Interpolation System for Traffic Condition Using Estimation/Learning Agents

Tetsuo MORITA*, Junji YANO and Kouji KAGAWA

We propose a multi-agent based interpolation system for traffic condition, which includes estimation and learning agents. These agents are assigned to all the road links. Estimation agents renew the velocity of each road link, and learning agents renew the weight values for estimation. Estimation and learning agents alternately calculate the results to improve the interpolation accuracy. The standard deviation of the estimated velocity is 8.53 km/h, which is obtained from the probe car data from 01/11/2007 to 31/10/2008.

Keywords: probe car, multivariate analysis, coefficient of determination, mean square error

1. Introduction

The role of traffic information services in reducing the consumption of fossil fuels and carbon dioxide is important. Traffic information is classified into two types: temporal information and spatial information. Temporal information denotes forecast technologies, and there is no effective way to forecast future traffic congestion. Spatial information corresponds to traffic congestion mapping. The multi-agent based interpolation system we propose can estimate traffic conditions based on small amounts of information.

The Vehicle Information and Communication System (VICS) is well-known as a traffic information service. VICS gathers the traffic information from roadside sensors and provides it to drivers. VICS is very useful to provide traffic information, but a huge capital investment in roadside sensors is essential. The probe car system is an effective method of reducing this capital investment. Probe cars measure the travel time along road links using Global Positioning System (GPS) sensors and other methods. With the probe car system, it is unnecessary to install sensors at the roadside. However since probe cars are very few in number, it is difficult to estimate traffic congestion only from Probe Car Data (PCD).

One commonly-used method to interpolate traffic conditions is by statistical analysis. This method uses statistical data for the time-sliced average of past PCD from road links. At present, the number of probe cars providing data that covers the same time and conditions is very few, and the sampling errors are very large.

The pheromone model⁽¹⁾⁻⁽³⁾is used to make up the deficit in PCD, with deposit, propagation, and evaporation as the pheromone parameters. While the pheromone model is normally used as a forecast technology, it can be used for interpolation. The pheromone intensity depends on the velocity of the traffic, and changes through a mechanism of propagation and evaporation. Thus traffic congestion can be estimated from the pheromone intensity. But the pheromone parameters are determined by human experience and are difficult to determine objectively.

The Feature Space Projection (FSP) method^{(4),(5)} is

proposed as a method to interpolate the traffic conditions, with the feature being obtained by Principal Component Analysis (PCA) with missing data. The PCA method without missing data is commonly used for multivariate analysis. But in this case, the probability of getting simultaneous PCD from two road links is very small, and so a method of using PCA with missing data is essential for this calculation. The size of this model depends on the product of the number of day factors and time resolution, and the required calculation volume is beyond the capacity of ordinary PCs.

We propose a method of solving these problems ^{(6),(7)}. Learning agents calculate the weight value corresponding to the pheromone parameters, and estimation agents make up the deficit in PCD. In other words, the interpolation accuracy can be improved by collaboration between the estimation and learning agents. The Coefficient of Determination (CD) and Mean Square Error (MSE) are used to evaluate the progress of learning.

2. Traffic Information Interpolation System

Figure 1 shows the multi-agent based interpolation system for traffic conditions. This system consists of both estimation and learning agents that are assigned to all the road links. Estimation agents renew the velocity for each

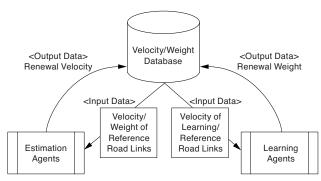


Fig. 1. Multi-agent Based Interpolation System for Traffic Conditions

road link, and learning agents renew the weight values for estimation. The estimated velocities and the weight values are stored in the velocity/weight database. Estimation and learning agents alternate in calculating the results to improve the interpolation accuracy.

Figure 2 shows the normalized velocity (NV) of road links. The normalized velocity used in this system is given by y = 1 - x/100. The x denotes the velocity (km/h), and the y denotes the NVs. When the x is more than 100 (km/h), the y is assigned 0.

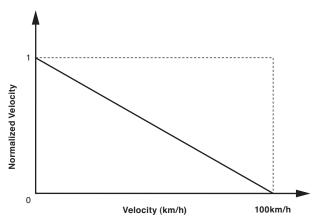


Fig. 2. Definition of Normalized Velocity

2-1 Estimation Agents

Figure 3 shows the example of the junction of road links. Road link C is connected with road links A and B. The linear operation is used in this system. The velocity of road link C is increased by simple addition of the velocity of road link A and B as the velocity is used. When the velocity of road link A and B is 80 km/h, velocity of road link C becomes 160 km/h. In general way, the velocity is decreased by junctions. The velocity of road link C is decreased by simple addition of the NV of road link A and B. When the NV of road link A and B is $0.2 \ (= 80 \text{ km/h})$, the NV of road link C becomes $0.4 \ (= 60 \text{ km/h})$. The NV is suitable for this calculation because the slope of the line in **Fig. 2** is negative.

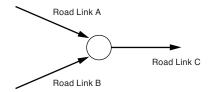


Fig. 3. Example of Road Link Junction

Figure 4 shows an example of the road link connections. The travel time for each road link is converted to an NV, and the NVs at time t are $t^{-1}v^{(1)}$, \cdots , $t^{-1}v^{(7)}$. For ex-

ample, ${}^{t-1}v^{(1)}$ denotes the NV of road link 1 at time t-1.

Estimation agents calculate the NV for the road link being estimated using the NVs for the reference road links and the weight values at time t, and the reference road links that are adjacent to the road link being estimated. The initial NV for each road link is 0, and the initial weights are $w_0^{(i)}=0, w_0^{(1)}, \cdots, w_n^{(1)}=1/n$. n denotes the number of adjacent (reference) road links for the road link i, and each road link has a different value of n. For example, the reference road links for road link 1 are 2, 3, 4, and 5 in **Fig. 4**, and the number of reference road link $n^{(1)}$ is 4. While the notation $n^{(i)}$ is appropriate, the superscript i is omitted in this section. Without initial values, the multi-agent based interpolation system cannot break the deadlock.

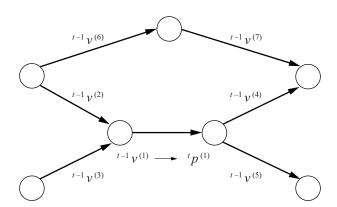


Fig. 4. Example of Road Link Connection

 $V^{(i)}$ denotes the NV vector for the reference road links associated with the road link i, and $w_0^{(i)}$ is the weight vector of the i-th road link at time t. $V^{(i)}$ consists of n NVs for the reference road links and a constant value 1. **Equation (1)** shows the definition of the estimated NV ${}^i\widetilde{E}^{(i)}$ for the road link i at time t. In other words, the estimated NV is the inner product of the NV vector and the weight vector. Occasionally, the weight value ${}^tw_0^{(i)}$ is referred to as the threshold.

$${}^{t}\widetilde{\widetilde{E}}^{(i)} = V^{(i)} \cdot {}^{t-1}w^{(i)} \quad \cdots \qquad (1)$$

$$V^{(i)} = \begin{pmatrix} 1 & V_1^{(i)} & \cdots & V_n^{(i)} \end{pmatrix} \qquad \cdots$$
 (2)

$$^{t-1}\boldsymbol{w}^{(i)} = \begin{pmatrix} t^{-1}w_0^{(i)} & t^{-1}w_1^{(i)} & \cdots & t^{-1}w_n^{(i)} \end{pmatrix}$$
(3)

When the NV for the PCD at time t is ${}^tp^{(1)}$, the component of the NV vector $V^{(i)}$ is renewed.

Estimation agents are assigned to all the road links, and they iteratively calculate the NV for the road links being estimated. First, the new PCD assigned to road link 1 and the NVs for adjacent road links 2, 3, 4, and 5 are calculated. Next, the NVs for road links 6 and 7 are calculated. When the road link being estimated is 2 and the reference road links are 1, 3, and 6, ${}^{\prime}\bar{E}^{(2)}$ can be calculated from **Equation (4)**,

$${}^{t}\tilde{E}^{(2)} = V^{(2)} \cdot {}^{t-1}w^{(2)}$$
(4)

$$V^{(2)} = \begin{pmatrix} 1 & {}^{t}p^{(1)} & {}^{t-1}v^{(3)} & {}^{t-1}v^{(6)} \end{pmatrix} \cdots (5)$$

$${}^{t-1}\boldsymbol{w}^{(2)} = \left({}^{t-1}w_0^{(2)} {}^{t-1}w_1^{(2)} {}^{t-1}w_2^{(2)} {}^{t-1}w_3^{(2)} \right) \cdots \cdots \cdots (6)$$

Using the same method, ${}^{t}\widetilde{E}^{(3)}$ can be calculated from **Equation (7)**.

$${}^{t}\widetilde{E}^{(3)} = V^{(3)} \cdot {}^{t-1}w^{(3)}$$
 (7)

$$V^{(3)} = \begin{pmatrix} 1 & {}^{t}p^{(1)} & {}^{t}\widetilde{E}^{(2)} \end{pmatrix}$$
(8)

$${}^{t-1}\boldsymbol{w}^{(3)} = \left({}^{t-1}w_0^{(3)} {}^{t-1}w_1^{(3)} {}^{t-1}w_2^{(3)} \right) \quad \cdots \qquad (9)$$

Then, when the new PCD is assigned to the road link, the NVs for the adjacent road links are calculated, and finally the NVs for all the road links can be calculated. We should note that the NV vector $V^{(i)}$ includes both the NV components at time t-1 and at time t.

2-2 Learning Agents

Learning agents are assigned to all the road links, and they calculate the weight vector w for learning the road link i referring to the probe NVs for the road link i and the NVs for the reference road links. The superscript for the time t and the road link number i is omitted in the notation. **Equation** (10) is m simultaneous equations with n+1 unknowns. P denotes the probe vector for the road link i with m NVs, and $V_{m \times (n+1)}$ denotes the matrix consisting of m NV vectors for the n reference road links and m constant values of 1. The subscript m denotes the number of PCD values, and each road link has a different value of m. While the notation $m^{(i)}$ is appropriate, the superscript i for m is also omitted.

$$\mathbf{P} = \mathbf{V}_{m \times (n+1)} \cdot \mathbf{w} \qquad \dots \tag{10}$$

$$\mathbf{P} = \left(P_1 \cdots P_m\right)^T \quad \dots \tag{11}$$

$$V_{m\times(n+1)} = \begin{pmatrix} 1 & V_{11} & \cdots & V_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & V_{m1} & \cdots & V_{mn} \end{pmatrix} \qquad \cdots (12)$$

When m is less than n+1, the solutions to the simultaneous equations in **Equation** (10) are not fixed. When the rank of $V_{m\times(n+1)}$ is n+1, **Equation** (10) can be solved. If the number of independent equations is greater than n+1, **Equation** (10) cannot be solved. In this case, the least mean squares method can be used to minimize the MSE. E denotes the estimated NV vector for the road link i with m NVs, which is the product of the NV matrix $V_{m\times(n+1)}$ and the weight vector w.

$$\boldsymbol{E} = \boldsymbol{V}_{m \times (n+1)} \cdot \boldsymbol{w} \qquad \dots \tag{13}$$

$$\boldsymbol{E} = \begin{pmatrix} E_1 & \cdots & E_m \end{pmatrix}^T \quad \dots \tag{14}$$

 ε_k denotes the residuals of the k-th component of P and E. The sum of the squares of the errors Q is given by **Equation (15)**. The sum of the squares of the errors Q divided by m is the MSE.

$$Q = \sum_{k=1}^{m} \varepsilon_k^2 = \sum_{k=1}^{m} (P_k - E_k)^2 \qquad (15)$$

The MSE for the road link i has a minimum value when the partial differential equations in **Equations** (16) and (17) equal 0.

$$\frac{\partial Q}{\partial w_0} = 0 \qquad (16)$$

$$\frac{\partial Q}{\partial w_u} = 0 \ (u = 1, \dots, n) \quad \dots \tag{17}$$

Since **Equation** (17) is linear, multivariate analysis can be used. The dependent variables are the probe NVs, and the independent variables are the NVs for the reference road links. **Equation** (18) is the transformed **Equation** (17). **Equation** (18) can be solved by Gaussian elimination and the regression coefficients, w_1, \dots, w_n , can be calculated.

$$\begin{pmatrix} s_{11} & s_{12} & \cdots & s_{1n} \\ s_{21} & s_{22} & \cdots & s_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ s_{n1} & s_{n2} & \cdots & s_{nn} \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} = \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{pmatrix}$$
 (18)

$$s_{qr} = \frac{1}{m-1} \sum_{i=1}^{m} \left(V_{jq} - \overline{V}_{q} \right) \left(V_{jr} - \overline{V}_{r} \right) \quad \dots \tag{19}$$

$$p_{q} = \frac{1}{m-1} \sum_{i=1}^{m} (V_{jq} - \bar{V}_{q}) (P_{j} - \bar{P}) \qquad (20)$$

$$\overline{V}_q = \frac{1}{m} \sum_{j=1}^m V_{jq}, \ \overline{P} = \frac{1}{m} \sum_{j=1}^m P_j$$
 (21)

The weight value w_0 can be also calculated from **Equation (22)**, which is transformed from **Equation (16)**.

$$w_0 = \overline{P} - \sum_{j=1}^n w_j \cdot \overline{V}_j \qquad (22)$$

3. Evaluation of Interpolation System

PCD for Nagoya taxis is used in this evaluation. The evaluation area for this system is the Nagoya district including Nagoya Station, an area of approximately 10 kilometers by 10 kilometers on a longitude from 136°52'30" to 137°00'00" and a latitude from 35°10'00" to 35°15'00". The taxi company that took part in this experiment has approximately 1200 taxis, and the total number of road links is 1128. The PCD can be obtained every 15 minutes (if probe cars exist at the road link).

The evaluation period was the one year from 01/11/2007 to 31/10/2008. We should note that the number of PCD values m must be not less than the number of unknowns n+1 for learning.

3-1 Evaluation by CD and MSE

In this simulation, the CD and MSE are used to monitor the progress of learning. R denotes the CD for the road link i, which is given by **Equation (23)**. While the notation $R^{(i)}$ is appropriate, the superscript i is omitted. p and e denote the unbiased variance of the PCD and the estimated NV, respectively. R denotes the ratio of the two variances.

$$R = e/p \qquad \cdots \qquad (23)$$

$$e = \frac{1}{m-1} \sum_{j=1}^{m} (E_j - \overline{E})^2 \qquad (24)$$

$$p = \frac{1}{m-1} \sum_{j=1}^{m} (P_j - \bar{P})^2 \qquad (25)$$

$$\bar{E} = \frac{1}{m} \sum_{j=1}^{m} E_j \qquad (26)$$

Figure 5 shows the fluctuations in the CD for several road links in the evaluation area. The x-axis denotes the number of PCD values, and the y-axis denotes the CD. The fifth digit in the road link number (ex. 30281) denotes the following:

- 1: Inter-city Expressways
- 2: Inner-city Expressways
- 3: Local roads

The rest of the digits denote the number assigned to the road link. The road link number in **Fig. 5** denotes a local road.

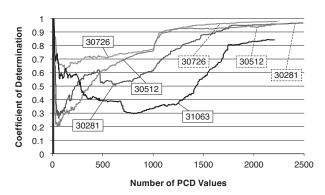


Fig. 5. Fluctuations in the CD

When the number of the PCD is less than that of the unknowns (m < n + 1), the CD is assigned 0. When **Equa**tion (10) can be solved, the CD becomes 1. As the number of PCD additionally increases, the CD abruptly decreases, and then increases gradually. The minimum value of the CD of the road link 30512 is approximately 0.2 after learning, and its value gradually increases to 0.95. The CDs of the road links 30512 and 30726 abruptly increases when the number of the PCD is approximately 1000, and increases gradually. The reason, that the CD abruptly changes, is that the upper limit on the number of PCD values stored in the velocity/weight database is 1000 for each road link. As the old PCD values are deleted when the number of PCD is more than 1000, CD decreases. The reason, that CD increases gradually, is considered that the PCD values are concentrated at the hyper-plane of the linear regression model.

The weight values can be calculated by multivariate analysis, which minimize the MSE. The MSE is the essen-

tial parameter for the evaluation. **Figure 6** shows the fluctuations in the MSE. The x-axis denotes the number of PCD values, and the y-axis denotes the MSE. When learning had finished (31/10/2008), the MSE for road link 30726 had a minimum value, and the MSE for road link 31063 had a maximum value. The trend in the MSE is opposite to the CD.

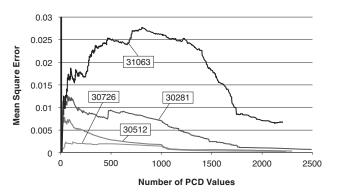


Fig. 6. Fluctuations in the MSE

To investigate the reasons, the numbers of PCD at the reference road links are measured. **Table 1** shows the results of the numbers of the PCD values at the reference road links. For the road link 30726, the number of the reference road links is 13, and the number of the PCD values greater than 400 is 12. For the road link 31063, the number of the reference road links is 11, and the number of the PCD values less than 200 is 8.

From these results, the CD is considered to depend on the number of the PCD at the reference road link.

Table 1. Frequency Distribution of Number of PCD Values

Road Link Number		Number of PCD Values	Road Link Number		Number of PCD Values
Learning	30726	2233	Learning	31063	2203
Reference	30720	4091	Reference	31062	1127
	30724	510		31076	72
	30729	475		31082	62
	30733	1028		30907	33
	31047	534		31070	47
	31121	2487		31065	124
	30721	468		31077	10
	30725	510		31083	22
	30728	478		31064	1507
	30727	3377		30906	1012
	30732	423		31069	102
	31050	1168			
	31124	52			

When the number of the PCD at the reference road links is a few, one PCD value greatly changes the weight value of the reference road link. Therefore the CD abruptly decreases, because the PCD are plotted far from the hyperplane. The slight fluctuation in the CD is considered as the same reason.

Figure 7 shows the fluctuations in the CD for road link 30587. The x-axis denotes the number of PCD values, and the y-axis denotes the CD. The evaluation period was from 01/11/2007 to 31/10/2008. The upper limits on the number of PCD values stored in the learning database are 1000 and 4901, respectively.

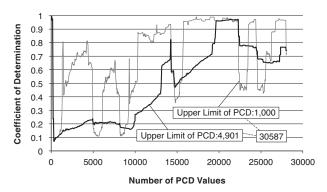


Fig. 7. Fluctuations in the CD

To investigate the reason of the fluctuation, the relationship between the number of PCD values and date is checked. The CD abruptly decreased during New Year's Day, in mid February (national holidays), at the end of April (successive national holidays), in mid August (Bon Festival), and in mid September (national holidays). During these periods, unexpected traffic congestion is considered to occur in this area.

When the upper limit on the number of PCD values stored in the learning database is 1000 and the number of PCD values exceeds 1000, the oldest item of PCD is deleted from the leaning database. When the PCD values associated with the traffic congestion don't exist in the database, the CD decreases abruptly in the case of traffic congestion. When the number of PCD values stored in the database is 4901, the database includes the PCD associated with traffic congestion. Therefore the fluctuation of the CD becomes moderate.

Figure 8 shows the fluctuations in the MSE. The *x*-axis denotes the number of PCD values, and the *y*-axis denotes the MSE. **Figure 8** indicates that the MSE follows the opposite trend of the CD. These results indicate that the estimation accuracy is improved, when the number of PCD values stored in the database increases. However when the number of the PCD values excessively increases, the response of this system decreases. In other word, the number of PCD values is considered to be the time constant of this system. It is important to optimize the time constant of this system.

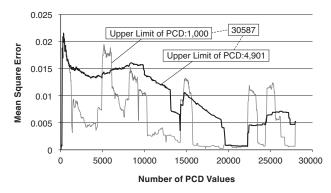


Fig. 8. Fluctuation in the MSE

3-2 Evaluation by Average CD and MSE

Figure 9 and **10** show the relationship between the NV of the PCD values P_k and the estimated NV E_k calculated by the interpolation system, when the number of PCD values is 4901. The x-axis denotes E_k , and y-axis denotes P_k . When the PCD values are plotted on the ideal line with a slope of 1, E_k coincides with P_k .

Figure 9 shows the relationship between E_k and P_k before the successive national holidays (25/04/2008-27/04/2008). Many PCD values are plotted on the ideal line. **Figure 10** shows the relationship between E_k and P_k in the successive national holidays (28/04/2008-30/04/2008). The plotted data of E_k are far from the ideal line.

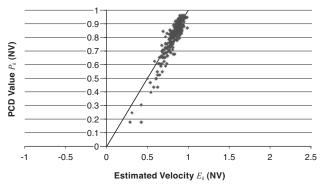


Fig. 9. Relationship between E_k and P_k (25/04/2008-27/04/2008)

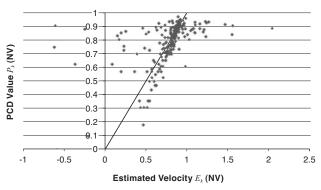


Fig. 10. Relationship between E_k and P_k (28/04/2008-30/04/2008)

Whereas the NV is the range from 0 to 1, the data of E_k plotted in **Fig. 10** exceed this range. When the learning agents calculate MSE, E_k exceeded this range is used. When E_k exceed this range, the actual velocity doesn't exist. For example, when E_k calculated by the interpolation system is 1.8, the stored E_k in the velocity/weight database becomes 1.

 R_{ave} is the average CD for all the road links in the evaluation area, which is given by **Equation (27)**. N denotes the number of road links in the evaluation area. The average MSE Q_{ave} is given by **Equation (28)**, and the notation $R^{(i)}$, $Q^{(i)}$ and $m^{(i)}$ are used in this equation.

$$R_{ave} = \frac{1}{N} \sum_{i=1}^{N} R^{(i)}$$
 (27)

$$Q_{ave} = \frac{1}{N} \sum_{i=1}^{N} \frac{Q^{(i)}}{m^{(i)}}$$
 (28)

Figure 11 shows the average CD and MSE for all the road links in the evaluation area. The x-axis denotes date, and the y-axis denotes the average CD and MSE. The evaluation period was from 01/11/2007 to 31/10/2008. The CD and MSE abruptly improved during New Year's Day and in mid February (national holidays). During these periods, Nagoya taxis choose different routes from normal days.

The final value for the average MSE is 5.5149×10^{-2} , and this value includes the road links without PCD. As the MSE of the road links without PCD is 1, this value can legitimately be subtracted. The number of road links without PCD is 54 on 31/10/2008, and the total number of links N is 1128. Therefore, the actual average MSE in this area is 7.2767×10^{-3} (= $5.5149 \times 10^{-2} - 54/1128$), and the standard deviation of the estimated NV error is 8.530×10^{-2} . The standard deviation of the estimated velocity error is 8.53 km/h, because the slope is 1/100. The average CD and MSE can be improved on by repeating the estimation/learning.

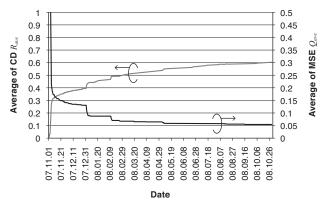


Fig. 11. Average CD and MSE for all the Road Links

4. Conclusions

We propose the interpolation system by estimation and learning agents. The interpolation accuracy is improved by the collaboration of estimation/learning agents. To evaluate the progress of learning, the CD and MSE are used. The CD and MSE can be improved by the repetition of estimation/learning. The fluctuation of CD and MSE depends on the number of the PCD at the reference road links. The standard deviation of the estimated NV error in the evaluation area can be decreased to 8.53 km/h. Furthermore, the CD and MSE could be improved on by repeating the estimation/learning.

* VICS is a trademark of Vehicle Information and Communication System Center.

References

- Ando, Y., Fukazawa, Y., Masutani, O., Iwasaki, H., Honiden, S.: Performance of Pheromone Model for Predicting Traffic Congestion: Proc. of the 5th International Joint Conference on Autonomous Agents and Multiagent Systems, Hakodate, pp. 73-80, May 2006.
- (2) Kurihara, S., Tamaki, H., Numao, M., Kagawa, K., Yano, J., and Morita, T.: Traffic Congestion Forecasting based on Pheromone Communication Model for Intelligent Transport Systems: IEEE Congress on Evolutionary Computation, Trondheim, pp. 2879-2884, May 2009.
- (3) Kurihara, S., Tamaki, H., Numao, M., Kagawa, K., Yano. J., and Morita, T.: Traffic Congestion Forecasting based on Ant Model for Intelligent Transport Systems: The 3rd International Workshop on Emergent Intelligence in Networked Agents (WEIN 2009) (Workshop at AAMAS 2009), Budapest, pp. 42-47, May 2009.
- (4) Kumagai, M., Fushiki, T., Kimita, K., Yokota, T.: Long-range Traffic Condition Forecast using Feature Space Projection Method: Proc. of 11th World Congress of ITS, Nagoya, CD-ROM, Oct. 2004.
- (5) Kumagai, M., Fushiki, T., Kimita, K., Yokota, T.: Spatial Interpolation of Real-time Floating Car Data Based on Multiple Link Correlation in Feature Space: Proc. of 13th World Congress of ITS, London, CD-ROM, Oct. 2006.
- (6) Morita, T., Yano, J., Kagawa, K.: Interpolation System of Traffic Condition by Estimation/Learning Agents: Proc. of 12th international Conference on Practice in Multi-Agent Systems, Nagoya, pp. 487-499, Dec. 2009.
- (7) Morita, T., Yano, J., Kagawa, K.: Multiagent Based Interpolation System for Traffic Condition by Estimation/Learning: Proc. of the 9th International Conference on Autonomous Agents and Multiagent Systems, Toronto, pp. 1697-1704, May 2010.

Contributors (The lead author is indicated by an asterisk (*)).

T. MORITA*

Doc. of Eng.
 Assistant General Manager, Applied ICT Systems Department, Information & Communications Laboratories
 Engaged in the development of traffic information systems



I VANO

• Assistant Manager, Applied ICT Systems Department, Information & Communications Laboratories

K. KAGAWA

 Manager, Applied ICT Systems Department, Information & Communications Laboratories