Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2015, Article ID 828493, 8 pages
http://dx.doi.org/10.1155/2015/828493

Research Article

Link Quality Prediction via a Neighborhood-Based Nonnegative Matrix Factorization Model for Wireless Sensor Networks

Yuxin Zhao,[1] Shenghong Li,[1] and Jia Hou[2]

[1] *Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China*
[2] *School of Electronic and Information Engineering, Soochow University, Suzhou 215006, China*

Correspondence should be addressed to Shenghong Li; shli@sjtu.edu.cn

Received 5 November 2014; Accepted 12 December 2014

Academic Editor: Qilian Liang

A core factor to consider when designing wireless sensor networks is the reliable and efficient transmission of massive data from source to destination. In practical situations, data transmission is often disrupted by link interference and interruption resulting in the data losses. Link quality prediction is an important approach to solve this problem. By estimating the link quality based on the past knowledge and information, link quality prediction is essential for routing decisions of future data transmission. Traditional link quality prediction algorithms are simply based on the statistical information of the links in the wireless sensor network. By introducing complex network theory and machine learning techniques, we propose a neighborhood-based nonnegative matrix factorization model to predict link quality in wireless sensor networks. Our model learns latent features of the nodes from the information of past data transmissions combing with local neighborhood structures of the underlying network topology and then estimates the link quality depending on the common latent features of the two nodes between the link. Extensive experiments on both real-world networks and simulation networks demonstrate the effectiveness and efficiency of our proposed model.

1. Introduction

Wireless sensor network (WSN) is a special kind of mobile communication network, which exhibits the properties of dynamic network topology, multihop communication, and limited energy [1, 2]. In wireless sensor network, stable links have great significance since many critical applications fundamentally rely on efficient and reliable data transmission from source node to sink node [3, 4]. However, in practical situations, data transmission is often disrupted by link interference and interruption resulting in the data losses [5, 6]. When the link quality deteriorates, data losses are inevitable since the size of internal buffers of intermediate nodes is limited. In the presence of poor link quality, the sender could easily fill up the buffer so that there is no sufficient time and space to transmit all packets reliably. This problem is especially severe for those applications which strongly require real-time data transmission or high-fidelity data transmission [7]. The main reason for this problem is the lack of feedback of link layer information to the upper layer applications and protocols.

Link quality prediction is an important approach to decrease the probability of data losses in wireless sensor network. For most applications in wireless sensor networks [8–10], it is necessary that each node has thorough knowledge about its direct neighbors. This information is collected and provided by neighborhood management protocols. One important criterion used by neighborhood management protocols to determine the importance of a node is the quality of the communication between nodes, which is provided by link quality prediction [11]. By estimating the link quality based on the past knowledge and information, link quality prediction is essential for routing decisions of future data transmission.

However, link quality prediction remains a challenging task due to the dynamic nature of wireless links. Firstly, the correlations between nodes cannot be directly obtained from the network. Secondly, in some situations, network data is quite sparse that the links used for data transmission are only a small proportion of all possible links.

Motivated by the great practical significance, many algorithms for link quality prediction have been proposed recently. The most existing algorithms simply use statistical information of the links to evaluate the link quality. In this paper, we proposed a nonnegative matrix factorization model for predicting link quality in wireless sensor networks, by introducing complex network theory and machine learning techniques. Our model is a supervised learning model. In our model, we associate the probability of a link with a nonnegative strength variable, which is related with the latent features of the nodes between the link. We also consider the influence of the neighborhood structure and make the latent features of each node learnt from the past data transmissions combing with local neighborhood structures of the underlying network topology.

The rest of the paper is organized as follows. In Section 2, we introduce the related background of link quality prediction in wireless sensor networks. Our proposed neighborhood-based nonnegative matrix factorization model for link quality prediction is described in detail in Section 3. The experimental results and discussions are reported in Section 4. Finally, Section 5 gives the conclusion of this paper.

2. Related Work

In recent years, various approaches have been proposed to provide a meaningful metric describing the actual link-quality and then predicting its future behavior. In this section, an overview of popular link quality prediction algorithms is presented at first. Then, we make an introduction of the link prediction theory in complex networks.

2.1. Link Quality Prediction for Wireless Sensor Networks. Basically, the measurements of link quality can be classified into two categories: physical metrics and logical metrics. Physical metrics depend on the radio hardware and evaluate link quality by the signal strength of a received packet. No additional costs are required by physical metrics, since the measurement is performed by the receiver hardware every time, a packet is received. Common physical metrics include the received signal strength indication (RSSI), the link-quality indication (LQI), and signal-to-noise ratio (SNR). On the other hand, logical metrics estimate the link quality by keeping track of message losses. They do not depend on specific hardware so that they are not influenced by the characteristics of hardware. Typical examples of logical metrics are packet success rate (PSR) [12], required number of packets (RNP) [13], and expected transmission count (ETX) [14].

Woo et al. [12] made the first attempt to estimate the link quality by proposing the window mean with exponentially weighted moving average (WMEWMA). This metric computes the average success rate of the link over a time period and smoothens the average with an EWMA. WMEWMA predicts the PSR of the link and has been widely adopted in WSNs. de Couto et al. [14] proposed ETX which uses the number of expected transmissions for a successful packet transmission as the evaluation of link quality. The ETX of a link is calculated using the forward and reverse delivery ratios of the link. The number of received packets within a fixed time period is counted and compared to the number of expected packets that are periodically broadcasted by each node. RNP introduced by Cerpa et al. [13] incorporates the distribution of losses within the time period. This metric is based on their observation fact that a link with consecutive losses should be rated lower than links with discrete losses.

Weyer et al. [15] proposed adaptive link estimator (ALE), which is an EWMA filter of the measurement PSR. Links with different qualities have different weights in the EWMA filter. For good links, ALE uses a higher weight for more stable estimation, while links with a lower quality are estimated in an agile fashion for a faster reaction.

A Kalman Filter Based Link-Quality Estimation was proposed by Senel et al. [16]. A Kalman filter is used to smoothen the RSSI of successfully received packets and the noise floor is subtracted to obtain an estimation of the SNR. The final PSR is derived by applying a hardware specific SNR-PSR mapping for the transceiver. This approach is very complicated and only applicable to the cases in which SNR and PSR are strictly correlated.

Chen et al. [17] proposed a model to predict link interruption and route interruption in wireless sensor networks by the historical link information and channel state obtained by periodic detection. The periodicity of environmental changes is utilized to help predict link interruption.

LISP proposed by Ma et al. [18] is based on the premise that, in order to achieve the best performance, the application layer behavior should be aware of the link layer conditions and adjust its behavior accordingly. They used the state space model to predict link quality and then provided these estimates as a system level service to application developers.

Wang et al. [19] introduced the supervised learning techniques to predict link quality in wireless sensor network. In their approach, link quality prediction is modeled as a classification problem with the features of the link, including forward probability, backward probability, channel load, and node depth from the source node.

2.2. Link Prediction for Complex Networks. In complex network theory, the research field of link prediction is very similar to the link quality prediction in wireless sensor network. Link prediction in complex networks focuses on estimating the likelihood of the existence of links in the network rather than the link quality.

Link prediction algorithms can be roughly classified into two classes: unsupervised models and supervised models. In unsupervised models, the probability of the existence of a link is measured by some specific similarity indices between the two nodes. Local similarity indices [20–24] only depend on the information of the neighborhoods, such as the common neighbors. Global similarity indices [25–27] require the entire topological information of the network from the perspectives of paths, random walks, and other properties.

The similarity in unsupervised models is predefined and is invariant to the specific structure of the input networks.

On the other hand, supervised models use the supervised learning approaches. They usually propose some patterns of the link behaviors and learn a series of parameters according to the observed links. Some popular approaches are hierarchical structure model (HSM) [28], stochastic block model (SBM) [29], and latent factor model (LFM) [30, 31]. The former two models use the explicit topological properties of the network, while latent factor models depend on latent features of the network, which can be viewed as an implicit representation of the network topological information. The main drawback of the former two models is the high calculation complexity, which makes them only applicable to small networks. By contrast, the latent factor models can be trained in linear time with the number of observed links.

3. Our Model

In our model, link quality prediction is formally modeled as a supervised learning problem. The information of past data transmissions is the training set of the model. Given a wireless sensor network, let the link between node i and node j be denoted by link (i, j). When the packet is successfully transmitted through a link (i, j), we have the label $A_{ij} = 1$, otherwise, the label $A_{ij} = 0$. The predicted score of link quality for link (i, j) obtained by our model is indicated by $\widehat{A}_{ij}$. Our model is learnt from the training set by minimizing the errors between the practical labels and the predicted scores. The final goal is to use the model to predict the link quality or successful transmission probability of all the possible links in the network.

3.1. Basic Latent Factor Model. In basic latent factor model (LFM), each node i is associated with a latent feature vector $F_i \in \mathbb{R}^k$, where k is the number of latent features. The latent features of all the nodes in the network constitute the latent feature matrix $\mathbf{F} \in \mathbb{R}^{n\times k}$, where n is the number of nodes in the network. Under the assumption that the link quality between two nodes is higher if they have more similar latent features, the predicted score of the link quality between node i and j can be written as

$$\widehat{A}_{ij} = L\left(\sum_{f=1}^{k} F_{if}F_{jf}\right) = L\left(F_iF_j^T\right), \quad (1)$$

where $L(\cdot)$ is a link function, which is monotonically increasing and is usually taken as identity function or sigmoid function.

The latent features of each node can be learnt by solving the following optimization problem:

$$\min_{\mathbf{F}} \sum_{(i,j)\in O} C\left(A_{ij}, L\left(F_iF_j^T\right)\right) + \Omega(\mathbf{F}), \quad (2)$$

where O is the set of the past successful and unsuccessful transmissions through the links in the network, $C(\cdot)$ is a loss function, and $\Omega(\cdot)$ is a regularization term that prevents overfitting. As suggested in [32], regularized square error loss function and L_2 norm regulation are especially suitable for latent factor models in practical applications. Then, the optimization problem can be rewritten as follows:

$$\min_{\mathbf{F}} \sum_{(i,j)\in O} \left(A_{ij} - L\left(F_iF_j^T\right)\right)^2 + \lambda\left(\|\mathbf{F}\|_F^2\right), \quad (3)$$

where λ is the regulation parameter and $\|\mathbf{F}\|_F^2$ is the Frobenius norm of matrix $\mathbf{F}$.

Stochastic gradient descent method is usually used to solve this optimization problem. The total training process exhibits linear time with the number of transmissions in the training set.

3.2. Neighborhood-Based Nonnegative Matrix Factorization Model. In basic latent factor model, the latent features of some nodes may have negative values, which may mislead the whole approach. Moreover, the influences of the neighborhood structure in the network are not considered in basic latent factor model.

We first assume that the latent feature matrix $\mathbf{F}$ is a nonnegative matrix and each pair of nodes in the network has a latent interaction of nonnegative strength variable X_{ij}. The transmission between the two nodes can be successful only if the corresponding strength variable $X_{ij} > 0$. Consider that nodes i and j generate an interaction of strength $X_{ij}^{(f)}$ with each latent feature f using a Poisson distribution with mean $F_{if} \cdot F_{jf}$:

$$X_{ij}^{(f)} \sim \text{Pois}\left(F_{if} \cdot F_{jf}\right). \quad (4)$$

The strength X_{ij} between nodes i and j is the sum of $X_{ij}^{(f)}$ for all the latent features:

$$X_{ij} = \sum_{f=1}^{k} X_{ij}^{(f)} \sim \text{Pois}\left(\sum_{f=1}^{k} F_{if} \cdot F_{jf}\right) = \text{Pois}\left(F_iF_j^T\right). \quad (5)$$

Then, the link quality between the pair of nodes (i, j) can be figured out:

$$P\left(X_{ij} > 0\right) = 1 - P\left(X_{ij} = 0\right) = 1 - \exp\left(-F_iF_j^T\right). \quad (6)$$

It is expected that the nodes with larger values in the same latent features have a higher-quality link between them.

Let latent feature matrix $\mathbf{F}$ be nonnegative and $1 - \exp(\cdot)$ be the link function we can reformulate the optimization problem (3) as

$$\min_{\mathbf{F}>0} \sum_{(i,j)\in O} \left(A_{ij} - 1 + \exp\left(-F_iF_j^T\right)\right)^2 + \lambda\left(\|\mathbf{F}\|_F^2\right). \quad (7)$$

Taking the node-specific biases into account, the optimization is then

$$\min_{\mathbf{F},b>0} \sum_{(i,j)\in O} \left(A_{ij} - 1 + \exp\left(b_i + b_j - F_iF_j^T\right)\right)^2 + \lambda\left(\|b\|_2^2 + \|\mathbf{F}\|_F^2\right), \quad (8)$$

where b is the node-specific bias vector, which is similar to the intercept terms in standard supervised learning.

Now, let us consider the influence of the neighborhoods on the link probability between the nodes. In unsupervised models, several link metrics are defined in the following form:

$$s_{ij} = \sum_{u\in\Gamma(i)\cap\Gamma(j)} w_u, \quad (9)$$

where $\Gamma(i)$ is the set of neighbors of node i and w_u is a measurement of the topological properties of node u. Common Neighbors Index (CN) [20] directly counts the neighborhood overlap of the two nodes so that $w_u = 1$. For Adamic-Adar Index (AA) [23] and Resource Allocation Index (RA) [24], w_u is related with the node degree k_u that $w_u = 1/\log(k_u)$ in AA index and $w_u = 1/k_u$ in RA index. Here, we extend this form by making each node have different influences upon the links:

$$\widehat{A}_{ij} = \sum_{u\in\Gamma(i)\cap\Gamma(j)} \left(w_{ui} + w_{uj}\right). \quad (10)$$

In order to reduce the number of parameters, we factorize the matrix $\mathbf{w} = \mathbf{X}^T\mathbf{Y}$, $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{n\times k}$. Then, the predicted score of link quality can be reformulated as

$$\widehat{A}_{ij} = \left(X_i + X_j\right) \sum_{u\in\Gamma(i)\cap\Gamma(j)} Y_u^T. \quad (11)$$

Finally, we combine the previous two models and predict the link quality by

$$\widehat{A}_{ij} = 1 - \exp\left(b_i + b_j - \alpha\cdot F_iF_j^T - \beta\cdot\left(F_i + F_j\right)\sum_{u\in\Gamma(i)\cap\Gamma(j)} Y_u^T\right), \quad (12)$$

where α and β are two strength coefficients. Here, we arbitrarily let $\mathbf{X} = \mathbf{F}$ to reduce the number of parameters. Thus, model parameters are learnt by solving the optimization problem associated with

$$\min_{\mathbf{F},b,\mathbf{Y}>0} \sum_{(i,j)\in O} \left(A_{ij} - 1 + \exp\left(b_i + b_j - \alpha\cdot F_iF_j^T - \beta\cdot\left(F_i + F_j\right)\sum_{u\in\Gamma(i)\cap\Gamma(j)} Y_u^T\right)\right)^2 + \lambda\left(\|b\|_2^2 + \|\mathbf{F}\|_F^2 + \|\mathbf{Y}\|_F^2\right). \quad (13)$$

An optimal solution of this optimization problem can be obtained using stochastic gradient descent method. Let the prediction error $A_{ij} - \widehat{A}_{ij}$ be denoted by e_{ij}. We loop through all observed links in the network. For a given observed transmission through link (i, j) in the training set, we modify the parameters by moving in the opposite direction of the gradient, yielding the following:

(i) $b_i \leftarrow \max(0, b_i + \gamma\cdot((1 - \widehat{A}_{ij})\cdot e_{ij} - \lambda\cdot b_i))$;

(ii) $b_j \leftarrow \max(0, b_j + \gamma\cdot((1 - \widehat{A}_{ij})\cdot e_{ij} - \lambda\cdot b_j))$;

(iii) $F_i \leftarrow \max(0, F_i + \gamma\cdot((1 - \widehat{A}_{ij})\cdot e_{ij}\cdot(\alpha\cdot F_j + \beta\cdot\sum_{u\in\Gamma(i)\cap\Gamma(j)} Y_u) - \lambda\cdot F_i))$;

(iv) $F_j \leftarrow \max(0, F_j + \gamma\cdot((1 - \widehat{A}_{ij})\cdot e_{ij}\cdot(\alpha\cdot F_i + \beta\cdot\sum_{u\in\Gamma(i)\cap\Gamma(j)} Y_u) - \lambda\cdot F_j))$;

(v) for all $u \in \Gamma(i)\cap\Gamma(j) : Y_u \leftarrow \max(0, Y_u + \gamma\cdot(\beta\cdot(1 - \widehat{A}_{ij})\cdot e_{ij}\cdot(F_i + F_j) - \lambda\cdot Y_u))$,

where γ is the learning rate.

Due to the link function, the predicted score of link quality in our model is in the range $[0, 1]$, which represents the probability of successful transmission through the link. A link with larger predicted score is more likely to complete a packet transmission in the network. For the links in the training set, the predicted score is approximate to the statistical success probability of past data transmissions. The link quality of the links without any transmission record is predicted according to the transmission information of the neighbor nodes combing with the topology of the underlying network.

4. Experimental Results

4.1. Evaluation on Complex Networks. We first apply our neighborhood-based nonnegative matrix factorization model to several real-world networks, which are widely used in link prediction literature. General information of these real-world networks is shown in Table 1. We also make comparisons with some unsupervised link prediction models, including Common Neighbors Index (CN) [20], Salton Index [22], Preferential Attachment Index (PA) [21], Adamic-Adar Index (AA) [23], and Karz Index [25]. The experiments are implemented by MATLAB 2009b running on a PC with a 3.0 GHz processor and 3 GB memory.

To evaluate the accuracy of different models, we adopt AUC proposed by Hanley and McNeil [33] as the basic measure for the experiments reported in this paper. The AUC value is defined as the probability in which a randomly chosen high-quantity link is assigned with a higher score than a randomly chosen low-quality link. If among n independent comparisons, there are n_1 times the high-quantity link having a higher score and n_2 times the scores are equal; the AUC value is

$$\mathrm{AUC} = \frac{n_1 + 0.5n_2}{n}. \quad (14)$$

If all the scores are randomly given, the AUC value should be approximated to 0.5. The degree to which the AUC value exceeds 0.5 indicates how much better the model performs than pure chance.

Table 1: General information of the real-world networks.

Network	Description	Node	Present link	Average degree
Karate	Zachary's karate club [35]	34	78	4.58
Dolphin	Social network of Lusseau's dolphins [36]	62	159	5.13
Usair	US air transportation system [37]	332	2126	12.81
Email	E-mail interchanges between members of the University of Rovira i Virgili [38]	1133	5451	9.62
Blog	Hyperlinks between blogs on US politics [39]	1222	16174	27.36
Protein	The interaction between proteins [40]	2473	6269	5.09
Powergrid	The topology of the power grid of the United States [41]	4941	6594	2.67
PGP	The interactions between users of pretty-good-privacy algorithm [42]	10680	24316	4.55

Table 2: The AUC values of different models on real-world networks.

	CN	Salton	PA	AA	Karz	Basic LFM	Our model
Karate	0.7035	0.6387	0.7461	0.7313	0.7877	0.8058	**0.8294**
Dolphin	0.7786	0.7076	0.6907	0.7851	0.8103	0.8109	**0.8275**
Usair	0.9368	0.8624	0.9017	0.9461	0.4136	0.9433	**0.9598**
Email	0.8541	0.8129	0.7814	0.8550	0.6415	**0.9131**	0.9105
Blog	0.9175	0.8469	0.8977	0.9205	0.4804	0.9292	**0.9384**
Protein	0.7624	0.6543	0.7232	0.7626	0.6322	0.8849	**0.8867**
Powergrid	0.5879	0.4411	0.4395	0.5878	**0.6587**	0.6344	0.6291
PGP	0.8371	0.6805	0.7117	0.8373	0.5245	0.8775	**0.8983**

For each network, the present links are partitioned into training set (90%) and test set (10%). The performances of different models on real-world networks are shown in Table 2.

As is shown in Table 2, our model performs the best among all the other models on most real-world networks and is only inferior to basic LFM on Email network and Karz Index on Powergrid network. For Karate network, Blog network, and PGP network, our model shows superiority over the unsupervised link prediction models and obvious improvement over basic LFM.

We also find out the reason why the performance of our model is inferior to that of Karz Index on Powergrid network. This is due to the fact that Powergrid network is a highly sparse network in which about 60% of the nodes only have one or two links connecting with other nodes. The sparsity makes many model parameters get insufficient training, which results in the fact that our model does not perform well on this network.

4.2. Simulation on Wireless Sensor Networks. To verify the prediction model, we also make simulations of our model on wireless sensor networks. Here, we use the Matlab as the simulation platform. The channel model is flat Rayleigh fading, carrier frequency is 20 kHz, the modulation is (QAM), and data transmission rate is set to 1000 bits/s. The packet size is set to 1000 bits. The topology of the wireless sensor network is randomly generated by the LFR approach [34]. An example network topology is shown in Figure 1, where the red node denotes the source node and the black node denotes the sink node. The training set contains all the packet transmission records in the wireless sensor network in a period of 1000 seconds. The packet transmission records in the next period of 200 seconds constitute the test set of the experiment.

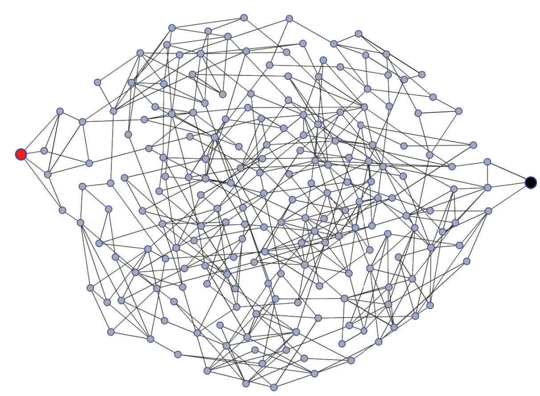

Figure 1: An example topology of wireless sensor network. The red node denotes the source node and the black node denotes the sink node.

Besides AUC, three other measures, known as precision, recall, and F-score, are also adopted to evaluate the performances of our model. For a given link, we use a threshold to determine whether the coming packet transmission through this link is successful or not. If the predicted link quality exceeds the threshold, the transmission is considered to be successful; otherwise, it is thought that the transmission would fail. Then, there are four possible situations. If the transmission is successful and prediction is success, it is counted as true positive. On the other hand, if the transmission is successful and prediction fails, it is counted as true negative. Similarly, If the transmission failed and prediction is successful, it is counted as false positive. And if the

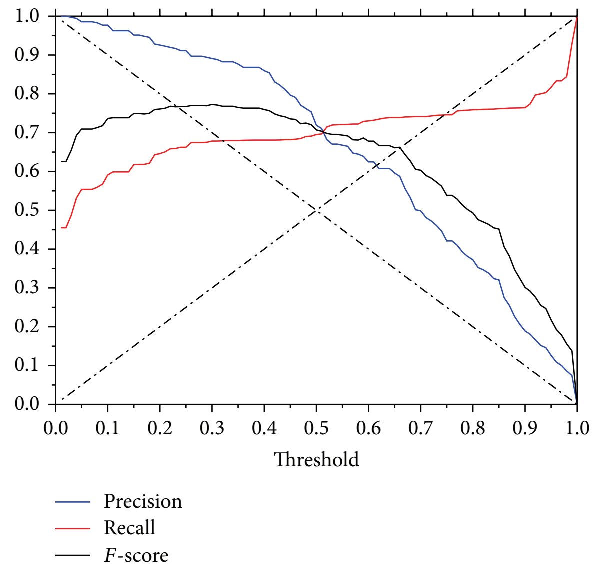

Figure 2: The precision, recall, and F-score of our model on wireless sensor networks with the threshold varying from 0 to 1.

transmission failed and prediction fails, it is counted as false negative. The precision is defined as the ratio of the number of true positives to total number of instances that are predicted to be positive:

$$\text{precision} = (\text{number of ture positives}) \cdot (\text{number of ture positives} + \text{number of negative positives})^{-1}. \tag{15}$$

The recall is defined as the ratio of the number of true positives to total number of instances that are actually positive:

$$\text{precision} = (\text{number of ture positives}) \cdot (\text{number of ture positives} + \text{number of false negatives})^{-1}. \tag{16}$$

F-score is an important measure of the model's accuracy by taking both the precision and recall into account. It can be written in the form of harmonic mean of precision and recall:

$$F = \frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}. \tag{17}$$

Figure 2 shows the precision, recall, and F-score of our model on wireless sensor networks with the threshold varying from 0 to 1. The two dashed lines are, respectively, the precision curve and recall curve by pure chance. Seen from the curves, the precision and recall of our model are significantly higher than pure chance in most situations. The F-score of our model reaches the highest value of 0.773 when the threshold is equal to 0.3. We also figure out that the AUC of our model is 0.827, which is far larger than 0.5. These measures indicate that our model is effective and promising for link quality prediction in wireless sensor networks.

Table 3: The improvements of data transmission with our model.

	Without our model	With our model
Successful transmission number	2289	4230
Failed transmission number	2326	1452
Successful transmission rate	49.60%	74.45%
Packet number received by sink node	19	43

We also use our model to select the next hop for the data transmission in the network. The neighbor node with larger predicted score has a higher probability to become the next hop of the packet transmission. The improvements of data transmission with our model in a certain period of time are shown in Table 3.

From the above experimental results, we can see that our model is effective for link quality prediction and very suitable for practical applications in wireless sensor networks.

5. Conclusion

In this paper, we propose a neighborhood-based nonnegative matrix factorization model for solving the problem of link quality prediction in wireless sensor networks. We extend link prediction model in complex networks to wireless sensor networks and use the supervised learning techniques to predict the link quality in wireless sensor networks. In our model, the quality of a link is associated with a nonnegative strength variable, which is related with the latent features of the nodes. The influence of the neighborhood structure is also taken into consideration. Thus, the latent features of each node are learnt from the overall topological structure combing with local neighborhood structures of the underlying network. We test our model on several real-world complex networks and also make simulations on wireless sensor networks. The experimental results demonstrate the effectiveness and efficiency of our model for link quality prediction in wireless sensor networks.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This research work is funded by the National Science Foundation of China (61271316), Shanghai Key Laboratory of Integrated Administration Technologies for Information Security, and Chinese National Engineering Laboratory for Information Content Analysis Technology.

References

[1] J. Yick, B. Mukherjee, and D. Ghosal, “Wireless sensor network survey,” *Computer Networks*, vol. 52, no. 12, pp. 2292–2330, 2008.

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

[3] J. Li and H. Gao, “Survey on sensor network research,” *Journal of Computer Research and Development*, vol. 45, no. 1, pp. 1–15, 2008.

[4] A. Cerpa, J. L. Wong, L. Kuang, M. Potkonjak, and D. Estrin, “Statistical model of lossy links in wireless sensor networks,” in *Proceedings of the 4th International Symposium on Information Processing in Sensor Networks (IPSN ’05)*, pp. 81–88, IEEE, Los Angeles, Calif, USA, April 2005.

[5] K. Akkaya and M. Younis, “A survey on routing protocols for wireless sensor networks,” *Ad Hoc Networks*, vol. 3, no. 3, pp. 325–349, 2005.

[6] C. Intanagonwiwat, R. Govindan, and D. Estrin, “Directed diffusion: a scalable and robust communication paradigm for sensor networks,” in *Proceedings of the 6th Annual International Conference on Mobile Computing and Networking (MOBICOM ’00)*, pp. 56–67, ACM, Boston, Mass, USA, August 2000.

[7] Q. Ren and Q. Liang, “Throughput and energy-efficiency-aware protocol for ultra wideband communication in wireless sensor networks: a cross-layer approach,” *IEEE Transactions on Mobile Computing*, vol. 7, no. 6, pp. 805–815, 2008.

[8] N. Xu, “A survey of sensor network applications,” *IEEE Communications Magazine*, vol. 40, no. 8, pp. 102–114, 2002.

[9] S. N. Pakzad, G. L. Fenves, S. Kim, and D. E. Culler, “Design and implementation of scalable wireless sensor network for structural monitoring,” *Journal of Infrastructure Systems*, vol. 14, no. 1, pp. 89–101, 2008.

[10] A. Arora, P. Dutta, S. Bapat et al., “A line in the sand: a wireless sensor network for target detection, classification, and tracking,” *Computer Networks*, vol. 46, no. 5, pp. 605–634, 2004.

[11] C. Renner, S. Ernst, C. Weyer, and V. Turau, “Prediction accuracy of link-quality estimators,” in *Wireless Sensor Networks*, vol. 6567 of *Lecture Notes in Computer Science*, pp. 1–16, Springer, Berlin, Germany, 2011.

[12] A. Woo, T. Tong, and D. Culler, “Taming the underlying challenges of reliable multihop routing in sensor networks,” in *Proceedings of the 1st International Conference on Embedded Networked Sensor Systems (SenSys ’03)*, pp. 14–27, November 2003.

[13] A. Cerpa, M. Potkonjak, J. L. Wong, and D. Estrin, “Temporal properties of low power wireless links: Modeling and implications on multi-hop routing,” in *Proceedings of the 6th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MOBIHOC ’05)*, pp. 414–425, May 2005.

[14] D. S. J. de Couto, D. Aguayo, J. Bicket, and R. Morris, “A high-throughput path metric for multi-hop wireless routing,” *Wireless Networks*, vol. 11, no. 4, pp. 419–434, 2005.

[15] C. Weyer, S. Unterschutz, and V. Turau, “Connectivity-aware neighborhood management protocol in wireless sensor networks,” in *Proceedings of 7th GI/ITG KuVS Fachgesprach ’Drahtlose Sensornetze’ (FGSN ’08)*, pp. 23–26, September 2008.

[16] M. Senel, K. Chintalapudi, D. Lal, A. Keshavarzian, and E. J. Coyle, “A Kalman Filter based link quality estimation scheme for wireless sensor networks,” in *Proceedings of the 50th Annual IEEE Global Telecommunications Conference (GLOBECOM ’07)*, pp. 875–880, Washington, DC, USA, November 2007.

[17] J. Chen, Y. Han, D. Li, and J. Nie, “Link prediction and route selection based on channel state detection in UASNs,” *International Journal of Distributed Sensor Networks*, vol. 2011, Article ID 939864, 11 pages, 2011.

[18] X. Ma, S. Djouadi, and Q. Cao, “LIPS: link prediction as a service for adaptive data aggregation in wireless sensor networks,” in *Proceedings of the 7th IEEE International Conference on Distributed Computing in Sensor Systems (DCOSS ’11)*, pp. 1–8, June 2011.

[19] Y. Wang, M. Martonosi, and L.-S. Peh, “Predicting link quality using supervised learning in wireless sensor networks,” *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 11, no. 3, pp. 71–83, 2007.

[20] M. E. J. Newman, “Clustering and preferential attachment in growing networks,” *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 64, no. 2, 2001.

[21] A.-L. Barabási and R. Albert, “Emergence of scaling in random networks,” *Science*, vol. 286, no. 5439, pp. 509–512, 1999.

[22] G. Salton and M. J. McGill, *Introduction to Modern Information Retrieval*, MuGraw-Hill, Auckland, New Zealand, 1983.

[23] L. A. Adamic and E. Adar, “Friends and neighbors on the Web,” *Social Networks*, vol. 25, no. 3, pp. 211–230, 2003.

[24] T. Zhou, L. Lü, and Y.-C. Zhang, “Predicting missing links via local information,” *European Physical Journal B*, vol. 71, no. 4, pp. 623–630, 2009.

[25] L. Katz, “A new status index derived from sociometric analysis,” *Psychometrika*, vol. 18, no. 1, pp. 39–43, 1953.

[26] P. Y. Chebotarev and E. V. Shamis, “The matrix-forest theorem and measuring relations in small social groups,” *Automation and Remote Control*, vol. 58, no. 9, pp. 1505–1514, 1997.

[27] W. Liu and L. Lü, “Link prediction based on local random walk,” *Europhysics Letters*, vol. 89, no. 5, Article ID 58007, 2010.

[28] A. Clauset, C. Moore, and M. E. J. Newman, “Hierarchical structure and the prediction of missing links in networks,” *Nature*, vol. 453, no. 7191, pp. 98–101, 2008.

[29] R. Guimerà and M. Sales-Pardo, “Missing and spurious interactions and the reconstruction of complex networks,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 52, pp. 22073–22078, 2009.

[30] A. Menon and C. Elkan, “Link prediction via matrix factorization,” in *Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases, Part II (ECML PKDD ’11)*, pp. 437–452, 2011.

[31] S. Gao, L. Denoyer, and P. Gallinari, “Link prediction via latent factor blockmodel,” in *Proceedings of the 21st Annual Conference on World Wide Web (WWW ’12)*, pp. 507–508, ACM, New York, NY, USA, April 2012.

[32] Y. Koren, R. Bell, and C. Volinsky, “Matrix factorization techniques for recommender systems,” *Computer*, vol. 42, no. 8, pp. 30–37, 2009.

[33] J. A. Hanley and B. J. McNeil, “The meaning and use of the area under a receiver operating characteristic (ROC) curve,” *Radiology*, vol. 143, no. 1, pp. 29–36, 1982.

[34] A. Lancichinetti, S. Fortunato, and F. Radicchi, “Benchmark graphs for testing community detection algorithms,” *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, vol. 78, no. 4, Article ID 046110, 2008.

[35] W. W. Zachary, “An information flow model for conflict and fission in small groups,” *Journal of Anthropological Research*, vol. 33, no. 4, pp. 452–473, 1997.

[36] D. Lusseau, “The emergent properties of a dolphin social network,” *Proceedings of the Royal Society B: Biological Sciences*, vol. 270, supplement 2, pp. S186–S188, 2003.

[37] V. Batageli and A. Mrvar, Pajek Datasets, http://vlado.fmf.uni-lj.si/pub/networks/pajek/data/gphs.htm.

[38] R. Guimerà, L. Danon, A. Díaz-Guilera, F. Giralt, and A. Arenas, “Self-similar community structure in a network of human interactions,” *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, vol. 68, no. 6, Article ID 065103, 2003.

[39] L. A. Adamic and N. Glance, “The political blogosphere and the 2004 US Election,” in *Proceedings of the Workshop on the Weblogging Ecosystem (WWW '05)*, pp. 36–43, August 2005.

[40] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, “Uncovering the overlapping community structure of complex networks in nature and society,” *Nature*, vol. 435, no. 7043, pp. 814–818, 2005.

[41] D. J. Watts and S. H. Strogatz, “Collective dynamics of small-world networks,” *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[42] M. Boguñá, R. Pastor-Satorras, A. Díaz-Guilera, and A. Arenas, “Models of social networks based on social distance attachment,” *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 70, no. 5, Article ID 056122, 2004.