Trans. of the Society of Instrument and Control Engineers
Vol.10, No.13, 113/122 (2011)

計測自動制御学会産業論文集
Vol.10, No.13, 113/122 (2011)

Estimation of Missing Traffic Volume by Genetic Algorithm Based on Traffic Flow Balance

Manoj KANTA MAINALI*, Shingo MABU* and Kotaro HIRASAWA*

Traffic information collected as a part of ITS has wide range of applications. However, the collected data might be incomplete due to the missing values caused by detector failures, communication problems or hardware/software related problems etc. To fully explore the potential of collected data, it is essential to fill the gaps caused by the missing values. In this paper, a method is proposed to estimate the missing traffic volume based on the traffic flow balance, i.e., the total incoming traffic volume is equal to the total outgoing traffic volume at intersections. The missing traffic volume is basically estimated using the traffic volume information of the neighboring road sections. But, since the traffic volumes of the neighboring road sections might be missing as well, genetic algorithm is applied to estimate multiple missing traffic volumes simultaneously. The proposed method is evaluated using the traffic volume information generated by a large scale traffic simulator SOUND/4U.

Key Words: Intelligent transportation systems, missing data, traffic volume estimation, genetic algorithm

1. Introduction

In order to alleviate the traffic congestion problem and facilitate the mobility in metropolises, a large amount of traffic information are collected as a part of intelligent transportation system (ITS) such as Vehicle Information and Communication System (VICS) [1] in Japan. Such information are collected via several kinds of traffic monitoring devices such as loop detectors set on the roads. These collected traffic data have wide range of applications. The real time traffic information is provided to travelers to support their decision making process on the optimal route choice [1]. As shown by the work of Kim et. al. [2], the real time information can contribute to reduce the operation cost and maximize resource utilization. In addition to these applications, the collected data could be applied to maximize the utilization of the infrastructure for smooth flow of the traffic. One such application of real time traffic data is the adaptive traffic signal control [3)~6)]. On the other hand, global routing strategies have been proposed by Yu et. al. [7)~9)] that utilize the traffic volume information to mitigate the traffic congestion. Since these applications make use of real-time traffic information, the archived traffic information become also valuable resources for their applications to be realized. The archived data could be used for several kinds of purposes such as network planning, congestion management, environmental analysis etc. [10),11)]. In addition, several researchers have applied data mining techniques to mine time related association rules from traffic databases and used its results for traffic prediction [12)~15)].

From the above, it is concluded that the collected traffic data, whether they are real-time data or archived data, are essential for many potential applications in ITS. However, the collected data might be incomplete due to the missing values caused by detector failures, communication problems or any other hardware/software related problems. The presence of missing data in the database would degrade the quality as well as reliability of the data and might impede the effectiveness of ITS applications. Therefore, it is essential to fill the gaps caused by missing data in order to fully explore the applicability of the data and realize the ITS applications.

While many different kinds of traffic data such as traffic volume, speed, occupancy etc. are collected, the focus of this research is on the missing traffic volume data. It is supposed that the detectors collecting traffic information are set up at road sections and the collected values represent the traffic volume for that road sections. The aim of this research is to fill the missing traffic volume data for road sections with their estimated values. Our approach is based on the simple concept of the traffic flow balance, i.e., the total incoming traffic volume is equal to the total outgoing traffic volume at intersections. More concretely, the feature of the proposed method is to estimate the missing data using the traffic volume information of the neighboring road sections. But, the problem is not so simple, because traffic volumes from many road sections might be missing simultaneously, in other words,

* Graduate School of Information, Production and Systems, Waseda University, Kitakyushu, Fukuoka 808-0135, Japan
(Received March 14, 2011)

the traffic volume estimation should be done considering the whole road networks, not intersection by intersection. In order to consider the multiple missing traffic volumes, Genetic Algorithm (GA) is applied for the overall estimation of the missing traffic volumes in the road networks.

The rest of the paper is organized as follows. Section 2 presents the related works and our contribution. Section 3 gives the definitions of notations used throughout the paper. In section 4, the proposed method to estimate the missing traffic volumes is described. Section 5 presents the performance evaluation of the proposed method. Finally, section 6 concludes the paper with the discussion on future works.

2. Related works and our contribution

Literature survey in the related field shows that several heuristic and statistical estimation techniques have been applied to estimate the missing traffic data [16)~18)]. Heuristic methods includes techniques such as historical average, average of surrounding time periods, average of surrounding detectors etc. Historical average simply takes the average of the past data of the same time of the day and the same day of the week as an estimate of the missing values. This approach assumes that the traffic flows through the same location are similar from day to day. However, the traffic flow through the same location may vary significantly from day to day and this approach cannot consider such variations. In the average technique of the surrounding time periods, the average of the data before and after the missing data in the same location is used to fill the missing values. But, the application of this method is limited to the cases where there is no consecutive missing data. While the above approaches considers only the data from the same location, the average of the surrounding detectors takes the relationship between the data from the adjacent detectors into account by replacing the missing value with the average of the adjacent detectors. This method assumes that the data from the adjacent detectors are available, and cannot be applied when the data in the adjacent detectors are missing too.

Statistical approaches are complex processes compared to the heuristic approaches and include techniques such as regression, expectation maximization, data augmentation etc [16)]. Depending on the statistical process and definition, they have different requirements for input data. Several works have applied and evaluated these methods to estimate the missing values in the transportation domain under different circumstances [16)~20)].

Other techniques such as Artificial Neural Networks have also been proposed by some researchers [21)~23)]. Ni et. al. [24)] has proposed a multiple imputation scheme to impute the missing values. Basically, this approach runs the expectation maximization/data augmentation approach multiple times and combines the results to replace the missing data.

Different techniques to estimate the missing data have different data requirements and their performance depends on the input data. Some researchers suggest using several estimation techniques based on some strategies instead of considering only one technique to estimate the missing values under all conditions [19), 22)].

Most of the previous studies focus on estimating the missing data for highways. In this paper, we do not focus on the highways, but consider estimating the missing traffic volumes for common road networks. Therefore, a new method is proposed in this paper to estimate the missing values based on the traffic flow balance, i.e., the total incoming volume is equal to the total outgoing volume at intersections. The proposed approach makes use of the traffic volume information from the neighboring road sections to estimate the missing traffic volume. However, there is no guarantee that the traffic volumes of the neighboring road sections are always available. To overcome this problem, the proposed method estimates a large number of missing traffic volumes in the road networks simultaneously using GA.

The unique features of the proposed method are,

- Estimate the missing values using only the traffic volumes of the neighboring road sections
- Estimate multiple missing values simultaneously using GA
- Since multiple missing traffic volumes are estimated at a time, the proposed approach can easily handle the situations where the traffic volumes of the neighboring road sections are missing as well

However, unlike the previous approaches applicable to estimate other missing data such as speed, occupancy etc., the proposed method is limited to the estimation of the traffic volumes.

3. Notations and Definitions

In this section we give the definitions of the notations used throughout the paper.

I : set of suffixes of intersections
S : set of suffixes of road sections
$i \in I$: intersection i

s_{ij} : road section from intersection $i \in I$ to intersection $j \in I$

$A(i)$: set of suffixes of intersections from intersection $i \in I$

$B(i)$: set of suffixes of intersections to intersection $i \in I$

v_{ij} : traffic volume of road section $s_{ij} \in S$

$\hat{v}_{ij}$: estimated traffic volume of road section $s_{ij} \in S$

4. Traffic volume estimation

Traffic volume gives the measurement of the number of cars that pass through a certain point in a given period of time. The traffic volume of the road section may be different depending on where it is measured. Under the assumption of the static traffic, for any intersection i we can consider the following equality and inequality equations in order to estimate the traffic volumes of the road networks (see **Fig. 1**),

$$\sum_{j \in B(i)} v_{ji} = \sum_{j \in A(i)} v_{ij} \qquad (1)$$
$$\forall i \in I$$

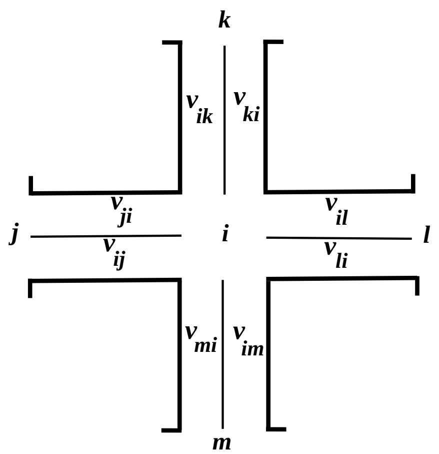


Fig. 1: Traffic volumes at intersection i

Eq. (1) defines the basic traffic flow balance, i.e., the total incoming traffic volume is equal to the total outgoing traffic volume at intersections. In addition, the following two inequality equations can also be considered,

$$v_{ji} \leq \sum_{k \in A(i)-\{j\}} v_{ik} \qquad (2)$$
$$\forall i \in I, \forall j \in B(i)$$

$$v_{ij} \leq \sum_{k \in B(i)-\{j\}} v_{ki} \qquad (3)$$
$$\forall i \in I, \forall j \in A(i)$$

Eq. (2) implies that the incoming volume to intersection i from intersection j should be less than or equal to the sum of the outgoing volume from intersection i except for intersection j. Similarly, Eq. (3) implies that the outgoing volume from intersection i to intersection j should be less than or equal to the sum of incoming volume to intersection i except for intersection j. While in practical cases these two inequality equations might be easily satisfied because the summation of the right hand of the equations include the traffic volumes coming from other intersections except intersection j or traffic volumes going to other intersections except intersection j.

For the traffic volume balance, the total incoming volume and total outgoing volume is only the information required. Therefore, the balance equation can be applied to any structure such as T-junction of intersections. Additionally, the restrictions such as one-way traffic exist in road networks. However, the proposed method can be applied to such situations by just restricting the direction, which has no influence on the traffic flow balance at intersections. In real road sections with multiple lanes, detectors are set for each lane [16] and the set of these detectors is called 'station' [16]. This paper does not consider the traffic volume for each lane, but only considers the traffic volume for the road section, in other words, the traffic volume from the station, which is obtained by summing up the traffic volumes of the detectors of each lane.

It is clear from Eq. (1) that the proposed method makes the estimation using the traffic volume information of the neighboring road sections. However, the traffic volume in many road sections might be missing and there is no guarantee that the traffic volumes of the neighboring road sections are always available. Therefore, to overcome this problem, GA is used to estimate multiple missing traffic volumes at a time. The details of the GA procedure to estimate the missing traffic volumes are described below. In the description, we intentionally skip the description on the general flow of GA as readers could refer to any literature in the evolutionary computation for the basic algorithm of GA [25), 26)].

4.1 Chromosome encoding

GA is a population based heuristic search method in which solutions to the problems are encoded as the chromosome of individuals. Usually, binary encoding is the most commonly used for the chromosome encoding. However, other encoding is also possible, and in this paper, we use the real number encoding. Suppose the number of missing traffic volumes to estimate is n. Then, an array of size n is used as the chromosome of individuals, where each gene represents an estimate of the missing traffic volume for a particular road section.

4.2 Initialization

An initial population of individuals is generated randomly. When using a binary encoding, the values for the genes are randomly initialized with 0 or 1. However, in this paper, a real value is used for encoding. Therefore, the lower bound and the upper bound are defined to randomly initialize the population. The lower bound is considered to be 0 and the upper bound is decided using the traffic volumes of the neighboring road sections. Let's assume that the gene represents the estimate of traffic volume $\hat{v}_{ij}$ for road section s_{ij}, then the upper bound $max\{S,E\}$ and initialized $\hat{v}_{ij}$ are calculated as follows,

$$S = \sum_{k\in B(i)} v_{ki} \ ,$$
$$E = \sum_{k\in A(j)} v_{jk} \ ,$$
$$\hat{v}_{ij} = random(0, max\{S,E\})$$

In the above upper bound calculation, if the traffic volume of the road section is missing, then the value represented by the individual is used for that road section if it is already initialized, otherwise, a value of 0 is used. However, when many values are missing, the initialization range is different. Let us consider an extreme case in which the traffic volumes of all the neighboring road sections of road section s_{ij} are missing. If the estimate of the neighboring road sections are not initialized yet, then the upper bound for s_{ij} becomes 0, and the same for the lower bound. As the previous example shows, when there are many missing values, the above upper bound could limit the initial values, which could also limit the search ability of GA. Because of this, only some of the individuals are initialized using the above upper bound (INIT_I) and the rest of the individuals are initialized using the following upper bound (INIT_II),

$$\hat{v}_{ij} = random(0, max\{S,E\} + \theta)$$
θ : parameter

An illustrative example of the above procedure is shown below. Let us assume that we are initializing the traffic volume of road section s_{ij} in **Fig. 2**. In road section s_{ij}, the traffic enters to the start intersection i and leaves from the end intersection j. The maximum traffic volume that passes through road section s_{ij} is the total incoming traffic volume (S) at intersection i or the total outgoing traffic volume (E) at intersection j. Thus, we define the upper bound by taking the maximum of S and E.

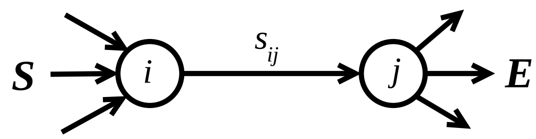


Fig. 2: Example of initializing traffic volume for road section s_{ij}

4.3 Fitness evaluation

The fitness function gives how the individual is fitted for a given problem and is used to guide the evolution process. The fitness function of individuals is defined as follow in this paper considering Eq. (1), Eq. (2) and Eq. (3).

$$F = F_1 + F_2 + F_3 + F_4$$

$$F_1 = \sum_{i\in I}\left(\sum_{j\in B(i)} v_{ji} - \sum_{j\in A(i)} v_{ij}\right)^2 ,$$

$$F_2 = \sum_{i\in I}\sum_{j\in B(i)} \begin{cases} (v_{ji} - \sum_{k\in A(i)-\{j\}} v_{ik})^2, \\ \qquad \text{if } v_{ji} > \sum_{k\in A(i)-\{j\}} v_{ik} \\ 0, \ \text{otherwise} \end{cases}$$

$$F_3 = \sum_{i\in I}\sum_{j\in A(i)} \begin{cases} (v_{ij} - \sum_{k\in B(i)-\{j\}} v_{ki})^2, \\ \qquad \text{if } v_{ij} > \sum_{k\in B(i)-\{j\}} v_{ki} \\ 0, \ \text{otherwise} \end{cases}$$

$$F_4 = \sum_{s_{ij}\in X}\left(\hat{v}_{ij} - \frac{1}{|B(i)-\{j\}|}\sum_{k\in B(i)-\{j\}} v_{ki}\right)^2 + \left(\hat{v}_{ij} - \frac{1}{|A(j)-\{i\}|}\sum_{k\in A(j)-\{i\}} v_{jk}\right)^2$$

X : Set of road sections whose traffic volumes are missing in both directions

Here, F_1, F_2 and F_3 evaluates the satisfaction of the Eqs. (1)(2)(3), respectively. In addition to these evaluations, one more evaluation F_4 is added, which is only defined for the special sections, where the traffic volumes of both directions in the sections are unknown. The fourth evaluation defines that the estimated traffic volume should be equal to the average of the incoming volumes except from the opposite direction at the start intersection and also should be equal to the average of the outgoing volumes except to the opposite direction at the end intersection. For example, let us consider **Fig. 3** in which the traffic volume v_{ij} and v_{ji} for road sections s_{ij} and s_{ji} are missing, respectively. These two road sections are opposite to each other, i.e., the outgoing and incoming road sections for intersection i, vice-versa for intersection j. In the traffic flow balance equation, these two values could easily offset each other as they are opposite in the direction. For example, lets consider the traffic flow balance at intersection i for which v_{ji} is incoming volume and v_{ij} is outgoing volume. In the traffic flow balance condition of Eq. (1), if the incoming traffic volume v_{ji} is subtracted by the outgoing volume v_{ij} as shown in Fig. 3, then any values of v_{ji} and v_{ij} could easily satisfy the balance condition, although Eq. (2) and Eq. (3) defines the upper bound. And, the above is also applied when considering the traffic flow balance for intersection j. In order to determine the estimate of the traffic volumes uniquely under such a condition, the evaluation criteria F_4 is added in the above fitness function.

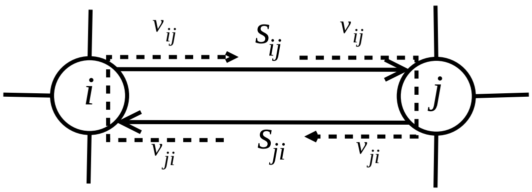


Fig. 3: Example of both directions traffic volume missing

In the fitness evaluation, F1 is the main evaluation factor and other evaluations are added as constraints. F2 and F3 are added only if the described conditions are not satisfied. Similarly, F4 is considered when the traffic volume of both directions in road sections is missing. We didn't consider any weight because F2, F3 and F4 are only added as constraints and if the constraints are satisfied, then these terms do not contribute to the fitness calculation, i.e., their values are 0. Our preliminary test of using weight parameters of F2, F3 and F4 didn't show any significant difference in the result, so we didn't consider any weight parameters.

Here, it should be noted that during the fitness evaluation, the estimate of missing values from the individuals is used for the missing traffic volumes. Also, the fitness is calculated using only the intersections that are associated with the road sections with missing traffic volumes.

GA is a heuristic search method which evolves to find optimal solutions. There is no guarantee that it always converge to the unique global optimal solution, but the obtained results in different trials are almost similar. Anyway, if there are many solutions which have the same fitness value, then one of them is adopted as the optimal solution.

4.4 Genetic operations

At each generation, genetic operators - selection, crossover and mutation - are applied to the current population to generate a new population.

4.4.1 Selection

In the selection operation, individuals are selected probabilistically based on their fitness values. While there are many different types of selection operations, we use elite selection, in which the elite individual is copied to the next generation, and tournament selection is also used.

4.4.2 Crossover

Crossover exchanges the gene information of two parents to create two offspring. The procedure of crossover operation is shown in Algorithm 1.

Algorithm 1 Crossover operation

```
Select two parents p_g and p_h, g ≠ h using tournament selection
Generate a random number x in (0,1)
if x < p_c , where p_c is crossover probability then
  for all i = 1 to n, where n is the length of chromosome do
    Generate a random number y in (0,1)
    if y < 0.5 then
      Exchange gene at i^th position of two parents
    end if
  end for
else
  Copy parents p_g and p_h to the next generation
end if
```

4.4.3 Mutation

Mutation operator modifies the individual by randomly changing the gene information based on mutation probability of P_m. Two types of mutation operations are used. The first one (MUT_I) randomly changes the gene information as shown in Algorithm 2.

Algorithm 2 Mutation operation

```
for all i = 1 to n do
  Generate a random number x in (0,1)
  if x < p_m then
    Calculate the following for road section s_ij that corresponds to i^th gene
    S = Σ_{k∈B(i)} v_ki
    E = Σ_{k∈A(j)} v_jk
    y = random(0, max{S, E})
    Replace the value of i^th gene with y
  end if
end for
```

In the second type of mutation (MUT_II), a part of the current value is added or subtracted with equal probability. The procedure is similar to the MUT_I, except the step of changing the value of i^{th} gene is replaced with the following,

$$\hat{v}_{ij} = \hat{v}_{ij} \pm \alpha \times \hat{v}_{ij}$$

α : parameter

When the GA procedure is finished, the values represented by the elite individual are regarded as the estimate of the missing traffic volumes for road sections.

5. Simulations

5.1 Study data and settings

The performance of the proposed method is evaluated using the traffic volume data generated by traffic simulator SOUND/4U [27]. Simulations run using 7941 major road sections around Kurosaki of Kitakyushu city (**Fig. 4**) by setting 20 OD and 100 OD pairs. The simulations run for 2 hrs and the traffic volume information for the road sections is collected every minute. The cars are generated randomly for the OD pairs and OD traffic volume is changed like the real traffics. An example of OD traffics is shown in **Fig. 5**.

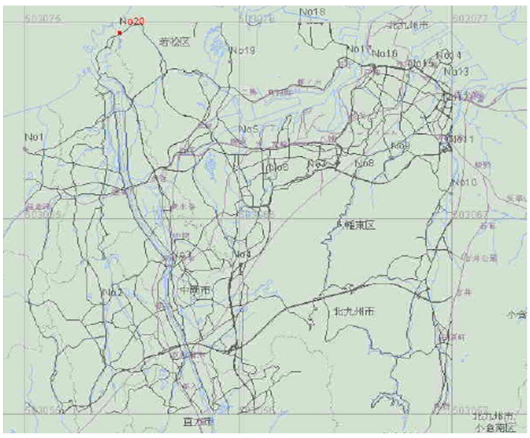


Fig. 4: Kitakyushu road network used in simulations

The traffic volume information was obtained at a randomly chosen time because only the static traffic volume estimation is studied in this paper, where we only selected the 2074 road sections with non-zero traffic volume. The traffic volume missing roads were selected randomly with the rates of 1, 2, 3, 5, 10 and 20% from these road sections. For the selected road sections, the traffic volume estimation is done using the proposed method. The proposed method is used to estimate the traffic volumes collected for 1 minute, 5 minutes, 10 minutes and 20 minutes. The traffic volume data with 1 minute is the one generated by the simulator. The traffic volume data for the remaining time periods is generated by aggregating the traffic volume data with 1 minute for the respective time period. In other words, we can regard the traffic volume collected for a small time period as a low traffic volume and that for a large time period as a high traffic volume.

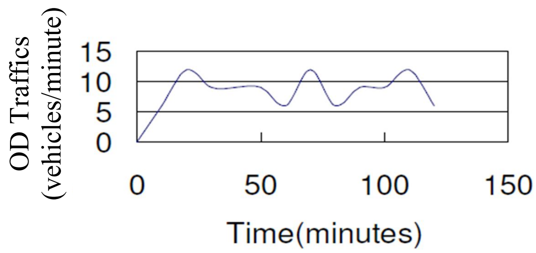


Fig. 5: Example of OD traffics for the OD pairs

Table 1: GA parameter settings

Parameter	Value
Population size	301
INIT_I size	150
INIT_II size	151
θ	10
Elite selection	1
Tournament size	4
Crossover rate	0.7
MUT_I rate	0.001
MUT_II rate	0.001
α	0.5
Maximum generation	2000

Simulations were carried out using GA under the parameter settings as shown in **Table 1**. We considered a population size of 301, out of which 150 individuals are initialized based on INIT_I initialization and the rest of the individuals are initialized using INIT_II. In the selection, elite selection and tournament selection of size 4 are applied. All evaluations are done by the average over ten runs of GA.

5.2 Evaluation indices

To evaluate the performance of the proposed method we use the following indices,

1. Mean Percentage Error (MPE) : This index gives the evaluation of the average estimation error in terms of percentage.

$$MPE = \frac{1}{n} \sum \frac{|v_{ij} - \hat{v}_{ij}|}{v_{ij}} \times 100$$

n : number of estimated traffic volumes

2. Mean Absolute Error (MAE) : This index gives the

evaluation of the average estimation error in terms of the magnitude, which is described in a unit of traffic volumes.

$$MAE = \frac{1}{n}\sum |v_{ij} - \hat{v}_{ij}|$$

3. Root Mean Square Error (RMSE) : This index also gives the evaluation of the average estimation error in terms of magnitude, but in this case more weight is given to large errors because the square of the error is considered. Therefore, this index can give the evaluation of the variance in the estimation errors.

$$RMSE = \sqrt{\frac{1}{n}\sum (v_{ij} - \hat{v}_{ij})^2}$$

5. 3 Simulation results

The evaluation of the proposed method to estimate the missing traffic volumes for different time periods is shown in **Fig. 6**–**9**. In the traffic volume with 1 minute (Fig. 6), the average of the percentage error of all the estimations were approximately within the range of 15% to 30% for different unknown rates. This means that, while the percentage error for each estimation may differ, on average, the proposed method estimates the missing traffic volumes within 15 $\sim$ 30% of actual values. The average of the percentage error of all the estimations in the traffic volume with 5 minutes (Fig. 7) were approximately between 4% to 6%. Similarly, in the traffic volume with 10 minutes and traffic volume with 20 minutes (Fig. 8 & Fig. 9), the average of the percentage error of all the estimations was approximately between 1.5% $\sim$ 3.5% and 0.5% $\sim$ 2%, respectively. We can see from these results that as the time period increases, the estimation error gradually decreases. It is because the traffic volumes in the longer time period tend to be filtered and become less noisy than the short time period. In addition, results show that the MAE values are very small in all the cases, which indicates that the estimated traffic volumes are very close to the actual traffic volumes on average. The RMSE evaluations show that the RMSE values are small and close to the MAE, which implies that the variance in the errors of estimated traffic volumes is small. The results of the above evaluation indices show that the proposed method gives the estimates of the missing traffic volumes very closely to the actual values on average. An example of the comparison of the actual and estimated traffic volumes is shown in **Fig. 10**, which shows the comparison of the traffic volumes of missing sections in the case of 1 minute and 20 minutes of 1% unknown rate. From Fig. 10 it is clear that the estimated traffic volumes are close to the actual traffic volumes.

In addition, we also evaluated the time required to estimate the missing traffic volumes by the proposed method. We measured the time required to estimate the missing traffic volumes with 20 minutes for the unknown rates of 1% and 20%. It approximately took about 20 (s) for 1% unknown rate and 180 (s) for 20% unknown rate when using a machine with Intel Core i5 2.53 GHz processor. It is obvious that the higher unknown rate requires high computational time because the problem size becomes big as the number of unknown traffic volumes is high.

5. 4 Discussion

The simulation results show that the proposed method can estimate the missing traffic volume data very closely to the actual values on average. The motivation behind estimating the missing traffic volume data is to supplement the ITS applications. Our perspective is that the results obtained can be used in ITS applications from the performance of the proposed method. However, the precision required for the practical use depends on the type of the applications and also on the policy the system adopts. In regard to the application of the estimated data to ITS, further research could be done to study the performance of the ITS applications quantitatively when using the estimated traffic volume data for missing data. However, such study is out of the scope of this paper and we leave it open for future work.

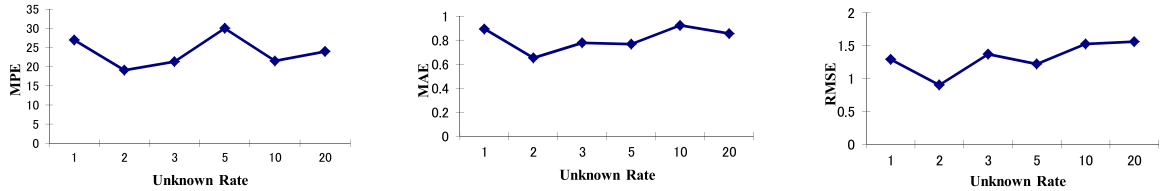
6. Conclusions

This paper has proposed a missing traffic volume estimation technique based on the traffic volume balance. It should be noted that Genetic Algorithm was applied to estimate the multiple missing traffic volumes considering the overall estimation of the missing data in the road networks. The proposed method was evaluated using the traffic volume information collected by a large scale road simulator SOUND/4U for different time periods with various unknown rates. It was clarified from the simulation results that the proposed method can estimate the missing traffic volume information very closely to the actual values. As a part of future work, further improvements of the proposed method could be studied for dynamic situations.

References

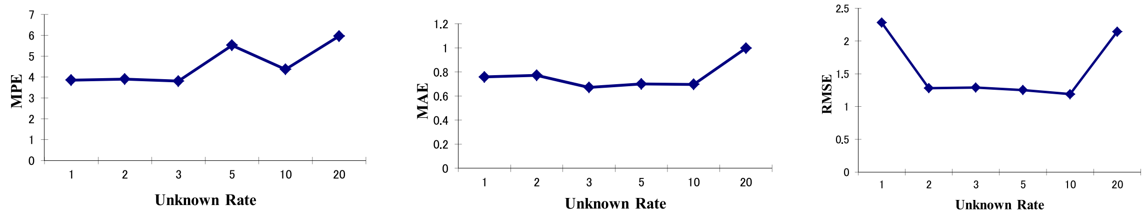
1) K. Tamura and M. Hirayama: Toward Realization of VICS - Vehicle Information and Communications System, *In Proc. of the IEEE Vehicle Navigation and Information Systems Conference*, 72/77, Oct. (1993)

2) S. Kim, M. E. Lewis and C. C. White, III: Optimal Vehicle Routing With Real-Time Traffic Information, *IEEE Transactions on Intelligent Transportation Systems*, **6**-2,



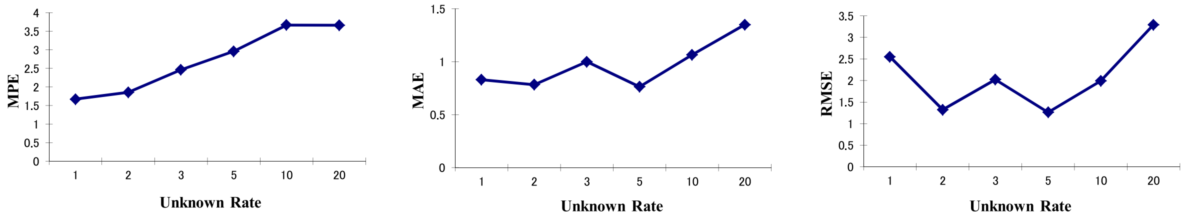
(a) MPE (b) MAE (c) RMSE

Fig. 6: Results in the traffic volumes with 1 minute



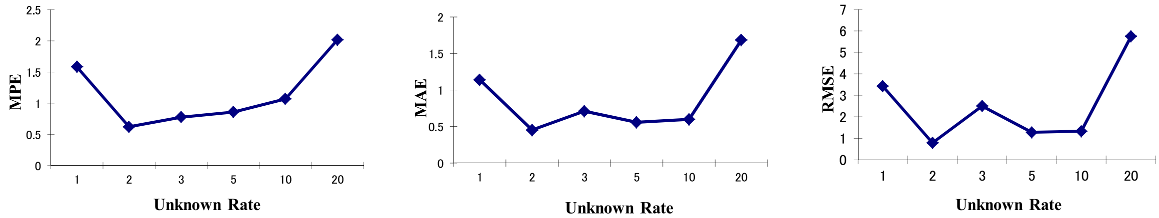
(a) MPE (b) MAE (c) RMSE

Fig. 7: Results in the traffic volumes with 5 minutes



(a) MPE (b) MAE (c) RMSE

Fig. 8: Results in the traffic volumes with 10 minutes

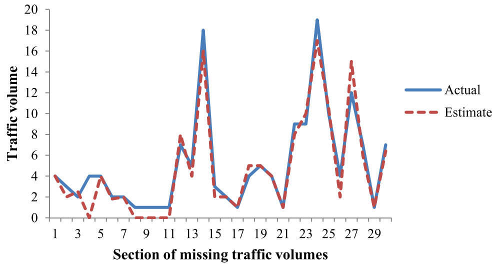


(a) MPE (b) MAE (c) RMSE

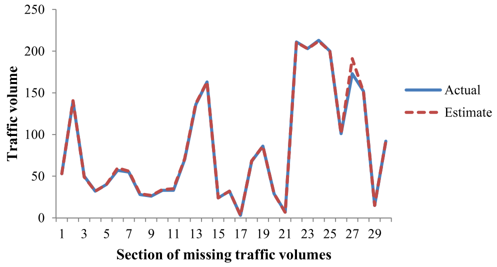
Fig. 9: Results in the traffic volumes with 20 minutes

178/188, June (2005)

3) S. Chiu: Adaptive Traffic Signal Control Using Fuzzy Logic, *In Proc. of the Intelligent Vehicles Symposium*, 98/107, July (1992)
4) P. Mirchandani and L. Head: A Real-time Traffic Signal Control System : Architecture, Algorithms, and Analysis, *Transportation Research Part C : Emerging Technologies*, **9**-6, 415/432, Dec. (2001)
5) J. Lee, B. Abdulhai, A. Shalaby and E.-H. Chung: Real-Time Optimization for Adaptive Traffic Signal Control Using Genetic Algorithms, *Journal of Intelligent Transportation Systems*, **9**-3, 111/122, July (2005)
6) C. Cai, C. K. Wong and B. G. Heydecker: Adaptive Traffic Signal Control Using Approximate Dynamic Programming, *Transportation Research Part C : Emerging Technologies*, **17**-5, 456/474, Oct. (2009)
7) S. Yu, S. Mabu, F. Ye, H. Wang, K. Shimada and K. Hirasawa: Q Value-Based Dynamic Programming with



(a) Result of the traffic volumes with 1 minute



(b) Result of the traffic volumes with 20 minutes

Fig. 10: Comparison of actual and estimated traffic volumes under 1% unknown rate

Boltzmann Distribution for Global Optimal Traffic Routing Strategy, *Journal of Advanced Computational Intelligence and Intelligent Informatics*, **13**-5, 581/591 (2009)

8) S. Yu, S. Mabu, M. K. Mainali, S. Eto, K. Shimada and K. Hirasawa: Multi-Routes Algorithm using Temperature Control of Boltzmann Distribution in Q value-based Dynamic Programming, *In Proc. of the IEEE International Conference on Systems, Man, and Cybernetics*, 3989/3994, Oct. (2009)

9) S. Yu, S. Mabu, M. K. Mainali, K. Shimada and K. Hirasawa: Various Temperature Parameter Control Methods in Q value-based Dynamic Programming with Boltzmann Distribution, *In Proc. of IEEE International Conference on Systems, Man and Cybernetics*, 2677/2683, Oct. (2010)

10) A. Stathopoulos and T. Tsekeris: Methodology for Processing Archived ITS data for reliability analysis in urban networks, *In Proc. of IEE Intelligent Transportation Systems*, **153**-1, 105/112, March (2006)

11) Y. Chen, J. Hu, Y. Zhang and X. Li: Traffic Data Analysis Using Kernel PCA and Self-Organizing Map, *In Proc. of Intelligent Vehicles Symposium*, 472/477, June (2006)

12) H. Zhou, W. Wei, K. Shimada, S. Mabu and K. Hirasawa: Time Related Association Rules Mining with Attributes Accumulation Mechanism and its Application to Traffic Prediction, *Journal of Advanced Computational Intelligence and Intelligent Informatics*, **12**-4, 393/403 (2008)

13) H. Zhou, S. Mabu, K. Shimada and K. Hirasawa: Backward Time Related Association Rule Mining with Database Rearrangement in Traffic Volume Prediction, *In Proc. of the IEEE International Conference on Systems, Man and Cybernetics*, 1047/1052, Oct. (2009)

14) H. Zhou, S. Mabu, K. Shimada and K. Hirasawa: Time Related Association Rule Mining with Accuracy Validation in Traffic Volume Prediction with Large Scale Simulator, *In Proc. of the IEEE International Conference on Systems, Man and Cybernetics*, 2249/2255, Oct. (2010)

15) X. Li, S. Mabu, H. Zhou, K. Shimada and K. Hirasawa: Genetic Network Programming with Estimation of Distribution Algorithms for Class Association Rule Mining in Traffic Prediction, *Journal of Advanced Computational Intelligence and Intelligent Informatics*, **14**-5, 497/509 (2010)

16) B. L. Smith, W. T. Scherer and J. H. Conklin: Exploring Imputation Techniques for Missing Data in Transportation Management Systems, *Transportation Research Record : Journal of the Transportation Research Board*, **1836**, 132/142 (2003)

17) J. H. Conklin and W. T. Scherer: Data Imputation Strategies for Transportation Management Systems, *A Research Project Report for the Virginia Department of Transportation (VDOT)*, Report No. UVACTS-13-0-80 (2003)

18) L. N. Ngyuyen and W. T. Scherer: Imputation Techniques to Account for Missing Data in Support of Intelligent Transportation Systems Applications, *A Research Project Report for the Virginia Department of Transportation (VDOT)*, Report No. UVACTS-13-0-78 (2003)

19) D. Maier, K. Tufte and R. J. Fernandez-Moctezuma: Improving Travel Information Products Via Robust Estimation Techniques, *Technical Report No. OTREC-RR-09-02*, Oregon Transportation Research and Education Consortium, March (2009)

20) L. Qu, Y. Zhang and J. Hu: PPCA-Based Mising Data Imputation for Traffic Flow Volume: A Systematical Approach, *IEEE Transactions on Intelligent Transportation Systems*, **10**-3, 512/522, Sept. (2009)

21) M. Zhong, S. Sharma and P. Lingras: Genetically Designed Models for Accurate Imputations of Missing Traffic Counts, *Transportation Research Record : Journal of the Transportation Research Board*, **1879**, 71/79 (2004)

22) M. Chen, J. Xia and R. Liu: Developing a Strategy for Imputing Missing Traffic Volume Data, *Journal of the Transportation Research Forum*, **45**-3, 57/75 (2006)

23) R. J. Fernandez-Moctezuma, K. A. Tufte, D. Maier and R. L. Bertini: Toward Management and Imputation of Unavailable Data in Online Advanced Traveler Information Systems, *In Proc. of IEEE Intelligent Transportation Systems Conference*, 24/29, Oct. (2007)

24) D. Ni, J. D. Leonard II, A. Guin and C. Feng: Multiple Imputation Scheme for Overcoming the Missing Values and Variability Issues in ITS Data, *Journal of Transportation Engineering*, **131**-12, 931/938, Dec. (2005)

25) J. H. Holland: Adaptation in Natural and Artificial Systems, University of Michigan Press, Ann Harbor (1975)

26) D. E. Goldberg: Genetic Algorithms in Search, Optimization, and Machine Learning, Addison Wesley (1989)

27) SOUND/4U, i-Transport Lab. Co. Ltd., http://www.i-transportlab.jp/products/sound/index.html.

Manoj Kanta mainali

He received his M.E. degree from the Graduate School of Information, Production and Systems, Waseda University, Japan, in 2009. Since 2009 he is a Ph.D. candidate at the Graduate School of Information, Production and Systems, Waseda University.

Shingo Mabu (Member)

He received the B.E. and M.E. degrees in Electrical Engineering from Kyushu University, Japan in 2001 and 2003, respectively and Ph.D. degree from Waseda University, Japan in 2006. Since April 2007, he has been an Assistant Professor in the Graduate School of Information, Production and Systems of Waseda University. Dr. Mabu is a member of the Society of Instrument and Control Engineers and IEEE.

Kotaro Hirasawa (Member)

He received the B.E. and M.E. degrees from Kyushu University, Japan in 1964 and 1966, respectively. From 1966 to 1992, he was with Hitachi Ltd., where he served as a Vice President of Hitachi Research Laboratory. From December 1992 to August 2002, he was a Professor in the Graduate School of Information Science and Electrical Engineering of Kyushu University. Since September 2002, he has been a Professor in the Graduate School of Information, Production and Systems of Waseda University. Dr. Hirasawa is a member of the Society of Instrument and Control Engineers and IEEE.