

Multiagent Based Interpolation System for Traffic Condition by Estimation/Learning

Tetsuo Morita
Information & Communication
Laboratories
Sumitomo Electric Industries
1-1-3, Shimaya, Konohana-ku
Osaka, 554-0024 Japan
morita-tetsuo@sei.co.jp

Junji Yano
Information & Communication
Laboratories
Sumitomo Electric Industries
1-1-3, Shimaya, Konohana-ku
Osaka, 554-0024 Japan
yano-junji@sei.co.jp

Kouji Kagawa
Information & Communication
Laboratories
Sumitomo Electric Industries
1-1-3, Shimaya, Konohana-ku
Osaka, 554-0024 Japan
kagawa-kouji@sei.co.jp

ABSTRACT

We propose a multiagent based interpolation system for traffic conditions that includes estimation and learning agents. These agents are allocated to all the road links. The Normalized Velocity (NV) is used in this system. Estimation agents renew the NV for each road link, and learning agents renew the weight values for estimation. The weight values can be calculated by multivariate analysis. 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. Vehicle Information and Communication System (VICS) data and Probe Car Data (PCD) are usually used for traffic information systems, but we have confirmed that the estimation accuracy without VICS data (only PCD) is higher than with VICS data. The standard deviation of the estimated NV error can be improved to 0.1312, and the standard deviation of the estimated velocity error is 6.56 km/h in the mid velocity region. It was possible to improve the CD and MSE by repeated estimation and learning.

Categories and Subject Descriptors

I.2.11 [Artificial Intelligence]: Distributed Artificial Intelligence – Coherence and coordination, Intelligent agents, Multiagent systems.

General Terms

Design, Algorithms.

Keywords

Probe car, Multivariate analysis, Coefficient of determination, Mean square error.

Cite as: Multiagent Based Interpolation System for Traffic Condition by Estimation/Learning, Tetsuo Morita, Junji Yano and Kouji Kagawa, *Proc. of 9th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2010)*, van der Hoek, Kaminka, Lespérance, Luck and Sen (eds.), May, 10–14, 2010, Toronto, Canada, pp. 1697-1704
Copyright © 2010, International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.

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 multiagent 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][2][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 [5] 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 proposed a method of solving these problems.[6] 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.

VICS data and PCD are usually used for traffic information systems. In this paper, we compare the interpolation accuracy with and without VICS data (Probe Only). We also investigate the relation between the interpolation accuracy and the number of PCD values for the adjacent (reference) road links. In addition, we investigate the effect of (1) the same PCD values used more than twice, and (2) removal of the old PCD values in the database.

2. INTERPOLATION SYSTEM FOR TRAFFIC CONDITIONS

Figure 1 shows the multiagent based interpolation system for traffic conditions.[6] This system consists of both estimation and learning agents that are allocated to all the road links. Estimation agents renew the velocity for each 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. In this study, the upper limit on the number of PCD values stored in the velocity/weight database is 1000 for each road link.

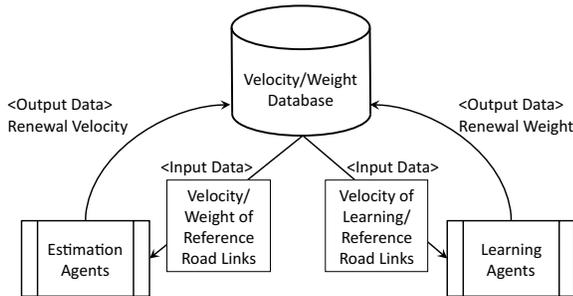


Figure 1: Multiagent Based Interpolation System for Traffic Conditions.

The normalized velocity (NV) is used in this system, defined as shown in Figure 2. The x -axis denotes the velocity, and the y -axis denotes the NV. This chart consists of three lines: (1) $y = 1 - x/1000$, (2) $y = 1.2 - x/50$, (3) $y = 0.1 - x/1000$.

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 increases 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 50km/h, velocity of road link C

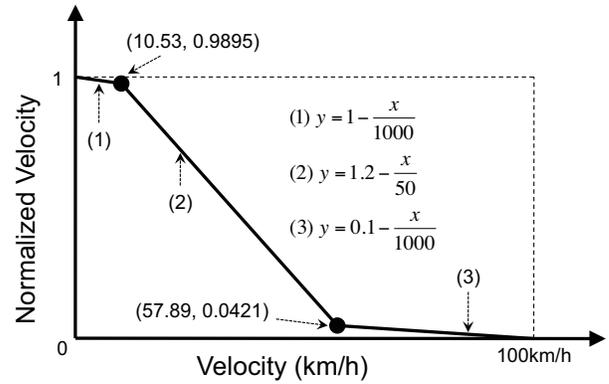


Figure 2: Definition of Normalized Velocity.

becomes 100km/h. In general way, the velocity decreases by junctions. The velocity of road link C decreases by simple addition of the NV of road link A and B. When the NV of road link A and B is 0.2 (= 50km/h), the NV of road link C becomes 0.4 (= 40km/h). The NV is suitable for this calculation because the slope of the line in Figure 2 is negative.

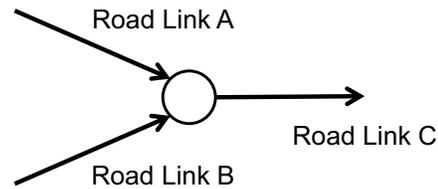


Figure 3: Example of Road Link Junction.

The slope in the mid velocity region is steeper than in the high and low velocity regions. In the mid velocity region, the interpolation accuracy is assumed to be better than in the other regions. The estimated NVs and the weight values are stored in the database, structured as shown in Figure 1.

2.1 Estimation Agents

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 ${}^t v^{(1)}, \dots, {}^t v^{(7)}$. For example, ${}^t v^{(1)}$ denotes the NV of road link 1 at time t .

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 $w_0^{(i)} = 0$, $w_1^{(i)}, \dots, w_n^{(i)} = 1/n$. The subscript 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 Figure 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 multiagent based

interpolation system cannot break the deadlock.

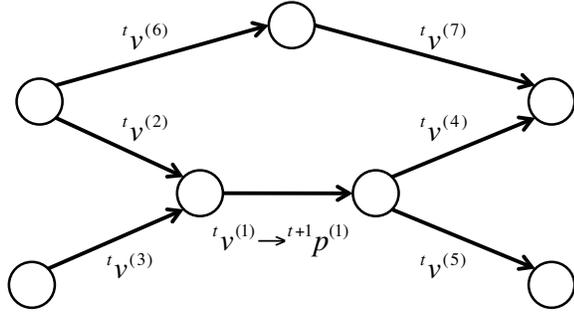


Figure 4: Example of Road Link Connections.

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

$${}^{t+1}\tilde{E}^{(i)} = \mathbf{V}^{(i)} \cdot {}^t\mathbf{w}^{(i)} \quad (1)$$

$$\mathbf{V}^{(i)} = \left(1 \quad V_1^{(i)} \quad \dots \quad V_n^{(i)} \right) \quad (2)$$

$${}^t\mathbf{w}^{(i)} = \left({}^t w_0^{(i)} \quad {}^t w_1^{(i)} \quad \dots \quad {}^t w_n^{(i)} \right)^T \quad (3)$$

When the NV for the PCD (or VICS data) at time $t+1$ is ${}^{t+1}p^{(1)}$, the component of the NV vector $\mathbf{V}^{(i)}$ is renewed. When both VICS data and PCD are assigned to the same road link, PCD overwrite the VICS data.

Estimation agents are allocated 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, ${}^{t+1}\tilde{E}^{(2)}$ can be calculated from Equation (4).

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

$$\mathbf{V}^{(2)} = \left(1 \quad {}^{t+1}p^{(1)} \quad {}^t v^{(3)} \quad {}^t v^{(6)} \right) \quad (5)$$

$${}^t\mathbf{w}^{(2)} = \left({}^t w_0^{(2)} \quad {}^t w_1^{(2)} \quad {}^t w_2^{(2)} \quad {}^t w_3^{(2)} \right)^T \quad (6)$$

Using the same method, ${}^{t+1}\tilde{E}^{(3)}$ can be calculated from Equation (7).

$${}^{t+1}\tilde{E}^{(3)} = \mathbf{V}^{(3)} \cdot {}^t\mathbf{w}^{(3)} \quad (7)$$

$$\mathbf{V}^{(3)} = \left(1 \quad {}^{t+1}p^{(1)} \quad {}^{t+1}\tilde{E}^{(2)} \right) \quad (8)$$

$${}^t\mathbf{w}^{(3)} = \left({}^t w_0^{(3)} \quad {}^t w_1^{(3)} \quad {}^t w_2^{(3)} \right)^T \quad (9)$$

Then, when the new probe data 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 $\mathbf{V}^{(i)}$ includes both the NV components at time t and at time $t+1$.

2.2 Learning Agents

Learning agents are allocated to all the road links, and they calculate the weight vector \mathbf{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. \mathbf{P} denotes the probe vector for the road link i with m NVs, and $\mathbf{V}_{m \times (n+1)}$ 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.

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

$$\mathbf{P} = \left(P_1 \quad P_2 \quad \dots \quad P_m \right)^T \quad (11)$$

$$\mathbf{V}_{m \times (n+1)} = \begin{pmatrix} 1 & V_{11} & \dots & V_{1n} \\ 1 & V_{21} & \dots & V_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & V_{m1} & \dots & V_{mn} \end{pmatrix} \quad (12)$$

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

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

$$\mathbf{E} = \left(E_1 \quad E_2 \quad \dots \quad E_m \right)^T \quad (14)$$

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

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

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

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

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

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.

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

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

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

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

$$\bar{P} = \frac{1}{m} \sum_{j=1}^m P_j \quad (22)$$

The weight value w_0 can be also calculated from Equation (23), which is transformed from Equation (16).

$$w_0 = \bar{P} - \sum_{j=1}^n w_j \cdot \bar{V}_j \quad (23)$$

3. EVALUATION OF INTERPOLATION

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 four months from 01/11/2007 to 29/02/2008. First, the PCD from the 1128 road links was checked (see Table 1). The subtotal for road links with single-figure PCD is 149. We should note that the number of PCD values m must be not less than the number of unknowns $n+1$ for learning, therefore it is difficult to calculate the weight value of a road link from single-figure PCD.

The subtotal for road links with four-figure PCD values is 179. In this range, the road links have collected a sufficient number of PCD values. The number of PCD values obtained per day was 96, and the evaluation period was 121

Table 1: Frequency Distribution of Number of PCD Values

Number of PCD Values	Count	Number of PCD Values	Count
1	31	10 To 19	82
2	31	20 To 29	56
3	18	30 To 39	39
4	13	40 To 49	31
5	23	50 To 59	24
6	12	60 To 69	28
7	5	70 To 79	17
8	8	80 To 89	19
9	8	90 To 99	16
Subtotal		Subtotal	
149		312	

Number of PCD Values	Count	Number of PCD Values	Count
100 To 199	118	1000 To 1999	89
200 To 299	104	2000 To 2999	31
300 To 399	53	3000 To 3999	23
400 To 499	48	4000 To 4999	13
500 To 599	31	5000 To 5999	6
600 To 699	17	6000 To 6999	5
700 To 799	22	7000 To 7999	3
800 To 899	15	8000 To 8999	6
900 To 999	14	9000 To 9999	3
Subtotal		Subtotal	
422		179	

Sum Total	1062	No PCD	66
-----------	------	--------	----

days, giving a maximum number of PCD values of 11,616. In the four-figure range, the NV for the road links could be estimated by traditional methods without using this interpolation system. At the end of the evaluation period (29/02/2008), the number of road links with PCD was 1062 at most. The rest of the 66 road links had no PCD. A suitable range of PCD for this interpolation system is considered to be from double figures to three figures.

3.1 Evaluation by CD and MSE

VICS is a well-known traffic information service. VICS gathers traffic information from roadside sensors and provides it to drivers. Probe cars measure the travel time along road links using GPS sensors and by other methods. It is unnecessary to install sensors at the roadside with the probe car system. In this study we compare the following two cases:

- (Case A) Using only PCD (Probe Only)
- (Case B) Using both VICS data and PCD (VICS+Probe)

The estimation order of the road link refers to the distance from the road link with VICS data or PCD. In Figure 4, the estimation order of road link 1 is 0, the order of road links 2, 3, 4, and 5 is 1, and the order of road links 6 and 7 is 2.

Table 2 shows the estimation order in this evaluation. The total number of road links was 1128. PCD for 10/11/2007 at 12:00 were used. When only PCD was used, the maximum estimation order was 8. However, the estimation order was

3 at most when both VICS data and PCD were used. In this case, more than 50% of road links had VICS data or PCD, and more than 40% of the road link had an estimation order of 1. Therefore, when VICS data and PCD were used together, almost all estimated NVs were calculated from the adjacent VICS data.

Table 2: Estimation Order

Order	Probe Only	VICS+Probe
0	65	578
1	251	471
2	265	68
3	211	11
4	117	0
5	107	0
6	65	0
7	29	0
8	18	0

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 (24). p and e denote the variance of the PCD and the estimated NV, respectively. R denotes the ratio of the two variances. While the notation $R^{(i)}$ is appropriate, the superscript i is omitted.

$$R = \frac{e}{p} \quad (24)$$

$$e = \sum_{j=1}^m \frac{(E_j - \bar{E})^2}{m-1} \quad (25)$$

$$p = \sum_{j=1}^m \frac{(P_j - \bar{P})^2}{m-1} \quad (26)$$

$$\bar{E} = \frac{1}{m} \sum_{j=1}^m E_j \quad (27)$$

3.2 Comparison between with and without VICS Data

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 Figure 5 denotes a local road.

When the number of PCD 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 PCD values increases, the CD decreases abruptly, and then increases gradually. The minimum value of the CD for road link 31002 (Probe Only) is approximately 0.75 after learning, and its value gradually increases

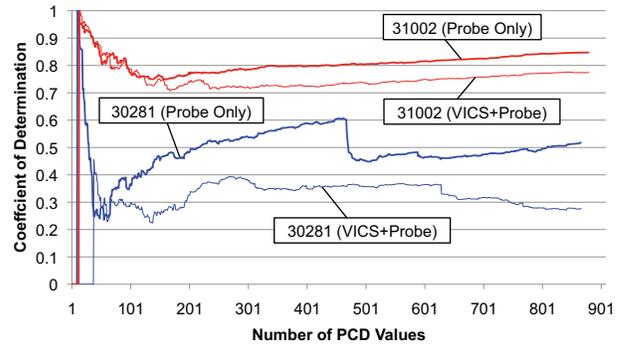


Figure 5: Fluctuation in the Coefficient of Determination.

to 0.85. The CD for road link 30281 (Probe Only) abruptly decreases when the number of PCD values is approximately 450, and increases gradually thereafter. The reason why the CD changes abruptly is thought to be that the weight value of the reference road link changes. The reason why the CD increases gradually is thought to be that the PCD values are concentrated in a hyperplane in the multivariate analysis.

From Figure 5, the CD for the Probe Only case is greater than the CD for the VICS+Probe case. Firstly, the PCD is taken as true even if the VICS data is true in reality. As the VICS data differs from the PCD, the VICS data is thought to disturb the calculation. Secondly, the VICS data disturbs the diffusion of PCD. For example, the PCD for road link 1 in Figure 4 never diffuses to the other road links when road links 2, 3, 4, and 5 have VICS data. The VICS data acts as some kind of bulwark against the PCD.

The weight values can be calculated by multivariate analysis, which minimizes the MSE, which is the essential parameter in 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 no PCD 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 PCD values increases, the MSE increases abruptly, and then decreases gradually. The trend in the MSE is opposite to the CD.

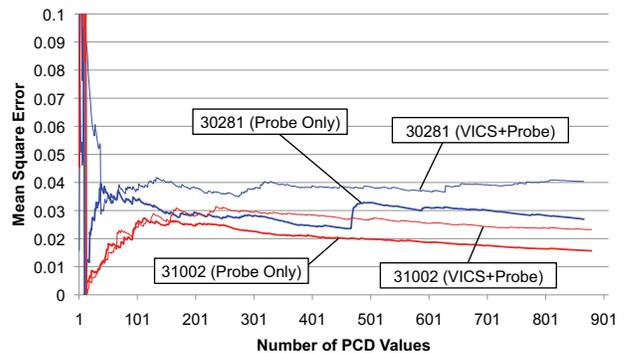


Figure 6: Fluctuation in the Mean Square Error.

Figure 7 shows the average CD and MSE for all the road links in the evaluation area. The x -axis denotes the elapsed time in days, and the y -axis denotes the average CD and MSE. R_{ave} is the average CD for all the road links in the evaluation area, which is given by Equation (28). N denotes the number of road links in the evaluation area. The average MSE Q_{ave} is given by Equation (29), and the notation $m^{(i)}$ is used in this equation. The evaluation period was from 01/11/2007 to 29/02/2008, and the same data was used twice. 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 on normal days. From these results, the CD and MSE can be improved when VICS data is not used.

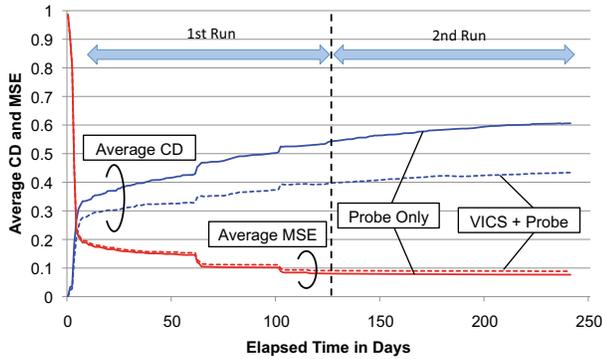


Figure 7: Average CD and MSE for All the Road Links.

$$R_{ave} = \frac{1}{N} \sum_{i=1}^N R^{(i)} \quad (28)$$

$$Q_{ave} = \frac{1}{N} \sum_{i=1}^N \frac{Q^{(i)}}{m^{(i)}} \quad (29)$$

3.3 Evaluation by Only PCD

Figure 8 shows fluctuations in the CD (Probe Only) for several road links in the evaluation area. The x -axis denotes the number of PCD values, and the y -axis denotes the CD. When learning finished (29/02/2008), the CD for road link 30255 had a minimum value, and the CD for road link 30544 had a maximum value. As the number of PCD values for road link 30255 is greater than for road link 30544, the CD does not depend on the number of PCD values.

Figure 9 shows fluctuations in the MSE (Probe Only). The x -axis denotes the number of PCD values, and the y -axis denotes the MSE. When learning had finished (29/02/2008), the MSE for road link 30544 had a minimum value, and the MSE for road link 30255 had a maximum value. The trend in the MSE is opposite to the CD.

To investigate the reason, the number of PCD values for the reference road links were measured, and these are shown in Table 3. For road link 30544, the number of reference road links was 13, and the number of PCD values greater than 200 was 12. For road link 30255, the number of reference road links was 13, and the number of PCD values less than 100 was 9.

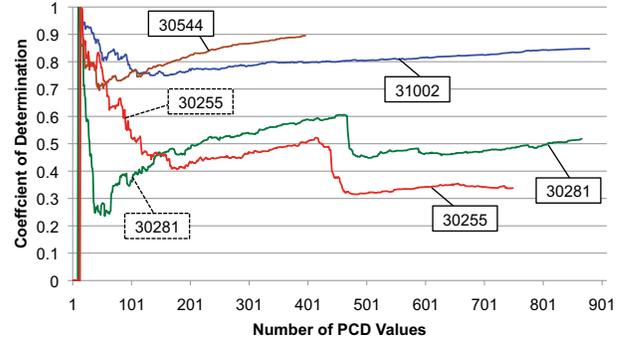


Figure 8: Fluctuation in the Coefficient of Determination.

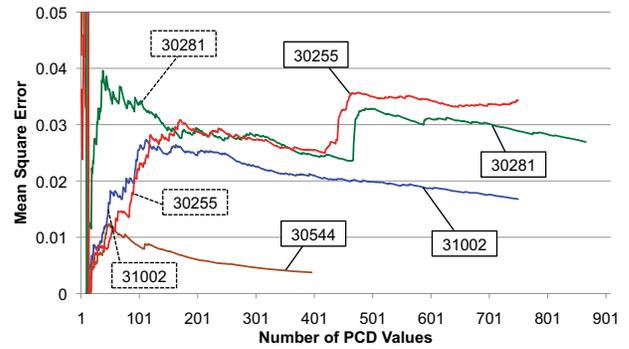


Figure 9: Fluctuation in the Mean Square Error.

Table 3: Number of PCD Values for the Reference Road Links

Road Link Number		Number of PCD Values	Road Link Number		Number of PCD Values
Learning	30544	396	Learning	30255	750
Reference	30543	671	Reference	30254	42
	30556	231		30259	757
	30941	250		31057	4
	30526	212		30163	139
	30528	1029		30251	14
	30545	457		31055	14
	30542	367		30253	46
	30555	147		30258	231
	30940	421		31058	18
	30541	302		30252	111
	30525	1488		30162	56
	30527	299		30250	11
	30540	449		31056	19

From these results, the CD is considered to depend on the number of PCD values for the reference road links. When the number of PCD values for the reference road links is few, one item of probe car data greatly changes the weight value for the reference road link. As the PCD values are plotted far from the hyperplane, the CD decreases abruptly. The slight fluctuation in the CD is thought to be for the same reason.

3.4 Repeated Use of Same PCD

When learning is in progress, the weight values of the reference road links are renewed. Even if the road link being learned gives the same PCD, estimation agents calculate the NVs for the reference road links with different weight values. Learning agents use different NVs in the multivariate analysis. Since the same PCD values are used twice, the number of PCD values increases substantially. When the number of PCD values is m and the same data is used twice, the actual number of PCD values becomes $2m$. The estimation errors during the first run are very large, and therefore iterative use of the same data is preferable.

Figure 10 shows fluctuations in the CD when the same PCD from 01/11/2007 to 29/02/2008 are used twice. The x -axis denotes the elapsed time in days, and the y -axis denotes the CD. The same PCD values are used for the 1st and 2nd runs, and the PCD values from 01/03/2008 to 30/06/2008 are used after that.

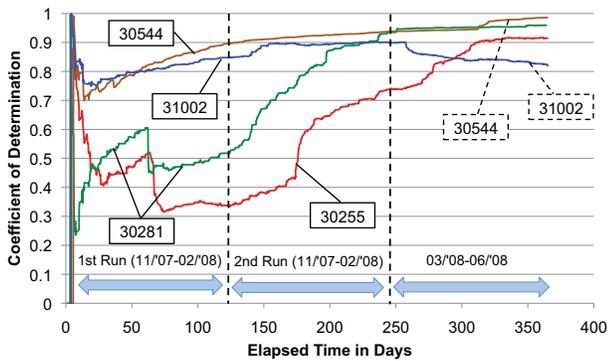


Figure 10: Fluctuation in the Coefficient of Determination.

Figure 10 indicates that the CD increases when the same PCD is used. The CD abruptly increases at several road links on the 2nd run, and this is considered to be because the number of PCD values exceeds 1000. The upper limit to the number of PCD values stored into the velocity/weight database is 1000 for each road link. When the number of PCD values exceeds 1000, the oldest item of probe car data is deleted from the database. As the old PCD, which is far from the hyperplane, is deleted, the CD can be improved. From these results, the CD can be effectively increased when the PCD is more than 1000.

The CD for road link 31002 decreases after the 2nd run. The reason is considered to be that the weight value of the reference road links changes abruptly. For road link 31002, the number of the referenced road link is 12, and the number of PCD values less than 100 is 4 after the 1st run. After the 2nd run, the weight value of the referenced road link changes

drastically, and therefore the CD decreases abruptly.

Figure 11 shows fluctuations in the MSE when the same PCD values are used. The x -axis denotes elapsed time in days, and the y -axis denotes the MSE. Figure 11 indicates that the MSE follows the same trend as the CD. The MSE of road link 31002 increases after the 2nd run, for the same reason as for the CD. The MSE of road link 30544 is 0.00048 at the end of the evaluation, and the standard deviation of the estimated NV error is 0.0219. The standard deviation of the estimated velocity error is 1.10 km/h in the mid velocity region as shown in Figure 2, because the slope in the mid velocity region is 1/50.

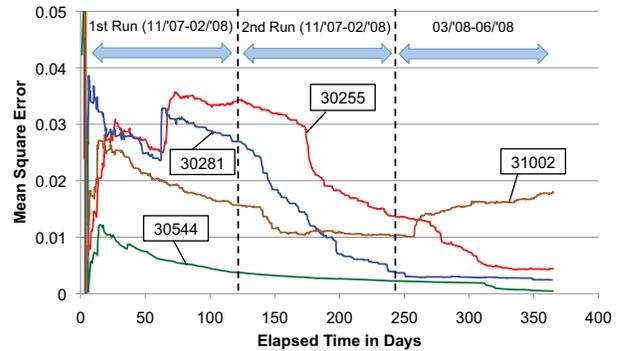


Figure 11: Fluctuation in the Mean Square Error.

Figure 12 shows the average CD and MSE for all the road links in the evaluation area. The x -axis denotes the elapsed time in days, and the y -axis denotes the average CD and MSE. The evaluation period is from 01/11/2007 to 30/06/2008, and the same PCD values are used twice. The average CD and MSE can be improved with the passage of time.

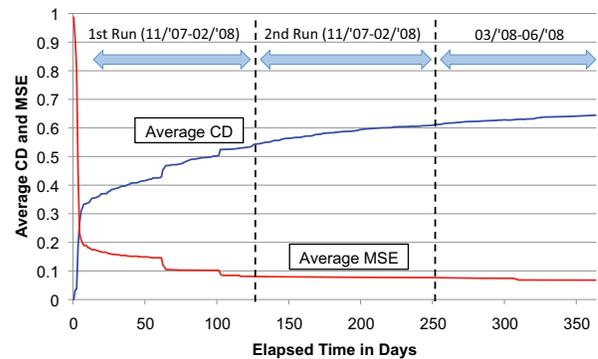


Figure 12: Average CD and MSE for All the Road Links.

The final value for the average MSE is 0.06774, 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 57 on 30/06/2008, and the total number of links N is 1128. Therefore, the actual average MSE in this area is 0.01721 ($= 0.06774 - 57/1128$), and the standard deviation of the

estimated NV error is 0.1312. The standard deviation of the estimated velocity error is 6.56 km/h in the mid velocity region shown in Figure 2, because the slope in the mid velocity region is 1/50. The standard deviation of the high and low velocity regions is 131.2 km/h. Therefore, this type of NV can be used for the mid velocity region only, not for the expressways. However, the average CD and MSE can be improved on by repeating the estimation/learning.

4. CONCLUSION

We propose a multiagent based interpolation system for traffic conditions. This system consists of two kinds of agents: an estimation agent and a learning agent, allocated to all the road links. The CD and the MSE are used to evaluate the interpolation accuracy. We compared the results for (A) Using only PCD (Probe Only), and (B) Using both VICS data and PCD (VICS+Probe). We confirmed that the interpolation accuracy without VICS data (Probe Only) is higher than with VICS data.

Under the traditional system, the interpolation accuracy of the road link depends on the number of PCD values for the road link itself. In this system, the interpolation accuracy depends on the number of PCD values for the adjacent (reference) road links. When the number of PCD values is low, the weight values of the reference road links are unstable. One item of probe data drastically changes the weight values of the reference road links, and the estimated NV for the reference road links also changes. Therefore the interpolation accuracy decreases when the number of PCD values for the referenced road link is low.

In addition, the interpolation accuracy can be improved when the same PCD values are used more than twice, because this increases the number of PCD values substantially. Removal of the old PCD values is a useful way to improve the interpolation accuracy, as well. The standard deviation of the estimated NV error can be improved to 0.1312, and the standard deviation of the estimated velocity error is 6.56 km/h in the mid velocity region. Furthermore, the CD and MSE can be improved by repeated estimation and learning.

5. REFERENCES

- [1] 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, pp. 73–80, 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.