MRF model based real-time traffic flow prediction with support vector regression

E.Y. Kim✉

A novel statistical model for real-time traffic flow prediction is proposed. In this study, a 3D Markov random field (MRF) is used to model the temporal dynamics of the traffic flow measured by VDS sensor network. Then, the spatial and temporal relations between roads at a given location are represented by the 3D graph using cliques, then its structure is determined by clique parameters. Here, a support vector regression is adopted for estimating the correlation parameters. The technique is applied to actual traffic flow data from Gyeongbu expressway, South Korea. The experiments demonstrate that the proposed method can predict the traffic flow with an accuracy of 85.6%, which improves 17.3% of the existing state-of-the-art method.

Introduction: The traffic prediction to provide the traffic flows of the next or several periods of time in the future is essential for providing traffic control in ITSs. During the past few decades, various algorithms have been proposed. An early approach used the statistical time series methods such as auto-regression and averaging. Recently, the methods that consider both spatial and temporal correlation across the road network have been intensively investigated. In [1], they represent the spatial relations by fixed set of matrix that depends on the distance between roads. Some generative models are proposed for spatio-temporal traffic prediction with a Gaussian process [2] and a Markov one [3], which can naturally express the spatial relations in well-formed formula. However, their main drawback is the computational cost for inference and optimising a large number of correlation parameters.

In this study, a novel statistical model for real-time traffic flow prediction is proposed. The model represents the evolution of the traffic flow rate, measuring the mean velocity of vehicles passing a given location per time unit. This flow rate is described using 3D Markov random field (MRF), as it can effectively represent complex spatial and temporal relations among adjacent roads. Then, unlike the existing methods using naïve dependency models based on the distance and connectivity, the relationship among roads is represented by 3D graph using cliques, and the clique parameters are obtained by a support vector regression (SVR). By using cliques, our model can automatically filter uninformative relations, so that can handle the dynamic interactions between roads over a space and time more efficiently, even though some sensor values are missing or noisy. Furthermore, as the prediction can be independently performed per each node in parallel, it can provide a real-time computation with less training time.

This Letter is an extension of our previous conference work [4]. Here, we clarify the proposed method using 3D graphical model and further validate its generalisation capability via comparing the existing works.

Spatiotemporal MRF modelling: To model the temporal dynamics of the traffic flow as measured from sensors, a spatio-temporal random field (STRF) is constructed. At one time stamp, the sensor network is represented by a spatial graph, G_t, then the graphs are connected to form a temporal chain that have same structures, G_1- G_2-…- G_T over time.

Then, the edges on G_i represent the spatio-temporal dependencies between adjacent roads, assuming the Markov property among them. In literature, the existing methods have used the naïve dependency model on the edges, that is, $E_{t,t+k} \neq 0$ whenever $k > 1$ for any t. In contrast, this study builds a novel dependency model using multi-level logistic model in MRF [5]. In the proposed model, the relationships between roads can be represented by the 3D graph using cliques, rather than by edges, and its structure is determined by the clique parameters.

Fig. 1 illustrates some cliques that generated on the second-order spatio-temporal neighbourhood system. A clique, c is defined as a set of nodes in which all pairs are mutual neighbours. As shown in Fig. 1, the cliques have their own parameters: $\{\alpha\}$ is the parameter set for two-pair cliques and $\{\beta\}$ is one for triple cliques. Then, the parameter set $A = \{\alpha, \beta\}$ represents the weights for the contribution to the next traffic flow prediction. The cliques with the larger weights have the greater influence on the prediction. Such clique parameters are estimated by example-based learning in the proposed method.

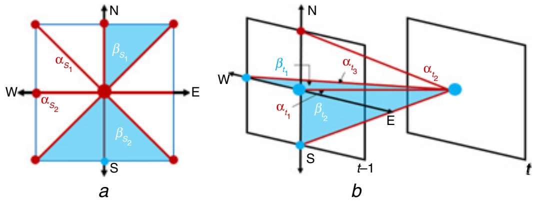

Fig. 1 *Cliques on second-order spatio-temporal neighbourhood*
a Two-pairwise and triple cliques in the spatial domain
b Two-pairwise and triple cliques in the temporal domain

Using cliques, our model can handle the dynamic interactions between adjacent roads over a space and time more efficiently.

Parameter estimation and traffic flow prediction: Our goal is to predict the G_{t+k} based on a stream of observed sensor values. Here, a maximum a posteriori (MAP) is used to estimate the future values. As a posterior has the Gibbs distribution (GD) according to Hammersley–Clifford's theorem [5], the MAP estimation is written as:

$$\begin{aligned} G^*_{t+k} &= \arg\max_{G_{t+k}} P\big(G_{t+k}|G_t, \mathrm{A}\big) \\ &= \arg\max_{G_{t+k}} \frac{1}{Z}\exp\big\{-E\big(G_{t+k}|G_t, A\big)\big\} \end{aligned} \quad (1)$$

The $E(G_{t+k}|G_t, A)$ is the energy function for predicting G_{t+k} given the model parameter A and G_t. Then, the normalising factor, Z can be discard in the MAP estimation. In the GD, the energy function is defined as the sum of the spatial clique potentials, $S_c\big(G_{t+k}|G_t, A\big)$, and the temporal clique potentials, $T_c\big(G_{t+k}|G_t, A\big)$; thus, MAP estimation in (1) is rewritten by minimising the energy function, as follows:

$$P\big(G_{t+k}|G_t, A\big) \propto \arg\min_{G_{t+k}} \sum_{c \in C} \big\{S_c\big(G_{t+k}|G_t, A\big) + T_c\big(G_{t+k}|G_t, A\big)\big\} \quad (2)$$

The C is all cliques defined in the spatio–temporal neighbourhood $\Gamma = \{\eta(r)\}$. As C is the same with the set of cliques at each node r in G, that is, $C = \sum_{r \in G_i} c_r$, the energy function in (2) can be computed at each location independently. When using the second-order neighbourhood, the energy function at a location r is given by the sum of its temporal clique potentials and spatial ones, as shown in the following equations:

$$\begin{aligned} T_c\big(r_{t+k}|r_t, A\big) = {}& \alpha_{\mathrm{T}} P\big(h_{r,t}, h_{r,t+k}\big) + \alpha_{\mathrm{T}} \sum_{q_1 \in \eta_s(r)} P\big(h_{r,t+k},\ h_{q_1,t}\big) \\ & + \beta_{\mathrm{T}} \sum_{(q_1, q_2) \in \eta_t(r) \& q_1 \neq q_2} P\big(h_{r,t},\ h_{q_1,t},\ h_{q_2,t+k}\big) \end{aligned} \quad (3)$$

$$\begin{aligned} S_c\big(r_{t+k}|r_t, A\big) = {}& \alpha_{\mathrm{S}} \sum_{q_1 \in \eta_s(r)} P\big(h_{r,t},\ h_{q_1,t}\big) \\ & + \beta_{\mathrm{S}} \sum_{(q_1, q_2) \in \eta_s(r) \& q_1 \neq q_2} P\big(h_{r,t},\ h_{q_1,t},\ h_{q_2,t}\big) \end{aligned} \quad (4)$$

In (3) and (4), $h_{r,t}$ denote the sensor value on site r at time t, and subscript S and T indicate the spatial domain and the temporal domain, respectively. $P(\cdot)$ denotes the potential functions for the cliques. Among several functions that widely used in GD, the function of $\log\big(1 + (\mathrm{dist}(x,\ k_1,\ \ldots,\ k_m))^2\big)$ was chosen through experiments.

The key component of the proposed model is to correctly estimate the clique parameters. Thus, we employ a SVR as nonlinear analysis to determine the clique parameters, because it has been successfully used in many non-linear problems and the large-scale predictions [6].

Given a training data $D = \{(x_t, y_t)|t = 1, 2, \ldots, n\}$, where each $x_t \subset R^n$ denotes the input dimension of the clique potentials and has a corresponding target value $y_t \subset R$ during n time indexes. The goal of the regression is to fit a function $g(x_t)$ which approximates the relation between the data set points and it can be used later to infer the output for a new input data point. Then, the $g(x_t)$ is represented as follows:

$$y_t = \mathrm{g}(x_t) = \sum_{j=1}^{r} \big(\theta_j - \theta_j^*\big) \kappa(x, x_t) + b \quad (5)$$

In (5), θ_j and θ_j^* are the Lagrange multipliers, $\kappa(x,\ x_t)$ is kernel function.

Thorough experiments, we adopt the radial basis function kernel, which is widely used in mapping nonlinear relationships.

As soon as the parameters are learned from the SVR, the prediction at each location r is computed via).

$$r^*_{t+k} = \arg\min_{r_{t+k}} \sum_{c \in C} \{ S_c(r_{t+k} | r_t, A) + T_c(r_{t+k} | r_t, A) \} \tag{6}$$

Using (6), the proposed model can predict the future sensor values by minimising the energy function, when given a stream of observed sensor measurements. Furthermore, it can apply to predict the sensor values for non-sensor locations by using only the spatial potentials.

Experiments: For quantitative performance comparison, two methods were adopted as baselines: Baseline 1 is Piatkowski *et al.*'s method [2], which employed Gaussian process that is one of popular approaches in traffic flow estimation. They used different dependency model in STRF to ours: the combinatorial Laplacian matrix is used to represent the relationship among roads; (2) Baseline 2 uses the same prediction model as the proposed method but it uses a different parameter estimation method such as multiple linear regression (MLR).

Three methods were applied to actual traffic flow data obtained from VDS sensors of Gyeongbu expressway, South Korea. The stream was collected from February and May 2015. The dataset includes 1076 sensors. The VDS sensors transmit information on traffic flow every thirty seconds. The data consists of the number of passing cars and their average velocity. In the pre-processing step, the sensor values are aggregated within fixed time intervals and some noisy data is filtered by interpolation. We tested various intervals and decided to five minutes, as lower aggregates are too noisy, which caused by sensor fidelity.

Thereafter, the spatial graph G_0 is constructed from the VDS sensor locations. A separate graph is constructed for a day, and each day is further partitioned into 288 graphs according to five-minute intervals. To predict the future traffic value at each location, the proposed method should identify the cliques from spatiotemporal neighbourhood and estimate their clique parameters. Here, a second-order neighbourhood was adopted, since the smaller can miss the correlation between adjacent roads and the larger can induce a high computational cost.

Fig. 2 describes means of some estimated parameters from 1076 sensors, when using a SVR and a MLR. Then, the parameters for five cliques that are commonly extracted from most of sensor locations are shown: $\{\alpha_1((r,t),(r,t+k)),\ \alpha_2((q_1,t),(r,t+k)),\ \alpha_3((q_2,t),(r,t+k)),\ \beta_1((q_1,t),(s,t),(r,t+k)),\ \beta_2((s,t),(q_2,t),(r,t+k))\}$. Although there are some differences between the methods, two common facts are observed: (i) the largest correlation is found in α_1 that represents the correlation between the same locations at different timestamps, which indicates that the temporal correlation has the most significant influence on predicting the next traffic; (ii) although there are no particular trends, the other parameters have the values of 0.3 to 0.7, which means both the spatial and temporal relations should be considered for prediction.

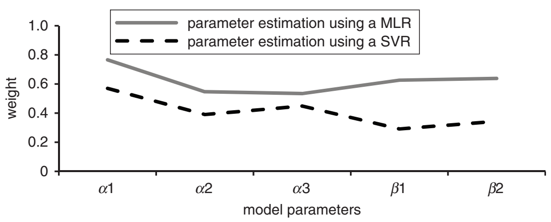

Fig. 2 *Clique parameters on three cone-zones of Fig. 2*

Fig. 3 presents a summary of the numerical comparison for the three methods. Evidently, the proposed method performs best in all durations. As can be seen from this data, the Baseline 1 using different traffic flow model exhibited the lowest accuracy than others. This implies that the proposed prediction model to represent the dependency using cliques is reasonable and practical. More importantly, the proposed method has superior performance to Baseline 2. This result demonstrates the effectiveness of the proposed parameter estimation method.

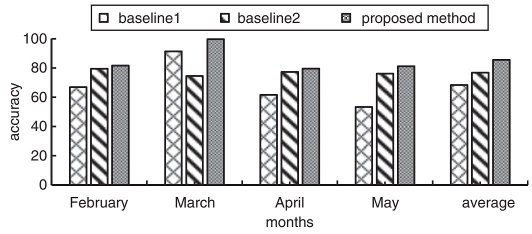

Fig. 3 *Performance comparison for three methods in terms of accuracy*

In addition, to be useful for the traffic flow estimation system, the proposed method should be operable in real-time. Table 1 shows the computational times in three methods for the parameter estimation and the prediction. The Baseline 1 took the longest, among three methods. The prediction time for the proposed traffic flow model took about 0.01 s. As mentioned above, in the proposed method, the prediction can be independently computed per a sensor location in parallel. Even if it is implemented on a serial computer, it can process the predictions for all roads on the Gyeongbu expressway within 11 s, which is smaller than the reading interval of the VDS sensor (30 s).

Table 1: Computational time in three methods (in seconds)

Method	Baseline 1	Baseline 2	Proposed method
Model parameter estimation	2800	30	4
Single run on one location	0.07	0.01	0.01

Consequently, the comparison results confirmed the efficiency and effectiveness of the proposed method in real-time traffic flow prediction.

Conclusion: The method for real-time traffic prediction should be fast and scalable. For this, a novel statistical model is proposed based on 3D MRF, where the spatial and temporal interactions at a location are represented by a 3D graph model using cliques. Then, the clique parameters is estimated by SVR. Experiments demonstrated that the proposed method can process with the accuracy of 85.6% in a real-time.

Submitted: *20 September 2016* E-first: *10 January 2017*
doi: 10.1049/el.2016.3472
One or more of the Figures in this Letter are available in colour online.

E.Y. Kim (*Department of Internet and Multimedia Engineering, Konkuk University, 1 Hawayang-dong, Gwangjin-gu, Seoul, 143-701, South Korea*)

✉ E-mail: eykim@konkuk.ac.kr

References

1 Kamirianakis, Y., Shen, W., and Wynter, L.: 'Real-time road traffic forecasting using regime-switching space-time models and adaptive LASSO', *Appl. Stoch. Models Bus. Ind.*, 2012, **28**, pp. 297–315, doi:10.1002/asmb.1937

2 Piatkowski, N., Lee, S., and Morik, K.: 'Spatio-temporal random fields: Comprehensive representation and distributed estimation', *Mac. Learn.*, 2013, **93**, pp. 115–139, doi:10.1007/s10994-013-5399-7

3 Schnitzler, F., Mannor, S., Liebig, T., and Morik, K.: 'Combining a Gauss-Markov model and Gaussian process for traffic prediction in Dublin city center'. Workshops of the EDBT/ICDT, 2014, pp. 24–28, doi:10.1.1.688.513

4 Ahn, J., Ko, E., and Kim, E.Y.: 'Highway traffic flow prediction using support vector regression and Bayesian classifier'. IEEE BigComp, 2016, pp. 239–244, doi:10.1109/BIGCOMP.2016.7425919

5 Li, S.: 'Markov random field models in computer vision' (Springer, 1995, 2nd edn.)

6 Asif, M.T., Dauwels, J., Goh, C.Y., Oran, A., Xu, M., Dhanya, M.M., Mitrovic, N., and Jaillet, P.: 'Spatiotemporal patterns in large-scale traffic speed prediction', *IEEE Trans. Intell. Transp.*, 2014, **15**, (2), pp. 794–804, doi:10.1109/TITS.2013.2290285