On-demand High Mobility Vehicular Connectivity Prediction using Random Forest Classification

Farooque Hassan Kumbhar, Soo Young Shin*
farooque.kumbhar@gmail.com, * wdragon@kumoh.ac.kr

Kumoh National Institute of Technology, Gumi, South Korea.

Abstract

Innovations and advances in the communication technologies for vehicular networks enable high data rates and efficient network access. However, the stochastic nature of high mobility vehicular networks creates challenges in ad-hoc communication and efficient route selection. We believe that pro-actively identifying the connectivity between two communicating devices can help in efficient route selection for higher successful message delivery. In this paper, we design a comprehensive dataset using OpenStreetMap and SUMO traffic trace to evaluate three major machine learning techniques on high mobility vehicles. Our results identified that the random forest classification method outperforms logistic regression and support vector machine in terms of accuracy and F1-score metrics.

Ⅰ. Introduction

Intelligent transportation systems (ITS) offer a variety of services such as adaptive cruise, automatic braking systems, blind spot, congestion control, and many more [1]. These next-generation services require flawless connectivity between commuting vehicles, enabling which, is a tedious task. State-of-the-art communication technologies of dedicated short-range communication (DSRC) 802.11p and cellular vehicle-to-any (C-V2X), and long-range cellular networks with 5G and 6G provide connectivity to the vehicular networks [2]. However, often these networks face packet drops and degraded service due to high mobility environment [3], [4]. Considering the root cause of these network challenges, we believe that a predictive solution that proactively identifies connectivity between two moving devices can help in reducing network degradation. In this paper, we contribute to the novelty, as follows: 1) We observe and design a comprehensive dataset using real-world map through OpenStreetMap and simulation for urban mobility (SUMO) traces, 2) We design and model connectivity time between two devices spanning over to five predicted classes, 3)We train and evaluate three machine learning (ML) techniques (random forest classification (RFC), logistic regression (LR) and support vector machine (SVM)), and 4) Our thorough results and observation highlight that the RFC outperforms other ML techniques in vehicular dataset..

Ⅱ. Trust Establishment and SP Selection

We propose to train a fog node-based ML model using vehicular mobility dataset so that it can be used to predict connectivity time between two devices. Each connected vehicle shares its information using basic safety messages (BSM) containing information such as direction, velocity, location, etc. Our dataset utilizes this data to create smart features that are used to predict connectivity between two vehicles. We identify the relation of BSM data with pairwise communication time. An immediate use case of predicted connectivity time between vehicles is to choose the best possible route; however, predictive resource allocation, efficient clustering, and many others are possible. Nevertheless, a trained supervised learning model (ML category) can predict various classes for the target label. We designed our dataset to have five features and one target label for prediction, i.e., $\lambda_{ij}, \Delta V_{ij}, \Delta\theta_{ij}, \Delta\theta^L_{ij}, Tend_{ij}$and Γ^L_{ij}. All six features (five

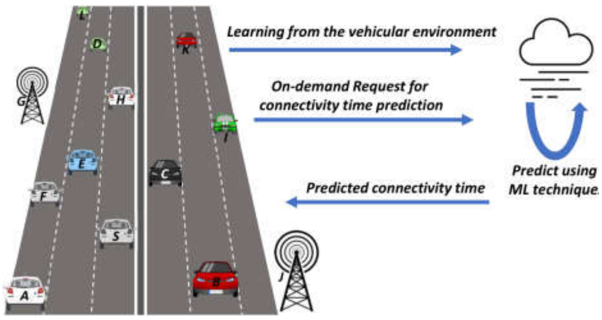
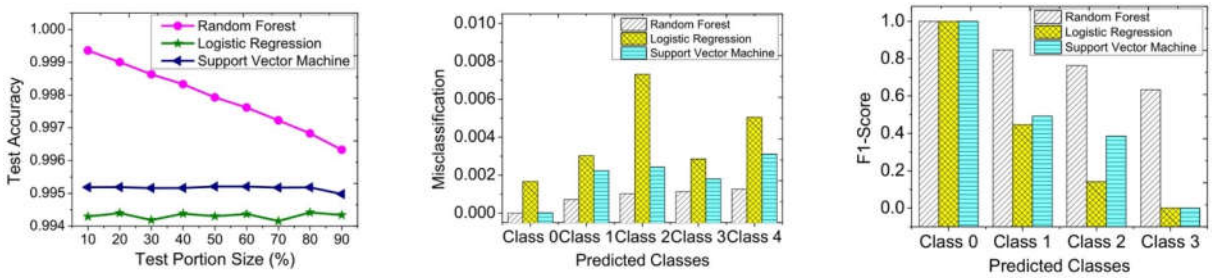


Figure 1 Proposed scheme

features and one target label) are, as follows: 1) Euclidean distance is the distance between two moving devices and calculated as, $\lambda ij = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$. 2) Relative velocity is a measure of how one device is moving with respect to another, $\Delta V_{ij} = \sqrt{(\varphi_i - \varphi_j)^2} + 1$. 3) Direction difference ($\Delta\theta_{ij}$) between devices can be calculated using vehicles current movement directions (θ_i and θ_j), where north points to 90° always, as given: $\Delta\theta_{ij} = \left|\left((\theta_i - \theta_j + 180)\%360 - 180\right)\right|$ 4) We



(a) Proposed fog node prediction (b) Proposed fog node prediction (c) Proposed fog node prediction

Figure 2 Performance evaluation of the proposed scheme

designed three labels for direction difference between two moving devices represented as ($\Delta\theta^{L}_{ij}$) using the direction of vehicles i and j, as follows:

$$\Delta\theta^{L}_{ij} = \begin{cases} 0 \text{ (Same)} & \text{if } \Delta\theta_{ij} \leq 60 \\ 2 \text{ (Opposite)} & \text{if } \Delta\theta_{ij} \geq 120 \\ 1 \text{ (Neither)} & \text{otherwise} \end{cases} \quad (1)$$

5) The tendency label ($\text{Tend}^{\,L}_{ij}$) is used to identify vehicles movement (*i,j*) is in the same direction or in opposite directions, as given below:

$$\text{Tend}^{\,L}_{ij} = \begin{cases} 0 & \text{if } \Delta\theta^{L}_{ij} == 2\lambda_{ij}(t2) - \lambda_{ij}(t1) < 0 \\ 1 & \text{if } \Delta\theta^{L}_{ij} == 2\lambda_{ij}(t2) - \lambda_{ij}(t1) > 0 \end{cases} \quad (2)$$

where $\lambda_{ij}(t1)$ and $\lambda_{ij}(t2)$ represent the vehicle's inter-distances, at times t1 and t2, respectively.

6) Connectivity time (Γ^{L}_{ij}) between two devices is designed with five classes as given below:

$$\Gamma^{L}_{ij} = \begin{cases} L0 & \text{if } \Gamma_{ij} == 0 \\ L1 & \text{if } \Gamma_{ij} > 2 \quad \Gamma_{ij} \leq 5 \\ L2 & \text{if } \Gamma_{ij} > 5 \quad \Gamma_{ij} \leq 10 \\ L3 & \text{if } \Gamma_{ij} > 10 \, \Gamma_{ij} \leq 15 \\ L4 & \Gamma_{ij} > 15 \end{cases} \quad (3)$$

Fig. 1 illustrate the proposed scheme where a fog node utilizes environment data to train supervised learning models. A vehicle from the environment can request using REST API for connectivity prediction between two devices.

Ⅲ. Performance Evaluation and Conclusion

We extract a real-world road scenario of central Seoul, South Korea using OpenStreetMap and populate the vehicular devices with DSRC connectivity using the SUMO simulator in python simulation. The dimension of the simulation model is 2.5 × 1.5 km with 30 intersections, having simulated 200 to 2500 vehicles for 3600 seconds. We train our fog node using three supervised learning models, i.e., RFC, LR, and SVM. Each model was trained with various test/train splits and expected to predict a class out of a total of five classes for connectivity time. Fig. 2a shows that the RFC outperforms LR and SVM by achieving more than 99.9% accuracy, whereas LR and SVM gain 99.4% and 99.5%, respectively. However, it can be seen that the increase in the test portion (which is a reduction in the training portion) impacts RFC negatively. On the other hand, Fig. 2b highlights that all three supervised learning methods have very small misclassification errors. Misclassification shows that how a predicted class is wrongly assigned a class label. Moreover, Fig. 2c shows the F1-score which is a ratio of recall and precision. The F1-score shows the true positive for a predicted class where a high value is good. The RFC performs well as compared to LR and SVM. However, use case evaluation of the proposed scheme is still an open challenge.

In this paper, we present a novel fog node-based mechanism to predict connectivity time between two moving devices. Moreover, we evaluate three well-known supervised learning methods for connectivity time prediction, i.e., RFC, LR, and SVM. The RFC outperforms its counter-part methods by achieving 99.9% accuracy with negligible misclassification and better F1-score.

ACKNOWLEDGMENT

This work was supported by Brain Pool Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (2019H1D3A1A01102978).

References

[1] F. Tang, Y. Kawamoto, N. Kato, and J. Liu, "Future Intelligent and Secure Vehicular Network Toward 6G: Machine-Learning Approaches," Proceeding IEEE, 2020, 108(2), pp. 292-307.

[4]. Irfan Azam, Muhammad Basit Shahab, Soo Young Shin." Role Switch-ing and Power Allocation Technique for Mobile Users in Non-orthogonal Multiple Access", Physical Communication, 2020, 43, p. 101179.

[3]. A. Ullah, X. Yao, S. Shaheen, and H. Ning, "Advances in Position Based Routing Towards ITS Enabled FoG-Oriented VANET-A Survey," IEEE Transaction Intelligent Transportation System, 2020, 21(2), pp. 828-840.

[4]. F. H. Kumbhar, S. Y. Shin," DT-VAR: Decision Tree Predicted Compatibility based Vehicular Ad-hoc Reliable Routing" IEEE Wireless Communications Letters, 2021, 21(1), pp. 87-91.