Aggregation of data from different sources in traffic flow tasks

V. M. Starozhilets, U. V. Chehovich

Dorodnicyn Computing Centre FRC CSC RAS, Moscow, Russia

Barcelona
2016

Motivation

Motivation

Traffic flow mathematical models require accurate data for its initialisation and solving.

Problems with traffic data:

- Traffic detectors data are accurate, but do not cover all considered parts of transport network
- GPS-track data has low accuracy, but covers all considered parts of transport network

Considered environments:

- highway itself
- highway entrances and exits

Example of initialisation

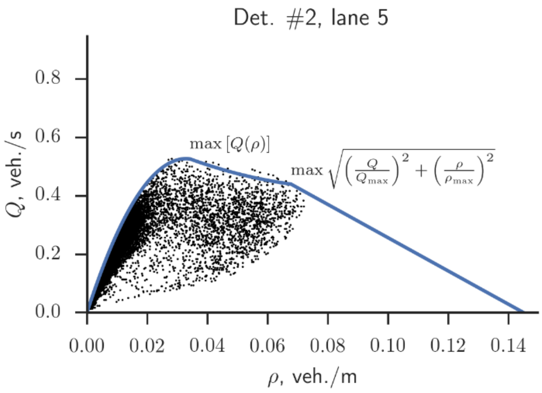


Figure: Fundamental diagram for Moscow Ring Road segment

Main assumption

Let $N_{\text{track},i} \in \mathbb{N}$, $V_{\text{track},i} \in \mathbb{R}_+$ be a number and speed of vehicles extracted from GPS-tracks at moment i.
Denote by $N_{\text{est},i} \in \mathbb{R}$ estimation of the real number of vehicles for the moment of time i, which is detected by traffic detectors.

Main assumption

$$N_{\text{est},i} = f(\mathbf{a} | N_{\text{track},i}, V_{\text{track},i}),$$

where $f : \mathbb{R}^m \times \mathbb{N} \times \mathbb{R}_+ \to \mathbb{R}$, $\mathbf{a} \in \mathbb{R}^m$ — parameters vector.

Problem statement

Let $N_{\text{det}} \in \mathbb{N}$ be a number of vehicles detected by traffic detectors, which considered as true number of vehicles.

Optimization problem

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(f(\mathbf{a}|N_{\text{track},i}, V_{\text{track},i}) - N_{\text{det},i})^2} \to \min_{\mathbf{a}},$$

where n is a number of two-minutes gaps in a chosen time interval.

Function f representation is dependent on data and is discussed below.

Speed transformation

Denote by $V_{\text{est},i} \in \mathbb{R}$ estimation of the real average speed of vehicles for the moment of time i, which is detected by traffic detectors.

Speed transformation

$$V_{\text{est},i} = b_1 + b_2 V_{\text{track},i},$$

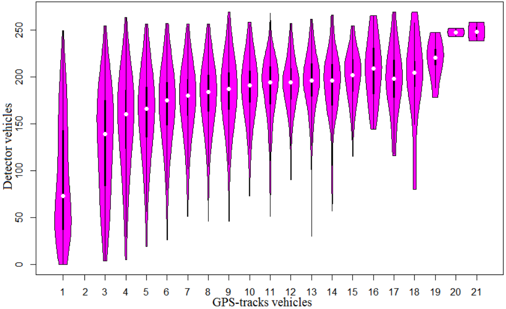
where b_1 and b_2 is a solution of the following problem:

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(b_1 + b_2 V_{\text{track},i} - V_{\text{det},i})^2} \to \min_{b_1, b_2},$$

where $V_{\text{det},i} \in \mathbb{R}_+$ is a average speed of vehicles detected by traffic detectors.

Function f representation

Plot dependence N_{det} vs. N_{track} and observe dependence similar to log function.



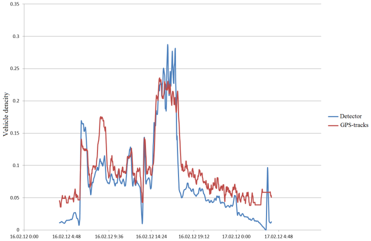
Therefore,

$$f(\mathbf{a}|N_{\text{track},i}, V_{\text{est},i}) = a_0 + a_1 N_{\text{track},i} + a_2 \log(N_{\text{track},i}) + \\ + a_3 V_{\text{est},i} + a_4 N_{\text{track},i} / V_{\text{est},i}$$

Gain from speed transformation

$$V_{\text{est}} = 12.34 + 0.639\, V_{\text{track}}$$

	V_{est}	V_{track}
Mean squared error	**0.03**	0.042
Pearson correlation	**0.787**	0.672



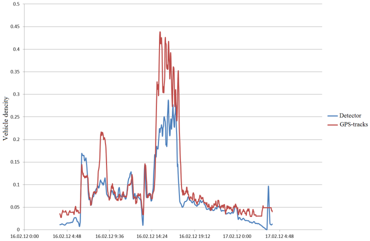


Figure: Plot with vehicle density calculated with (left) and without (right) speed transformation.

Parameter optimization

For vehicle density higher than 0.05:

$$N_{\text{est}} = 157.78 + 4.54 N_{\text{track}} - 4.59 \log(N_{\text{track}}) + 0.153 V_{\text{est}} - \\ -85.069 N_{\text{track}} / V_{\text{track}}.$$

For vehicle density less than 0.05:

$$N_{\text{est}} = 117.75 + 2.11 N_{\text{track}} + 41.55 \log(N_{\text{track}}) - 0.327 V_{\text{est}} - \\ -128.89 N_{\text{track}} / V_{\text{est}}$$

P-values for all items $\leq 10^{-5}$ and therefore every item is significant.

	Train	Test_1	Test_2	Test_3	Test_4
Mean squared error	**0.03**	0.0363	0.0382	0.0339	0.0393
Pearson correlation	0.787	0.823	0.80	**0.85**	0.65

Vehicle density estimation on train data

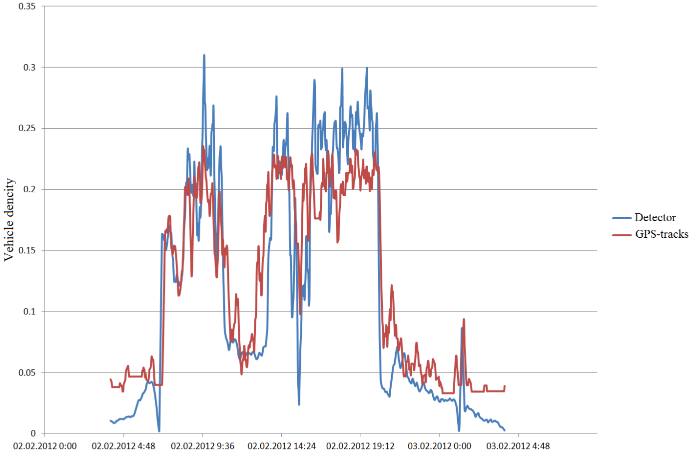


Figure: Vehicle density averaged on 10-minutes obtained after train and ground truth.

Vehicle density estimation on test data

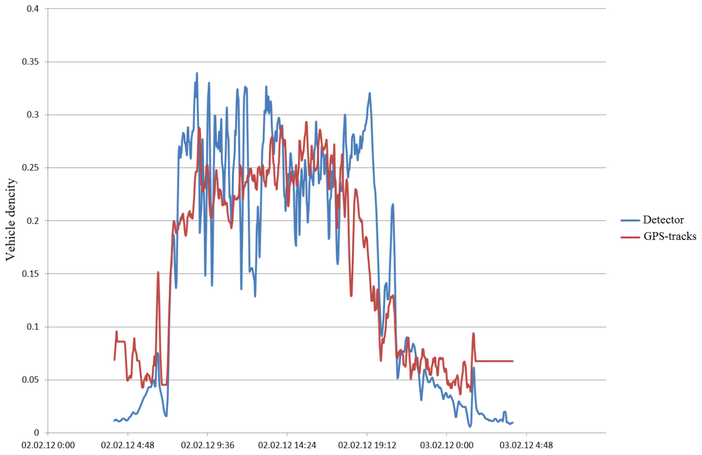


Figure: Vehicle density averaged on 10-minutes obtained after test_1 and ground truth.

Quality of model for on-line prediction

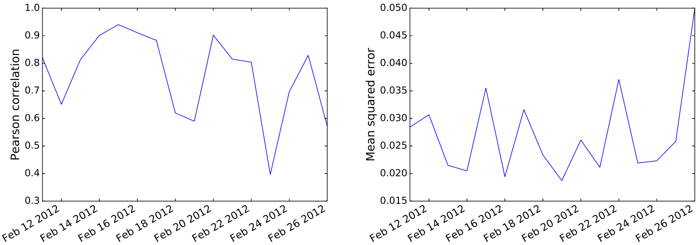


Figure: Correlation (left) and mean squared error (right) averaged on 10-minutes obtained after 7-days learning experiment.

Entrances and exits properties

Specific issues for entrances and exits:

- extremely small amount of data
- data from traffic detectors is not ground truth

Let $N_{\text{ain}} \in \mathbb{R}_+$, $N_{\text{aout}} \in \mathbb{R}_+$ be in and out vehicles estimation in highway crossroad.
Denote by $N_{\text{in}} \in \mathbb{R}$ total amount of vehicles entered the highway and $N_{\text{out}} \in \mathbb{R}$ total amount of vehicles leave the highway.

Balance equation

$$N_{\text{ain}} + N_{\text{in}} = N_{\text{aout}} + N_{\text{out}}$$

Computation of N_{in} and N_{out} is discussed below.

Entrances partition

Let $K_{\text{in}} = \{1, \ldots, K\}$ be a set of entrance indexes.
Denote by N^k_{det} value of N_{det} on entrance k.

$$N_{\text{in}} = \sum_{k \in K_{\text{in}}} N^k_{\text{det}}$$

Entrances partition

Let $K_{\text{in}} = \{1, \dots, K\}$ be a set of entrance indexes.
Denote by N^k_{det} value of N_{det} on entrance k.

$$N_{\text{in}} = \sum_{k \in K_{\text{in}}} N^k_{\text{det}}$$

Problem

There exists a set $K_{\text{intrack}} \subset K_{\text{in}}$ such that $\forall k \in K_{\text{intrack}}$ N^k_{det} is undefined.

Therefore, $K_{\text{in}} = K_{\text{indet}} \cup K_{\text{intrack}}$, such that $K_{\text{intrack}} \cap K_{\text{indet}} = \varnothing$ and

- for $k \in K_{\text{intrack}}$ we do not know N^k_{det}
- for $k \in K_{\text{indet}}$ we do know N^k_{det}

N_{in} computation

Assumption

For $k \in K_{\text{intrack}}$ $N^k_{\text{det},i} = f(\mathbf{a}|N^k_{\text{track},i}, V^k_{\text{est},i})$

Denote by I^k_{in} a set of time indexes i such that we have both $N^k_{\text{det},i}$ and $N^k_{\text{track},i}$ data for k-th entrance.

Optimization problem

$$\sqrt{\frac{1}{|I^{k^*}_{\text{in}}|}\sum_{i \in I^{k^*}_{\text{in}}} (f(\mathbf{a}|N^{k^*}_{\text{track},i}, V^{k^*}_{\text{est},i}) - N^{k^*}_{\text{det},i})^2} \to \min_{\mathbf{a}},$$

where $N^{k*}_{\text{track},i}, V^{k^*}_{\text{track},i}, N^{k^*}_{\text{det, i}}$ is $N_{\text{track}}, V_{\text{track}}, N_{\text{det}}$ for entrance $k^* \in K_{\text{indet}}$, which has the large amount of GPS-track data in the i-th moment of time, $i \in I^{k^*}_{\text{in}}$.

Problem statement for entrances and exits

Let N_{estin}, N_{estout} be estimation of N_{in}, N_{out}. Then to find them we propose to solve the following optimization problem

Optimization problem

$$(N_{ain} + N_{estin} - N_{aout} - N_{estout})^2 \to \min_{N_{estin}, N_{estout}}$$

$$\text{s.t.} \sum_{i \in I_{in}} |N_{estin,i} - N_{in,i}| + \sum_{i' \in I_{out}} |N_{estout,i'} - N_{out,i'}| < \delta,$$

where $I_{in} = \bigcap_{k \in K_{intrack}} I_{in}^k$, $I_{out} = \bigcap_{k \in K_{outtrack}} I_{out}^k$ and δ is appropriate approximation error.

Entrances and exits data recovery algorithm

- Choose crossroad, segments related to entrances, exits and segments from which we take data about N_{ain}, N_{aout}.
- Determine entrances and exits from $K_{intrack}, K_{outtrack}$. Available small amount of data we use to determine parameters of random Poisson process for chosen entrances and exits.
- To initialize proposed algorithm we use Poisson process with obtained parameters and data from traffic detectors if they are available. The proposed algorithm targets to satisfy balance equation.

Data recovery visualization

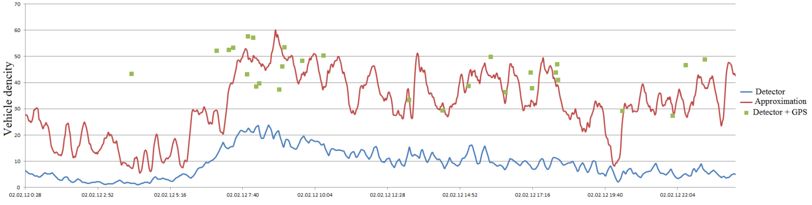


Figure: Blue line — data from traffic detector for one of the entrances, green dots — summary of data from data detector and GPS-tracks. Red line — recovered total number of entered vehicles N_{estin}.

Summary

- We propose algorithms for data recovery on highway using GPS-track data and traffic detectors data.
- We visualize given data to represent target function in the most appropriate way.
- We extend algorithm for highway to highway enters and exits.
- We perform computational experiments for every proposed algorithm.

Thank you for your attention!