International Journal of Engineering Trends and Technology
ISSN: 2231 – 5381 / https://doi.org/10.14445/22315381/IJETT-V70I6P205

Volume 70 Issue 6, 42-46, June 2022

Original Article

Utilizing Hybrid Machine Learning Models to Predict Quality of Service (QoS) in Multi-Channel Wireless Networks

Aisha M. Mashraqi

College of Computer Science and Information Systems, Najran University, Najran, 61441, Saudi Arabia.
aishamashraqi@gmail.com

Received: 18 March 2022 Revised: 11 May 2022 Accepted: 25 May 2022 Published: 27 June 2022

Abstract *- Understanding the behaviors of wireless networks and monitoring network flows are critical to network performance. This research combined unsupervised and supervised learning methods to develop a hybrid model for forecasting QoS in multi-channel wireless networks based on network traffic data. K-means clustering techniques were employed in the model to detect similarity in the aggregated collected data and label the dataset. The findings suggest that two is the optimal number of classes. The classified dataset was then fed into various machine learning classifiers, with the Decision Tree approach outperforming the others with a 0.971 accuracy. As a result, the Decision Tree algorithm was found to be the most effective method of forecasting QoS in multi-channel wireless networks.*

Keywords *- Clustering, Classification, Machine Learning, Wireless Networks, QoS.*

1. Introduction

Wireless networks serve an important role in daily lives, such as education, payment, and entertainment, allowing us to live comfortably. Furthermore, the four industrial revolutions, particularly the Internet, smart devices, and IoT (Internet of Things) [1] applications, have increased the need to deliver a good quality of services (QoS) [2] to clients. As a result, understanding network behavior has recently been a target of concerted research.

A variety of factors have an impact on network performance. Monitoring network traffic and analyzing performance is a complicated process. As the complexity of networks grows, network monitoring becomes increasingly difficult. Furthermore, wireless networks are the mediums shared between network devices, affecting network performance. As a result, as the number of dropped packets increases, network performance will become less satisfactory [3]. Because the networks will be congested and the signals will interfere with each other [4], particularly for video streaming applications that require a high data rate and are time-sensitive. One way is using a multi-channel approach to improve network performance in terms of throughput [5].

The primary goal of this research is to better understand wireless network behavior by identifying patterns in data and predicting risk factors that affect the quality of services provided by wireless networks. The following are the paper's targets of study:

- Data collection: The dataset for this study was gathered using the Network Simulator to perform wireless network scenarios (NS).
- Clustering: Clustering is an unsupervised learning method for identifying different traffic clusters based on dataset similarity.
- Feature selection: The feature selection approach was used to reduce irrelevant attributes in the dataset by using the filter selection method and discovering correlations between the attributes. Reduced attributes lead to
higher-quality data, which improves the quality of machine learning algorithms.
- Classifications: Four machine learning algorithms are used to help anticipate the aspects that affect network performance. They will assist professionals in planning and deploying networks and provide good QoS in wireless networks.

The following is a breakdown of the paper's structure. The paper's second section discusses related work that employs machine learning to improve network performance. The research approach is introduced in Section 3. The experimental data and their interpretation are presented in Section 4. Finally, Section 5 concludes with a discussion of future work.

2. Related Works

Machine learning methods have been considered for increasing network performance in wireless networks, either by predicting performance or finding ideal settings. One study [6] tried to predict throughput and latency using a neural network (NN) with MAC layer parameters as input data. Another study used a Decision Tree model to choose the appropriate MAC protocol for a certain application [7]. Qiao et al. [8] examined the accuracy of different algorithms such as Naïve Bayes (NB), Random Forest (RF), decision trees, and SMO [9] to select between DCF and TDMA protocols in dynamic networks.

On the other hand, machine learning tools can anticipate connection quality based on physical parameters and enhance routing [10]. The authors in [11] used machine learning approaches to predict the packet loss rate in wireless sensor networks. Another study [12] used average user throughput, number of active users, and channel quality metrics to estimate the quality of experience in cellular networks using the NN approach.
Applying supervised learning to network traffic is not always feasible, so another machine learning method that is used is unsupervised learning. Because the training data in unsupervised learning is not labeled, some machine learning algorithms rely on discovering similarities and patterns to cluster the data. Liu et al. [13] utilized a K-Means method with unsupervised data using TCP flows. Another study [14] clustered the data and found the optimal number of clusters by applying the Expectation-Maximization (EM) algorithm.

3. Hybrid Machine Learning Models

This section discusses the strategies that were utilized to forecast QoS. The method flowchart is shown in fig. 1, and the following subsections discuss the model in depth.

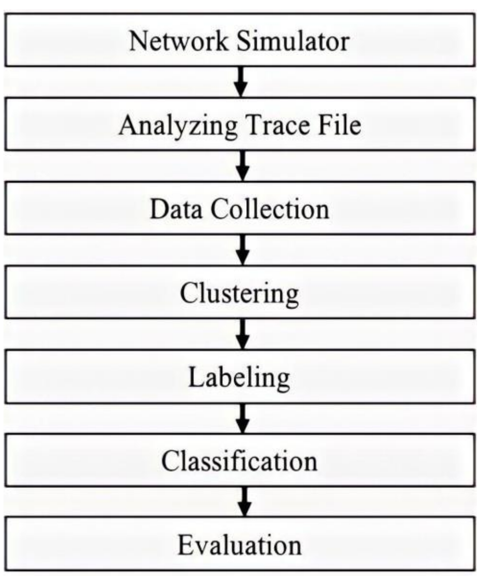

Fig. 1 Hybrid machine learning models flowchart

3.1. Data Collection

The dataset used in this study was gained from the ns-2 network simulator [15]. The following procedures were used to generate the dataset: the network topology was generated randomly with 40 nodes, and then the network was simulated for a fixed period. The trace files were analyzed using the AWK tool. The network simulation was run 700 times in this study, and the findings were gathered and used as input to machine learning algorithms. Table 1 shows the dataset attributes.

Table 1. Dataset attributes

Attributes	Description
A1	Total received packet size
A2	Total throughput (kbps)
A3	Average throughput (kbps)
A4	Flow1 throughput (kbps)
A5	Flow2 throughput (kbps)
A6	Flow3 throughput (kbps)
A7	Flow4 throughput (kbps)
A8	Flow transmission rate (Mbps)
A9	The end time for simulation (s)
A10	Number of dropped packet flow 1
A11	Number of dropped packet flow 2
A12	Number of dropped packet flow 3
A13	Number of dropped packet flow 4
A14	Packet loss ratio
A15	Total number of sending packets
A16	Total number of receiving packets
A17	Packet delivery ratio

3.2. Preprocessing

The preprocessing stage is critical for cleaning and preparing the dataset for use as input to the classification algorithm. This model used two preprocessing methods: the clustering algorithm for labeling the data set and standardization.

3.2.1. K-means

The K-means technique separates data into K clusters based on similarity [16]. It is one of the oldest and most efficient clustering algorithms. The Euclidean metric from each sample to the cluster point is calculated, and each cluster has a center point M_K. Each cluster has a center point, which is used to calculate the Euclidean metric from each sample to the cluster point. The sample is then assigned to the cluster with the shortest distance between it and the cluster center. One of the advantages of the K-means algorithm is that it allows the number of K clusters that must be pre-selected. The Silhouette Coefficient (SC), which assesses how similar a point is to its cluster relative to other clusters, is used in this study to determine the appropriate K. Because the SC ranges from –1 to 1, the SC that is closest to 1 indicates a superior clustering outcome [17].

3.2.2. Standardization

Another important process in data preprocessing is standardization, which gives equal weight to all dataset attributes. One of the most widely used standardization techniques is the mean [18] applied in this study.

3.3. Features Selection

Selecting the relevant attributes from the dataset is a vital step in the machine learning process. There are various techniques for feature selection. One common method is filtering, which was applied in this study. The filtering method assigns a score to each attribute based on the relationship between the features and the target variable. Only features in the dataset with higher K scores are considered [19].

3.4. Classifiers

During this phase, the following four machine learning classification techniques were considered:

3.4.1. Support Vector Machine (SVM)

The SVM is one of the promising classifiers for the supervised machine learning algorithm. The classification method in the SVM is based on statistical learning theory and generates a hyperplane to classify data points. In addition, the SVM is the most robust and exact classification technique [20].

3.4.2. Decision Tree (DT)

Another common machine learning method in supervised learning algorithms is the Decision Tree (DT), which studies classification and regression issues. The main idea behind the DT algorithm is that class prediction is based on the understanding of training data [21].

3.4.3. K-nearest neighbors (KNN)

In machine learning, KNN is one of the most fundamental classification methods. It assigns points to the dataset among *N* neighbors. Classification decisions are based on the computed distance between the closest data points to the samples [22].

3.4.4. Artificial neural network (ANN)

The ANN algorithm works by simulating the natural process of learning. The ANN structure is separated into layers with many nodes. The ANN structure is divided into layers, each with many nodes. Data is transformed from the input to the output layer [23,27].

3.5. Performance Evaluation

There are various measurement metrics and accuracy, precision, and recall [24,25] to consider when evaluating the performance of different machine learning algorithms.

4. Results and Discussion

Using Anaconda Navigator [26], machine learning models were implemented and examined in Python 3.8. The models' performance was evaluated using measurement variables such as accuracy, precision, and sensitivity. It was done using several machine learning classifiers and clustering approaches to predict QoS and risk indicators in wireless networks. This section explains how to use a correlation-based features selection method to find the most significant characteristics in a dataset, how to use clustering results to label the dataset, and how to use a classification technique to create a prediction model.

Pearson's correlation coefficients of dataset variables were screened to find highly associated variables in wireless networks. Correlations between dataset features are depicted as a heat map. The linear relationship between variables was described as follows: r = 0.01–1.0 shows a positive correlation (1.0 is a strong positive correlation), and r< 0 indicates a negative correlation (with –1.0 considered a strong negative correlation). Fig. 2 shows the study results as a heat map with correlation coefficient threshold values for a positive correlation, r = 0.5, and for a negative correlation, r = 0.5. A large number of missed packets resulted in a substantial negative correlation of r = –0.8 in the packet delivery ratio, lowering network throughput and decreasing the QoS. Furthermore, there is a strong positive correlation between the total number of sending packets and the packet loss ratio, implying that when it comes to improving QoS, the flow data should be carefully chosen to avoid link congestion and increase network throughput.

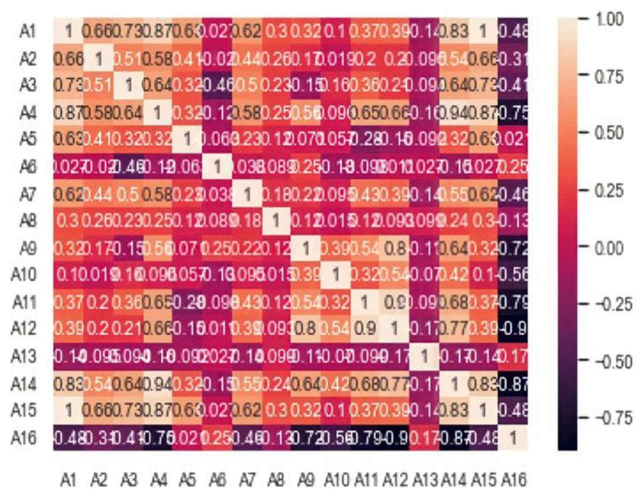

Fig. 2 Heatmap illustrating the features correlations

Clustering is a common way to study unsupervised learning, with many different clustering techniques. Clustering was used in this study to predict the class of raw data in the dataset based on similarity. The K-means approach was used to cluster the dataset, and SC determined that the best number of classes is two, as shown in fig. 3. The findings from this stage were used to label the dataset and prepare it for classifier input.

Fig. 3 Optimum number of classes in the K-means clustering algorithm.

Fig. 4 and Table 2 show the performance of different classifiers used in machine learning: Decision Tree, Nave Bayes, and Vector Machine (SVM), K-Nearest Neighbors (KNN), and Artificial Neural Network (ANN). The system is tested to see if it works properly during the validation phase; the models were tested using the train-test split method. The model was trained with 70% of the dataset, and the remaining data were used for testing. Therefore, the results show the Decision Tree achieved high accuracy, precision, and sensitivity: 0.971,0.972, and 0.986, respectively. In comparison to other algorithms, the SVM classifiers produced the worst results. Consequently, as shown in fig. 4 and table2, the Decision Tree algorithm is the best algorithm for predicting QoS and risk factors on wireless network performance.

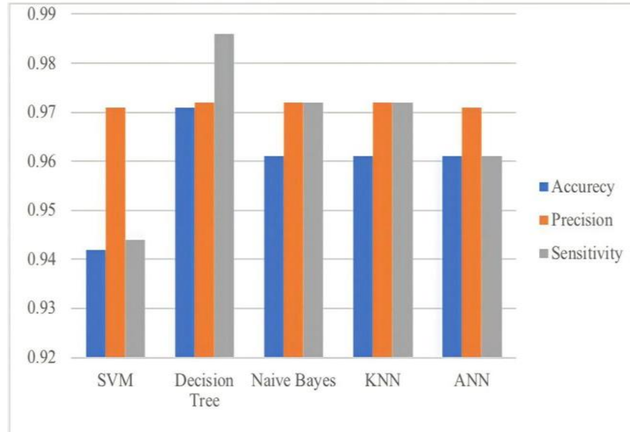

Fig. 4 Classifier performance results

Table 2. Evaluation parameters of different classifiers models

Predictive Models	Accuracy	Precision	Sensitivity
SVM	0.942	0.971	0.944
Decision Tree	0.971	0.972	0.986
Naive Bayes	0.961	0.972	0.972
KNN	0.961	0.972	0.972
ANN	0.961	0.971	0.961

5. Conclusion

Optimizing wireless network performance is a key part of QoS and ensures that customers are satisfied with the services they receive. The number of dropped packets, for example, is one aspect that affects network performance. Machine learning methods were used in this study to comprehend network behavior better. The K-means approach was utilized to label the data and serve as input for the classifiers due to the nature of the acquired data, which is unsupervised learning. Out of five classification algorithms tested, the results showed that DT is the best classifier to use in forecasting the quality of service in multi-channel wireless networks. In the future, machine learning techniques should be utilized to predict the best routing algorithms for different network scenarios to improve network performance.

References

[1] J. Gubbi, R.Buyya, S.Marusic and M. Palaniswami, Internet of Things (IoT): A Vision, Architectural Elements and Future Directions, Future Generation Computer. 29(7) (2013) 1645–1660.
[2] S. Sehrawat, R. Bora, D. Harihar, Performance Analysis of QoS supported by Enhanced Distributed Channel Access (EDCA) Mechanism in IEEE 802. 11 e. IMECS. Citeseer. (2006) 943–948.
[3] I. Yeom, A. Reddy, Realizing Throughput Guarantees in a Differentiated Services Network, In Proc IEEE International Conference on Multimedia Computing and Systems, IEEE, Italy. (1999) 372–376.
[4] K. Jain, J. Padhye, V. Padmanabhan, L. Qiu, Impact of Interference on Multi-Hop Wireless Network Performance, in Proc, MobiCom '03. 11 (2005) 471–487.
[5] J. Tang, G. Xue, W Zhang, Interference-Aware Topology Control and Qos Routing in Multi-Channel Wireless Mesh Networks, in Proc, the 6th ACM International Symposium on Mobile Ad Hoc Networking and Computing. (2005) 68–77.
[6] B. Al-Kaseem, H. Al-Raweshidy, Y. Al-Dunainawi, K. Banitsas, A New Intelligent Approach for Optimizing 6 Lowpan Mac Layer Parameters, IEEE Access. 5 (2017) 16229–16240.
[7] M. Sha, R. Dor, G. Hackmann, C. Lu, T. Kim, T. Park, Self-Adapting Mac Layer for Wireless Sensor Networks, Inproc 34th Real-Time Systems Symposium, IEEE. (2013) 192–201.
[8] M. Qiao, H. Zhao, S. Wang, J. Wei, Mac Protocol Selection Based on Machine Learning in Cognitive Radio Networks, In Proc 19th International Symposium on Wireless Personal Multimedia Communications WPMC, IEEE. (2016) 453–45.
[9] S. Keerthi, S. Shevade, C. Bhattacharyya, K. Murthy, Improvements to Platt's SMO Algorithm for SVM Classifier Design, Neural Computation. 13 (2001) 637–649.
[10] T. Liu, A. Cerpa, Foresee 4C: Wireless Link Prediction Using Link Features, in Proc the 10th ACM/IEEE International Conference on Information Processing in Sensor Networks, IEEE. (2011) 294–330.
[11] M. Kulin, E. De Poorter, T. Kazaz, I. Moerman, Towards a Cognitive MAC Layer: Predicting the MAC-Level Performance in Dynamic WSN Using Machine Learning, In Proc. EWSN2017, the International Conference on Embedded Wireless Systems and Networks. (2017) 1–2.
[12] L. Pierucci, D. Micheli, A Neural Network for Quality of Experience Estimation in Mobile Communications, IEEE Multi Media. 23 (2017) 42–49.
[13] J. Liu, I. Matta, M. Crovella, End-to-End Inference of Loss Nature in a Hybrid Wired/Wireless Environment, WiOpt'03: Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks. (2003) 9.
[14] A. McGregor, M. Hall, P. Lorier, J. Brunskill, Flow Clustering Using Machine Learning Techniques, International Workshop on Passive and Active Network Measurement, Springer. (2004) 205–214.
[15] (2021). Network Simulator. [Online]. Available: https://www.isi.edu/nsnam/ns/Last checked on 29.10.2021.
[16] A. Jain, M. Murty, P. Flynn, Data Clustering: A Review, ACM Computing Surveys CSUR. 31 (1999) 264–323.
[17] P. Rousseeuw, Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis, Journal of Computational and Applied Mathematics. 20 (1987) 53–65.
[18] W. Mining, Data Mining: Concepts and Techniques, Morgan Kaufinann. 10 (2006) 559–569.
[19] G. Chandrashekar, F. Sahin, A Survey on Feature Selection Methods, Computers & Electrical Engineering. 40 (2014)16–28.
[20] M. Hearst, S. Dumais, E. Osuna, J. Platt, B. Scholkopf, Support Vector Machines, IEEE Intelligent Systems and their Applications. 13 (1998) 18–28.
[21] F. Yang, An Extended Idea about Decision Trees, In Proc. International Conference on Computational Science and Computational Intelligence CSCI. (2019) 349–354.
[22] G. Guo, H. Wang, D. Bell, Y. Bi, K. Greer, KNN Model-Based Approach in Classification, OTM Confederated International Conferences on the Move to Meaningful Internet Systems, Springer. (2003) 986–996.
[23] T. Kohonen, An Introduction to Neural Computing, Neural Networks. 1 (1988) 3–16.
[24] W. Peng Zhang, Statistical Inference on Recall, Precision and Average Precision under Random Selection, In Proc. 9th International Conference on Fuzzy Systems and Knowledge Discovery. (2012) 1348–1352.
[25] M. Junker, R. Hoch, A. Dengel, On the Evaluation of Document Analysis Components by Recall, Precision, and Accuracy, In Proc. the Fifth International Conference on Document Analysis and Recognition, ICDAR '99 Cat. No.PR00318. (1999) 713–716.
[26] (2021). Anaconda, I. Anaconda Navigator. [Online]. Available: https://docs.anaconda.com/anaconda/navigator/index.html.
[27] A. Bhatnagar, R. Varshney and K. Manu, A Review: Energy Efficiency using Various Protocols in Wireless Sensor Networks, SSRG International Journal of Electronics and Communication Engineering SSRG-IJECE. 4(4) (2017) 26-29.