



Learning-Based Evaluation of Routing Protocol in Vehicular Network Using WEKA

Amal Hadrich, Amel Meddeb Makhlouf, Faouzi Zarai

► To cite this version:

Amal Hadrich, Amel Meddeb Makhlouf, Faouzi Zarai. Learning-Based Evaluation of Routing Protocol in Vehicular Network Using WEKA. 17th IFIP International Conference on Network and Parallel Computing (NPC), Sep 2020, Zhengzhou, China. pp.55-67, 10.1007/978-3-030-79478-1_5. hal-03768737

HAL Id: hal-03768737

<https://inria.hal.science/hal-03768737>

Submitted on 4 Sep 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



This document is the original author manuscript of a paper submitted to an IFIP conference proceedings or other IFIP publication by Springer Nature. As such, there may be some differences in the official published version of the paper. Such differences, if any, are usually due to reformatting during preparation for publication or minor corrections made by the author(s) during final proofreading of the publication manuscript.

Learning-based evaluation of routing protocol in Vehicular Network using WEKA

Amal Hadrich¹, Amel Meddeb Makhoul^{2,3}, and Faouzi Zarai³

National School of Electronics and Telecommunications, Sfax, Tunisia

Abstract. Internet of things connects any object to another and communicate with them using the most performed routing protocol. But in vehicular networks, topology and communication links frequently change due to the high mobility of vehicles. So, the key challenge of our work is to choose the best routing protocol using machine learning algorithms. When choosing routing protocol, most research focuses on the improvement of the performance of specific routing protocol using one machine learning algorithm. In this paper, we propose a solution in order to find the best routing protocol in such critical condition using machine learning algorithms in order to maximize the precision of the true positive rate. After the use of a specific algorithms such as Artificial Neural Network, Random Forest and Naive Bayes, we found that the last one is the best algorithm when it have all the true positive rate with a precision equal to 0.9987 to select the best routing protocol.

Keywords: Vehicular Networks - Routing Protocol - Machine Learning Algorithms.

1 Introduction

Recent advances in networks have given rise to the emergence of vehicular networks. The research work has started from the past few years on the area of Mobile Ad-hoc Networks also called as MANETS (Mobile Adhoc NETWORKS). It allows the mobile nodes to communicate in one to one and one to many without any predefined infrastructure. The protocols which are required to support MANETS are more complex when compared to other non-mobile networks because of the mobility and of the non-existence predefined infrastructure or topology for MANETS. New technologies are used to provide more and more facilities including safety applications. Among these networks, Vehicular networks are a subclass of MANETs, where vehicles are simulated as mobile nodes. The vehicular network is a self organizing network to provide Intelligent Transport System(ITS). Users in vehicles connect to the Internet at any time in order to obtain the desired services. There are three types of common communication models in vehicular networks: vehicle-to-infrastructure (V2I), vehicle-to-roadside (V2R) and vehicle-to-vehicle (V2V) [15]. The routing protocol is an important aspect in communication between networks. It is a set of rules that are framed for exchanging the information in a network from one node to another node. MANET routing protocols are divided into three categories: Proactive routing

protocol means that the routing information is maintained in the background irrespective of communication requests like the protocol OLSR (Optimized Link State Routing Protocol), reactive routing protocol opens the route only when it is necessary for a node to communicate with each other such as AODV (Ad-hoc On-demand Distance Vector), DSR (Dynamic Source Routing) and ZRP (Zone Routing Protocol) is an example of hybrid routing protocols that is introduced to reduce the control overhead of proactive routing protocols and decrease the initial route discovery delay in reactive routing protocols. The main difference between the MANET and vehicular routing is network topology, mobility patterns, demographics, density of nodes at different timings, rapid changes in node arriving and leaving the network. Vehicular networks was divided into three categories [1, 2]:

* Routing information: It is also divided into two categories like topology-based routing protocols that use link information that exists in the network and position-based routing protocols use GPS devices.

* Transmission strategies: It is divided into three categories, where broadcast routing protocols are frequently used in vehicular network, multicast such as cluster-based routing protocols and geocast routing protocols and unicast.

* Delay tolerant: It is useful to help balance traffic loads and reduces traveling time.

Figure 1 represents vehicular routing protocol chosen in our work and their categories.

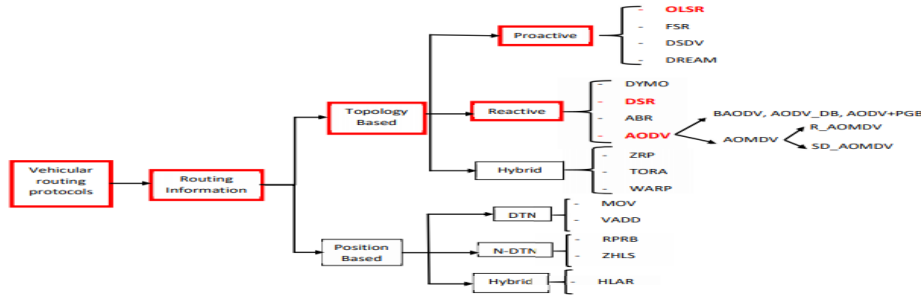


Fig. 1: Vehicular routing protocol chosen in our work.

Vehicles in their network have high mobility, which causes a change in the network topology and the connections instability. Recently, the nature-inspired intelligence algorithms have been widely used as direct search and optimization tools in many different problems where machine learning algorithms are used for the choice of the best route for the devices. For this reason, we propose a solution to choose the best routing protocol depending on the network topology using machine learning algorithms. The remainder of this paper is organized as follows. Section 2 reviews the related works. Section 3 introduces our proposed solution in order to choose the perfect routing protocol. Section 4 shows the

evaluation results. Section 5 concludes this paper and gives some possible future perspectives.

2 Related work

Researchs focus to design an efficient routing algorithms that is adaptable to the mobility characteristics of vehicular network.

Marzak *et al.* [3] proposed a new vehicular clustering algorithm based on the ANN (Artificial Neural Network) system to select cluster-head. This algorithm increases network connectivity and improves the stability of the cluster structure. Moreover, research in [4] introduced a Modified AODV routing protocol of VANET communications using fuzzy neural network algorithm to determine the best route from the source node to the destination node through several intermediate nodes. Because of the importance of stability and the security in vehicular networks, research in [6] presents a secure and stable AODV, named GSS-AODV, based on the fuzzy neural network to ensure that the calculated node stability is in accordance with the actual situation. Zaho *et al.* [5] proposed a greedy forwarding algorithm based on SVM (Support Vector Machine) to process the vehicle data and to generate routing metric. This algorithm reduces the packet loss and network delay.

The above-mentioned literatures [3–6] mainly use machine learning methods to increase the network connectivity, determine the best route, improve the performance of the network and applications, increase the stability of the protection and reduce the packet loss and network delay. But, they still do not applicated this approach only in one protocol or a cluster and use only one algorithm named ANN or SVM. However, to determine the best protocol, we have to use several machine learning algorithms according to the high value of accuracy to avoid the problem of the accidents and improve the performance and the protection. The most related work is presented by Marzak *et al.* [3] because it selects the best vehicular routing protocol using machine learning algorithm according to the feature specification, but the last mentionned autor focused on cluster routing protocol and use only ANN algorithm and in our solution, we focus in all routing protocol with the apply of machine learning algorithms.

Table 1 summarizes the studied related works and shows their limits.

3 Proposed solution

Our proposed solution set based on the parameter of routing protocol inserted in the dataset and estimates the optimal protocol for data transmission using machine learning algorithms in order to minimize the cost of network maintenance. The proposed approach must ensure optimal operation of the network by maximizing the stability, minimizing the delivery delay of information and bandwidth.

For this research, we propose to study three vehicular routing protocols based on three specifications, which are "AODV", "DSR" and "OLSR" with three

Table 1: Comparison with the existing and proposed solution

Solution	Algorithm	Protocol	Limits
Clustering in vehicular ad-hoc network using artificial neural network [3]	Artificial Neural Network	Clustering routing protocols	Focus on cluster routing protocol and use only ANN algorithm.
AODV Routing Protocol Modification With Dqueue (dqAODV) and Optimization With Neural Network For VANET In City Scenario [4]	Fuzzy neural network algorithm	A modified AODV routing protocol	Focus only on AODV routing protocol using fuzzy neural network.
Improving Security and Stability of AODV with Fuzzy Neural Network in VANET [6]	Fuzzy neural network	GSS-AODV	Focus only on AODV routing protocol using fuzzy neural network.
SVM based routing scheme in VANETs[5]	SVM neural network	Vehicular routing protocol	Use only SVM neural network.
Proposed solution	All algorithms in Neural Network	All routing protocols	————

common parameter "TTL", "QUEUE LENGTH" and "HELLO INTERVAL". There is more routing protocol with their categories like position based, delay tolerant and transmissions strategies but we choose this three routing protocol because we find that this have some common carectrestic and features based for the select the best routing protocol.

A particular mobile node may wish to change certain of the parameters, in particular the HELLO INTERVAL. In the latter case [10], the node should advertise the HELLO INTERVAL in its Hello messages, by appending a Hello Interval Extension.

The routing protocol uses minimum hop count criteria to establish routes between sources and destinations. However [11], selecting the routing protocol with minimum hop is not usually the best decision. A node which might not have sufficient available buffer size could be leading to have high probability of packet loss and network congestion and link breakage. Thus, the routing algorithm consider available node queue length during routing protocol selection process.

To control congestion by applying efficient local route repair method we should use the value of TTL [12]. A TTL value is only decreased by one at each node. When the TTL value declines to zero, the route request packet should be dropped even though it didn't find the destination.

To choose the best routing protocol based on the three parameters, we base our analysis on the following Machine Learning Algorithms:

- * ANN, which are computational algorithms in order to simulate the behavior of biological systems composed of neurons [7],
- * Random Forest, which randomly creates and merges multiple decision trees into one forest in order to create a collection of decision models to improve accuracy [9]
- * Naive Bayes, which are a collection of classification algorithms based on Bayes theorem, it is not a single algorithm but a family of algorithms where all of them share a common principle, i.e. every pair of features being classified is independent of each other[8].

In section 3.2 there are more description about this machine learning algorithm mentioned above. These three machine learning algorithms show interesting results to select the perfect vehicular routing protocol. Figure 2 shows the proposed algorithm of our solution in order to minimize the cost of network maintenance.

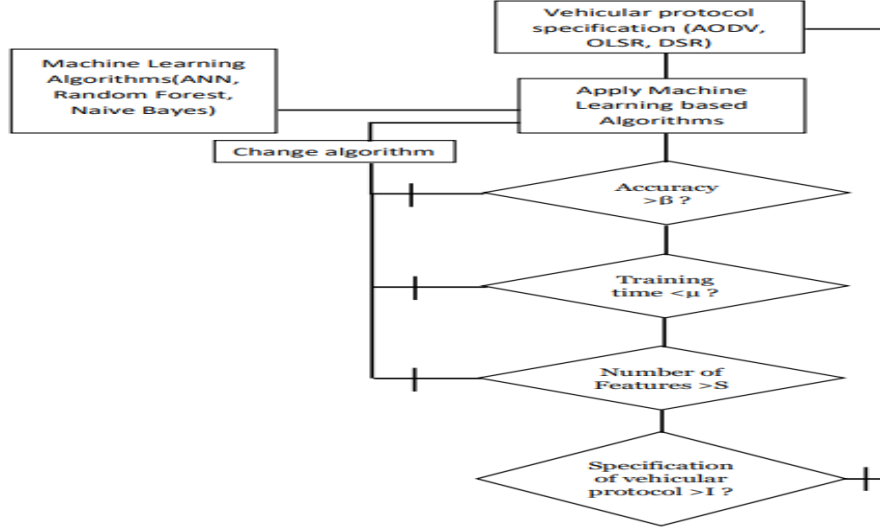


Fig. 2: Choose of the perfect vehicular routing protocol with the apply of machine learning based algorithms.

First, we choose three specific vehicular routing protocol named AODV, OLSR and DSR then we apply one machine learning algorithm like ANN, Random Forest and Naive Bayes. To test our solution, we must compare the accuracy with a threshold value beta flows the equation below 1.

$$\text{beta} = \frac{2 * \text{precision} * \text{recal}}{\text{precision} + \text{recall}} \quad (1)$$

where recall and precision show in equations 2 and 3.

$$recall = \frac{TP}{TP + FN} \quad (2)$$

$$precision = \frac{TP}{TP + FP} \quad (3)$$

where TP, FN and FP represent respectively True Positive , False Negative and False Positive. If the accuracy down to beta, we must change the machine learning algorithm. And if not, we compare the training time to a fixed value. If the time upper to this value, we must switch the machine learning algorithm. If not, we move to the next step in order to select the best routing protocols with the perfect feature. So, we calculate the number of features to decide. If the number of features is down to S, then we change the machine learning algorithm and if not, we test the specification of the routing protocol. If this result down to a fixed value, representing the specification of the previous result, then we rebuild the algorithm and if not then we set this algorithm as the best vehicular routing protocol using the perfect machine learning algorithm. The table 2 shows the different features used in our proposed algorithm.

Table 2: Definition of the used features

Feature Name	Definition
Time To Live (TTL)	It's the number of hops that a packet is permitted to travel before being discarded by a router.
QUEUE LENGTH	It is used to hold packets while the network interface is in the process of transmitting another packet.
HELLO INTERVAL	A node may offer connectivity information by broadcasting local Hello messages. A node should only use hello messages if it is part of an active route.

First, we need a description about the routing protocol and the algorithms chosen.

3.1 Improved protocol description

Security protocols are optimized for routing and routing maintenance protocol processes.

AODV routing protocol: The Ad hoc On-Demand Distance Vector (AODV) algorithm enables dynamic, self-starting, multihop routing between participating mobile nodes wishing to establish and maintain an ad hoc network. AODV allows mobile nodes to obtain routes quickly for new destinations, and does not require nodes to maintain routes to destinations that are not in active communication. AODV allows mobile nodes to respond to link breakages and changes in network topology in a timely manner. Route Requests (RREQs), Route Replies (RREPs), and Route Errors (RERRs) are the message types defined by AODV [11].

OLSR routing protocol: The Optimized Link State Routing Protocol (OLSR) operates as a table driven, proactive protocol, i.e., ex-changes topology information with other nodes of the network regularly [13]. Each node selects a set of its neighbor nodes as MultiPoint Relays (MPR). In OLSR, only nodes, selected as such MPRs, are responsible for forwarding control traffic, intended for diffusion into the entire network. MPRs provide an efficient mechanism for flooding control traffic by reducing the number of required transmissions [13].

DSR routing protocol: The Dynamic Source Routing protocol (DSR) is a simple and efficient routing protocol designed specifically for use in multi-hop wireless ad hoc networks of mobile nodes [14]. Using DSR, the network is completely self-organizing and self-configuring, requiring no existing network infrastructure or administration. Network nodes cooperate to forward packets for each other to allow communication over multiple hops between nodes not directly within wireless transmission range of one another.

3.2 Machine learning algorithms description

Machine learning is a method of data analysis that automates analytical model building. It is a branch of technology that allows systems to learn from data, identify patterns and make decisions with minimal human intervention. It has revolutionized the world of computer science by allowing learning with large datasets, which enables machines to change, restructure and optimize algorithms by themselves [16]. After analysis, we choose to use three machine learning based algorithms, namely, Artificial Neural Network (ANN), Naive Bayes and Random Forest, because they have the most interesting results compared to other machine learning algorithms.

Naive Bayes: Classification is a fundamental issue in machine learning and data mining. Naive Bayes is the simplest form of Bayesian network, in which all attributes are independent given the value of the class variable [8]. Figure 3 shows Naive Bayes of our solution where QL and HI represent respectively QUEUE LENGTH and Hello intervall. In naive Bayes, each attribute node has no parent except the class node.

Random Forest: Random Forests in machine learning is an ensemble learning technique about classification, regression and other operations that depend on a multitude of decision trees at the training time. They are fast, flexible, represent a robust approach to mining high-dimensional data [9]. Figure 4 shows Random Forest algorithm of our solution.

Artificial Neural Network: Artificial neural network (ANN) is a computational model that consists of several processing elements that receive inputs and deliver outputs based on their predefined activation functions. The ANN

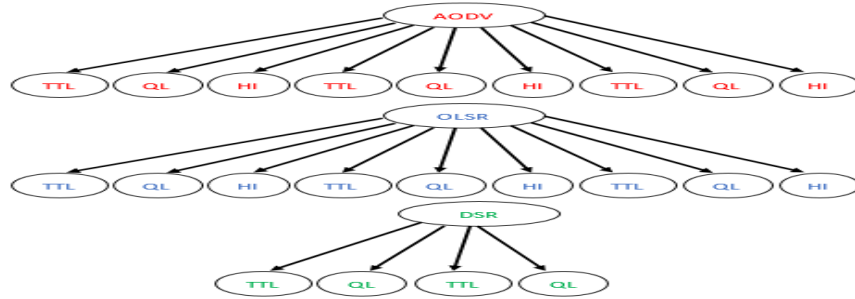


Fig. 3: Naive Bayes algorithm of our solution.

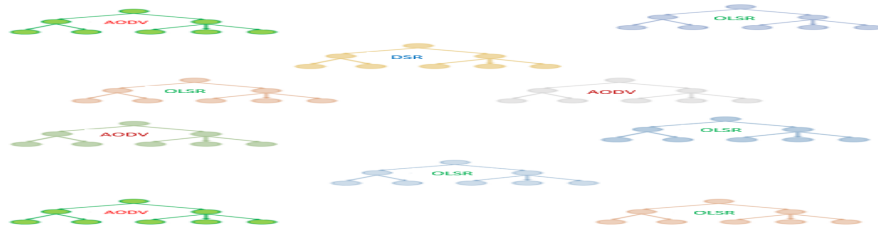


Fig. 4: Random Forest algorithm of our solution.

consists of a set of processing elements, also known as neurons or nodes, which are interconnected [7]. Figure 5 represents the ANN algorithm to our solution.

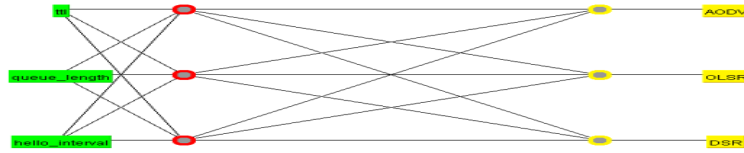


Fig. 5: ANN algorithm for our solution

4 Evaluation results

For the evaluation procedure, we propose to use WEKA, which is a machine learning workbench that supports many activities of machine learning practitioners. The main features of WEKA are [17]:

* Data preprocessing: As well as a native file format (ARFF), WEKA supports various other formats (for instance CSV, Matlab ASCII files), and database connectivity through JDBC.

* Classification: One of WEKA's drawing cards is the more than 100 classification methods it contains.

* Clustering: Unsupervised learning is supported by several clustering schemes.

* Attribute selection: The set of attributes used is essential for classification performance.

* Data visualization: Data can be inspected visually by plotting attribute values against the class, or against other attribute values.

To test our solution, we choose k-folds cross-Validation model. It's a re-sampling procedure used to evaluate machine learning models on a limited data sample. The procedure has a single parameter called k that refers to the number of groups that a given data sample is to be split into. The general procedure is as follows:

- 1- Shuffle the dataset randomly.
 - 2- Split the dataset into k groups
 - 3- For each unique group:
 - A- Take the group as a hold out or test data set.
 - B- Take the remaining groups as a training data set.
 - C- Fit a model on the training set and evaluate it on the test set.
 - D- Retain the evaluation score and discard the model.
 - 4- Summarize the skill of the model using the sample of model evaluation scores.
- Importantly, each observation in the data sample is assigned to an individual group and stays in that group for the duration of the procedure. This means that each sample is given the opportunity to be used in the hold out set 1 time and used to train the model k-1 times.

The main objective of our work is to choose the best routing protocol based on the calculation of the performance for each protocol and algorithm in order to minimize the cost of network maintenance. First, we set k value to 10. The retrieved value of the accuracy follows the matrix below 4.

$$Accuracy = \begin{pmatrix} RandomForest & NaiveBayes & ANN \\ 82.22\% & 88.88\% & 80\% \end{pmatrix} \quad (4)$$

The simulation results show that the Random Forest Algorithm takes a delay time equal to 0.15 s, ANN algorithm takes about a 0.29 s and Naive Bayes algorithm takes a delay time equal to 0.1 s to build their model. Figures 6a, 6b and 6c show the results of our machine learning algorithms, where the abscissa axis is the protocol name and y-axis is the predicted protocol name. Figure 6a shows the result of the Random Forest Algorithm. We notice that the random forest presents errors in DSR and OLSR routing protocol, because the two protocols represent some common interval of the value of TTL and queue length. Figure 6b shows the result of the application of the ANN Algorithm. We notice that ANN presents errors in DSR routing protocol predicted as OLSR protocol and OLSR protocol predicted as AODV and DSR protocols. Figure 6c shows the result of Naive Bayes Algorithm. We notice in this case that Naive Bayes presents errors in DSR routing protocol predicted as OLSR protocol.

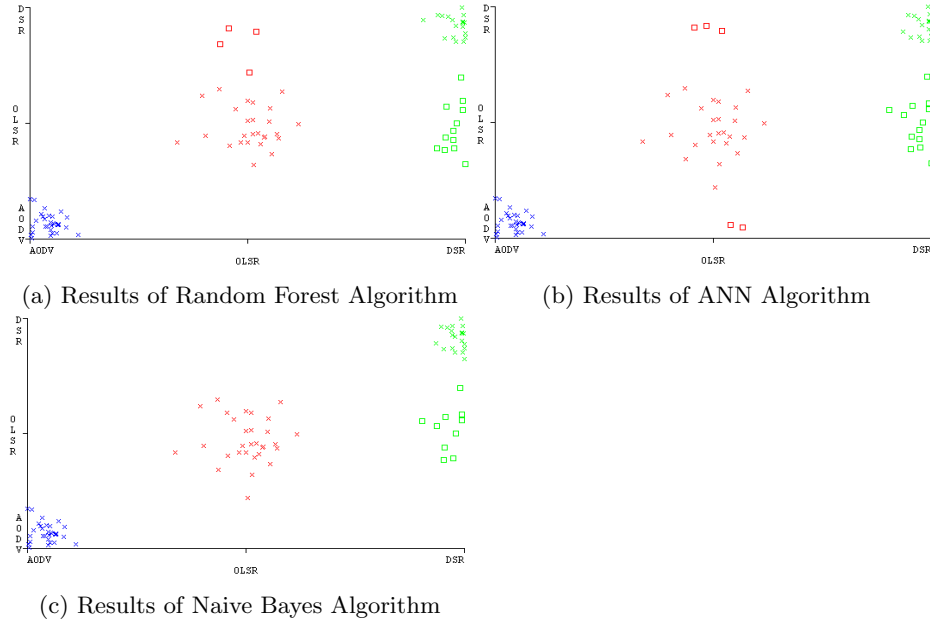


Fig. 6: Results of the chosen machine learning algorithms.

To evaluate the performance of our solution, we should choose the algorithm with the accuracy ≥ 0.8 . Thus, the best three algorithms are Naive Bayes, ANN and Random Forest. Then, we move to fix the training time, where we choose to eliminate the algorithm with training time > 6 s. For the algorithm Random Forest, the training time is 3s, naive Bayes is 1s and ANN algorithm is 7s. Thus, the best two algorithms, according to the time metric, are Random Forest and Naive Bayes. Finally, we should have one feature to chose the best routing protocol. For the Naive Bayes the OLSR and AODV have a predicted margin between 0.93 and 1 for the three features TTL, HELLO INTERVAL and QUEUE LENGTH. The OLSR routing protocol has the lowest predicted margin than AODV routing protocol for the two features TTL and QUEUE LENGTH. Thus, the best routing protocol for the naive bayse is AODV. For Random Forest, the routing protocol AODV has a predicted margin between 0.91 and 1 for the three features TTL, HELLO INTERVAL and QUEUE LENGTH. Then we should compare the margin prediction between AODV routing protocol for the two algorithms. The predicted margin in Naive Bayes is 0.99 and in Random Forest equal to 0.91 for the two features for TTL and QUEUE LENGTH. Than, we can conclude, using these conditions, that best routiong protocol is AODV the algorithm Naive Bayes. Table 3 resumes the results about the best algorithm.

To verify our resolution, Figure 7 shows the threshold curve of the OLSR routing protocol using the Naive Bayes algorithm where the abscissa axis is the

Table 3: Summarize of the choose the best algorithm

Algorithms \ Steps	Accuracy	Training time	Number of Features
ANN	AODV,OLSR	✗	NaN
Random Forest	AODV,OLSR	AODV	✗
Naive Bayes	AODV,OLSR	AODV,OLSR	AODV

True Positive Rate and y-axis is the precision. We notice that the true positive rate is less than 0.7, which is an interesting result of our simulations.

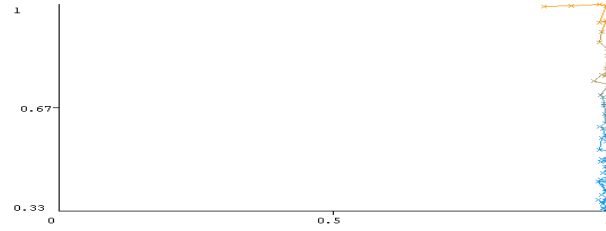
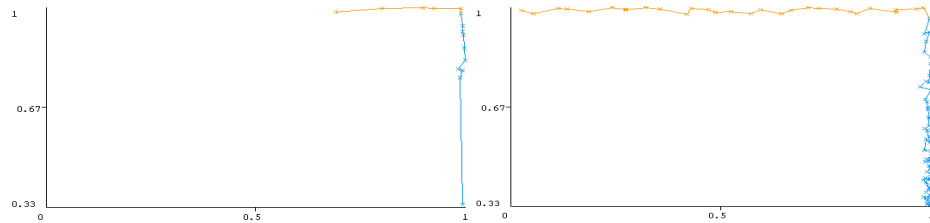


Fig. 7: The threshold curve for OLSR routing protocol with Naive Bayes algorithm

Figure 8a shows the threshold curve of AODV routing protocol for the Random Forest algorithm, to show the variation of the precision with respect to the True Positive Rate. We notice that the true positive rate values are between 0.7 and 1. Finally, Figure 8b shows the threshold curve of AODV routing protocol for the Naive Bayes algorithm. We notice that all the true positive rates are evaluated with a precision equal to 0.9987.



(a) Threshold curve using Random Forest algorithm (b) Threshold curve using Naive Bayes algorithm

Fig. 8: Threshold curve of AODV routing protocol.

In conclusion, under specific conditions, the best routing protocol chosen is AODV using Naive Bayes Algorithm with precision of the true positive rate equal to 0.9987. Naive Bayes is the most suitable algorithm of our proposed solution since it has the most greater precision, where the AODV routing protocol is simple to be determined.

5 Conclusion

As the population is becoming more and more larger in the world, so vehicles are increasing per day and simultaneously accidents are rising. There is no protection for human life and the vehicles. Vehicular network is the mostly utilized connection that communicates among the vehicles. It is not only utilized for the communication purpose but also could be structured for navigation and traffic control. In this manuscript, a routing based solution has been lightened that would help in the reduction of accidents based on the choose of the best routing protocol using specific machine learning algorithms. The performance parameters are computed to depict the effectiveness of the proposed work. The threshold curve indicates that the AODV routing protocol when using Naive Bayes algorithm is the best vehicular routing protocol, based on simulation hypothesis. In our future work, we plan to use more features to choose the best routing protocol. These features can be based on security and network characteristics, which increase the efficiency of our solution.

References

1. Chakkaphong Suthaputchakun and Zhili Sun.: "Routing protocol in intervehicle communication systems: a survey". In: IEEE Communications Magazine **49**(12), 150–156 (2011)
2. Yun-Wei Lin, Yuh-Shyan Chen, and Sing-Ling Lee.: "Routing protocols in vehicular ad hoc networks: A survey and future perspectives." In: J. Inf. Sci. Eng. **26**(3), 913–932 (2010)
3. Bouchra Marzak et al.: "Clustering in vehicular ad-hoc network using artificial neural network". In: International Review on Computers and Software (IRECOS) **11**(6), 548–556 (2016)
4. Saha. , Soumen and Roy. , Utpal and Sinha. , DD. : MATEC Web of Conferences.AODV Routing Protocol Modification With Dqueue (dqAODV) and Optimization With Neural Network For VANET In City Scenario.EDP Sciences, Vol.57, p. 02001. (2016)
5. Zhao., Liang and Li., Yujie and Meng., Chao and Gong., Changqing and Tang., Xiaochun.: 2016 16th International Symposium on Communications and Information Technologies (ISCIT).A SVM based routing scheme in VANETs. IEEE, pp. 380–383 (2016)
6. Huang., Baohua and Mo., Jiawei and Cheng., Xiaolu.: International Conference on Wireless Algorithms, Systems, and Applications. Improving Security and Stability of AODV with Fuzzy Neural Network in VANET. Springer, pp. 177–188 (2018)

7. Prieto., Alberto and Prieto., Beatriz and Ortigosa., Eva Martinez and Ros., Eduardo and Pelayo., Francisco and Ortega., Julio and Rojas., Ignacio.:Neural networks: An overview of early research, current frameworks and new challenges. *Neurocomputing* **214**,242–268 (2016)
8. Zhang., Harry.: The optimality of naive Bayes. *AA* **1**(2),3 (2004)
9. Segal., Mark R.:Machine learning benchmarks and random forest regression (2004)
10. Park., Neung-Um and Nam., Jae-Choong and Cho., You-Ze.:2016 International Conference on Information and Communication Technology Convergence (ICTC). Impact of node speed and transmission range on the hello interval of MANET routing protocols.IEEE, pp. 634-636 (2016)
11. Ahmed., Bisengar and Mohamed., Rziza and Mohamed., Ouadou.:Queue length and mobility aware routing protocol for mobile ad hoc network. *International Journal of Communication Networks and Information Security* **4**(3), 207 (2012)
12. Bindra., Puneet and Kaur., Jaswinder and Singh., Gurjeevan.: Effect of TTL parameter variation on performance of AODV route discovery process. *International Journal of Computer Applications* **70**(4), (2013)
13. Clausen., Thomas and Jacquet., Philippe.:RFC3626: Optimized link state routing protocol (OLSR). RFC Editor, (2003)
14. Johnson., David B and Maltz., David A and Broch., Josh and others: DSR: The dynamic source routing protocol for multi-hop wireless ad hoc networks. *Ad hoc networking* **5**(1), 139–172 (2001)
15. Ksouri., Chahrazed and Jemili., Imen and Mosbah., Mohamed and Belghith., Abdelfettah.:International Workshop on Distributed Computing for Emerging Smart Networks.VANETs Routing Protocols Survey: Classifications, Optimization Methods and New Trends.Springer, pp.3–22 (2019)
16. Ayodele., Taiwo Oladipupo.:Machine learning overview. *New Advances in Machine Learning*, 9–19 (2010)
17. Bouckaert, Remco R and Frank, Eibe and Hall, Mark A and Holmes, Geoffrey and Pfahringer, Bernhard and Reutemann, Peter and Witten, Ian H.: The Journal of Machine Learning Research.WEKA—Experiences with a Java Open-Source Project **11**, 2533–2541 (2010)