High Performance Spatial Interpolation System for Traffic Conditions by Floating Car Data

Tetsuo MORITA

The performance of spatial interpolation system for traffic conditions by Floating Car Data (FCD) can be improved. This system includes estimation and learning agents, which are assigned to all the road links. Estimation agents renew the Normalized Congestion Level (NCL) for each road link, and learning agents renew the weight values for estimation. The weight values can be calculated by a data mining method. Estimation and learning agents alternately calculate the results to improve the interpolation accuracy. The Coefficient of Determination (CD) and Mean Square Error (MSE) are used to evaluate the interpolation accuracy. The obtained FCD is stored in a learning database. A hard disk drive-based learning database is changed into a memory-based one to improve the performance. The processing time is reduced to 1/30, when the memory-based learning database is used. The average standard deviation of the estimated velocity error is 7.34 km/h in the evaluation area.

Keywords: floating car, multivariate analysis, data mining, coefficient of determination, mean square error

1. Introduction

The role of traffic information services in reducing the consumption of fossil fuels and emissions of carbon dioxide is important. Traffic information is classified into two types: temporal information and spatial information. Temporal information denotes forecast technologies, and spatial information corresponds to traffic congestion mapping.

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

One commonly-used method to interpolate traffic conditions is by statistical analysis. This method uses statistical data for the time-sliced average of past FCD from road links. At present, the number of floating 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 FCD, 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 therefore it is difficult to determine them objectively.

The Feature Space Projection (FSP) method [(4),(5)] has been 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 FCD from two road links is very small, and so a method of using PCA with missing data is essential for this calculation.

A method of solving these problems has been proposed. [(6),(7)] Learning agents calculate the weight value corresponding to the pheromone parameters, and estimation agents make up the deficit in FCD. 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.

In this study, the performance of the spatial interpolation system has been improved. Learning agents calculate the weight value by the velocity of the reference road links, which are stored in the learning database. As a Hard Disk Drive (HDD) is used for the learning database, the HDD access is the bottleneck in this system. The size of the learning database is too large to store it in the memory of a PC. The velocities for the road links, however, can be stored in the memory of the PC when compressed, which greatly improves the performance of the system.

2. Spatial Interpolation System for Traffic Condition

2-1 Concept of spatial interpolation

Figure 1 shows the definition of the Normalized Congestion Level (NCL). NCL is used for the calculation of the interpolation system. Whereas the word of normalized velocity is used in the previous papers [(6),(7)], the definition of the normalized velocity has the same of the NCL. The NCL used in this system is given by y=1-x/100. The x-axis denotes the floating car velocity (km/h), and the y-axis denotes the NCL. When the x is more than 100 (km/h), the y is as-

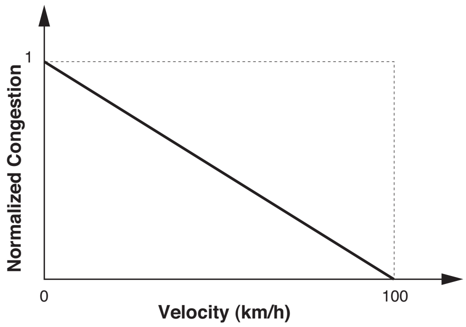

Fig. 1. Definition of Normalized Congestion

signed 0. When the floating car velocity at the road link is 100 km/h, the NCL is 0, and when that is 0 km/h, the NCL is assigned 1.

The NCL for the road link is the function of the time and space. **Figure 2** shows the image of traffic flow, such as the traffic flow continuum. The NCL for the specific road link is determined by the NCL of the present adjacent road link and the past road link itself.

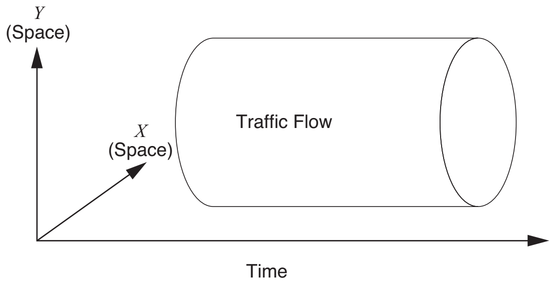

Fig. 2. Image of the Traffic Flow Continuum

In the traditional system, the NCL can be calculated using the only one road link. For example, to estimate the traffic congestion at 9 a.m. on Monday, the average traffic congestion at 9 a.m. on previous Mondays for several months is used as the estimated congestion. As the traffic flow continuum is cut including the time axis shown in **Fig. 3**, the congestion is the (periodic) non-linear function of time.

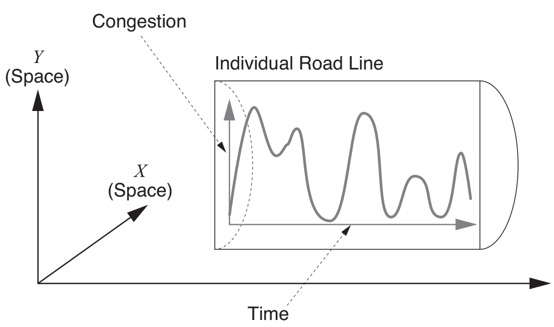

Fig. 3. NCL Fluctuation of the Specific Road Link

Data mining is difficult for non-linear systems.

The Fourier series is the well known method to describe the periodic function, and the relation between the NCL and time can be described by Fourier series, which consist of sine and cosine functions. ω is the angular frequency and a_h, b_h are coefficients. Whereas the infinite number of coefficients should be calculated in **Equation (1)**, FCD value can be obtained every several minutes (e.g. 15 minutes) in actual traffic information systems. Therefore actual NCLs can be described by the limited number of coefficients. For example, period is 1 week and the interval of FCD acquisition is 15 minute (4 FCDs an hour), h is from 1 to 336 (multiple of 7 days, 24 hours, 4, and 1/2). As the period increases, the number of coefficient increases.

$$g_0(t) = \frac{a_0}{2} + \sum_{h=1}^{\infty} \{a_h \cos(h\omega t) + b_n \sin(h\omega t)\} \qquad \cdots\cdots\cdots\cdots (1)$$

Figure 4 shows the schematic view of the traffic flow continuum cut including the space axes. Cutting like this, the NCL can be considered as the linear combination of the NCL of the adjacent road links. As the NCL is linear function of space, data mining can be used for interpolation.

The congestion model can be described by the register network shown in **Fig. 5**. The nodes in **Fig. 5** denote road links, not (road link) junctions. The register network denotes the dual graph of the actual road link connections.

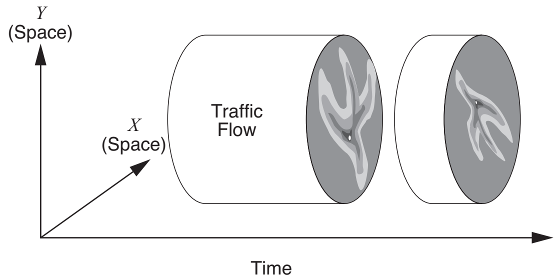

Fig. 4. NCL Fluctuation along the Space Axes

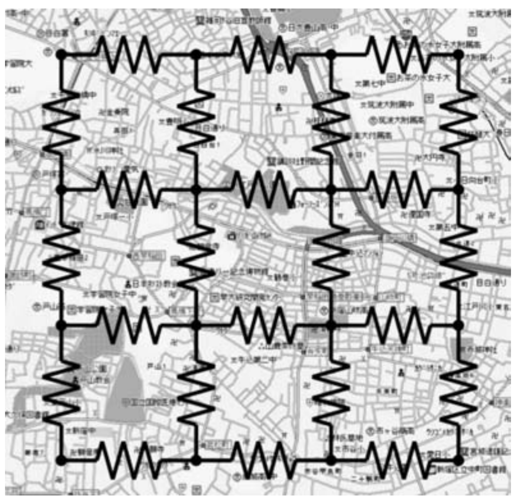

Fig. 5. Dual Graph for the Road Link Network

Figure 6 shows the actual road link connections. These two-way roads are divided into 3 parts and consists of 6 road links. The road links are assigned the road link number from 1 to 6.

Figure 7 shows the topological view for the road link connections. Road links 3 and 4 are connected to the same nodes.

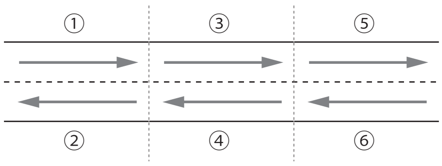

Fig. 6. Actual Road Link Connections

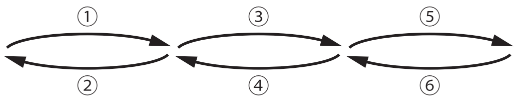

Fig. 7. Topological Connections for the Road Links

Figure 8 shows the dual graph expression of the topological road link connections in **Fig. 7**. The nodes denote the road links and the road link 3 is connected to road link 1, 2, 4, 5, and 6. $V_1, \cdots, V_6$ denote the NCLs for road link 1, ···, 6, which correspond to the voltage of the register network. As road link 3 is connected to the all the other road links (1,2,4,5,6), the number of the edge connected to road link 3 is 5. Road link 3 is connected to all the other road links by the register.

The congestion flow (current) from rode link 1 to 3 can be calculated by the multiple of the conductance and the difference of the NCL (voltage). Therefore, the NCL for road link 1 spreads to adjacent road link 3. In other word, the congestion of road link 1 causes the congestion of adjacent road link 3. The weight value $w'_{(1\rightarrow3)}$ denotes the conductance between road link 1 and 3. As the unisotropic weight values (conductance) are used in this system, weight $w'_{(1\rightarrow3)}$ and $w'_{(3\rightarrow1)}$ have the different values.

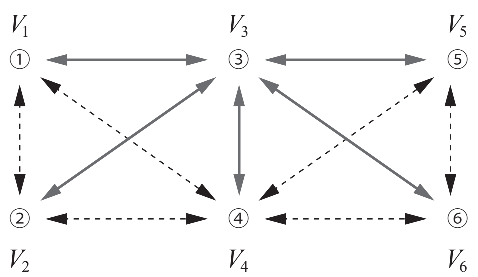

Fig. 8. Road Link Connections (Dual Graph)

As Kirchhoff's First Law is applied to this model, the total congestion flow (current) to road link 3 is 0. Therefore **Equation (2)** can be obtained.

$$w'_{(1\rightarrow3)}(V_1-V_3)+w'_{(2\rightarrow3)}(V_2-V_3)+w'_{(4\rightarrow3)}(V_4-V_3)+w'_{(5\rightarrow3)}(V_5-V_3)+w'_{(6\rightarrow3)}(V_6-V_3)=0 \quad \cdots\cdots\cdots\cdots (2)$$

When the weight value is large, the congestion easily spreads to the adjacent road links. Road links 3 and 4 have the opposite direction and there is no traffic flow from road link 4 to 3. As drivers look aside to the opposite road when accident occurs, congestion spread may occur from road link 4 to 3.

Equation (3) is the transforme of **Equation (2)**. The coefficients of **Equation (3)** are given by **Equations (4) (5)**.

$$V_3=w_{(1\rightarrow3)}V_1+w_{(2\rightarrow3)}V_2+w_{(4\rightarrow3)}V_4+w_{(5\rightarrow3)}V_5+w_{(6\rightarrow3)}V_6 \quad \cdots\cdots\cdots\cdots (3)$$

$$w_{(1\rightarrow3)}=\frac{w'_{(1\rightarrow3)}}{5\overline{w}_{(3)}},\ w_{(2\rightarrow3)}=\frac{w'_{(2\rightarrow3)}}{5\overline{w}_{(3)}},\ w_{(4\rightarrow3)}=\frac{w'_{(4\rightarrow3)}}{5\overline{w}_{(3)}},\ w_{(5\rightarrow3)}=\frac{w'_{(5\rightarrow3)}}{5\overline{w}_{(3)}},\ w_{(6\rightarrow3)}=\frac{w'_{(6\rightarrow3)}}{5\overline{w}_{(3)}} \quad \cdots\cdots\cdots (4)$$

$$5\overline{w}_{(3)}=w'_{(1\rightarrow3)}+w'_{(2\rightarrow3)}+w'_{(4\rightarrow3)}+w'_{(5\rightarrow3)}+w'_{(6\rightarrow3)} \quad \cdots\cdots\cdots\cdots (5)$$

By applying Kirchhoff's First Law to this system, the voltage for road link 3 can be expressed by the linear combination of the voltage for other road links 1, 2, 4, 5, and 6. In other word, the NCL for the road link can be expressed by the linear combination of the NCL for the adjacent road links.

Figure 9 shows the example for the road link connections. $g_1(t)$, $g_2(t)$, $g_3(t)$, $g_4(t)$ are the NCL functions for road link 1, 2, 3 and 4, respectively. When the road link connections are shown in **Fig. 9**, the NCL for road link 0 can be given by **Equation (6)**. Whereas $g_1(t)$, $g_2(t)$, $g_3(t)$, $g_4(t)$ are non-liner functions of time, $g_0(t)$ can be expressed by the liner combination of the non-liner function. Therefore the data mining method can be used for this system.

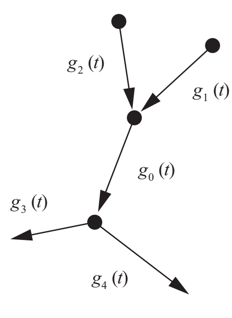

Fig. 9. Example for the Road Link Connections

$$g_0(t)=w_{(1\rightarrow0)}\,g_1(t)+w_{(2\rightarrow0)}\,g_2(t)+w_{(3\rightarrow0)}\,g_3(t)+w_{(4\rightarrow0)}\,g_4(t) \quad \cdots\cdots\cdots (6)$$

The number of the items in **Equation (6)** is much less than that in **Equation (1)**. In **Equation (1)**, $g_0(t)$ is expressed by the sine and cosine functions, which waveforms differ from the waveform of road link 0. As the waveform of $g_0(t)$ is similar to that of $g_1(t)$, $g_2(t)$, $g_3(t)$, $g_4(t)$, the number of terms of **Equation (6)** is much less than that of **Equation (1)**. From the other point of views, when the traffic flow continuum is cut including the space axes, the parameter t becomes implicit (omitted).

By the way, when the waveform of $g_0(t)$ is much similar to that of $g_1(t)$, $g_2(t)$, $g_3(t)$, $g_4(t)$, the calculation accuracy degraded by the well-known problem of multi-collinearity. To solve this problem, when the correlation coefficient of the two road links is more than 0.8, a pair of these road links should be deleted.

3. Traffic Information Interpolation System

Figure 10 shows the spatial 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 NCL (velocity) for each road link, and learning agents renew the weight values for estimation. The estimated NCL and the weight values are stored in the NCL/weight database. Estimation and learning agents alternate in calculating the results to improve the interpolation accuracy. The NCLs for learning agents are extracted from the NCL/weight database and stored in the learning database.

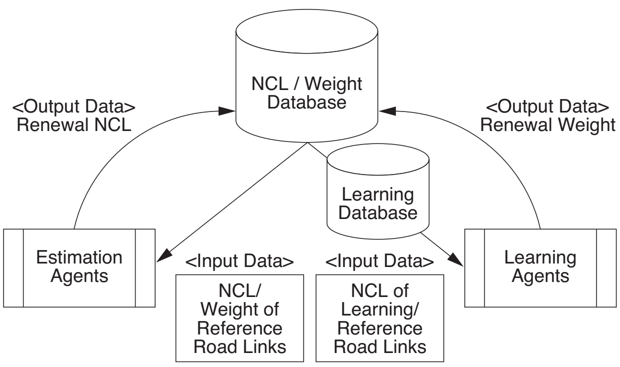

Fig. 10. Spatial Interpolation System for Traffic Conditions

3-1 Estimation agents

Figure 11 shows an example for the road link connections. The travel time for each road link is converted to an NCL, and the NCLs at time t-1 are ${}^{t\text{-}1}v^{(1)}, \cdots, {}^{t\text{-}1}v^{(7)}$. For example, ${}^{t\text{-}1}v^{(1)}$ denotes the NCL for road link 1 at time t-1.

Estimation agents calculate the NCL for the road link being estimated using the NCLs for the reference road links and the weight values at time t-1, and the reference road links that are adjacent to the road link being estimated. The initial NCL for each road link is 0, and the initial weights are $w_0^{(i)}=0$, $w_1^{(i)}= \cdots = w_n^{(i)}=1/n$ denotes the number of the adjacent (reference) road links for the road link i, and each road link has a different value of $n^{(i)}$. For example, the reference road links for road link 1 are 2, 3, 4, and 5 in **Fig. 11**, 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 spatial interpolation system cannot break the deadlock.

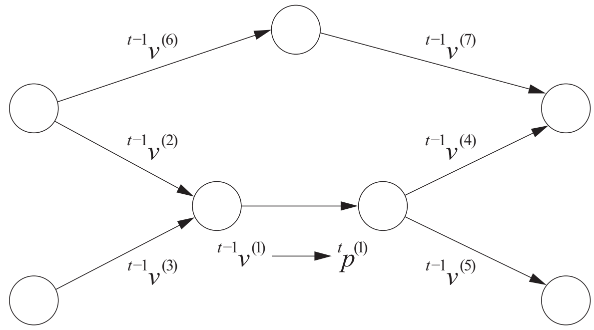

Fig. 11. Example of Road Link Connections

$\boldsymbol{V}^{(i)}$ denotes the NCL vector for the reference road links associated with the road link i, and ${}^{t\text{-}1}\boldsymbol{w}^{(i)}$ is the weight vector of the i-th road link at time t-1. $\boldsymbol{V}^{(i)}$ consists of n NCLs for the reference road links and a constant value 1. **Equation (7)** shows the definition of the estimated NCL ${}^{t}\tilde{E}^{(i)}$ for the road link i at time t. In other words, the estimated NCL is the inner product of the NCL vector and the weight vector. Occasionally, the weight value ${}^{t}w_0^{(i)}$ is referred to as the threshold.

$$ {}^{t}\tilde{E}^{(i)} = \boldsymbol{V}^{(i)} \cdot {}^{t\text{-}1}\boldsymbol{w}^{(i)} \quad \cdots\cdots\cdots\cdots (7)$$

$$ \boldsymbol{V}^{(i)} = \begin{pmatrix} 1 & V_1^{(i)} & \cdots & V_n^{(i)} \end{pmatrix} \quad \cdots\cdots\cdots\cdots (8)$$

$$ {}^{t\text{-}1}\boldsymbol{w}^{(i)} = \begin{pmatrix} {}^{t\text{-}1}w_0^{(i)} & {}^{t\text{-}1}w_1^{(i)} & \cdots & {}^{t\text{-}1}w_n^{(i)} \end{pmatrix}^{\mathrm{T}} \quad \cdots\cdots\cdots\cdots (9)$$

When the floating car NCL of road link 1 at time t is ${}^{t}p^{(1)}$, the component of the NCL vector $\boldsymbol{V}^{(i)}$ is renewed. In this case, the estimation order is 2, 3, 4, 5, 6, 7. When the floating car NCL for road link 7 is obtained, the estimation order is 4, 6, 1, 5, 2, 3.

3-2 Learning agents

Learning agents are assigned to all the road links, and they calculate the weight vector ${}^{t}\boldsymbol{w}^{(i)}$ for learning the road link i at time t referring to the floating car NCL (velocity) for the road link i and the NCLs for the reference road links. The superscripts for time t and the road link number i are omitted in the notation. **Equation (10)** is m simultaneous equations with n+1 unknowns. $\boldsymbol{P}$ denotes the floating car NCL vector for the road link i with m NCLs, and $\boldsymbol{V}_{m\times(n+1)}$ denotes the matrix consisting of m NCL vectors for the n reference road links and m constant values of 1. The subscript m denotes the number of the NCL vectors, 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.

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

$$\boldsymbol{P} = \begin{pmatrix} P_1 & \cdots & P_m \end{pmatrix}^T \qquad (11)$$

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

When m is less than n+1, the solutions to the simultaneous equations in **Equation (10)** are not fixed. When the rank of $\boldsymbol{V}_{m\times(n+1)}$ 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. $\boldsymbol{E}$ denotes the estimated NCL vector for the road link i with m NCLs, which is the product of the NCL matrix $\boldsymbol{V}_{m\times(n+1)}$ and the weight vector $\boldsymbol{w}$.

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

$$\boldsymbol{E} = \begin{pmatrix} E_1 & \cdots & E_m \end{pmatrix}^T \qquad (14)$$

ε_k denotes the residuals of the k-th component of $\boldsymbol{P}$ and $\boldsymbol{E}$. The sum of the squares of the errors Q is given by **Equation (15)**. The sum of the square 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) (17)** equal 0.

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

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

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

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

$$s_{qr} = \frac{1}{m-1} \sum_{j=1}^{m} \left(V_{jq} - \bar{V}_q\right)\left(V_{jr} - \bar{V}_r\right) \qquad (19)$$

$$p_q = \frac{1}{m-1} \sum_{j=1}^{m} \left(V_{jq} - \bar{V}_q\right)\left(P_j - \bar{P}\right) \qquad (20)$$

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

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

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

4. Performance of Processing

FCD 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 for road links is 1128. The FCD can be obtained every 15 minutes. The FCD from 01/11/2007 to 31/10/2008 was used for this evaluation.

First, the number of FCD values from the 1128 road links was checked (see **Table 1**). The subtotal for road links with single-figure FCD values is 87. We should note that the number of FCD values m must be not less than the number of unknowns n+1 for learning; therefore it is difficult to cal-

Table 1. Frequency Distribution of Number of FCD Values

Number of FCD	Count	Number of FCD	Count
1	19	10 To 19	59
2	7	20 To 29	32
3	15	30 To 39	34
4	14	40 To 49	23
5	8	50 To 59	19
6	7	60 To 69	20
7	9	70 To 79	23
8	5	80 To 89	18
9	3	90 To 99	14
Subtotal	87	Subtotal	242
Number of FCD	Count	Number of FCD	Count
100 To 199	107	1000 To 4999	237
200 To 299	52	5000 To 8999	58
300 To 399	53	9000 To 12999	22
400 To 499	38	13000 To 16999	12
500 To 599	32	17000 To 20999	5
600 To 699	34	21000 To 24999	6
700 To 799	30	25000 To 28999	7
800 To 899	23	29000 To 32999	1
900 To 999	28	33000 To 35136	0
Subtotal	397	Subtotal	348
Sum Total	1074	No FCD	54

culate the weight value of a road link from single-figure FCD values. The subtotal for road links with four- and five-figure FCD values is 348. In this range, the road links have collected a sufficient number of FCD values. The maximum number of FCD values obtained per day was 96, and the evaluation period was 366 days, giving a maximum number of FCD values of 35136.

In the four- and five-figure ranges, the NCL for the road links could be estimated by traditional method (the time-sliced average of past FCD) without using this interpolation system. At the end of the evaluation period (31/10/2008), the number of road links with FCD was 1074 at most. The rest of the 54 road links had no FCD. A suitable range of FCD for this interpolation system is considered to be three figures.

The floating car NCL vector $\boldsymbol{P}$ and the NCL matrix $\boldsymbol{V}_{m\times(n+1)}$ are stored in the learning database. The size of the learning database increases in proportion to the number of FCD values m. The size of the learning database was approximately 500 MB at the end of the evaluation (31/10/2008). When new FCD is obtained (1 added to m), the values of $\bar{V}_q$ and $\bar{P}$ in **Equation (21)** are changed. Therefore the values of s_{qr} and p_q in **Equations (19) (20)** should be recalculated.

The number of coefficients s_{qr} and p_q in **Equation (18)** is n^2+n. As the matrix in **Equation (18)** is symmetrical, the number of coefficients that should be calculated is $n(n+1)/2+n$. When the number m increases, the number of coefficients no longer depends on m. **Equations (23) (24)** are the transforms of **Equations (19) (20)**. The coefficient of **Equation (18)** can be calculated by the sum and product of V_{jq}, V_{jr}, P_j. The size of the learning database drastically decreases when **Equations (23) (24)** are used.

$$s_{qr} = \frac{m\sum_{j=1}^{m}(V_{jq}\cdot V_{jr}) - \sum_{j=1}^{m}V_{jq}\sum_{j=1}^{m}V_{jr}}{m(m-1)} \quad \cdots\cdots\cdots\cdots (23)$$

$$p_q = \frac{m\sum_{j=1}^{m}(P_j\cdot V_{jq}) - \sum_{j=1}^{m}P_j\sum_{j=1}^{m}V_{jq}}{m(m-1)} \quad \cdots\cdots\cdots\cdots (24)$$

Figure 12 shows the processing time of the interpolation system from 01/11/2007 to 31/10/2008. The interpolation system with HDD-based learning database takes approximately two weeks for calculation, but that with memory-based one takes only 12 hours. The processing time was reduced to 1/30, when the memory-based learning database was used. For calculation, an Intel Core2 Duo CPU (clock speed 1.2 GHz, 1 GByte RAM) was used.

5. Evaluation by CD and MSE

In this simulation, the CD and MSE are used to monitor the progress of learning. [(6),(7)] **Figure 13** shows the fluctuations in the CD for several road links in the evaluation area. The x-axis denotes the number of FCD values, and the y-axis denotes the CD. The FCD used for this evaluation was from 01/11/2007 to 31/10/2008. The same FCD was used twice. Firstly, the FCD from 01/11/2007 to 31/10/2008 was used, and the same FCD from 01/11/2007 to 31/03/2008 was used thereafter. The road links for which the number of FCD values is between 2500 and 3000 are selected. The upper limit on the number of FCD values stored in the learning database is 32000 for each road link. Therefore no FCD values are removed in this evaluation.

The fifth digit in the road link number (e.g. 30567) 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. 13** denotes a local road.

When the number of FCD values is less than the number of unknowns ($m < n + 1$), 0 is assigned to the CD. When **Equation (10)** can be solved, the CD becomes 1. As the number of FCD values increases, the CD decreases abruptly, and then increases gradually. The minimum value of the CD for road link 30245 is approximately 0.2 after learning, and its value gradually increases to 0.82. The CD for road link 31132 abruptly decreases when the number of FCD values is approximately 1500, and increases gradually thereafter. The reason why the CD changes abruptly is thought to be that the weight values of the reference road links change. The reason why the CD increases gradually is thought to be that the FCD is concentrated in

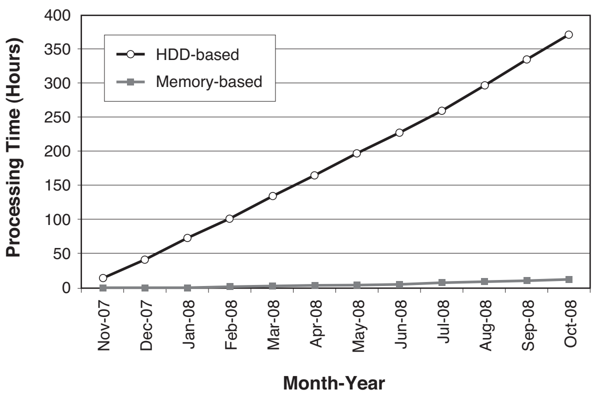

Fig. 12. Processing Time of Interpolation System

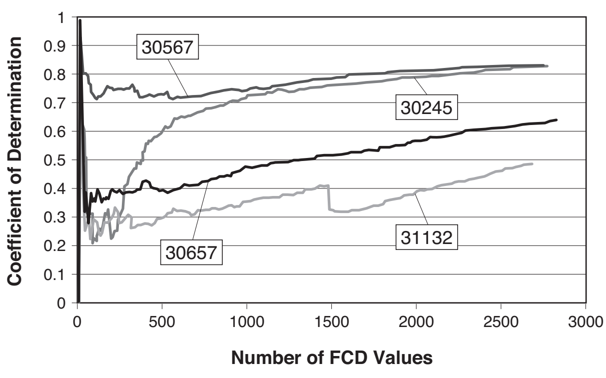

Fig. 13. Fluctuations in the Coefficient of Determination

a hyperplane in the multivariate analysis.

The weight values can be calculated by multivariate analysis, which minimizes the MSE, which is the essential parameter in the evaluation. **Figure 14** shows the fluctuations in the MSE. The *x*-axis denotes the number of FCD values, and the *y*-axis denotes the MSE. When no FCD exists for the road link, 1 is assigned to the MSE. When **Equation (10)** can be solved, the MSE becomes 0. As the number of FCD values increases, the MSE increases abruptly, and then decreases gradually. The trend in the MSE is opposite to the CD.

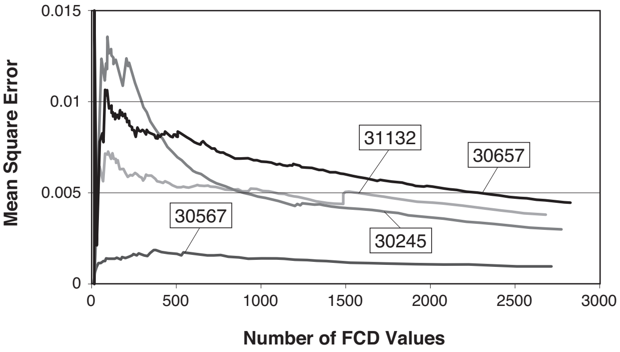

Fig. 14. Fluctuations in the Mean Square Error

Figure 15 shows the frequency distribution of the estimated velocity error of all the road links in the evaluation area. The *x*-axis denotes the standard deviation of the estimated velocity error, and the *y*-axis the count. The total number of road links is 1128, and the number of road links without FCD is 65 at 31/03/2008 in the first run. The number of road links without FCD is 54 at 31/03/2008 in the second run. All road links without the FCD are removed from **Fig. 15**.

The dashed line denotes the frequency distribution for the first run (01/11/2007 to 31/03/2008), and the solid line denotes the frequency distribution for the first (01/11/2007 to 31/10/2008) and second runs (01/11/2007 to 31/03/2008). At a position on the *x*-axis of 2, the standard deviation x of the estimated velocity error is $1 < x \leq 2$. The count is 13 for the first run when the position on the *x*-axis is 0. This means that the standard deviation of the estimated velocity error of 13 road links is 0, because **Equation (10)** can be solved.

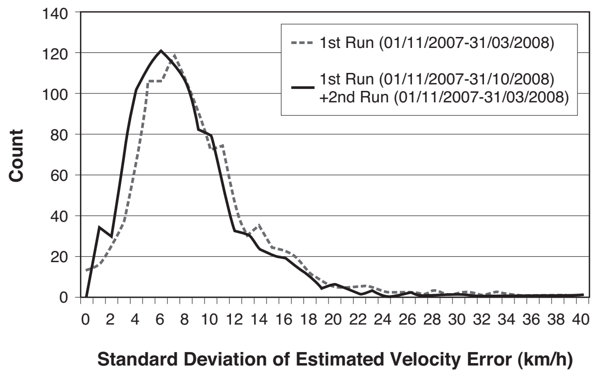

Fig. 15. Frequency Distribution of Estimated Velocity Error

The average of the standard deviation of the estimated velocity error is 8.37 (km/h) at 31/03/2008 in the first run. The average of the standard deviation of the estimated velocity error becomes 7.34 (km/h) at 31/03/2008 in the second run. The standard deviation of the estimated velocity error decreases when the number of FCD values increases.

6. Conclusions

A spatial interpolation system for traffic conditions that includes estimation and learning agents is proposed. These agents are allocated to all the road links. Estimation agents renew the NCL for each road link, and learning agents renew the weight values for estimation. The weight values can be calculated by the data mining method. The interpolation accuracy can be improved when the number of FCD values increases. An HDD-based learning database is changed into a memory-based one to improve the performance. When the memory-based learning database is used, the processing time becomes 1/30. The average standard deviation of the estimated velocity error becomes 7.34 km/h in the evaluation area.

References

(1) Y. Ando, Y. Fukazawa, O. Masutani, H. Iwasaki, S. Honiden, "Performance of Pheromone Model for Predicting Traffic Congestion," Proc. of the 5th International Joint Conference on Autonomous Agents and Multiagent Systems, pp. 73-80 (2006)

(2) S. Kurihara, H. Tamaki, M. Numao, K. Kagawa, J. Yano, and T. Morita, "Traffic Congestion Forecasting based on Pheromone Communication Model for Intelligent Transport Systems," IEEE Congress on Evolutionary Computation, Trondheim, pp. 2879-2884 (May 2009)

(3) S. Kurihara, H. Tamaki, M. Numao, K. Kagawa, J. Yano, and T. Morita, "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) M. Kumagai, T. Fushiki, K. Kimita, T. Yokota, "Long-range Traffic Condition Forecast using Feature Space Projection Method," Proc. of 11th World Congress of ITS, Nagoya, CD-ROM (Oct. 2004)

(5) M. Kumagai, T. Fushiki, K. Kimita, T. Yokota, "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) T. Morita, J. Yano, K. Kagawa, "Interpolation System of Traffic Condition by Estimation/Learning Agents," in Proceedings of 12th International Conference on Practice in Multi-Agent Systems, Nagoya, pp. 487-499 (Dec. 2009)

(7) T. Morita, J. Yano, K. Kagawa, "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)

Contributor

T. MORITA

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