model", *Proc. of International Conference on Ubiquitous Intelligence & Computing & International Conference on Autonomic & Trusted Computing*, Xi'an, China, pp.175–179, 2010.

[12] J. Luo, L. Yu, D. Zhang, *et al.*, "A new link quality estimation mechanism based on LQI in WSN", *Information Technology Journal*, Vol.12, No.8, pp.1626–1631, 2013.

[13] X.Y. Peng, D.W. Pan, Y. Peng, *et al.*, "On multiple time scales link estimation in wireless sensor network", *Acta Electronica Sinica*, Vol.39,No.3A, pp.80–85, 2011.(in Chinese)

[14] Z.Q. Guo, Q. Wang, M.H. Li, *et al.*, "Fuzzy logic based multi-dimensional link quality estimation for multi-hop wireless sensor networks", *IEEE Sensors Journal*, Vol.13, No.10, pp.3605–3615, 2013.

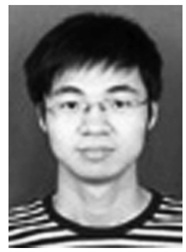
[15] F. Wang, K. He, Y. Liu, *et al.*, "Research on the selection of kernel function in SVM based facial expression recognition", *Industrial Electronics and Applications*, Vol.34, No.9, pp.789–798, 2013.

[16] C. Burges, "A tutorial on support vector machines for pattern recognition", *Data Mining and Knowledge Discovery*, Vol.2, No.2, pp.121–167, 1998.

[17] M.X. Chu, A.N. Wang and R.F. Gong, "Improvement on least squares twin support vector machine for pattern classification", *Acta Electronica Sinica*, Vol.42, No.5, pp.998–1003, 2013. (in Chinese)

[18] C.C. Chang and C.J. Lin, "LIBSVM: A library for support vector machines", *ACM Transactions on Intelligent Systems and Technology*, Vol.2, No.3, 2011.

SHU Jian was born in 1964. He is a professor and M.S. supervisor at the School of Software in Nanchang Hangkong University. His main research interests include wireless sensor network, embedded system and software engineering. (Email: shujian@nchu.edu.cn)



LIU Song was born in 1990. He received the B.S. degree in 2013 from Nanchang Hangkong University, where he is currently pursuing the M.S. degree. His main research interest is wireless sensor network. (Email: liusong090451@163.com)

LIU Linlan (corresponding author) was born in 1968. She is a professor and M.S. supervisor at the School of Information Engineering in Nanchang Hangkong University. Her main research interests include wireless sensor network and distributed system. (Email: liulinlan@nchu.edu.cn)

ZHAN Liqin was born in 1990. She received the B.S. degree in 2013 from Nanchang Hangkong University, where she is currently pursuing the M.S. degree. Her main research interest is software engineering.

HU Gang was born in 1991. He received the B.S. degree in 2012 from Nanchang Hangkong University, where he is currently pursuing the M.S. degree. His main research interest is wireless sensor network.