Naive Bayes Classifier Based Driving Habit Prediction Scheme for VANET Stable Clustering

Tong Liu(✉), Shuo Shi, and Xuemai Gu

Harbin Institute of Technology, Harbin 150001, China
liutongsasa@hotmail.com

Abstract. Vehicular ad hoc networks (VANETs) is a promising network form for future application on road, like arriving automatic driving and in-vehicle entertainment. Compare with traditional mobile ad hoc networks (MANETs), its advantages are multi-hop communication without energy restriction and relative regular moving pattern. However, the high mobility of nodes raises many challenges for algorithm designers such as topology changing, routing failures, and hidden terminal problem. Clustering is an effective control algorithm provides efficient and stable routes for data dissemination. Efficient clustering algorithms became challenging issues in this kind of distributed networks. In this paper, a novel machine learning based driving habit prediction scheme for stable clustering is proposed, briefly named NBP. In the scheme, vehicles are divided into two alignments with opposite driving habit from which stable cluster design could benefit. Naive Bayes classifier is introduced to estimate the alignment of vehicles by several factors, such as relative speed, vehicle type, number of traffic violations and commercial vehicle or not. Combined with clustering design, the proposed method has been proven effective for stable clustering in VANET.

Keywords: Naive Bayes classifier · VANET clustering · Driving habit

1 Introduction

As a research focus of intelligent transportation system (ITS), auto-driving attracts tremendous attention. It is an enormous and complicated project supported by stable communication system, quality sensors, deep data mining and so on. The most important part is stable communication which provide a foundation for safety application, assistant to the drivers and emergency warning. VANET is an architecture design for vehicles to exchange data with other neighbor vehicles. It is derived from MANET whose characteristic is multi-hop communications and no-infrastructures. Generally, communication in VANET is divided into two parts: vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I). On board units (OBUs) are installed on vehicles to make vehicles as a message sender, router and receiver. Road side units (RSUs) are assembled along the roads to provide Internet access and relaying. Compare with traditional MANET, nodes in VANET have relative regular mobility patterns on account of road restrictions, and energy consumption is no longer taken into

S. Han et al. (Eds.): AICON 2019, LNICST 286, pp. 445–452, 2019.
https://doi.org/10.1007/978-3-030-22968-9_40

consideration [1]. At the same time, the high speed makes network topology change rapidly than before.

Such characteristics raised new challenge in VANET, like hidden terminal in high dynamic topology and message congestion in high density network. Associating mobile vehicles into groups is a reasonable solution, that is clustering. According some rule set, it groups nodes to enhance communication efficiency. Cluster head (CH) is selected as a center of the cluster to host the whole cluster. In order to avoid congestion and meet quality of service (QoS) criteria, cluster also can be used for frequency reuse [2]. Recent years, widely study and discuss are focus on it. [4] provide a comprehensive and comparative survey of clustering technique. Vehicle communication benefit a lot from clustering, such as high communication efficiency, routing scalability and frequency resource sharing [5].

Dynamic topology of VANET make stable cluster a hard stuff. Therefore, the stability is one of the most important performance criteria. It means a stable cluster needs a long lifetime of CH and cluster members (CM) to decrease overhead. Fewer numbers of changes in vehicles states and fewer CH changes also can be take into consideration [6]. The recent designs of clustering algorithms are based on velocity of nodes, running lanes, nodes density, moving pattern and communication range. These methods are aim to ensure cluster stability by analyzing dynamic elements around vehicle. A well-design clustering algorithm need keep stable cluster and have integrated member joining and leaving procedures, and control overheads at the same time [7, 8]. Recently Seyhan Ucar proposed a new stable cluster based message dissemination method in [9]. VMaSC choosing relative mobility as the metric for CH selection. Relative mobility is calculated between neighboring vehicles. In order to reduce overheads, it introduces a direct connection from CM to CM. Periodic hello packets and CM information broadcasting maintain the cluster structure.

In this paper, the proposed novel machine learning based scheme aims to predict driving habit for stable clusters constructions. In this scheme, vehicles are divided into two alignments with opposite driving habit from which stable cluster design could benefit. Naive Bayes classifier is introduced to estimate the alignment of vehicles by several factors, such as relative speed, vehicle type, number of traffic violations and commercial vehicle or not.

The rest of the paper is arranged as follows: Sect. 2 introduces how the prediction scheme works, including vehicle alignments definition, Naive Bayes classifier and driving habit prediction system. Then, it describes CH selection method with the result from proposed scheme. In Sect. 3, simulation set-up scenarios, results and discussions are presented. At last, we conclude our algorithm and raise some future works in Sect. 4.

2 Driving Habit Prediction Scheme

2.1 Vehicle Alignments

Driving habit of drivers are decided by many personal factors, such as character, education, and life experience. The habit is developed from long time accumulation and will be relatively stable. It is believed that certain vehicles have fixed drivers on certain roads in most cases. We can infer that a vehicle keeps same driving habit in a certain route. Based on this conclusion, vehicles are divided into opposed alignments: Law Vehicles (LVs) and Chaos Vehicles (CVs). Originally the law and chaos axis were defined as the distinction between "the belief that everything should follow an order, and that obeying rules is the natural way of life", as opposed to "the belief that life is random, and that chance and luck rule the world". In this Scheme, the characteristics of alignments are only restricted in driving habit category. Vehicles in different alignments follow totally opposite driving patterns.

Most of the time, LVs keep moving at a constant speed which is around the average speed running on the road. When meeting vehicles ahead with similar speed on the same lane, LVs will slow down and stay uniform speed without changing lane. The rate of changing lanes is low for LVs. On the contrary, CVs always keep moving with higher speed than average, even beyond the speed limits on road sometimes. When meeting vehicles ahead, CVs will change lane immediately and overtake them with acceleration. Therefore, the rate of changing lanes is high for CVs.

2.2 Naive Bayes Classifier

In machine learning, naive Bayes classifiers are probabilistic classifier based on applying Bayes theorem with strong independence assumptions. Despite the simple design of naive Bayes classifier, naive Bayes classifiers have worked well in many area, such as medicine, economics and so on.

Bayes' theorem is stated as follows:

$$P(b|x) = \frac{P(b)P(x|b)}{P(x)} \tag{1}$$

where $P(b|x)$ is a conditional probability: the likelihood of event b occurring given that x is true, $P(b)$ and $P(x)$ is the prior probability.

When the "naive" conditional independence assumptions come into play, we assume that each feature x_i is conditionally independent of every other feature x_j for $i \neq j$, given the category b. The formula can be rewritten as follows:

$$P(b|x) = \frac{P(b)}{P(x)} \prod\nolimits_{i=1}^{d} P(x_i|b) \tag{2}$$

The naive Bayes classifier combines this model with a decision rule. One common rule is to pick the hypothesis that is most probable; this is known as the maximum a posteriori rule.

$$h_{nb}(x) = \arg\max\nolimits_{c \in y}\ P(b) \prod\nolimits_{i=1}^{d} P(x_i|b) \tag{3}$$

2.3 Driving Habit Prediction Framework

It is easy to judge the alignment of a vehicle on a specific road by monitoring its driving record. The number of lanes changing and the proportion of acceleration time can be used to draw a conclusion. In the process of vehicle clustering, we need to predict the vehicle motion pattern to establish a stable cluster. However, vehicles did not complete the driving on this section when choosing CHs. It is impossible to judge their camp according to the method just mentioned. It means that we need to judge which alignment the vehicle belongs to according to its current inherent properties before the vehicle completes the whole driving process on the road. As introduced before, Naive Bayes Classifier is suitable for this occasion. Relative independent features of vehicles are required to design the prediction method. Here four vehicle characteristics that may be relevant to driving habit have been selected: relative speed, vehicle size, number of violations and whether commercial vehicles.

- Relative speed: the speed difference between the vehicle and the traffic. Traffic speed is defined as the average speed of neighbor vehicles. This feature represents whether the speed of vehicles in this section is higher than the average level. The higher the relative speed is, the higher possibility of belonging to the chaotic alignment is. The formula of relative speed is stated as follows:

$$v_i^r = v_i - \frac{\sum_{j \in N} v_j}{n} \tag{4}$$

  where v_i^r means the relative speed of vehicle i, N is the set of neighbor vehicles, and v_j is the speed of neighbor vehicle j.
- Vehicle size: distinguished by the length of vehicles. In this scheme, vehicles are divided into two categories, small and large. A small car is less than six meters long, such as cars, jeeps, minivans, light buses, and light trucks. A large car is with a length of 6 m or more, such as ordinary buses, medium-sized and large trucks.
- Number of violations in the past three years: we believe that the more violations, the more barbaric the driving habits of drivers, that is, the more inclined to the chaotic camp.
- Whether commercial vehicles or not: commercial vehicle means a vehicle engaged in road transport business activities for the purpose of making a profit. It is believed that the commercial vehicles focus on driving safety, because profit risk need to be taken into account. That is, a higher probability of belonging to the law alignment.

It is assumed that the velocity of vehicles has a normal distribution [10]. According the formula 4, we can easily infer that the relative speed also follows a normal distribution.

$$p(x = v_{\mathrm{i}}^{\mathrm{r}} | c) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp(-\frac{(v_{\mathrm{i}}^{\mathrm{r}} - \mu_i)^2}{2\sigma_i^2}) \tag{5}$$

Where μ_i and σ_i^2 denote the average value and the variance of relative velocity for the vehicle i.

The other three characteristics are all discrete variable. The conditional probability of these features can be calculated as follows:

$$P(x_i|c) = \frac{|D_{c,x_i}|}{|D_c|} \tag{6}$$

where D_c denotes the training set of category c, and $|D_{c,x_i}|$ denotes the elements number of the set. D_{c,x_i} is the set of elements whose value is x_i in category c.

Combined with formulas 3, 5 and 6, the vehicles' alignment can be judged. Then the driving pattern of vehicles can be predicted when a clustering process initiate.

2.4 Stable Clustering

The stable clustering process consists of several steps: start, joining, CH selection, leaving procedure and merging procedure. The proposed scheme is working on the CH selection step. For the reason that the speed of CVs changes frequently, CVs are excluded from CH candidates in the algorithm. In traditional method, the metric is unique without an evolving view. The following moving pattern of vehicles make great difference to the cluster stability. For example, when the CH election start, a CV just happen to meet the requirement of CH. After becoming a CH, CV overtake the front car and keep acceleration as its moving pattern. The cluster will expire after CV driving out of the communication range. It can be avoided with the proposed Naive Bayes Classifier based driving prediction method.

The procedures of cluster formation are described below:

- Start: Every node is marked as an uncertain node. It does not receive any joining message (JM) or hello message (HM) after running on the street from beginning.
- Join: CHs disseminate JM periodic when they running on the road. If an uncertain node which is not belong any cluster receives the JM, and it is running on the same direction with this CH, this node will send a reply message (RM) to the relevant CH. After the CH receives the RM, it will allocate the node a unique cluster member number and send it to the new CM. If the uncertain node does not receive any reply from the CH after a certain period, it will broadcast a joining request (JR) for seeking a cluster.
- CH selection: If an uncertain node still does not receive any (JM) after several period, CH selection procedure will start. It will calculate relative speed between neighbor nodes, and select the center of the graph as the new CH. Then using the result of NBP, if the new CH is LV, continue to build the cluster. Or else, restart the CH selection procedure and exclude the CVs.
- CM Leaving: When a CM fails to receive any packets from CH over some periods, this node will change itself to an uncertain node role. It means this CM may be out of the communication range of its CH. At the same time, the CH will also remove this CM from the membership list.
- Cluster Merging: When a CH receive a JM from another CH, that means two CHs are within communication range. Cluster merging procedure need be executed.

Both of the CHs will give up the role of CH and come to CH selection procedure to build a new cluster.

3 Simulation Results

The simulations are running on the Network Simulator 2 platform. In order to simulate realistic mobility of vehicles, Simulation of Urban Mobility (SUMO) is employed combined with ns2. It is an open-source traffic simulator which can mode drivers' behavior. Driving habit of LVs and CVs are adjusted in the SUMO. For example, the acceleration and overtaking decision of the vehicles is determined by the driving habit which is obtained from designed prediction scheme. The road is designed two lanes and two-way. The simulation time is 500 s. The maximum speed of the vehicles is a variable ranging from 60 to 120 km/h.

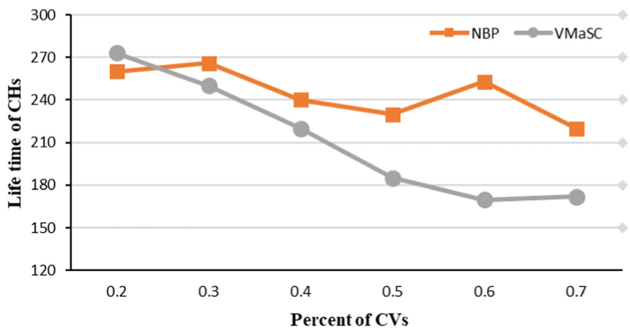

Fig. 1. Lifetime of CHs with the percent of CVs

A simplified VMsSC which only adopts its CH election strategy is compared with proposed scheme in this paper. Figure 1 states relationship between the lifetime of the CHs with the percent of CVs. With larger proportion of CVs, the CH lifetime decrease rapidly. That is because there is more CVs selected as CH, which is an unstable factor. The outcome of simulation also illustrates that NBP has longer CHs lifetime than VMaSC.

Figure 2 shows the lifetime of the cluster head between NBP and VMaSC with transmission range. The simulation result shows the lifetime of the clusters is prolonged by using NBP. That is because CVs is exclude from CH candidates. The unstable factor of node moving pattern is almost from CVs. On another hand, the result demonstrated wide transmission range will provide stability of clusters. Cluster stability benefits from longer transmission range.

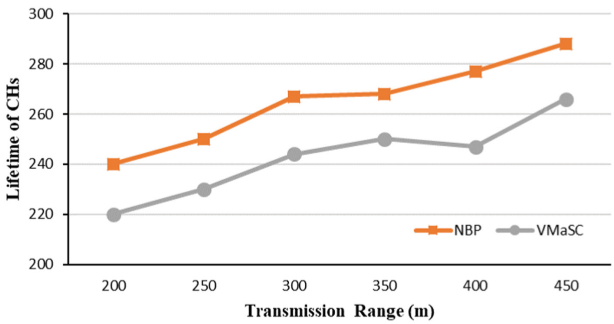

Fig. 2. Lifetime of CHs with its transmission range

4 Conclusions

In order to construct stable clusters, a novel machine learning based driving habit prediction scheme is proposed in this paper. In this scheme, vehicles are divided into two alignments with opposite driving habit from which stable cluster design could benefit. Naive Bayes classifier is introduced to estimate the alignment of vehicles by several factors, such as relative speed, vehicle type, number of traffic violations and commercial vehicle or not. The proposed technique has been proven effective for stable clustering in VANET.

References

1. Karagiannis, G., et al.: Vehicular networking: a survey and tutorial on requirements, architectures, challenges, standards and solutions. IEEE Commun. Surv. Tutor. **13**(4), 584–616 (2011)
2. Eze, E.C., Zhang, S., Liu, E.: Vehicular ad hoc networks (VANETs): current state, challenges, potentials and way forward. In: Proceedings of the 20th International Conference on Automation and Computing: Future Automation, Computing and Manufacturing, ICAC 2014, pp. 176–181 (2014)
3. Shea, C., Hassanabadi, B., Valaee, S.: Mobility-based clustering in VANETs using affinity propagation. In: GLOBECOM - IEEE Global Telecommunications Conference, pp. 1–6 (2009)
4. Cooper, C., Franklin, D., Ros, M., Safaei, F., Abolhasan, M.: A comparative survey of VANET clustering techniques. IEEE Commun. Surv. Tutor. **19**(1), 657–681 (2017)
5. Azizian, M., Cherkaoui, S., Hafid, A.S.: DCEV: a distributed cluster formation for VANET based on end-to-end realtive mobility. In: 2016 International Wireless Communications and Mobile Computing Conference, IWCMC 2016, pp. 287–291 (2016)
6. Ren, M., Khoukhi, L., Labiod, H., Zhang, J., Vèque, V.: A mobility-based scheme for dynamic clustering in vehicular ad-hoc networks (VANETs). Veh. Commun. **9**, 233–241 (2017)
7. Rawashdeh, Z.Y., Mahmud, S.: A novel algorithm to form stable clusters in vehicular ad hoc networks on highways. EURASIP J. Wirel. Commun. Netw. **2012**(1), 15 (2012)

8. Kannekanti, S.Y., Nunna, G.S.P.: An efficient clustering scheme in vehicular ad-hoc networks. In: 2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON), pp. 282–287 (2017)
9. Ucar, S., Ergen, S.C., Ozkasap, O.: Multihop-cluster-based IEEE 802.11p and LTE hybrid architecture for VANET safety message dissemination. IEEE Trans. Veh. Technol. **65**(4), 2621–2636 (2016)
10. Eiza, M.H., Ni, Q.: An evolving graph-based reliable routing scheme for VANETs. IEEE Trans. Veh. Technol. **62**(4), 1493–1504 (2013)