New Methods of Density Estimation for Vehicle Traffic

Fevzi Yasin Kababulut[1], Damla Kuntalp[1], Timur Duzenli[1]

[1]Dokuz Eylul University, Izmir, TURKEY
fyasin.kababulut@udhb.gov.tr, damla.kuntalp@deu.edu.tr, timur.duzenli@deu.edu.tr

Abstract

This study is prepared to produce solutions for estimating vehicle traffic density which is one of the biggest problems of urban life today. Four different algorithms are proposed for density estimation problem with different perspectives. All these proposed algorithms have been tried to estimate next state of the road by looking at history of density data. Algorithms are inspired by the methods used to estimate spectrum holes in cognitive radio channels. A similar approach is used for estimation of traffic density. In this study, Istanbul Metropolitan Municipality Traffic Control Center data received from the busiest roads of Istanbul in 2013, have been used as traffic data. Different simulations have been performed using these algorithms and results are evaluated based on several performance criteria.

1. Introduction

The majority of vehicle density estimation studies are based on traffic flow model of Lighthill and Whitham [1] and Richards [2] which actually started from the model of flow in the river. While different methods use different parameters obtained from loop detectors and probe vehicles such as traffic flow, speed, previous density, lane number, length of the section between upstream and downstream detectors, etc., the proposed density estimation methods use previous density information. A similar estimation problem, i.e. estimation of primary user existence in the channel, is an important problem in cognitive radio applications. In these cognitive radio applications, hidden Markov model, autocorrelation and linear regression methods are used as well as earlier studies use Kalman filter, particle filter, switching-mode model, second-order macroscopic model, filter-based heuristic method and Newtonian relaxation method, etc [3]-[6].

2. Proposed algorithms

Cognitive radio application can estimate existence of the primary user in a slot using previous slot history to avoid simultaneous usage of primer user (PU) and second user.

In these methods, the status of spectrum and location of spectral holes for future periods are guessed by analysing the previous spectral decisions or spectral information. According to this sceme, traffic parameters (state transition probability, arrival rate, etc.) are estimated first. Then, using these estimated parameters, a decision is made as to whether the PU will be in the channel or not for the next sensing period.

In the design stage of the proposed methods two new parameters are introduced:

History Window (Wh): This window contains the binary decisions (0 or 1) obtained in the previous sensing periods where 0 refers to channel is idle and 1 refers to channel is being used by primary user. The length of this windows varies as 5, 10, 20, and 30 as stated in [7].

Prediction Window (Wp): This window shows that for how many time slots the prediction is carried out.

According to first two methods proposed in this report, the traffic parameters are obtained from the history window by using the parameters defined below:

- **arrival_rate:** Number of 0-1 transitions in the history window/window length,
- **departure_rate:** Number of 1-0 transitions in the history window/window length,
- **last_active_slot:** Index of the last sample of history window where the decision is 1,
- **last_idle_slot:** Index of the last sample of history window where the decision is 0.

The reciprocals of the arrival rate and departure rate represent the mean busy and idle durations, respectively [8]. Using this knowledge, it can be said that the channel will incline to be busy if last_active_slot is greater than last_idle_slot and the length of the samples between last_active_slot and last_idle_slot is smaller than mean busy duration of the PU. On the other hand, if last_idle_slot is greater than last_active_slot and the length of the samples between last_idle_slot and last_active_slot is smaller than mean idle duration of the PU, than the decision for the next period will incline to be idle again. This scheme is summarized in Algorithm 1.

Algorithm 1

```
if last_active_slot > last_idle_slot then
    if last_active_slot − last_idle_slot < arrival_rate^−1
    then
        decision = 1
    else
        decision = 0
    end if
else
    if last_idle_slot − last_active_slot > depart._rate^−1
    then
        decision = 1
    else
        decision = 0
    end if
end if
```

Collected data from Istanbul Metropolitan Municipality at 10 roads are used. First, it must be converted to binary or ternary form to be used by algorithms. Raw data consist of average vehicle speed at one road at certain time interval. Converted binary data is binary state of average vehicle speed (change between 0-121 km/h) labelled with 0 for above 50 km/h speed and

1 for below 50 km/h. We have defined vector cluster which is named “temp” and used for taking samples at history window length from data. True positive parameter is the value which shows how much “zero” we have estimated truely. True negative is the value which shows how much “one” we have estimated truely. Converted ternary data is ternary state of average vehicle speed (change between 0-121 km/h) labelled with 0 if data is above 50 km/h (free flow). It is labelled with 1 if data is between 50 km/h-30 km/h (less congested). If data is below 30 km (high congested), it is labelled with 2.

When Algorithm 1 is used with converted binary data, we observed that true negative rate is lower than true positive ratio. So we have revised our algorithm. In our revised algorithm, if there are at least three same data between last five data, our decision is that binary value. This scheme is summarized in Algorithm 2 below.

Algorithm 2

if $last\_active\_slot > last\_idle\_slot$ **then**
  if $last\_active\_slot - last\_idle\_slot < arrival\_rate^{-1}$ **then**
    $decision = 1$
  else if $(sum(temp(6:10))) > 2$ **then**
    $decision = 1$
  else
    $decision = 0$
  end if
else
  if $last\_idle\_slot - last\_active\_slot > depart.\_rate^{-1}$ **then**
    $decision = 1$
  else if $(sum(temp(6:10))) > 2$ **then**
    $decision = 1$
  else
    $decision = 0$
  end if
end if

The other two methods suggested from this study, use some probability parameters which are obtained by the analysis of decisions in history window. First, state transition probabilities are calculated, after that decision slot is estimated by comparing these probabilities [9].

In algorithm 3, we have calculated probability of transitions (P_{00}, P_{01}, P_{10}, P_{11}). Also when we have divided number of 1 and 0 (In ternary estimation additionally 2) to history window length, we have obtained probability of being one (P_1) and zero (P_0) in history window.

We have calculated unconditional probability of 0 ($P(0)$) and 1 ($P(1)$) (in ternary estimation additionally 2) by

$$P(0) = (P_{00}P_0) + (P_{10}P_1)$$
$$P(1) = (P_{01}P_0) + (P_{11}P_1)$$

Depending on the highest unconditional probability value in binary or ternary case, estimation slot is assigned the corresponding value. If unconditional probability values are equal, estimation slot is assigned as the majority state. Nevertheless if number of transition state in history window are equal, the decision slot will be the last data of history window. This scheme is summarized in Algorithm 3 below.

In Algorithm 4, we have evaluated probability values individually which are used for calculating unconditional probability in Algorithm 3. In which transition probability of binary or ternary state is high, we have estimated that last data is passing to this value. If the last data is zero, we have checked 0-0 transition probability ($\hat{P}_{00}$) and 0-1 transition probability ($\hat{P}_{01}$) (in ternary estimation additionally 0-2 transition probability). If the last data is one, we have checked 1-0 transition probability ($\hat{P}_{10}$) and 1-1 transition probability ($\hat{P}_{11}$) (in ternary estimation additionally 1-2 transition probability). In ternary estimation if the last data is two, we have checked 2-0 transition probability, 2-1 transition probability, and 2-2 transition probability. If transition probabilities are equal, estimation slot is assigned as the majority state. Nevertheless if number of transition state in history window are equal, the decision slot will be the last data of history window. This scheme is summarized in Algorithm 4 below.

Algorithm 3

$P(0) = (P_{00}P_0) + (P_{10}P_1)$
$P(1) = (P_{01}P_0) + (P_{11}P_1)$
if $P(0) > P(1)$ **then**
  $decision = 0$
else
  $decision = 1$
end if

Algorithm 4

if $temp(Wh) == 0$ **then**
  if $\hat{P}_{00} > \hat{P}_{01}$ **then**
    $decision = 0$
  else
    $decision = 1$
  end if
else
  if $\hat{P}_{10} > \hat{P}_{11}$ **then**
    $decision = 0$
  else
    $decision = 1$
  end if
end if

3. Simulation results

Sample accuracy (acc) is the most frequently used performance criterion of a multiclass classifier which is defined as the number of correct predictions across all classes, k, divided by the number of examples, n [10].

For overcoming the limitation of sample accuracies, another performance evaluation parameter the Kappa coefficient is used to measure the degree of overall agreement within a given matrix $\mathcal{C} \in N^{lxl}$ [11]

$$K_c = \frac{p_0 - p_e}{1 - p_e} \quad (1)$$

$$p_0 = \frac{1}{l}\sum_{i=1}^{l} k_i \quad (2)$$

$$p_e = \frac{1}{l^2}\sum_{i=1}^{l} \mathcal{C}_{i+} \times \mathcal{C}_{+i} \quad (3)$$

where k_i is the number of correct predictions in class i and l is the number of classes. $\mathcal{C}_{i+}$ and $\mathcal{C}_{+i}$, are the row-wise and column-wise sums of row and column i in the confusion matrix, respectively [10]. The balanced accuracy (bac) is alternative parameter which is defined as the average accuracy obtained on all classes. In the case of a multi class classification problem, its formula is given by

$$\hat{\lambda} = \frac{1}{l} \sum_{i=1}^{l} \frac{k_i}{n_i} \tag{4}$$

where n_i is the number of examples in class i. The balanced accuracy is commonly used in these problems and has several advantages to other criteria because of its simplicity [10].

We have evaluated our algorithms by these criteria which can be used on performance analysis. The algorithms have been tested by simulating two different scenarios. In the first scenario, the algorithms have estimated binary traffic state in the 60^{th} minute of road 529 named Caglayan SSK by looking at past 50 minutes of density data of the same road. In the second scenario, the algorithms have estimated ternary traffic state in the 60^{th} minute of the third adjacent road (529 Caglayan SSK road) by looking at past 50 minutes of density data of the preceding roads (528 Okmeydani Koprusu road and 297 Perpa road).

Table 1. Binary Estimation Results of Algorithm 1 (529 Caglayan SSK Road)

	0	1
0	**10222**	135
1	135	**1593**
Predicted	**10357**	**1728**

Table 2. Binary Estimation Results of Algorithm 4 (529 Caglayan SSK Road)

	0	1
0	**10140**	245
1	217	**1483**
Predicted	**10357**	**1728**

Table 3. Ternary Estimation Results of Algorithm 2 (528-297-529 Adjacent Roads)

	0	1	2
0	**1654**	23	35
1	7	**4**	12
2	8	12	**183**
Predicted	**1669**	**39**	**230**

When we have considered results separately binary and ternary, we have expected that binary estimation's performance is higher. Because we have made a more general acceptance with one threshold in binary estimation. When we considered algorithms, we have observed that Algorithm 1 is the best at binary estimation of 529 named Caglayan SSK road and Algorithm 2 is the best at ternary estimation of adjacent roads according to all performance criteria.

4. Conclusion

In this study, four different algorithms are proposed in order to estimate traffic density. First two algorithms have taken into account four parameters; arrival rate, departure rate, last active slot and last idle slot, last two algorithms have taken into account density transition probabilities. These parameters have been used for estimating future density state by considering history density data. For measuring success, sample accuracy, balanced accuracy and kappa coefficient have been utilized as performance criteria. While at binary estimation of 529 named Caglayan SSK road, performance values of Algorithm 1 are the highest one, at ternary estimation, performance values of Algorithm 2 are the highest one. We have observed that ternary estimation of proposed algorithms have low success rate at balanced accuracy and kappa coefficient criteria because of effects of unexpected traffic events. In future studies, by making unexpected traffic events as additional system parameter, it is planned to increase the success rate.

Table 4. Ternary Estimation Results of Algorithm 3 (528-297-529 Adjacent Roads)

	0	1	2
0	**1657**	30	81
1	0	**1**	2
2	12	8	**147**
Predicted	**1669**	**39**	**230**

Table 5. Perfermonce Evaluation of Best Algorithms

Rtms ID	**Est.**	**acc**	**bac**	**K$_c$**	**Algo.**
529	Binary	**0.9777**	**0.9544**	**0.9128**	Algo.1
529	Binary	**0.9618**	**0.9186**	**0.8430**	Algo.4
528-297-529	Ternary	**0.9499**	**0.6297**	**0.7791**	Algo.2
528-297-529	Ternary	**0.9314**	**0.5525**	**0.6637**	Algo.3

5. References

[1] M. J. Lighthill and G. B. Whitham, "On kinematic waves II: A theory of traffic flow on long crowded roads", *Proceedings of The Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 229, no.1178, pp: 317-345, 1955.

[2] P. I. Richards, "Shockwaves on the highway", *Operations Research,*, vol. 4, no. 1, pp: 42-51, 1956.

[3] L. Yang, "Stochastic Traffic Flow Modeling and Optimal Congestion Pricing", Ph. D. Thesis, *Industrial and Operations Engineering Dept.*, University of Michigan, Michigan, USA, 2012.

[4] L. Munoz, X. Sun, R. Horowitz and L. Alvarez, "Traffic Density Estimation with the Cell Transmission Model", in *Proceedings of the American Control Conference*, Denver, Colarado, 2003, pp. 3750-3755.

[5] Z. T. Qiu, "Macroscopic Traffic State Estimation For Large Scale Freeway Network Using Wireless Network Data", Ph. D. Thesis, *Civil and Environmental Engineering Dept.*, University of Wisconsin-Madison, Madison, USA, 2007.

[6] J. C. Herrera, "Assessment of GPS-enabled smart phone data and its use in traffic state estimation for highways", Ph. D. Thesis, *Civil and Environmental Engineering Dept.*, University of California Berkeley, California, USA, 2009.

[7] G. S. Uyanik, B. Canberk and S. Oktug, "Predictive spectrum decision mechanisms in cognitive radio networks", in

IEEE Globecom Workshops (GC Wkshps), Anaheim, CA, 2012, pp. 943-947.

[8] S. Kay, “Intuitive probability and random processes using MATLAB”, *Springer,* New York, USA, 2006.

[9] T. Duzenli and O. Akay, “Channel State Prediction for Cognitive Radios with Stochastically Varying Primary User Traffic Density”, in *IEEE 23. Signal Processing and Communication Applications Conference (SIU)*, Malatya, 2015, pp. 1232 - 1235.

[10] H. Carrillo, K. H. Brodersen and J. A. Castellanos, “Probabilistic performance evaluation for multiclass classification using the posterior balanced accuracy”, *ROBOT2013: First Iberian Robotics Conference*, vol. 252, 2013, pp: 347-361.

[11] J. Cohen, “A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement”, *Measurement*, vol. 20, no.1, pp: 37-46, 1960.