December 4, 2009
RT0884
Computer Science 6 pages

Research Report

Estimating the amount of CO2 emission from probe-car data

Takayuki Osogami

IBM Research - Tokyo
IBM Japan, Ltd.
1623-14 Shimotsuruma, Yamato
Kanagawa 242-8502, Japan

Estimating the amount of CO2 emission from probe-car data

Takayuki Osogami
IBM Research - Tokyo
1623-14 Shimotsuruma, Yamato
Kanagawa 242-0011, Japan

December 4, 2009

Abstract

A mixed integer programming is formulated for estimating the speed and the acceleration of a vehicle as functions of the time based on the data from a probe-car system. It is assumed that the data from the probe-car system include the statistics (specifically, the maximum, the minimum, and the average) of the speed and the acceleration, respectively, of the vehicle. The estimated functions have the statistics that agree with those measured by the probe-car system. Also, the estimated function of the acceleration is smooth, which agrees with the observation that drivers are less likely to change the acceleration radically. The estimated functions of the speed and the acceleration of a vehicle can be used to estimate the amount of CO2 emission from the vehicle.

1 Introduction

The manuscript considers the problem of estimating the amount of the CO2 (carbon dioxide) emission from a vehicle based on the statistics of the speed and the acceleration of the vehicle. The statistics include the maximum values, the minimum values, and the average values. Such statistics are available from a probe-car system (also known as a floating-car system). The estimated amount of CO2 emission from each vehicle can be used for example to understand the total amount of CO2 emission in a particular area. The understanding allows one to take appropriate actions for reducing CO2 emissions.

A mixed integer programming (MIP) will be formulated to estimate the speed and the acceleration as functions of the time such that the statistics of the estimated speed and acceleration approximately agree with those of the measured speed and acceleration. It is assumed that the amount of CO2 emission can be estimated from the estimated functions of speed and acceleration.

The rest of the manuscript is organized as follows. Section 2 considers the case where the statistics in a single interval are given. Section 3 extends the approach to the case with multiple intervals.

2 Single interval

Let $a^{(\min)}$, $a^{(\max)}$, and $a^{(\text{ave})}$, respectively, be the minimum acceleration, the maximum acceleration, and the average acceleration in a given interval, $(0, T]$. Let $b^{(\min)}$, $b^{(\max)}$, and $b^{(\text{ave})}$ be the corresponding statistics for the speed. The goal is to find the approximate functions $a(t)$ and $b(t)$ for $0 < t \leq T$, where $a(t)$ denotes the acceleration at time t and $b(t)$ denotes the speed, such that the functions are consistent with the given statistics and $a(\cdot)$ is smooth. The smoothness of the acceleration is desirable, since drivers are less likely to change the acceleration radically.

The interval $(0, T]$ is divided into k intervals, where the j-th subinterval is $(T(j-1)/k, Tj/k]$ for $j = 1, \ldots, k$. For each subinterval, the acceleration will be uniquely determined from m candidates, $a_1, \ldots, a_m$, and the speed from n candidates, $b_1 \ldots, b_n$. Without loss of generality, it is assumed that $a^{(\max)} = a_1 > \cdots > a_m = a^{(\min)}$ and $b^{(\max)} = b_1 > \cdots > b_n = b^{(\min)}$. Let $x_{i,j}$ be the indicator such that $x_{i,j} = 1$ if the acceleration at the j-th subinterval is a_i and $x_{i,j} = 0$ otherwise. Let $y_{i,j}$ be the indicator such that $y_{i,j} = 1$ iff the speed at the j-th subinterval is b_i. Note that each of $a(t)$ and $b(t)$ will be estimated as a step function such that the function is constant for each subinterval.

Since the estimated maximum acceleration should agree with the measured maximum acceleration, it is required that $\max_{1 \leq j \leq k} A_j$ agree with $a^{(\max)}$. This requirement can be expressed as

$$\sum_{j=1}^{k} x_{1,j} \quad \geq \quad 1. \tag{1}$$

For analogous reasons for the minimum and for the speed, it is required that

$$\sum_{j=1}^{k} x_{m,j} \quad \geq \quad 1 \tag{2}$$

$$\sum_{j=1}^{k} y_{1,j} \quad \geq \quad 1 \tag{3}$$

$$\sum_{j=1}^{k} y_{n,j} \quad \geq \quad 1. \tag{4}$$

Observe that the acceleration at the j-th subinterval is

$$A_j \quad = \quad \sum_{i=1}^{m} a_i x_{i,j}, \tag{5}$$

and the speed at the j-th subinterval is

$$B_j \quad = \quad \sum_{i=1}^{n} b_i y_{i,j} \tag{6}$$

for $1 \leq j \leq k$.

Since the estimated average acceleration should agree with the measured average acceleration, it is required that the average A_j over $1 \leq j \leq k$ agree approximately with $a^{(\mathrm{ave})}$. This requirement can be expressed as

$$\frac{1}{k}\sum_{j=1}^{k} A_j \quad = \quad a^{(\mathrm{ave})} + \varepsilon^{(a)}, \tag{7}$$

where $\varepsilon^{(a)}$ is a small constant. The small error of $\varepsilon^{(a)}$ is allowed, since the acceleration is determined from the m discrete points, $a_1, \ldots, a_m$. For analogous reasons for the speed, it is required that

$$\frac{1}{k}\sum_{j=1}^{k} B_j \quad = \quad b^{(\mathrm{ave})} + \varepsilon^{(b)}, \tag{8}$$

where $\varepsilon^{(b)}$ is a small constant.

Since the speed and the acceleration are related with the equation of motion, it should hold that

$$B_j \quad = \quad B_{j-1} + \frac{T}{k} A_{j-1} + \varepsilon_j, \tag{9}$$

for each j, where ε_j is a small constant. Again, the small error of ε_j is allowed, since the speed and the acceleration are determined from the discrete points.

To make the estimated acceleration smooth, one can minimize $\max_j |A_j - A_{j-1}|$, which together with (1)-(8) results in the following MIP formulation:

$$
\begin{array}{lll}
\text{min.} & d & \\
\text{s.t.} & x_{i,j} \in \{0,1\}, \ \text{for } i = 1, \ldots, m, \ j = 1, \ldots, k & (A.1) \\
& \sum_{i=1}^{m} x_{i,j} = 1, \ \text{for } j = 1, \ldots, m & (A.2) \\
& \sum_{j=1}^{k} x_{1,j} \geq 1 & (A.3) \\
& \sum_{j=1}^{k} x_{m,j} \geq 1 & (A.4) \\
& A_j = \sum_{i=1}^{m} a_i x_{i,j}, \ \text{for } j = 1, \ldots, k & (A.5) \\
& \frac{1}{k}\sum_{j=1}^{k} A_j = a^{(\mathrm{ave})} + \varepsilon^{(a)} & (A.6) \\
& \sum_{j=2}^{k} A_j - A_{j-1} \geq -d & (A.7) \\
& \sum_{j=2}^{k} A_j - A_{j-1} \leq d & (A.8) \\
& y_{i,j} \in \{0,1\}, \ \text{for } i = 1, \ldots, n, \ j = 1, \ldots, k & (A.9) \\
& \sum_{i=1}^{n} y_{i,j} = 1, \ \text{for } j = 1, \ldots, n & (A.10) \\
& \sum_{j=1}^{k} y_{1,j} \geq 1 & (A.11) \\
& \sum_{j=1}^{k} y_{m,j} \geq 1 & (A.12) \\
& B_j = \sum_{i=1}^{n} b_i y_{i,j}, \ \text{for } j = 1, \ldots, k & (A.13) \\
& \frac{1}{k}\sum_{j=1}^{k} B_j = b^{(\mathrm{ave})} + \varepsilon^{(b)} & (A.14) \\
& B_j = B_{j-1} + \frac{T}{k} A_{j-1} + \varepsilon_j, \ \text{for } j = 2, \ldots, k & (A.15) \\
& \varepsilon^{(a)} \leq \delta & (A.16) \\
& \varepsilon^{(a)} \geq -\delta & (A.17) \\
& \varepsilon^{(b)} \leq \delta & (A.18) \\
& \varepsilon^{(b)} \geq -\delta & (A.19) \\
& \varepsilon_j \leq \delta, \ \text{for } j = 2, \ldots, k & (A.20)
\end{array}
$$

(A.1) and (A.2) guarantee that, for each j, exactly one of $x_{1,j}, \ldots, x_{m,j}$ is one, and the others are zero. (A.5) defines the acceleration at the j-th subinterval for each j. (A.3) guarantees that the maximum acceleration agrees with that measured, and (A.4) guarantees that the minimum acceleration agrees with that measured. (A.6) guarantees that the average acceleration matches with that measured with the error of $\varepsilon^{(a)}$. (A.7) and (A.8) states that $|A_j - A_{j-1}| \leq d$ for any j. That is, the maximum difference is within d. The objective function is to minimize d, so that the maximum difference is minimized.

For the speed, (A.9)-(A.14) are the conditions analogous to (A.1)-(A.6). (A.15) states the equation of motion with the error of ε_j for the j-th interval. (A.16)-(A.20) guarantee that the errors are within δ, where δ is a specifiable parameter.

The optimal solution to the MIP gives the estimated acceleration, $A_j^\star$, and the estimated speed, $B_j^\star$, for each j. Let $f(a,b)$ be the amount of CO2 emissions per unit time when the acceleration is a and the speed is b. Then the amount, C, of CO2 emission in the interval is estimated with

$$C = \sum_{j=1}^{k} \frac{T}{k} f(A_j^\star, B_j^\star). \qquad (10)$$

Example

Let the measured statistics for the interval $(0, 60]$ be $a^{(\min)} = -1.0$, $a^{(\max)} = 1.0$, $a^{(\text{ave})} = 0.5$, $b^{(\min)} = 0.0$, $b^{(\max)} = 18.0$, and $b^{(\text{ave})} = 6.0$. Let $m = 5$ and $n = 7$. Specifically, $a_i = -1.5 + 0.5i$ for $i = 1, \ldots, 5$ are the representative values of the acceleration, and $b_i = 3i - 3$ for $i = 1, \ldots, 7$ and the representative values of the speed. Let $k = 10$, so that the interval is divided into 10 subintervals. The error is allowed within $\delta = 0.5$.

Then the optimal solution to the MIP gives the speed and the acceleration shown in Figure 1. The optimal solution is found with GLPK (GNU Linear Programming Kit).

3 Multiple intervals

This section considers the case with N intervals for $N \geq 2$. The speed and the acceleration will be estimated as functions of the time for each interval by taking into account the statistics in the other intervals. The extended formulation differs in three cases: the first interval, the last interval, and the other intervals.

Consider the first interval, $(0, T)$. Let $\hat{b}^{(\min)}$ be the minimum speed in the second interval and $\hat{b}^{(\max)}$ be the corresponding maximum. The MIP formulation in Section 2 is extended by incorporating the following two additional constraints:

$$B_k + \tfrac{T}{k} A_k \geq \hat{b}^{(\min)} - \delta \quad (A.21)$$
$$B_k + \tfrac{T}{k} A_k \leq \hat{b}^{(\max)} + \delta \quad (A.22)$$

The new constrains ensure that there exist consistent feasible solutions for the MIP of the first interval and the MIP of the second interval for a

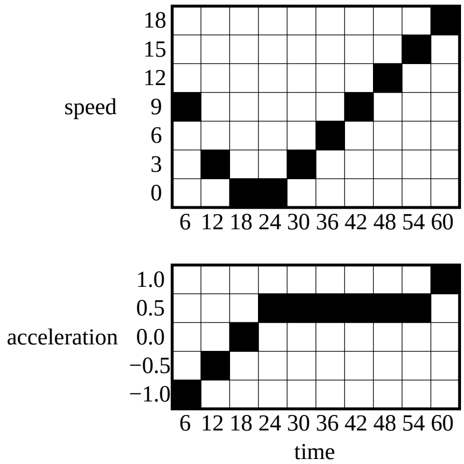

Figure 1: The speed and the acceleration estimated with the MIP as functions of the time, when $a^{(\min)} = -1.0$, $a^{(\max)} = 1.0$, $a^{(\text{ave})} = 0.5$, $b^{(\min)} = 0.0$, $b^{(\max)} = 18.0$, $b^{(\text{ave})} = 6.0$, and $\delta = 0.5$.

sufficiently large δ. Note that the equation of motion implies that the left hand side of the new constraint expresses the speed at the first subinterval in the second interval.

Consider the ℓ-th interval for $2 \le \ell \le N-1$. It is assumed that the speed and the acceleration in the $(\ell-1)$-st (i.e., preceding) interval have been estimated as functions of the time. Let $\bar{a}$ be the acceleration at the last subinterval in the preceding interval and $\bar{b}$ be the corresponding speed. Let $\hat{b}^{(\min)}$ be the minimum speed in the $(\ell+1)$-st (i.e., succeeding) interval and $\hat{b}^{(\max)}$ be the corresponding maximum. When the ℓ-th interval is considered, let $(0, T]$ be the interval, so that the MIP in Section 2 is well defined for the ℓ-th interval. The MIP formulation is extended by incorporating the following five additional constraints:

$$
\begin{array}{ll}
B_k + \frac{T}{k} A_k \ge \hat{b}^{(\min)} - \delta & (A.21) \\
B_k + \frac{T}{k} A_k \le \hat{b}^{(\max)} + \delta & (A.22) \\
B_1 = \bar{b} + \frac{T}{k} \bar{a} + \bar{\varepsilon} & (A.23) \\
\bar{\varepsilon} \le \delta & (A.24) \\
\bar{\varepsilon} \ge -\delta & (A.25).
\end{array}
$$

Note that (A.21) and (A.22) are equivalent to the two constraints incorporated for the case of the first interval, establishing the consistency between the ℓ-th interval and the succeeding interval. (A.23) establishes the consistency between the ℓ-th interval and the preceding interval. Specifically, (A.23) ensures that the speed and the acceleration at the last subinterval in the preceding interval and the speed at the first subinterval in the ℓ-th interval are consistent with the equation of motion, allowing the error of $\bar{\varepsilon}$. (A.24) and (A.25) ensure that the error is within δ.

Consider the last (i.e., N-th) interval. It is assumed that the speed and the acceleration in the preceding (i.e., $(N-1)$-st) interval have been estimated as functions of the time. Let $\bar{a}$ be the acceleration at the last subinterval in the preceding interval and $\bar{b}$ be the corresponding speed. When the last interval is considered, let $(0, T]$ be the interval, so that the MIP in Section 2 is well defined for the last interval. The MIP formulation is extended by incorporating the following three additional constraints:

$$B_1 = \bar{b} + \frac{T}{k}\bar{a} + \bar{\varepsilon} \quad (A.23)$$
$$\bar{\varepsilon} \leq \delta \quad (A.24)$$
$$\bar{\varepsilon} \geq -\delta \quad (A.25),$$

which are equivalent to the three constraints incorporated for the case with the ℓ-th interval for $2 \leq \ell \leq N-1$.

4 Concluding remarks

The presented approach can be generally applicable to the time-series data from sensors, where the speed is understood as the quantity of interest, and the acceleration is understood as the differential of the quantity. However, the presented approach is most effective when there is one of the following three difficulties. First, it is expensive to send the data from the sensor to the analyzer. For example, satellite communication is expensive. The probe-car system is another example, since the users are sensitive to the charge for the packet communication. Second, it is infeasible to store all of the data. For example, the space of a storage might be limited. Third, it is infeasible to record all of the data. For example, the rate of transmission might be faster than the speed of a storage. With the presented approach, one can store only the statistics of the data from sensors and send the statistics to the analyzer, who can restore the time-series data as needed for analysis. Note that the maximum, the minimum, and the average are the statistics that can be calculated very efficiently.

The presented approach allows the analyzer to select the intervals where the time-series data should be restored for analysis. For example, the analyzer might want to know whether the difference between the data from one sensor and the from another is greater than a threshold at some moment for anomaly detection. Then the analyzer can restore the data only in the interval where the maximum from one sensor and the minimum from another differ by more than the threshed.

Acknowledgments

This work was supported by "Promotion program for Reducing global Environmental loaD through ICT innovation (PREDICT)" of the Ministry of Internal Affairs and Communications, Japan.