Energy Efficient Routing using Machine Learning based Link Quality Estimation for WMSNs

Kirubasri G¹, Kiruthika S², TYJ Naga Malleswari³, Jayareka KS⁴

¹Assistant Professor, Department of Computer Science and Engineering, Sona College of Technology, Salem, Tamil Nadu, India (Corresponding author)*

²Assistant Professor, Department of Computer Science and Engineering, Sona College of Technology, Salem, Tamil Nadu, India

³Assistant Professor, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu, Chennai

⁴Assistant Professor, Department of Computer Science and Engineering, Sona College of Technology, Salem, Tamil Nadu, India

Article History: Received: 11 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 10 May 2021

Abstract

In Wireless Multimedia Sensor Networks (WMSNs), Energy constraint battery-operated sensors collect surplus data from the target area, process them and transmit to the end server using the adopted routing approach. During this transmission, the quality of the multimedia data such as audio, scan images, recordings, live video streaming are get diminished by unbalanced, insecure low quality wireless links due to various causes. The design and implementation of energy efficient routing with suitable link quality estimation (LQE) model is an essential and challenging job in resource constrained WMSNs for upholding the energy of the network and affords Quality of Service.Many sophisticated LQE approaches for WMSNs have been proposed in recent days. In this paper an Energy Efficient Routing using Machine Learnig (EERML-LQE) technique for link quality prediction is proposed to hit the impact of Link Quality Analysis (LQA) using Gaussian Naive Bayes classification for predicting the link accuracy. Experimental results of the proposed routing protocol with machine learning based link quality estimation reveals that the EnRoML outperforms current routing approaches in terms of various network performance parameters like energy utilization, packet drop ratio, life time of the network and average packet latency

Keywords: Energy Efficiency, Link Quality prediction, Remaining Energy, Machine Learning, Gaussian Naive Bayes classification.

1. INTRODUCTION

Broadcasting features of wireless links vary suggestively by time and space, which affects the radio links quality abundantly. To guarantee the reliability in networks, substantial LQE technique is essential for get used the radio link parameters efficiently and also for selecting the alternate route in case of link drop due to poor link conditions. In late 90s, data driven LQE using real data measurements started[1], [2].Models developed using synthesized data are published in [3]-[8]. Plethora of models for Machine Learning based LQE has been developed in recent years [9]–[10]. Data collected from numerous sources by using various methodsand the technological improvement helps the researches to develop ML models more easy and have vital control on wireless networking procedures.

Planning the routing design has crucial task of predicting the link quality in terms of reliability for supporting the applications that deal with real time data transmission. Even though the link prediction has been introduces in many analysis [11]-[13], stillfew works on robust routing using LQEexists. Predicting link quality can be used in many fields includes recognizing incorrect interactive networks, mining absent information [14],[15]. Presently, around fourcentral methods for different link forecasting process, which areconstructed based on network topology information, random walk, machine learning, and probability analysis.

Preferably, with the use of ML, a wireless link with high quality data transmission with least energy utilization is possible. An instinctive method for study and analyze any kind of transceiver links and better technology to support current working aspects of wireless networks helps in designing Optimized wireless networks [16],[17]. Mechanisms for understanding radio frequency (RF) spectrum usage are presently suggested in [18], [19]. To make sure stable and sustainable Wireless networks, there is a need of better link estimation method, so that the parameters of the radio link can be modified and an unconventional channel can be chosen for transmitting data wirelessly. To put it succinctly, the enhanced the quality of the link, the higher the effective reception ratio and thus the more efficient communication.

Differences in link efficiency can have a huge effect on the overall network's connectivity. Efficient link quality prediction can lead to better network enactment includes increased work done, least packet depletion during network transmission, imperfect network lifetime, and limited route calculations, by the wise link estimation algorithm.

1.1 MACHINE LEARNING PRELIMINARIES

Machine Learning approaches are classified broadly as Supervised Learning (SL), Un-supervised Learning and Reinforcement Learning (RL). In SL, a sample of input with expected output is given before train the model. In order to predict an unknown function that maps the inputs of example training to the outcome. Classification and regression are the two methods to be followed in SL to determine the output data is numerical or categorical type. These methods have strict constraint of their data scalability. In Unsupervised Learning (UL), only a

sample of unlabeled inputs is given, not their outputs. The aim is to identify an invisible trend in data that has been mislabeled. Despite the fact that such an approach is becoming more popular, it could be revealing a pattern that is of little significance. In RL, no input data is provided and they acquire knowledge by mingle with the rest of the world by taking action and winning rewards. When the input data is delayed, such methods are more efficient. It takes long time to converge as its disadvantage [20].

In ML, an approach called Feature Extraction (FE) is used for segregating a subclass of data which intern used for finding the attractive patterns. FE includes feature identification and reduction. The effectiveness of the ML system is determined by the characteristics selected. Complex characteristics necessitate more memory, computing power, and training time. The reduction of features is a technique for shrinking the scale of a collection of features. Some of the important techniques used for dimensionality reduction in machine learning are Principal Component Analysis [21], Factor Analysis, Prediction Pursuit and Independent Component Analysis [22].

1.2 MACHINE LEARNING – A NETWORK PERSPECTIVE

In recent days ML techniques play a vital role in conventional network specifically these techniques are used to effectively manage available resource, work allocation and prediction of energy consumption and traffic classification. There are also some attempts at incorporating approaches to machine learning into WSNs. The aims of using machine learning vary from optimizing QoS [23] like network delay [24],[25],[26], traffic congestion, and transmission reliability and for protect the network from security attack by using the various SL algorithms. For complex network problems, RL is used widely. Figure 1 shows the statistics of the growth of ML approaches in WSN optimizations.



Fig.1. The development of ML models for link quality prediction

Contributions of this work are as follows:

- 1. PRR based link quality estimation for intermediary link holding with precise estimate of strong links.
- 2. A prediction model is created for the assessment of link quality of resource constrained nodes of WMSN. The model is evaluated with the required set of network parameters.
- 3. Evaluating the operational process of the prediction model from the receiver. The model output is used by the routing approaches to predict the quality of the temporal links as week, medium and high-quality.

Remaining sections of this paper is organised as follows:

Section 2, describes the network model as well as it confers link quality estimation and routing metrics used by the system, Section 3 explains the data collection and model training process. In section 4, the proposed link quality estimation method is given which includes physical layer parameters, model evaluation. Section 5details the experimental results and section 6 concludes the proposed analysis and includes the future work.

2. NETWORK MODEL AND ASSUMPTIONS

The proposed network model is formed as a directed graph G_r with multimedia sensors and sink node. G_r is represented as $G_r = (N, E, LQ)$ where V signifies the set of sensor nodes, E stands for the set of edges that represent node connections and E is each edge's nonnegative Link Quality value. A non-overlapping path between source to Cluster Head (CH) is represented as $P = {path_1, path_2... path_n}$. Adopt a network with uniform sensor characteristics with few are having high processing capabilities and energy and the entire sensors are immovable after their deployment. All sensors can calculate its present energy level and keep the record of link quality status between the neighbors and hop count of each route used. The positions of nodes are not predetermined; rather, they are distributed at random.

All of the sensor nodes in the monitoring area are used for data collection and do not move after deployment. Figure 2 depicts the network architecture. Every cluster has a head node called CH which is having high power and coverage, processing capability, allocation of time period for the source node and data aggregation etc., every sensor nodes are having uniform initial energy and sensing, computing, and communication capacities. Each multimedia sensor node can change the transmission power to save energy, and the sink node is not constrained by energy or capacity. In order to hold network energy for a long time, each cluster's gateway node must choose the best path for its sending node based on dynamic connection behavior. In our algorithm, CH calculates the cost of the routing route.



Fig. 2. Network Architecture of the proposed system

2.1 LINK QUALITY ESTIMATION

Wireless communication of data gets affected by plenty of reasons that worsen the data delivery mechanism and lead to failure of data transmission. The prominent element for influencing the routing of the cluster based routing algorithms in wireless multimedia sensor networks is the quality of the link. This varies over time depending on various factors such as diffusion power and node hop length. Poor-quality links have an effect on the data path's reliability and the network's lifetime. (Baccour N et al. 2012). Link instability is caused by constantly changing network topology, which has an effect on data packet flow across week links.

Link Quality Estimation (LQE) is a technique for evaluating dynamic link behaviour and assisting lowpower wireless sensor routing protocols in selecting the best route for packet transfer. (Cerpa A et al.2005). When compared to conventional ad-hoc networks, the nodes are densely deployed and the radio powers of the nodes are very low and irreplaceable. Due to these, estimation of wireless link quality is highly difficult. As a result, it is important for WMSN routing designers to precisely measure the quality of low-power links. (Gomez C et al. 2010). LQE is a tool that analyses the statistical characteristics of low-power links and serves as a basis for a number of mechanisms and network protocols.

Snooping the link, recoding the link parameters, and computing the link quality metric (shown in figure 3) are all part of the LQE method.

The LQE theory divides connection quality (LQ) variance into spatial and temporal categories. The network can save a lot of energy by selecting high-quality links in wireless transmission, particularly in energy-constrained WMSNs using the LQE technique.



Fig. 3. Phases of Link Quality Estimation

Research Article

Data transmission over weak links can result in frequent packet loss, necessitating packet retransmission, depleting the nodes' energy. As a consequence, for resource limited WMSNs, LQE techniques that account for accurate connection measures are required to highlight good quality links in order to achieve efficient data transmission by preventing packet loss through bad LQ links.

Hard Ware-based Estimators (HWE) and Soft Ware-based Estimators (SWE) are two forms of link quality estimators using for wireless multimedia sensor networks (SWE). HWE obtains relevant data of the link from the hardware. During transmission, it only considers the packets that were successfully received, not the packets that were lost. SWE, on the other hand, is measured using both sent and received packets. In addition, various methods for measuring SWE have been used. SWE are a simple and reliable connection estimator based on packet loss information obtained from neighbouring layers. The metric values are determined by software-based LQ estimators on either the sender or receiver side. PRR, RNP and score based estimators are the few examples under SWE (NouhaBaccour et al. 2015).

2.2 ROUTING METRICS

The factors that asses the performance of the route. A metric is a value that is connected with a node and the route path through which the packets are transmitted. In this paper, the metric is used to choose the most energy-efficient and stable path.

2.2.1. LINK QUALITY INDICATOR (LQI)

The Physical layer of the IEEE 802.15.4 standard contains a component called Connection Consistency Indicator. This factor is used to describe the efficiency of a link as interpreted by a route packet receiver at the time of packet reception. The MAC sublayer receives an integer value ranging from 0 to 255 as the resultant link quality value. The lowest and highest quality IEEE 802.15.4 reception observable by the receiver are correlated with the minimum and maximum LQI values (0 and 255), respectively, and the LQI values in between are distributed between these two limits. The values can be obtained directly from its radio module since it is a Hardware-based Link quality estimator, which reduces computation overhead. As a consequence, the LQI is a strong candidate for use as a routing metric input.



Fig. 4. LQI of the path.

The AvgLQI metric is defined in this paper as the average LQI values of all the links between the source and destination. AvgLQI values are used to classify sensors across their coverage performance ranges. This metric may be useful in a setting where wireless connections are notoriously unreliable. For the current network scenario, Figure 4 shows the average LQI estimate of a route. Via a flooding method in which, route discovery is initiated by source with a request message including RE and node ID. This the process continues till the packet arrived at the destination node.

2.2.2. REMAINING ENERGY LEVEL

Figure 5 depicts the remaining energy calculation of a path. The Average RE of the path should be determined in this work by taking into account the RE of all nodes in the path. The equation (1) is used to calculate the RE of a node n at a given time.

 $RE_n = E_{initial} - E_{current}$ (1)



Fig. 5. RE of the Path.

If the remaining energy of all nodes in the path exceeds the network's Energy threshold (E_{th}), the Average RE for that path will be determined. Otherwise, it will be labeled as invalid and removed from the route selection algorithm.

3. DATA COLLECTION AND MODEL TRAINING

To train the model, we used packet traces from Motelab [27]. The physical layer information of each wireless link packet received was documented using collection software. During one data collection experiment (Tx-power=0dB, channel 26), source node sends the packets to destination node with an inter-packet interval of 100 milliseconds. When a packet is sent, the destination node records the sequence number, RSSI, and LQI of the packet.

With the sample of ambient noise taken for 15 times after 1 ms interval helps to evaluate SNR. To avoid the clash between inter-nodes, data collection module is executed among 68 senders – receivers pairs in different timeslots. Over an aggregate of 160 hours of information assortment, we enlisted data for around 5.4 million parcels. For every 68 packet drops, there are 80,000 packets were recorded. Out of 68 links, around 12 weak links (PRR < 10%), 14 average quality links (10% < PRR < 90%), and 42 strong links (PRR > 90%) were recorded. In order to train the model, from each 10 links, packet traces were received for Motelab with the interval of 64 packets per second.

Parameters	Values
Input Feature	$PRR + \{RSSI, SNR, LQI\}$
Number of Links (L)	20, 7, 5, 3, 2, 1
Number of Packets (P)	36000, 10000, 5000, 1000, 500
Input Window (W)	1, 2, 3, 4, 5, 10
Packet Interval (I)	0.1, 0.2, 0.5, 1, 10, 60 (seconds)

Table 1. Trigger Parameters

Parameters used for training of our model with respect the input dataset are shown in Table 1. Lack of data addressed by the use of domain information. The missing RSSI values have been replaced by 0. The possible RSSI values are integer values ranging from 0 (signal-free bad link) to 127 (signal-free nice link), while 128 is an invalid value or an error value.Figure 6 shows the RSSI accuracy. Input function is created by combining PRR ,RSSI, SNR and LQI. From the data obtained, pick the L links. WMEWMA filter is used for PRR calculation in which the window size of the filter is set to 5.In order to create an input vector, PHY parameters for W packets are combined with recent PRR and the steps are repeated for all packets.





During the process, the target vector is also constructed by testing the receipt status of each input vector's next packet: we mark the input vector's target (desired output) as 1 if the next packet is received and 0 otherwise. From the input data set, 58% is selected for training the input data and the residual 58% is used for testing the model. Prediction output are matched with the expected target value.

When the training data does not contain all possibilities, Naive bayes does very well, so it can be quite successful with low data volumes. It takes advantage of its simplicity in terms of processing speed.

4. PROPOSED LINK QUALITY ESTIMATION METHOD

For link quality assessment, this paper proposes a data driven methodology. The first step is to prepare the dataset for LQE model that comprises of the time series value and RSSI value of each packet. Packet transmission factors like PRR and PSR derived from mathematical calculation and RSSI value is derived from transceiver's data. A significant task before developing and training the model is to clear the impure data from the dataset. In the proposed system, Gaussian based interpolation method is used for replacing the missing values by a random data selected from the next close data from the front and back. The range of RSSI value is between 0 to 127.

The proposed approach uses supervised learning to classify the result into one of the three categories like best, medium and worst. RSSI value measured for a link and the packet's sequence number are provided in the dataset is used for creating the new features. To classify the link quality as best, medium, and worst, collective values of RSSI like $RSSI_{std}$, and $RSSI_{avg}$ are used. Bayes classification is applied for categorising the new instance. Bayes theorem-based probabilistic classification method works based on the premise that features are strictly different from the class variable.

For a given instance of function $G(x1 \dots xn)$ and the class variable y then the relationship can be as per the Bayes theorem.

$$P(b|a_{1,\dots}a_n) = \frac{P(b)P(a_{1,\dots}a_n|b)}{P(a_1,\dots a_n)} \quad (2)$$

In this work, the Gaussian Naïve Bayes classification algorithm is used, which assumes the probability of characteristics as a Gaussian distribution and the probability of xi for a given y can be mathematically expressed as follows:

$$P(a_i|b) = \frac{1}{\sqrt{2\pi\sigma^2 y}} \exp\left(-\frac{(a_i - \mu_y)^2}{2\sigma^2 y}\right) \quad (3)$$

Sensor node's Energy utilization for data communication for any available path can be calculated as given below.

$$EC_{path} = \sum_{i=1}^{NH_p} EC_i (4)$$

Where, Number of hops in each path, energy utilization by individual link is used as parameters for path energy calculation.

5. EXPERIMENT AND RESULTS

In general research utilities for LQE using ML algorithms are nonlinear models with strict limitations and computational energy conditions. As there are fewer features compared to the larger number of samples, the naive bayes classifier is taken into account. The model has to be fully retrained when real-time sensors are used and when the network is reconstructed with some heterogeneous nodes. A simple probabilistic model is favored with these constraints and drawbacks of non-linear models. Better results were shown by a combination of RSSI (raw), RSSIavg, RSSIstd. The modifications were integrated into the NS-3 and simulated traditional routing with and without LQE and routing with LQE using ML (LQEuML).



Fig. 7. Comparison of routing models

The comparison of the simulation results showed that with the proposed link quality estimation, LQEuML showed higher PRR with lower latency. Using the Guasian Naive Based classifier, the link quality classification was analysed. The Figure 7 presents the accuracy of the different routing classes in which for the Naive Bayes model, the precision of the class level for good and poor links is greater. It misclassifies the intermediate links wrongly as either good or bad.



Fig. 8. Confusion matrix

The uncertainty matrix graph of the best models selected from the k-fold cross validation process is shown in Figure 8.



Fig. 9.PDR comparison with increasing simulation time

The stratified cross validation used during training ensures that the model has been conditioned to use the same proportion of different groups. In terms of nodes, the Packet Drop Ratio (PDR) and Network Life of the models are shown in Figure 9 and 10. It is evident from the graph that the suggested LQE based on Machine Learning outperforms the conventional LQE routing approach.



Fig. 10. Nodes alive (%) comparison against time

8. CONCLUSION

Wireless sensor networks (WSNs) are becoming an important field of research with advancements in sensor applications. WSNs contain several small sensor nodes, depending on the given application, to sense an area. Though the plethora of energy-efficient protocols has been enforced, it is described as an ill-posed issue to select an optimal path between base station and sensor nodes due to the power restrictions on sensor nodes. To enhance inter-cluster communication within WSNs. In order to measure the shortest path between cluster heads and sink, machine learning based LQE is considered. We have trained the network in such a way that it takes into account different WSN features and is able to determine which sensornode will be chosen as the next hop to construct the shortest path between the chosen cluster heads and the sink. It is also possible to consider the noncluster head nodes when choosing the shortest path. Hardware metrics are used in the proposed method including the RSSI, LQI, sequence number to assess the quality of the links on the receiver side. The research comprises the following series of interpolations for the management of missing values like feature engineering, data set sampling and machine learning model construction. Based on the comparison of the findings reported in the literature and the experiments conducted, it is clear that the proposed approach performed better than traditional routing without LQE and ML when classifying link quality for the selected representative dataset. Extensive experimental findings indicate that the methodology proposed outperforms the protocols for competitive energy performance. The overall results showed that the link quality estimate was more reliable based on data traffic approach.

REFERENCES

- G. T. Nguyen, R. H. Katz, B. Noble, and M. Satyanarayanan, "A tracebased approach for modeling wireless channel behavior," in Proceedings of the 28th conference on Winter simulation. IEEE Computer Society, 1996, pp. 597–604.
- [2] H. Balakrishnan, R. H. Katz et al., "Explicit loss notification and wireless web performance," in IEEE Globecom Internet Mini-Conference. IEEE, 1998, pp. 1–5.
- [3] H.-J. Aud'eoud and M. Heusse, "Quick and efficient link quality estimation in wireless sensors networks," in Wireless On-demand Network Systems and Services (WONS), 2018 14th Annual Conference on IEEE, 2018, pp. 87–90.
- [4] R. Fonseca, O. Gnawali, K. Jamieson, and P. Levis, "Four-bit wireless link estimation." in HotNets, 2007.
- [5] C. A. Boano et al., "The triangle metric: Fast link quality estimation for mobile wireless sensor networks," in International Conference on Computer Communication Networks, 2010, Zurich, Switzerland, 2010.
- [6] N. Baccour et al., "F-lqe: A fuzzy link quality estimator for wireless sensor networks," in European Conference on Wireless Sensor Networks. Springer, 2010, pp. 240–255.
- [7] Z.-Q. Guo et al., "Fuzzy logic based multidimensional link quality estimation for multi-hop wireless sensor networks," IEEE Sensors Journal, vol. 13, no. 10, pp. 3605–3615, 2013.
- [8] S. Rekik et al., "Low-power link quality estimation in smart grid environments," in Wireless Communications and Mobile Computing Conference (IWCMC), 2015 International. IEEE, 2015, pp. 1211– 1216.
- [9] S. Manikandan and M. Chinnadurai, "Evaluation of Students' Performance in Educational Sciences and Prediction of Future Development using TensorFlow", International Journal of Engineering Education Vol. 36, No. 6, pp. 1783–1790, 2020, 0949-149X/91, TEMPUS Publications, Printed in Great Britain
- [10] W. Sun et al., "Wnn-lqe: Wavelet neural- network-based link quality estimation for smart grid wsns," IEEE Access, vol. 5, pp. 12 788–12 797, 2017.

- [11] Manikandan, S, Chinnadurai, M, "Effective Energy Adaptive and Consumption in Wireless Sensor Network Using Distributed Source Coding and Sampling Techniques", Wireless Personal Communication (2021), Springer, https://doi.org/10.1007/s11277-021-08081-3.
- [12] M. Nickel et al., ``A review of relational machine learning for knowledge graphs," Proc. IEEE, vol. 104, no. 1, pp. 11_33, Jan. 2016.
- [13] H. Hu et al., ``LPI-ETSLP: LncRNA_protein interaction prediction using eigenvaluetransformation-based semi-supervised link prediction," Mol. BioSyst., vol. 13, no. 9, pp. 1781_1787, 2017.
- [14] C. Fu et al., "Link weight prediction using supervised learning methods and its application to yelp layered network," IEEE Trans. Knowl. Data Eng., vol. 30,no. 8, pp. 1507_1518, Aug. 2018.
- [15] D. M. Dunlavy et al., "Temporal link prediction using matrix and tensor factorizations", ACM Trans. Knowl. Discovery from Data, vol. 5, no. 2, pp. 1_27, Feb. 2011.
- [16] C. Fortuna et al., "Data driven wireless network design: a multi-level modeling approach," Wireless Personal Communications, vol. 88, no. 1, pp. 63–77, 2016.
- [17] J. Jiang et al., "Unleashing the potential of data-driven networking," in International Conference on Communication Systems and Networks. Springer, 2017, pp. 110–126.
- [18] M. Z. Zheleva et al., "Enabling a nationwide radio frequency inventory using the spectrum observatory," IEEE Transactions on Mobile Computing, vol. 17, no. 2, pp. 362–375, 2018.
- [19] S. Rajendran et al., "Electrosense: Open and big spectrum data," IEEE Communications Magazine, vol. 56, no. 1, pp. 210–217, 2018.
- [20] R. S. Sutton and A. G. Barto, Reinforcement learning: An introduction. MIT press, 2018.
- [21] Manikandan, S & Chinnadurai, M 2019, 'Intelligent and Deep Learning Approach OT Measure E-Learning Content in Online Distance Education', The Online Journal of Distance Education and e-Learning, vol.7, issue 3, July 2019, ISSN: 2147-6454.
- [22] I. K. Fodor, "A survey of dimension reduction techniques," Lawrence Livermore National Lab., CA (US), Tech. Rep., 2002.
- [23] L. Yanjun, L. Xiaobo, and Y. Osamu, "Traffic engineering framework with machine learning based metalayer in software-defined networks," in Network Infrastructure and Digital Content (IC-NIDC), IEEE International Conference on. IEEE, 2014, pp. 121–125.
- [24] A. Azzouni, R. Boutaba, and G. Pujolle, "Neuroute: Predictive dynamic routing for software-defined networks," CoRR, vol. abs/1709.06002, 2017.
- [25] S. Sendra, A. Rego, J. Lloret, J. M. Jimenez, and O. Romero, "Including artificial intelligence in a routing protocol using software defined networks," in IEEE International Conference on Communications Workshops (ICC Workshops), May 2017, pp. 670–674.
- [26] G. Stampa et al., "A deep-reinforcement learning approach for software-defined networking routing optimization," arXiv preprint arXiv:1709.07080, 2017.
- [27] G. Werner-Allen et al., Motelab: a wireless sensor network testbed. In IPSN '05, page 68. IEEE, 2005.