SCJ SOFT COMPUTING JOURNAL
ISSN: 2322-3707
University of Kashan

Investigating the Performance of Machine Learning Methods for Link Quality Estimation in Wireless Networks

Mohammad Javad Nasri-Lowshani [1], Salman Goli-Bidgoli [2,*]

[1] Department of Artificial Intelligence, Faculty of Electrical and Computer Engineering, PhD Student, University of Kashan, Kashan, Iran.

[2] Department of Software Engineering, Faculty of Electrical and Computer Engineering, Assistant Professor, University of Kashan, Kashan, Iran.

ABSTRACT. With the widespread use of the internet and the development of wireless networks that transfer large data streams, the importance of assessing and controlling the quality of communication links in wireless networks has gained significant attention. By predicting link quality, energy consumption of network nodes and the overall stability of the network can be improved. One category of methods used for predicting the quality of wireless links is machine learning techniques. This paper examines the performance of ensemble methods, a type of supervised machine learning approach that has previously received less focus in the context of wireless link quality prediction. Additionally, due to the advantages of unsupervised methods that can be trained on unlabelled datasets, the performance of the k-means algorithm is also evaluated. The results show that ensemble algorithms are highly effective in predicting the quality of communication links in wireless networks. Among the ensemble methods, Gradient Boosting achieved the best performance with an F1 score of 95.79, while the k-means method demonstrated superior performance in the recall metric, achieving a value of 96.47 compared to other methods.

KEYWORDS: Ensemble Methods, K-Means Clustering, Link Quality Classification, Principal Component Analysis, Received Signal Strength Indicator.

---

1 INTRODUCTION

Wireless networks constitute a fundamental pillar of modern communications. By eliminating the need for wired infrastructure, they enable data connection and exchange in dynamic, mobile, and distributed environments. These networks find application in diverse forms, including Wireless Sensor Networks, Vehicular Ad-hoc Networks, and Cloud-Edge systems. Their use extends to fields such as environmental monitoring, smart healthcare, transportation, and the Internet of Things.

In wireless networks, radio signal propagation channel conditions can vary significantly over time and location, impacting radio link quality. Given that some links inherently possess lower quality, data retransmission can partially maintain data integrity when packets traverse poor-quality links. However, this approach reduces transmission efficiency, increases energy consumption, and introduces delays detrimental to real-time applications [1]. Conversely, employing accurate LQE can prevent packet loss and enhance Quality of Service (QoS) [2]. Consequently, utilizing LQE methods is essential to identify and select optimal paths for reliable data delivery.

Machine learning (ML) methods represent a distinct category of approaches for LQE. Leveraging ML for LQE offers substantial improvements in wireless network performance, as these techniques excel at processing large volumes of data traces, learning from them, and developing a comprehensive, high-level understanding of wireless link characteristics [3]. ML-based LQE methods can be implemented through two primary approaches: continuous numerical value prediction (regression) or discrete value prediction (classification).

Accepted SCJ

Recent years have witnessed substantial research in the field of LQE for wireless networks. Extensive ML approaches have been developed for LQE in wireless networks. Within supervised learning approaches, a hybrid method integrating Support Vector Machines (SVM) with decision trees has been developed to estimate wireless link quality [1]. This technique analyses key features—Received Signal Strength Indicator (RSSI) and Link Quality Indicator (LQI)—to perform five class quality predictions based on Packet Reception Rate (PRR) metrics. The study demonstrates that this ML implementation significantly reduces network energy consumption while extending overall network lifetime.

Another supervised learning approach introduces a comprehensive framework for developing wireless link quality classifiers [4]. This methodology highlights how design decisions—across data pre-processing, feature engineering, and learning algorithm selection—critically impact ML-based LQE. The study demonstrates that resampling to balance training classes and generating synthetic features substantially enhances both overall classification accuracy and minority class detection. Implemented using the Rutgers dataset [5], the framework employs a decision tree classifier with optimal feature combination selection to predict three level link quality (bad, medium, good) from RSSI values. It further evaluates random forest performance within ensemble methods.

Separately, a distinct random forest-based technique also leverages the Rutgers dataset for LQE [6], but implements alternative classification thresholds for its three quality tiers. This approach formulates hyperparameter tuning as a search optimization problem, resolved through an enhanced sparrow search algorithm.

A distinct approach implements a six-layer CART decision tree (employing Gini impurity) for link quality classification in wireless sensor networks [7]. Utilizing the Rutgers benchmark dataset, this method incorporates resampling to balance training classes and generates synthetic features through RSSI value summation and multiplication. The research further demonstrates that tree depth critically impacts model efficacy, with six layers yielding optimal performance.

Contrastingly, a separate supervised learning technique also leverages the Rutgers dataset for evaluation [8] but introduces a novel four class classification model (very bad, bad, medium, good) diverging from conventional three class model. Implemented via gradient boosting for IoT device link quality assessment, this method similarly enhances performance through synthetic feature creation and resampling techniques.

Addressing the scarcity of unsupervised approaches for LQE in software-defined wireless mesh networks, a novel framework enables real-time anomaly detection [9]. This methodology employs an enhanced clustering algorithm leveraging elastic similarity metrics to effectively characterize wireless link reliability. Further, it introduces a specialized change point detector that minimizes overestimation errors through a dual-mechanism approach: a rank-based statistical test coupled with a recursive maximization procedure. When triggered by a potential anomaly, this recursive mechanism identifies peak fluctuation positions – serving as heuristic indicators of true change point locations.

Alternatively, several approaches leverage unsupervised clustering to generate annotated datasets for training LQE models. Specifically, In [10], They employs hierarchical clustering for data labelling, subsequently applies resampling to augment minority class representation, and ultimately trains a LightGBM (Light Gradient Boosting Machine) classifier.

Current literature analysis reveals that ensemble techniques remain underutilized for LQE compared to other supervised learning approaches. Moreover, comparative studies confirm supervised methods' predominant adoption, despite their inherent dependency on labelled datasets – a requirement that poses significant data acquisition challenges. This limitation consequently positions unsupervised learning methods as a viable alternative for LQE applications.

This paper examines the efficacy of ML methods in classifying the quality levels of links within wireless networks. Specifically, the contributions encompass:

- Implementation of ensemble methods for classifying wireless link quality levels.
- Application of unsupervised learning techniques for clustering wireless links by quality class.
- Utilization of Principal Component Analysis (PCA) to enhance the performance of the unsupervised method employed.

Accepted SCJ

2 FINDINGS

The proposed methodology investigates two distinct ML categories for predicting wireless link quality classes: ensemble techniques and unsupervised approaches. Feature extraction from data traces constitutes the essential pre-processing step for training these models. Specifically, features derived from RSSI measurements—stored by radio receivers—are utilized in this framework, with comprehensive details provided in Table 1.

Table 1: Features derived from RSSI measurements for training machine learning models.

Instantaneous Received Signal Strength	**Extracted Features**
Mean RSSI	$rssi$ ، $rssi^2$ ، $rssi^3$ ، $rssi^4$ ، $rssi^{-1}$ ، $rssi^{-2}$ ، $rssi^{-3}$ ، $rssi^{-4}$
Variance and Derivative of RSSI	$rssi_{avg}$ ، $rssi_{avg}^2$ ، $rssi_{avg}^3$ ، $rssi_{avg}^4$ ، $rssi_{avg}^{-1}$ ، $rssi_{avg}^{-2}$ ، $rssi_{avg}^{-3}$ ، $rssi_{avg}^{-4}$
Instantaneous RSSI	$rssi_{dr}, rssi_{std}$

Computation of key features—mean RSSI, RSSI variance, and RSSI slope—necessitates multiple signal strength measurements within a sliding observation window ($W_{history}$). Consistent with our objective of wireless link quality classification, target categories must be assigned to all dataset links. Following established methodologies [4, 7], the proposed framework implements three class classification (bad/medium/good) using PRR metrics derived from Equation (1).

$$y = \begin{cases} bad, \forall\ PRR \leq 0.1 \\ intermediate, \forall\ else \\ good, \forall\ PRR \geq 0.9 \end{cases} \quad (1)$$

Equation (1) defines PRR as the ratio of successful message receptions within transmission range to total potential receivers in that range. This metric is calculated within a dedicated observation window (W_{PRR}).

The proposed methodology employs three ensemble techniques—Bagging, Gradient Boosting, and AdaBoost—to predict link quality using extracted features. Concurrently, k-means clustering serves as the unsupervised approach for LQE assessment. Prior to clustering, all Table 1 features undergo PCA for dimensionality reduction and feature space optimization. The transformed PCA output subsequently trains the k-means model.

Missing values are replaced with zero, indicating that no packet was received. Following this replacement, the specified features are extracted from the trace dataset using two windows, $W_{history}$ and W_{PRR}, each of size 10. These features are then normalized using the standard normalization method.

The constructed dataset trains both the ML methods introduced in the proposed approach and baseline methods, enabling performance comparison. Crucially, not all features are fed directly into the ML models; instead, optimal features are selected. The three selected features are $rssi$, $rssi_{avg}$, and $rssi_{std}$, identified through exhaustive evaluation of all feature combinations [12].

This dataset is then used to train and compare the proposed and baseline methods Table 2. All results are reported using 3-fold cross-validation and evaluated across three metrics: accuracy, recall, and F1-score. Weighted averaging is applied during the calculation of both accuracy and recall.

Table 2: Experimental results comparing the performance of the proposed method's variants against baseline approaches when trained on the dataset.

ML Approach	**Features**	**Precision**	**Recall**	**F1-Score**
Decision Tree [12]	All	97.00	88.69	91.91
CART Decision Tree [17]	Selected	85.50	95.50	90.20
Based on Logistic Regression [13]	Selected	92.30	92.20	92.20

Accepted SCJ

Based on Decision Tree [13]	Selected	93.20	93.20	93.20
Random Forest	Selected	96.82	93.52	94.64
Decision Tree [12]	Selected	97.04	94.92	95.64
AdaBoost	Selected	96.10	92.64	93.88
Bagging	Selected	97.02	94.74	95.52
Gradient Boosting	Selected	**97.10**	95.15	**95.79**
K-means	Selected	85.84	66.31	73.24
K-means	Selected + PCA	85.28	74.56	76.06
K-means	All + PCA	95.03	**96.91**	94.91

Gradient boosting yielded the highest performance in both accuracy 97.10% and F1-score 95.79% metrics. High accuracy indicates the proportion of links correctly classified as having a target quality level. For recall, K-means combined with PCA achieved the top result 96.47%, reflecting its effectiveness in identifying all links possessing the target quality. The F1-score represents the harmonic mean of precision and recall. Notably, as unsupervised methods are not conventionally assessed using classification metrics, these measures were employed exclusively for comparative performance analysis against supervised approaches.

3 CONCLUSION

Recent advancements in high-throughput wireless networks have heightened the importance of measuring and controlling wireless link quality. Accurate link quality prediction enhances network stability and energy efficiency. ML offers state-of-the-art solutions for this task. This study investigated ensemble methods – previously underexplored for link quality prediction – demonstrating their strong predictive capability. Among ensemble techniques, gradient boosting achieved the highest F1-score 95.79%. We also evaluated unsupervised approaches (eliminating the need for labelled data), finding that K-means with PCA attained the best recall 96.47%, outperforming other methods in this metric.

REFERENCES

[1] J. Shu, S. Liu, L. Liu, L. Zhan, and G. Hu, "Research on link quality estimation mechanism for wireless sensor networks based on support vector machine," *Chinese Journal of Electronics*, vol. 26, no. 2, pp. 377-384, 2017, doi: 10.1049/cje.2017.01.013.

[2] A. Tabouche, B. Djamaa, and M. R. Senouci, "Link quality estimation in 6TiSCH networks: A review," in *2022 First International Conference on Computer Communications and Intelligent Systems (I3CIS)*, 2022: IEEE, pp. 99-104, doi: 10.1109/I3CIS56626.2022.10076145.

[3] G. Cerar, H. Yetgin, M. Mohorčič, and C. Fortuna, "Machine learning for wireless link quality estimation: A survey," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 2, pp. 696-728, 2021, doi: 10.1109/COMST.2021.3053615.

[4] G. Cerar, H. Yetgin, M. Mohorčič, and C. Fortuna, "On designing a machine learning based wireless link quality classifier," in *2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications*, 2020: Ieee, pp. 1-7, doi: 10.1109/PIMRC48278.2020.9217171.

[5] S. K. Kaul, M. Gruteser, and I. Seskar, "Creating wireless multi-hop topologies on space-constrained indoor testbeds through noise injection," in *2nd International Conference on Testbeds and Research Infrastructures for the Development of Networks and Communities, 2006. TRIDENTCOM 2006.*, 2006: IEEE, pp. 10 pp.-521, doi: 10.1109/TRIDNT.2006.1649191.

[6] Y. Wang, Z. Jia, R. Gan, J. Li, and Z. Xiao, "Evaluate Link Quality based on ISSA-BRF in Wireless Network," presented at the Proceedings of the 2024 5th International Conference on Computing, Networks and Internet of Things, Tokyo, Japan, 2024. [Online]. Available: https://doi.org/10.1145/3670105.3670109.

[7] M. Zha, L. Zhu, Y. Zhu, and T. Hu, "Link Quality Analysis of Wireless Sensor Networks Based on CART Decision Tree," in *2023 International Conference on Image Processing, Computer Vision and Machine Learning (ICICML)*, 2023: IEEE, pp. 600-605, doi: 10.1109/ICICML60161.2023.10424860.

[8] K.-H. Le-Minh, K.-H. Le, and Q. Le-Trung, "A lightweight Machine-learning based Wireless Link Estimation for IoT devices," in *2022 27th Asia Pacific Conference on Communications (APCC)*, 2022: IEEE, pp. 526-531, doi: 10.1109/APCC55198.2022.9943633.

[9] S. Skaperas, L. Mamatas, and V. Tsaoussidis, "A Link-Quality Anomaly Detection Framework for Software-Defined Wireless Mesh Networks," *IEEE Transactions on Machine Learning in Communications and Networking*, 2024, doi: 10.1109/TMLCN.2024.3388973.

[10] L. Liu, M. Niu, C. Zhang, and J. Shu, "Light Gradient Boosting Machine-Based Link Quality Prediction for Wireless Sensor Networks," *Wireless Communications and Mobile Computing*, vol. 2022, no. 1, p. 8278087, 2022, doi: 10.1155/2022/8278087.

Accepted SCJ