Estimation of urban arterial travel time based on dynamic bayesian network

1st Zhonghe He
Beijing Key Lab of Urban Intelligent Traffic Control Technology, North China University of Technology,
Beijing 100144, China
zhonghehe@ncut.edu.cn

2nd Renchi Shao
Beijing Key Lab of Urban Intelligent Traffic Control Technology, North China University of Technology,
Beijing 100144, China
1192769304@qq.com

3rd Jiaming Wang
Beijing Key Lab of Urban Intelligent Traffic Control Technology, North China University of Technology,
Beijing 100144, China
506305706@qq.com

2021 China Automation Congress (CAC) | 978-1-6654-2647-3/21/$31.00 ©2021 IEEE | DOI: 10.1109/CAC53003.2021.9727412

***Abstract*—Estimation and prediction of urban arterial travel time is important for the development of transportation field, especially measures and methods of traffic management and control. This article presents a modeling method for estimating and predicting arterial travel time based on Dynamic Bayesian Network. The probability distribution model of travel time conditioned on different traffic states is constructed by combining the traffic flow theory at intersections and Dynamic Bayesian Network. The Cell Transmission Model that was applied to accurately describe the evolution process of state variables and the travel time observations were obtained by trajectory data, then the relationship between the traffic state and travel time was established by Dynamic Bayesian Network. The EM algorithm and particle filter algorithm are used to iterate the static parameters of the road (such as traffic signal parameters) through machine learning, so as to estimate and predict the distribution of arterial travel time in real time under the current traffic state. The data of Changping District of Beijing were selected to verify the validity of the probability distribution estimation of travel time in this paper, then the direction and value of subsequent improvement and application were proposed.**

Keywords—Dynamic Bayesian Network, arterial travel time, machine learning

I. INTRODUCTION

As the backbone roads of urban road network, urban traffic arterial carries the main long-distance traffic of the city, and its operation state directly affects the traffic operation efficiency of the whole urban road network. Therefore, the state estimation of urban traffic arterial is of great significance to alleviate traffic pressure and improve traffic condition. Because urban traffic is complex and changes with time and space, it is important to research a modeling method to accurately estimate the traffic state of urban arterial, and provide the basis for traffic state discrimination, abnormal event detection, traffic control and so on.

Estimation and prediction of travel time as one of the core contents of traffic state estimation is an important basis for providing traffic guidance services and releasing traffic information, which can directly reflect the traffic state. Travel time distribution can show the dynamic characteristics of urban roads, and evaluate the effect of traffic measures on travel time of viechles, which is of great practical significance to urban management and control.

At present, a series of estimation and prediction models and methods that based on average travel time in traffic area have been put forward, such as model-based method and data-driven method [1][2][3][4][5][6]. However, these models have good application in expressway, which is different from urban roads due to intersection, pedestrian, vehicle types and other factors [7].

The research on the estimation and prediction of travel time distribution of urban roads depends on two kinds of methods: traffic flow model and statistical probability applied in the estimation. For example, in paper [8] [9], the average travel time of signal road is estimated by shock wave theory. In paper[10], CTM model is applied to urban road network with signalized intersections. In paper[11], evaluation index of estimating traffic state is proposed based on LWR model. In paper [12], the delay mode of intersection is approximately expressed using the travel time observed. Furthermore, the paper [13] built the model of the arterial travel time distribution based on the LWR model,. The paper [14] also analyzes the distribution of the travel time based on the relationship of the adjacent intersections under different states based on LWR model. The above method uses the traffic flow modeling method. However, due to the influence of intersections, traffic control, pedestrians, urban environment and other uncertain factors, it is difficult to establish an accurate and real traffic flow model for urban road network. The second type: in paper[15], the arterial travel time distribution is obtained by transition probability between the partial travel time distribution, obtained by Markov chain method. In paper [16], the vehicle evolution on each link is studied by probability method to estimate the travel time and difference. In paper [17], the Bayesian network model is constructed in order to estimate the probability of different sampling vehicles index in the process of random arrival and departure. The paper [18] proposed an arterial travel time distribution (TTD) method and model integrating characteristics of travel time, which provides a basis for quantifying the TTR of travel time reliability. Above all, these estimation methods are based on probability, ignoring the physical characteristics of traffic flow generally.

Vehicle trajectory data become more common with vehicle communication technology, which makes us combine traffic flow model with statistical probability to estimate traffic parameters and state variables together. It is very important for traffic signal control optimization based on integrated model. A famous example is Hofleitner et al. who successfully proposed a method of combining traffic model with Dynamic Bayesian Network (DBN) to estimate and predict arterial travel time distribution[19]. Besides, a framework combining traffic flow model with Bayesian network model is established to jointly estimate traffic parameters and travel time distribution by the sampling vehicle trajectory data in paper [20].

The method proposed in this paper follows the similar method of Hofleitner et al. [19], embedding the traffic flow model in the statistical probability to learn the key traffic parameters and probability distribution of travel time. The

This work is supported by Scientific Research Project of Beijing Educational Committee (No. KM201910009008).

remainder of the article is organized as follows. We relates the states and observations by Dynamic Bayesian Network and the evolution process of state variables is optimized by Variable Cell Transmission Model (VCTM) compared with paper[19]. The probability distribution of travel time conditioned different states is proposed. Then, the EM algorithm and particle filter in paper [19] are applied to iteratively optimize the link parameters which express the probability distribution of travel time. Then last sections verify the validity of the nethod in this paper and put forward the direction of improvement.

II. DYNAMIC BAYESIAN NETWORK MODEL

The arterial traffic condition changes dynamically with space and time. We considers Dynamic Bayesian Network similar to that in paper[19] that relates the variables (reflecting traffic states) and the observations, as shown in Fig. 1.

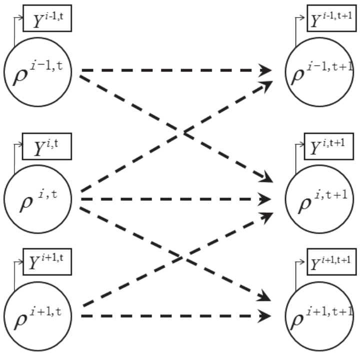

Fig. 1. Dynamic Bayesian Network

The random variable shown in this figure is: *1)* the circular node $\rho^{i,t}$ is the state variable, which represents outflow density of link *i* in each time interval *t*. The states of a link and adjacent links previously effect the current state of this link; *2)* the rectangular node is the observed variable $\boldsymbol{Y}^{i,t}$ representing the set of observations $y^{i,t}$ of current link *i*. The probability distribution of travel time is conditional on hidden variables in a link.

In article [19], state variable is the number of stopping vehicles in link *i* in a cycle at time interval *t*. In this paper, considering the following characteristics of the vehicles in each part of link and traffic states of each cycle in sampling period, Variable Cell Transmission Model (VCTM) is applied to the evolution process of state variables to describe the characteristics of traffic flow and the evolution process of more accurately.

III. TRAFFIC FLOW MODEL AND EVOLUTION

A. *Evolution of state variables*

The Cell Transmission Model (CTM) can describe the dynamics characteristics in traffic such as queuing formation and dissipation. However, CTM requires that the length of cell to be consistent, the lengths of the links are different, so the description accuracy of complex urban road network with dense intersections is not ideal [21]. Considering the characteristics and operation process of urban road traffic flow because of the signal configuration, Variable Cell Transmission Model (VCTM) suitable for intersection nodes in paper [21] is proposed applied in this paper. Fig. 2 shows the Variable Cell Transmission Model of multi node urban road network.

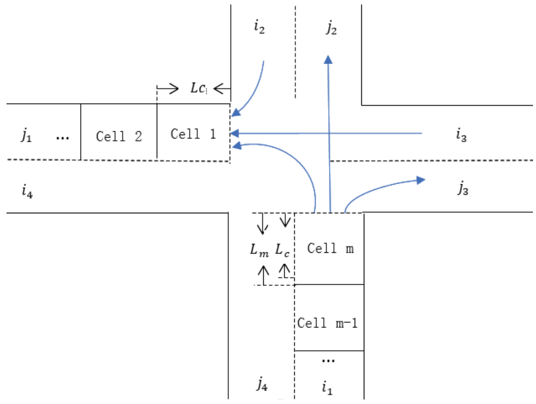

Fig. 2. Cell Transmission Model of multi node urban road network

During the green light duration of a cycle, the density of outflow and the critical density in link *i* are as follows:

$$\rho^i(t) = \frac{q^i(t)}{d_T \cdot v_f} \tag{1}$$

$$\rho_c^i(t) = \frac{Q^i(t)}{d_T \cdot v_f} \tag{2}$$

Where: $\rho^i(t)$ represents the outflow density of link *i* at time *t*, pcu/m; $q^i(t)$ is the outflow of link *i* during the green light duration at time *t* , pcu; $\rho_c^i(t)$ represents the the critical density of link *i* at time *t*; $Q^i(t)$ represents maximum outflow in link *i* at time *t*, pcu; v_f is free flow velocity, m/s; d_T is the time step, s;

The send capacity and the receptivity of cell directly affect the outflow of cell *m* at time *t*. VCTM vehicle transfer formula is changed as follows:

$$L_c = v_f \cdot d_T \tag{3}$$

$$q_m(t) = \min\{S_m(t), R_{m+1}(t)\} \tag{4}$$

Where: L_c is the effective length of basic cell, m; *m* is the cell, *m*-1 and *m*+1 are the upstream and downstream adjacent cells of cell *m*, respectively; $q_m(t)$ represents the number of vehicles leaving the cell *m* at time *t*, pcu; S_m represents the send capacity of upstream cell *m* at time *t*, pcu; $R_{m+1}(t)$ is the receptivity of cell *m+1* at time *t*, pcu.

VCTM defines that the send capacity $S_m(t)$ of upstream cell *m* at time *t* is related to the number of vehicles and the maximum outflow of upstream cell itself, and the receiving capacity $R_{m+1}(t)$ at time *t* is related to the carrying capacity of local cell *m*.

$$S_m(t) = \min\left\{\frac{L_c}{L_{m-1}} n_m(t), Q_m(t)\right\} \tag{5}$$

$$R_{m+1}(t) = \frac{L_c}{L_{m+1}} \cdot \frac{\omega}{v_f} \cdot \left[N_{m+1}(t) - n_{m+1}(t)\right] \tag{6}$$

Where: $Q_m(t)$ is the maximum outflow of upstream cell *m* into local cell *m* at time *t*, pcu; $N_m(t)$ represents maximum number of vehicles on cell *m* at time *t*, pcu; L_m represents actual length of cell *m*, m.

The flow conservation formula is as follows:

$$n_m(t+1) = n_m(t) + q_{m-1}(t) - q_m(t) \tag{7}$$

Where: $n_m(t)$ represents number of vehicles on cell *m* at time *t*, pcu;

It represents different road traffic states in a light cycle by comparing $\rho^i(t)$ to $\rho^i_c(t)$ during green light duration, namely undersaturated state and congested state. If all queuing viecles leave from link i (the density of outflow is less than the critical density) during a cycle, the current state is judged to be undersaturated state, otherwise it is congested state. For each sampling period $t \in (1,2,3...T)$, if cech cycle is undersaturated state, the state variable of this sampling period $\rho^{i,t}$ is undersaturaed, otherwise it is oversaturated.

B. Probability distribution of travel time

The traffic flow model in paper[22] is used to derive the probability distribution of travel time under two traffic conditions: undersaturated state and congested state, considering the sending data location probability of the sampling vehicles and the driver's behavior parameters.

Conditioned the condition of traffic state $\rho^{i,t}$, the probability density distribution of travel time $y_{L,0}$ between stop line for adjacent upstream link and link i is expressed as $g^i\left(y_{L,0} \mid \rho^{i,t}\right)$, where we assume that the free flow pace satisfies gamma distribution ,whose parameter is $\theta_p = (\alpha, \beta)$.

In this paper, we assumes that the most likely path is known and the probability distribution of travel time is parametered as a group of independent parameters. Static traffic model parameters (history learning) are signal cycle, red light duration and free flow pace distribution parameters, which expressed as $\left(R^i, C^i, \theta^i_p\right)$; traffic state (dynamic estimation) is outflow density $\rho^{i,t}$.

In order to specify the DBN model exactly, it is necessary to optimise the following variables, which is detailed in article [19]:

1) $P(\rho^{i,0})$ reflects probability of outflow density in link i at the initial time;

2) $v^{i,j}$ represents probability of assignment from link i to link j and λ^j represents intensity of the Poisson process;

3) the distribution of travel time g^i of link i is conditioned on the state $\rho^{i,t}$ and parameterized by the parameters $\left(R^i, C^i, \theta^i_p\right)$.

IV. MAXIMUM LIKELIHOOD ESTIMATION

In this part, we take the parameters $\left(R^i, C^i, \theta^i_p\right)$ to represent the probability distribution of travel time. Referring to the algorithm in article [19], the joint probability of the outflow density and travel time observations in each time interval is as follows:

$$P(\rho, \mathbf{y}) = \left(\prod_{t=0}^{T}\prod_{i\in I} P(\rho^{i,t})\right) \times \left(\prod_{t=0}^{T}\prod_{i\in I} P(\mathbf{y}^{i,t} \mid \rho^{i,t})\right) \tag{8}$$

Where: $\prod_{t=0}^{T}\prod_{i\in I} P(\mathbf{y}^{i,t} \mid \rho^{i,t})$ represents the probability of the observations $y^{i,t}$ conditioned on the state $\rho^{i,t}$ for the link i and time interval; $\prod_{t=0}^{T}\prod_{i\in I} P(\rho^{i,t})$ represents the probability of state variables for the link i and time interval.

A. EM algorithm

EM algorithm is applied to solve the problem when there are hidden variables that cannot be observed in the dynamic model.

1) Step E: particle filter

Given the travel time observations obtained from original data and the initial values of the traffic parameters and particle filter is used for processing. Algorithm in article [19] is applied in this paper and V particles are sampled, which represents instances of states, that is, the possible states of the link and time interval of the network.

In the time interval t, given the observed travel time $\mathrm{y}^{i,t}$ and parameters $\left(R^i, C^i, \theta^i_p\right)$, the probability of state E^t of the particles evolving with time is as follows:

$$p\left(E^t \mid \mathrm{y}^{i,t'}, R^i, C^i, \theta^i_p : t' \in \{0...t\}\right) \approx \sum_{v=1}^{V} \omega^t_v 1_{E^t}\left(E^t_v\right) \tag{9}$$

Where: E^t_v is the instance of the particles; ω^t_v is their importance weight; If the particle has state instance E^t, then $1_{E^t}\left(E^t_v\right)$ is 1, otherwise it is 0.

According to Jansen equation, the conversion (8) is as follows, which is detailed in article [19]:

$$\begin{aligned}\langle l(\rho, \mathrm{y})\rangle = &\left(\sum_{t=0}^{T}\sum_{j\in I}\sum_{\rho^{i,t}} a^{i,t}(\rho^{i,t}) \ln(P(\rho^{i,t}))\right) \\ &+\left(\sum_{t=0}^{T}\sum_{i\in I}\sum_{0}^{\rho^i_{\max}} a^{i,t}(\rho^{i,t})\left(\sum_{y_{L,0}\in y^{i,t}} \ln(\mathrm{g}^i(\mathrm{y}_{L,0} \mid \rho^{i,t}))\right)\right)\end{aligned} \tag{10}$$

Where: $a^{i,t}(\rho^{i,t})$ represents the probability of the state by (9).

2) Step M: maximize (10) to solve the parameters

$$\underset{C^i, R^i, \theta^i_p}{\text{maximize}} \sum_{t=0}^{T}\sum_{0}^{\rho^i_{\max}} a^{i,t}\left(\rho^{i,t}\right)\left(\sum_{y_{L,0}\in y^{i,t}} \ln\left(g^i\left(y_{L,0} \mid \rho^{i,t}\right)\right)\right) \tag{11}$$

The parameters in the above formula are obtained by genetic algorithm, and the travel time probability density distribution formula in [22] is substituted for real-time estimation and prediction.

B. Steps of algorithm

Step2-step6 is EM algorithm; *step2-step4* is *step E* and update state evolution through particle filtering; *step5* is *step M* and the expected complete logarithmic likelihood is maximized, and then travel time distribution can be estimated in real time based on the link parameters $\left(R^i, C^i, \theta^i_p\right)$, $v^{i,j}$ and λ^j by genetic algorithm.

step1. Initialize the parameters $\left(R^i, C^i, \theta^i_p\right)$, $v^{i,j}$, λ^j and the state probabilities at the initial moment $P\left(\rho^{i,0}\right)$.

step2. Simulate particles representing the initial states and having initial weight $\omega_v = 1/V$.

step3. Update and normalize the weights of the samples that are conditioned on the observations $y^{i,t}$.

step4. Update particle states according to traffic flow model.

step5. Compute $P\left(\rho^{i,0}\right)$, $v^{i,j}$, λ^j and $\left(R^i, C^i, \theta^i_p\right)$.

step6. Cycles *step2* to *step5*, iterating parameters to the optimum, estimating and predicting travel time distribution in real time.

V. VERIFICATION OF TRAVEL TIME DISTRIBUTION

In this section, in order to verify the validity of the travel time distribution of link in this paper, we selected Nanhuan Road in Changping District of Beijing as the experimental site, and adopted the trajectory data in the morning (10: 00am-11:00am) for one consecutive week. The trajectory data record the position of vehicles and other information every three seconds. The travel time observed is used to obtain the traffic parameters of the link through the maximum likelihood estimation, and then the travel time probability distribution is obtained. The experiments verify the validity of travel time distribution in this paper.

We compared travel time probability distribution of this paper(estimated pdf) with that in paper[19] (comparative model pdf), and valid the estimated travel time pdf with kernel smoothing density (ksdensity pdf). Fig.3 is a example of the road network that shows the estimated pdf, ksdensity pdf and comparative model pdf of the travel times on link of the network. In the diagram, the histogram is the actual travel time data.

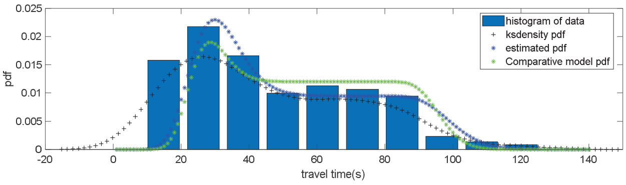

Fig. 3. Estimated pdf of travel times of a link

Histogram of travel times was used to verify the accuracy of the estimated travel time distribution in this paper, and three distributions accept K-S test with H0=0 conditioned significance level 0.05. However, according to the travel time distribution diagram, the travel times of more vehicles are between 10 and 45 s and the travel times of less vehicles are more than 45 seconds. The mean absolute percentage error(MAPE) is applied to evalute the estimation accuracy of average travel time. Table I.is the MAPE between estimated travel time distribution and data. The average travel time MAPE of estimated model in this paper reduce 4% than that of comparative model.

TABLE I. AVERAGE TRAVEL TIME MAPE

	Estimation	*Kernel smoothing density*	*Comparative model*
MAPE(%)	11.46	17.25	15.67

We use the assumption of uniform arrival in article [22] to make the delay uniformly distributed in this paper, which is unreasonable in practice, and this is the direction for future improvement.

VI. CONCLUSION

In this paper, a hybrid modeling framework based on dynamic Bayesian network is proposed to estimate and predict arterial traffic state. The intersection traffic flow model and dynamic Bayesian Network Model construct the probability distribution model of arterial travel time jointly, and obtain the road static parameters (such as traffic signal parameters) through machine learning, in order to estimate and predict the distribution of arterial travel time. Compared with the average travel time distribution, the method estimates the probability distribution of travel time of links. Compared with the model in article [19], this paper takes the improved VCTM in article [21] as the evolution process of state variables, and takes into account the state of each cycle, which describes the evolution process of traffic state more accurately.

However, this paper assumes that each lane of a link is in the same condition and arrivals on each link are consistent at a given time. This assumption does not hold true in controlled mains where signal synchronization is important. More accurate traffic flow model and signal control factors will be considered to construct more accurate travel time distribution of arterial.

REFERENCES

[1] Chen M , Chien S . Dynamic Freeway Travel-Time Prediction with Probe Vehicle Data: Link Based Versus Path Based[J]. Transportation Research Record Journal of the Transportation Research Board, 2001, 1768:157- 161.

[2] Chien I J , Kuchipudi C M . Dynamic Travel Time Prediction with Real-Time and Historic Data[J]. Journal of Transportation Engineering, 2003, 129(6):608-616.

[3] Kwon J , Petty K . Travel Time Prediction Algorithm Scalable to Freeway Networks with Many Nodes with Arbitrary Travel Routes[J]. Transportation Research Record Journal of the Transportation Research Board, 2005, 30(1935):p.147-153.

[4] Innamaa S.Short-Term Prediction of Travel Time using Neural Networks on an Interurban Highway[J]. Transportation, 2005, 32(6):649-669.

[5] Hinsbergen C P I V , Lint J , Zuylen H . Bayesian Training and Committees of State-Space Neural Networks for Online Travel Time Prediction[J]. Transportation Research Record: Journal of the Transportation Research Board, 2009, 2105(1):118-126.

[6] J.W.C. van Lint and S.P. Hoogendoorn and H.J. van Zuylen. Accurate freeway travel time prediction with state-space neural networks under missing data[J]. Transportation Research Part C: Emerging Technologies, 2005.13(5–6):347–369.

[7] Zheng F, Zuylen H V ,LiuX . A methodological framework of travel time distribution estimation for urban signalized arterial roads[J].TransportationScience,2017,51(3):893-917.

[8] Kwon J , Petty K . Travel Time Prediction Algorithm Scalable to Freeway Networks with Many Nodes with Arbitrary Travel Routes[J]. Transportation Research Record Journal of the Transportation Research Board, 2005, 30(1935):p.147-153.

[9] Skabardonis A , Ge Roliminis N . Real-Time Monitoring and Control on Signalized Arterials[J]. Intelligent T ransportation Systems:Tech., Planning, Oper. 12(2):64–74.

[10] A Cell-Based Traffic Control Formulation: Strategies and Benefits of Dynamic Timing Plans[J]. Transportation Science, 2001, 35(2):148-164.

[11] Lighthill M J , Whitham G B . On Kinematic Waves. II. A Theory of Traffic Flow on Long Crowded Roads[J]. Proceedings of The Royal Society A Mathematical Physical and Engineering Sciences, 1955, 229(1178):317-345.

[12] Zheng F , Zuylen H V , Liu X . A methodological framework of travel time distribution estimation for urban signalized arterial roads[J]. Transportation Science, 2017, 51(3):893-917.

[13] Applying variational theory to travel time estimation on urban arterials[J]. Transportation Research Part B, 2015, 78:169-181.

[14] Zheng F , Zuylen H V , Liu X . A methodological framework of travel time distribution estimation for urban signalized arterial roads[J]. Transportation Science, 2017, 51(3):893-917.

[15] Ramezani M , Geroliminis N.On the estimation of arterial route travel time distribution with Markov chains[J]. Transportation Research Part B: Methodological, 2012, 46(10):1576-1590.

[16] Lu Y, Chang GL. A stochastic model for estimating the time-varying arterial travel time and its variability using only linkdetector data. Proc. TRB 91st Annual Meeting, Washington, DC.

[17] Hao, P., Sun, Z., Ban, X.J., Guo, D., Ji, Q., 2013. Vehicle index estimation for signalized intersections using sample travel times[J]. Proc.-Soc. Behav. Sci. 2012,80, 473–490.

[18] Chen M , Yu G , Chen P , et al. A copula-based approach for estimating the travel time reliability of urban arterial[J]. Transportation Research Part C Emerging Technologies, 2017, 82(sep.):1-23.

[19] Hofleitner A ,Herring R , Bayen A . Arterial travel time forecast with streaming data: A hybrid approach of flow modeling and machine learning[J]. Transportation Research Part B, 2012, 46(9):1097-1122.

[20] Wang S, Huang W, Lo H K. Traffic parameters estimation for signalized intersections based on combined shockwave analysis and Bayesian Network[J]. Transportation research part C: emerging technologies, 2019, 104: 22-37.

[21] Chang Y L,Yuan C H,Sun C,et al.Calculation method for actual impendance of urban network based on improved cell transmission model[J].Journal of Jilin University(Engineering and Technology Edition),2020,50(01):132-139. (Chinese)

[22] Hofleitner A , Graduate Student Researcher. A hydrodynamic theory based statistical model of arterial traffic. 2011.