Bayesian Estimation of Traffic Lane State

Ivan Nagy[1,2], Miroslav Kárný[1], Petr Nedoma[1], Šárka Voráčová[2]

[1] Institute of Information Theory and Automation, AVČR,
P.O. Box 18, 182 08 Prague 8, Czech Republic, {school,nedoma}@utia.cas.cz
[2] Faculty of Transportation Sciences, Czech Technical University,
Na Florenci 25, 110 00 Prague 1, Czech Republic, {nagy,voraco}@fd.cvut.cz

Summary

Modelling of large transportation systems requires a reliable description of its elements that can be easily adapted to the specific situation. This paper offers mixture model as a flexible candidate for modelling of such element. The mixture model describes particular and possibly very different states of a specific system by its individual components. A hierarchical model built on such elements can describe complexes of big city communications as well as railway or highway networks.

Bayesian paradigm is adopted for estimation of parameters and the actual component label of the mixture model as it serves well for the subsequent decision making. As a straightforward application of Bayesian method to mixture models leads to infeasible computations, an approximation is applied. For normal stochastic variations, the resulting estimation algorithm reduces to a simple recursive weighted least squares.

The elementary modelling is demonstrated on a model of traffic flow state in a single point of a roadway. The examples for simulated as well as real data show excellent properties of the suggested model. They represent much wider set of extensive tests made.

Keywords

Mixture models, estimation, Bayesian approach, transportation systems.

Acknowledgement

This research was supported by EC project IST-9912058, ProDaCTool.

1 Introduction

Simple time-invariant linear models form typical building blocks of complex models. They may serve well even for description of varying systems when variations are caused either by aging or by inherent non-linearity and the need to change the model with the changing state. The changes have to be, however, "slow enough" and they must not call for a change of model structure, for instance, increase of model order. Then, we can cope with them by introducing some kind of more or less sophisticated forgetting, [11]. If the changes of the modeled system are too quick or (and) they even alter the model structure an alternative solution is needed.

Such cases occur typically in traffic-control area. Everyday experience confirms that traffic situations on roads of big towns are extremely different in different day times and use of a single permanently adapting model is very inefficient. Fortunately, several traffic patterns can be distinguished only, say morning and afternoon peaks, night silence etc. It indicates that a finite set of linear models (components) could be sufficient. Their validity domains overlap and thus it is reasonable to construct the overall model as their mixture [14]. The mixture components can be estimated recursively in parallel and their slow adaptation through forgetting is possible as well [8]. Information carried by newly measured data is distributed to each component with a weight given by the degree of membership of the component to the data.

The cited quasi-Bayes algorithm estimates both parameters of all components and the actual "component label" pointing at the most suitable component for the present moment. This classifies well traffic states needed for traffic monitoring and control. Parameter and state estimate may support modelling [4], simulation [2], [10], prediction [7] and traffic control [12], to plan efficient traffic monitoring [1], etc.

The control is necessary on highways as well as in towns. It consists in making free some preferred directions through specified traffic areas or it prefers public traffic to private cars [5], [6], [15]. The means of control can be signal lights on cross-roads or some more drastic actions as, for example, temporarily restricting traffic in chosen roads for private cars.

Solutions to all these problems require models similar to that treated here. The research focuses on identification intended for an automatic traffic control in big cities by means of traffic lights in cross-roads. We suppose that *traffic density* – the number of cars per unit length – and *traffic volume* – the number of cars per unit of time, are measured nearby cross-roads. The behaviour of cars is modelled by a mixture of regression-type models with a finite number of components. Each component reflects one state of the traffic. Structures of the components are supposed to be a priori defined up to a finite set of unknown parameters. These structures can be estimated, too [9], but this problem is out of scope of this paper.

Aims of the paper

The paper has the following two aims

- to present an outline of new approach to mixture model estimation based on Bayesian approach to identification,

- to demonstrate the theory on an elementary transportation application of modelling and estimation of traffic flow state in a specified point of a roadway.

The practically trivial application presented should serve as a building block for modelling and estimation of transportation miniregions and consequently large transportation areas. Nevertheless, even such a simple model has some practical sense. It may e.g. detect exceptional transportation situations.

2 Problem formulation

A system that switches among a finite number of states is considered. Further, it is supposed, that the system in each of his states can be described by a linear model; the transitions among the states can be nonlinear. A sequence of data d_t is measured on the system in discrete time instants $t = 1, 2, \dots, \mathring{t}$. The states of the system are denoted by labels f which are natural numbers $1, 2, \dots, \mathring{f}$, where $\mathring{f} < \infty$ denotes the number of different states. The current label of the active component (which is not measured and thus it is unknown) is denoted by f_t, $t = 1, 2, \dots, \mathring{t}$.

The **theoretical goal** is to describe such system by means of mixture model composed of $\mathring{f}$ components and to estimate parameters of the model. It turns out, that to solve this task it is necessary to estimate currently the active component label in an approximative way.

The **practical goal** is an application of the theory to a small two dimensional transportation problem of estimating the state of traffic flow on a single point of a roadway.

3 System model

In the paper a probabilistic description of variables, used in Bayesian approach to identification [13], is employed. As it is not so widely spread, the basic models are introduced here.

MIXTURE MODEL

Two variables are supposed to be modelled. They are measured data sequence d_t, and unmeasured sequence of actual component label f_t, $t = 1, 2, \dots, \mathring{t}$. Their full probabilistic description is given by the mixture model

$$p(d_t, f_t | d(t-1), \alpha, \Theta), \tag{1}$$

where the following notations have been used:

d_t	currently measured data item,
$d(t-1)$	data items measured up to the time instant $t-1$,
f_t	current component label, $f_t \in \{1, 2, \dots, \mathring{f}\}$, $\mathring{f} < \infty$,
α, Θ	collections of unknown parameters (α concerning component label and Θ data variable).

Using chain rule, the mixture model can be decomposed

$$p(d_t, f_t|d(t-1), \alpha, \Theta) = p(d_t|d(t-1), f_t, \alpha, \Theta)p(f_t|d(t-1), \alpha, \Theta). \quad (2)$$

Model of component

The first part of the decomposition (2) represents the description of the f_t th component. It can be simplified

$$p(d_t|d(t-1), f_t, \alpha, \Theta) = p(d_t|\varphi_{t-1}, f_t, \Theta_{f_t}) \quad (3)$$

when applying the following **assumptions**:

- the actual data item d_t depends only on a finite number of delayed data items collected in the regression vector φ_{t-1},
- the model of the f_t th component does not depend on parameters α (that enter the second part of the decomposition) and concerning the collection Θ, it depends only on its own sub-collection Θ_{f_t}, not on the rest.

This model represents general probabilistic description. For the possibility of realistic computations some specific distribution must be used. Here, normal regression model will be considered.

Model of switching

The second part of the decomposition (2) of mixture model describes the process of switching among individual components. It is considered in the form with the predefined distribution

$$p(f_t|d(t-1), \alpha, \Theta) = f(f_t|\alpha) = \alpha_{f_t} \quad (4)$$

which expresses the **assumptions**:

- the current component label f_t does not depend on the past data $d(t-1)$ and component parameters Θ,
- the probabilities of individual states are time invariant and equal to the items of the vector α, for which, indeed, it must hold $\alpha_i \in (0, 1)$, $i = 1, 2, \ldots, \mathring{f}$ and $\sum_{i=1}^{\mathring{f}} \alpha_i = 1$.

Final form of mixture model

After substituting the simplified expressions into (2), the mixture model gets the form

$$p(d_t, f_t|d(t-1), \alpha, \Theta) = p(d_t|\varphi_{t-1}, f_t, \Theta_{f_t})\alpha_{f_t}. \quad (5)$$

4 Parameter estimation

The models introduced depend on unknown parameters α and Θ. According to Bayesian principles, these parameters are considered random variables. They are described by probability density functions (pdfs) and estimated from measured data.

Description of parameters

The probabilistic description of parameters including the information from data $d(t)$ (measured up to and including time instant t) is given by the conditional pdf

$$p(\alpha, \Theta | d(t)). \tag{6}$$

Bayes rule

As the data are measured in time for $t = 1, 2, \ldots, \mathring{t}$, the description (6) can be sequentially developed according to the Bayes rule

$$p(\alpha, \Theta | d(t)) \propto \underbrace{p(d_t | d(t-1), \alpha, \Theta)}_{\text{model for Bayes update}} \; p(\alpha, \Theta | d(t-1)), \tag{7}$$

where $\propto$ denotes proportionality; the parameter description on the left-hand side has included information from d_t while the parameter description on the left-hand side is without this information and $p(d_t | d(t-1), \alpha, \Theta)$ is the model necessary for imbedding the new information.

Prior parameter description

The recursion starts with prior parameter description $p(\alpha, \Theta | d(0))$ which includes the prior expert knowledge or that extracted from so called prior parameters denoted by $d(0)$.

5 Approximation in estimation

The model $p(d_t | d(t-1), \alpha, \Theta)$ entering the Bayes rule (7) is not directly at disposal and it must be derived.

Precise but unfeasible way

The straightforward Bayesian solution is to compute marginal distribution from mixture model (5)

$$p(d_t | d(t-1), \alpha, \Theta) = \sum_{f_t=1}^{\mathring{f}} p(d_t, f_t | d(t-1), \alpha, \Theta),$$

but it can be clearly seen that after substituting this expression into (7), we obtain recursion with product of sums which is not numerically feasible. So, another approximative way has to be used.

Approximative way

It is based on the fact, that a variable entering a joint pdf can be "removed" either by marginalization (which is not possible here) or approximatively by substituting it by some its estimate as shown in the following scheme

$$p(d_t|d(t-1),\alpha,\Theta) \approx p(d_t, f_t = \hat{f}_t|d(t-1),\alpha,\Theta),$$

where $\hat{f}_t$ is an estimate of f_t. Nevertheless, the real approximation is a bit different and its description follows.

For the approximation intended we formally rewrite the mixture model (5) into a product form

$$p(d_t, f_t|d(t-1),\alpha,\Theta) \underbrace{=}_{(5)} p(d_t|\varphi_{t-1}, f_t, \Theta_i)\alpha_{f_t} = \prod_{i=1}^{\mathring{f}} \left[p(d_t|\varphi_{t-1}, i, \Theta_i)\alpha_i\right]^{\delta(f_t,i)} \tag{8}$$

in which, indeed, the unknown component label is still present, but it occurs only in a single location – the Kronecker delta function $\delta(i,j) = 1$ for $i = j$ and $\delta(i,j) = 0$ for $i \neq j$. The approximation consists in replacing this function by its point estimate, which is its conditional mean value

$$\delta(f_t, i) \approx E[\delta(f_t, i)|d(t)] = \Pr(f_t = i|d(t)) = w_{i,t}, \tag{9}$$

where $\Pr(\cdot|\cdot)$ denotes conditional probability.

Remark: *The effect of the approximation corresponds to the supposed uncertainty of the information about the true state carried by the noisy measured data. Both, the delta function and the vector of probabilities w_t can be considered a distribution of the component label f_t. While the delta function describes the component label deterministically (it points at its true value), the vector w_t distributes the probability among all possible component label values, using also the information from the current data item d_t.*

The approximative expression for the model from Byes rule (7) is then

$$p(d_t|d(t-1),\alpha,\Theta) \approx \prod_{i=1}^{\mathring{f}} \left[p(d_t|\varphi_{t-1}, i, \Theta_i)\alpha_i\right]^{w_{i,t}}. \tag{10}$$

Computation of weights w_t

The vector of probabilities w_t can be computed using Bayes rule with the result

$$w_{i,t} = \int\int p(\alpha,\Theta, f_t = i|d(t))\mathrm{d}\Theta\mathrm{d}\alpha \propto \int\int p(d_t|\varphi_{t-1}, i, \Theta_i)\alpha_i p(\alpha,\Theta|d(t-1))\mathrm{d}\Theta\mathrm{d}\alpha \tag{11}$$

where the integrations are performed along the whole ranges of the corresponding variables.

Approximated Bayes rule

The Bayes rule (7) performing parameters estimation directly unusable for mixture models can now be expressed in a feasible approximative form. Similarly to the model and on accord with the conjugate form we express also parameter pdfs as products

$$p(\alpha,\Theta|d(t)) = \prod_{i=1}^{\mathring{f}} p(\alpha_i,\Theta_i|d(t)). \tag{12}$$

By substituting (10) and (12) into (7) the approximative Bayes rule is

$$\prod_{i=1}^{\mathring{f}} p(\alpha_i,\Theta_i|d(t)) \propto \prod_{i=1}^{\mathring{f}} \left[p(d_t|\varphi_{t-1},i,\Theta_i)\alpha_i\right]^{w_{i,t}} p(\alpha_i,\Theta_i|d(t-1)) \tag{13}$$

6 Exponential class model

The functional recursion (13) recomputes parameter pdfs as functions and thus cannot be practically exploited. If the pdfs entering the relation are expressed through a finite number of parameters in reproducing form (i.e. the recursion does not increase the number of parameters), the functional recursion is transformed into algebraic recursion for statistics of the used distribution. A suitable and sufficiently broad system of such distributions is the exponential class [8] among witch also the normal distribution belongs.

Component model

Using the exponential class for expressing the models of components (3) in the product form (8) we obtain

$$p(d_t|d(t-1),i,\Theta_i) \propto \prod_{i=1}^{\mathring{f}} \exp\left\{q_i(\Theta)\ s(d_t,\varphi_{t-1}) + r_i(\Theta)\right\} \tag{14}$$

where q_i, s and r_i are specified vector functions of their arguments.

Self-reproducing parameter description

The corresponding self-reproducing form of parameter pdfs is (**??**)

$$p(\alpha,\Theta|d(t)) \propto \prod_{i=1}^{\mathring{f}} \left[\exp\{q_i(\Theta)S_{i,t} + \vartheta_{i,t} r_i(\Theta)\}\alpha_i^{\kappa_{i,t}-1}\right], \tag{15}$$

where $S_{i,t}$, $\vartheta_{i,t}$ and $\kappa_{i,t}$ are statistics of the i th component at the time instant t.

Bayes rule for exponential class model

After substitution of (14) and (15) into the general formula (13) we obtain the following recursion for statistics of the pdfs involved

$$\begin{aligned} \vartheta_{i,t} &= \vartheta_{i,t-1} + w_{i,t} \\ \kappa_{i,t} &= \kappa_{i,t-1} + w_{i,t} \\ S_{i,t} &= S_{i,t-1} + w_{i,t}\, s(d_t, \varphi_{t-1}). \end{aligned} \tag{16}$$

for $i = 1, 2, \ldots, \mathring{f}$ and $t = 1, 2, \ldots, \mathring{t}$, starting at the time $t = 1$ with prior statistics $\vartheta_{i,0}$, $\kappa_{i,0}$ and $S_{i,0}$ defining prior parameter distribution.

7 Algorithm of estimation

Let us summarize the results into an algorithm.

I. off-line initial part

– specify prior distribution $p(\alpha, \Theta|d(0))$, through its prior statistics $\vartheta_{i,0}$, $\kappa_{i,0}$ and $S_{i,0}$;
– gather prior data for filling the regression vector φ_0 in function $s(d_1, \varphi_0)$

II. on line part: for $t = 1, 2, \ldots, \mathring{t}$

1. **measure data** item d_t
2. **estimate component label** f_t (determine weights $w_{i,t}$ according to (11)) using new data measurement, i.e.
   • construct the function

   $$p(d_t|\varphi_{t-1}, i, \Theta_i)\alpha_i p(\alpha, \Theta|d(t-1))$$

   which (compare (11), (13) and (16)) corresponds to statistics recomputation with $w_{i,t} = 1$, i.e.

   $$\begin{aligned} \vartheta_{i,t} &= \vartheta_{i,t-1} + 1 \\ \kappa_{i,t} &= \kappa_{i,t-1} + 1 \\ S_{i,t} &= S_{i,t-1} + s(d_t, \varphi_{t-1}). \end{aligned}$$

   • perform integrations (see (11)) and determine weights $w_{i,t}$ (the component label with maximal weight can be considered a point estimate of the label of the current state).

3. **update statistics**
   recompute Bayes rule (13) with the newly measured data d_t

$$p(\alpha_i, \Theta_i | d(t)) \propto [p(d_t | \varphi_{t-1}, i, \Theta_{f_t}) \alpha_i]^{w_{i,t}} \, p(\alpha_i, \Theta_i | d(t-1)) \tag{17}$$

   which corresponds to recomputation of statistics (16)

$$\begin{aligned} \vartheta_{i,t} &= \vartheta_{i,t-1} + w_{i,t} \\ \kappa_{i,t} &= \kappa_{i,t-1} + w_{i,t} \\ S_{i,t} &= S_{i,t-1} + w_{i,t} \, s(d_t, \varphi_{t-1}). \end{aligned} \tag{18}$$

4. **compute point estimates** of parameters and component label, if required. The parameter estimates are standard [13], estimate of current component label has been mentioned above.

III. off-line concluding part

The result of the estimation is posterior pdf $p(\alpha, \Theta | d(\mathring{t}))$ from which all necessary objects can be determined:

– point estimates of parameters are mean values of the posterior pdf $p(\alpha, \Theta | d(\mathring{t}))$,

– data predictor can be expressed by the posterior pdf, too

$$p(d_t | d(t-1)) \propto \int \int p(\alpha, \Theta | d(t)) \mathrm{d}\alpha \mathrm{d}\Theta$$

.

8 Traffic-state identification

TRAFFIC PROBLEM FORMULATION

We demonstrate the power of the above algorithm on a simple but still practical example of estimation the service level in one point in a single traffic lane. This example could be a basic building block for modelling, estimation and prediction of traffic networks.

The data measured for identification are supposed to be "density" and "intensity" of traffic flow. They can be obtained by means of detectors placed under the roadway surface. The density $d_{1,t}$ is a number of vehicles per distance unit at time t. The intensity $d_{2,t}$ is a number of vehicles passing a specified point per time unit at time t. These two variables are considered to be sufficient for full description of the state of traffic flow in a specified point.

It is necessary to take into account both these variables as, for example, from zero intensity it is not clear whether the road is empty or it is completely stuffed with standing cars. Nevertheless, they are mutually dependent. It is well known that the relationship of intensity in dependence on density, has concave parabolic shape. It reflects the basic types of the traffic:

- *growing traffic:* The road is free, the speed of cars is determined by the upper allowed speed limit in cities. With increasing density, while the speed being still constant, the intensity grows linearly, up to a certain limit. Then the distances between cars become too small and safety reasons force the drivers to decrease speed.

- *stagnant traffic:* The further growth of density results in a continued decrease of the speed. The intensity is almost constant and the top of the reversed parabolic shape is reached.

- *decaying traffic:* The drivers are afraid to go on and they slow down or stop when the distances between cars fall below a certain critical level. The cars stack up and become a tail-back. This stage represents a collapse.

The parabolic shape of the function mentioned is common but its precise course varies in dependence of the road monitored and other factors like the kind of a weekday, season etc. In any case, the measured data pairs $[d_{1,t}, d_{2,t}] = [\text{"density"}, \text{"intensity"}]_t$ represent points in an "intensity" - "density" plane. They cluster there and form regions of increased data density. Each detected region reflects a specific state of the traffic area. A description of these regions and their intermediate stages is a useful model of the investigated system.

Traffic model

As discussed above, the data pair $d_t = [d_{1,t}, d_{2,t}]$ bears the necessary information about the traffic state at a time instant t. At a state described by a specific component label, we model their relationship by two dimensional static regression

$$\begin{bmatrix} d_{1,t} \\ d_{2,t} \end{bmatrix} = \begin{bmatrix} \theta_{1,i} \\ \theta_{2,i} \end{bmatrix} + \begin{bmatrix} e_{1,i,t} \\ e_{2,i,t} \end{bmatrix} = \theta_i + e_{i,t}, \qquad i = 1, 2, 3. \tag{19}$$

with zero mean Gaussian noise $p(e_{i,t}) = N(0, R_i)$, with the covariance matrix R_i. This model corresponds to the general model of the i th component (3), with

$d_t = [d_{1,t}, d_{2,t}]'$	modelled data variable,
$i \in \{1, 2, 3\}$	index of the modelled component,
$\Theta_i = \{\theta_i, R_i\}$	regression coefficients and noise covariance of the i th component, with $\theta_i = [\theta_{1,i}, \theta_{2,i}]'$
$\varphi_{t-1} = 1$	formally assigned regression vector to static model.

The model deserves few comments.

- Gaussian pdf is the prominent example of exponential class. For traffic applications with static models, other cases, like Poisson, log-normal or exponential pdfs may be of a direct interest.

- Dynamic model, involving dependence of actual data item d_t also on older ones $d(t-1)$, could be and generally is used.

- A finer modelling that distinguishes more than three basic states is simple and worth of considering.

- The physical meaning of the considered case allows us to select reasonable prior estimates of θ_i as guesses of mean values in respective traffic state. Also, prior estimates of covariance matrices R_i may serve for simple definition of the expected shape for particular traffic states.

- The estimates of component label $p(f_t|d(t))$ classify immediately the current traffic state. The conditional variants of the predictive pdf $p(d_{2,t}|d_{1,t}, d(t-1))$ or $p(d_{1,t}|d_{2,t}, d(t-1))$ models the dependence needed for control of one of those variables by influencing the other one. It is needed for optimization of some traffic criterion, say maximization of intensity of the traffic flow in a controlled area.

9 Experiments

In the experiments presented, we intend to demonstrate the basic ability of the mixture model to describe individual data clusters on both simulated and real data samples. The estimated data clusters describe the multi-modal situations. They can be used either for estimation of the actual traffic state or for prediction of the traffic variables.

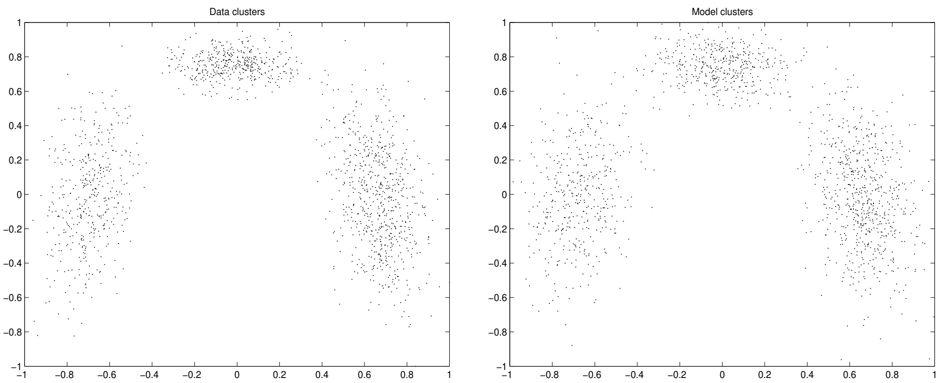
In the example, we will distinguish three different traffic conditions (growth, stagnancy and decay). Consequently, we expect three dominant clusters in our mixture model. For comparison and judging influence of this expectation, the referred tool for automatic estimation of a number of components as well as determining their initial setup [9] will be used. It works on the basis of successive splitting of one initial cluster, using repeatedly the same data sample with a special precaution (likelihood flattening) that prevents to get unjustified information on precision of estimates.

The global predictive pdf is the mixture of predictors obtained for individual components. Estimates of the mixing weights $p(f_t|d(t))$ serve for estimating of the actual traffic state. The value maximizing this probability is taken as the estimate of current component label. Comparison of the value of this weight to the rest of weights serve as a certainty measure classifying the state estimate. This makes us to show all these weights.

Results with simulated data sample

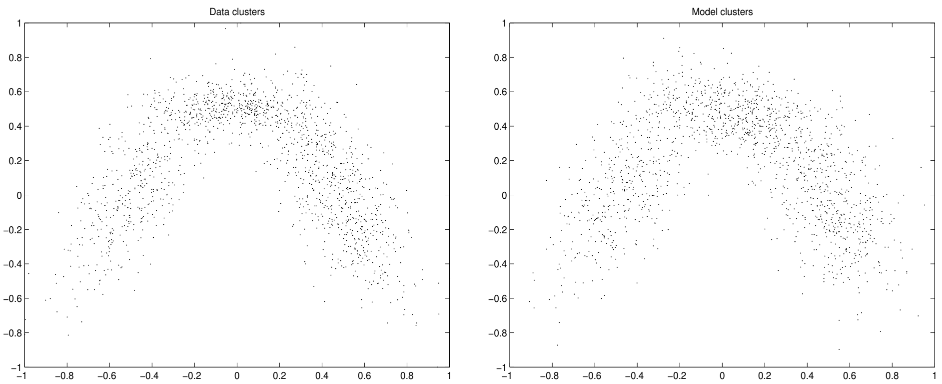
First we are going to show simulated case, nevertheless, similar to the real one. Here, the basic classification properties of the model and its ability to combine individual components to provide a most suitable mixture model for the present time instant can be clearly demonstrated. The first serial of figures shows data clusters and those generated by the identified model. The first experiment involves rather small noise, so that the clusters are easily distinguishable, the second one involves bigger noise and the clusters are rather overlapping.

Data and model clusters with small noise



Data sample   Data from estimated model

Data and model clusters with bigger noise



Data sample   Data from estimated model

In both cases, the clusters are correctly found. In the case of bigger noise, the estimated covariance of clusters are a bit larger.

The ability to distinguish among individual clusters can be demonstrated by a number of correctly estimated components during the simulation. In both cases 1500 data samples were generated, with a random choice of the active component at a time.

In the first, easier example only four samples were wrongly assigned to generating component. Only thirteen times the maximum of the probabilistic weight was smaller than 0.9. In the more difficult overlapping case, 158 estimates of active components ($\approx 10\%$) were wrong and 369 probabilistic weights had maximum smaller than 0.9. It means that in $369 - 158 = 211$ cases – when the situation was not taken as clear – the estimator assigned nonzero probability to more than one component. This corresponds well to the cluster overlap.

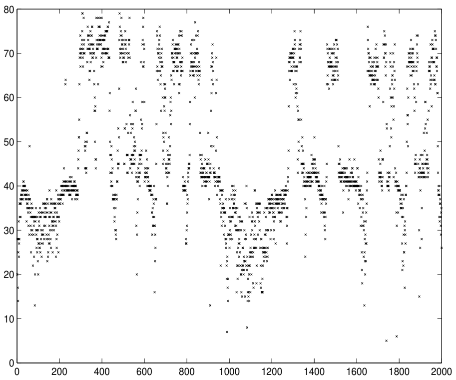
In order to judge predictive properties of the model, the second channel (simulated traffic flow intensity) was predicted from the first one (simulated traffic flow density). The ratios of standard deviation of prediction error and standard deviation of data channel are: 0.51 for the small noise and 0.49 for the bigger one.

Results with real data sample

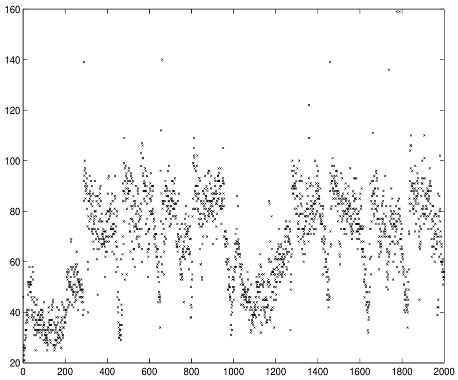
The experiments with real data are naturally more interesting and relevant. All experiments use 2000 data pairs describing normalized density and intensity of traffic flow measured in the centre of Prague. The detectors were located in Legerova street, which is a highway passing through Prague from south to north. In the location of measurements it has three lanes and some 300 metres is the main crossroads with Ječná street. As data pre-filtration, only removing the whole days with damaged data and normalization of the rest has been used.

The data are plotted in the following figures. First, time course of both data channels is drawn, then their scatter diagram indicating data clusters is shown.

Data channels in time



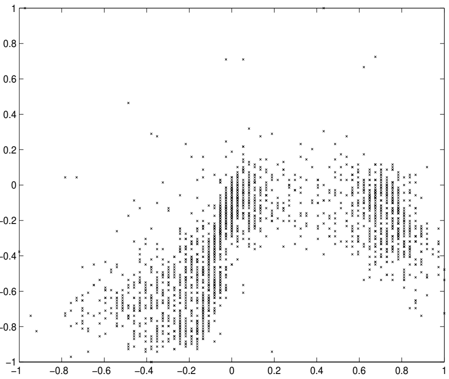
Density of traffic flow



Intensity of traffic flow

The data pairs are measured each 10 minutes. From it follows that the figures reflect approximately two weeks. It can be also seen from the periodicity of the signals – the shorter one represents daily variations and the larger one is given by alternating of weekdays and weekends.

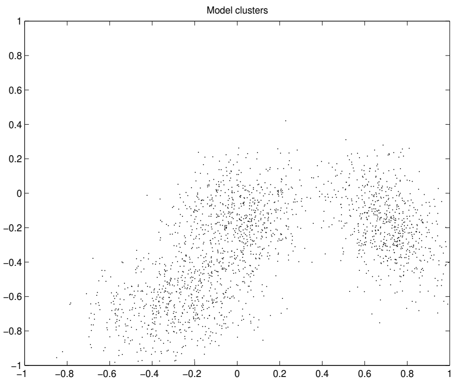
Data channels in scatter diagram



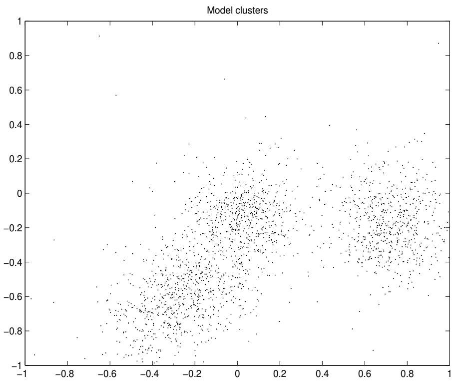
Density-intensity clusters

The figures show that the assumption of data clustering is generally correct but precise separation of particular clusters from a mere sight can be hardly done. The estimation algorithm does it much more reliably as it can be seen from the following pictures.

Estimates of data clusters



Model with 3 components



Model with initialization

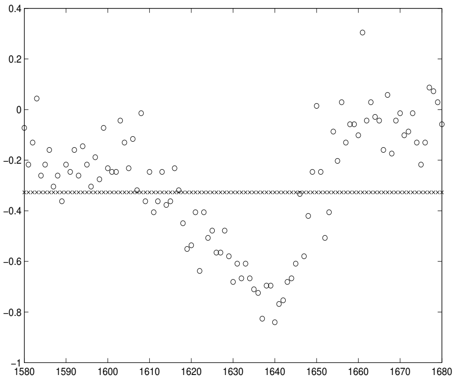
The first picture presents result of estimation for model with 3 components. The number 3 has been derived visually from the the picture with data clusters. The second one illustrates the case with initialization [9] that selects also the number of components. (10 in this example). The estimation quality is routinely compared according to logarithms of predictive pdf evaluated for the whole learning data sample. The found values 2109.0 and 2187.4 indicate that the estimation with automatic initialization is better.

Results with predicting of real data sample

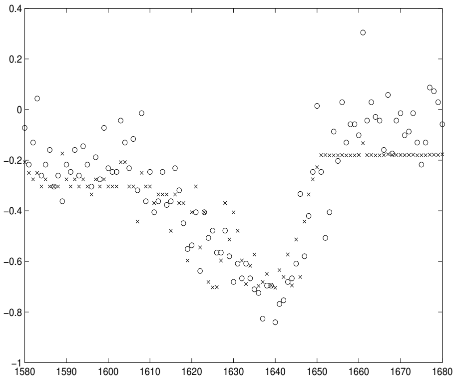
Real predictive abilities are essentially the only property that can be used for judging quality of the estimated model and consequently for clustering and traffic-state estimation.

The predictions were computed for the whole length of the data but, for clarity, the plots are done only for a short interval. In the plots, "o" are values of traffic intensity and "x" represent their predictions.

Predictions of channel 2 with static model



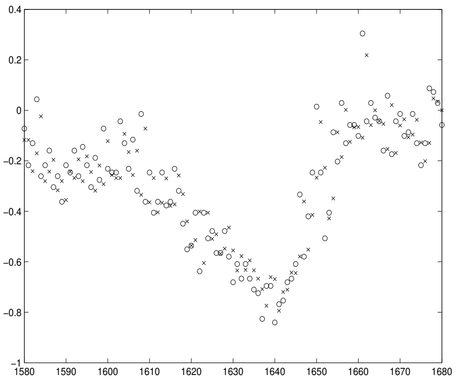
Marginal channel



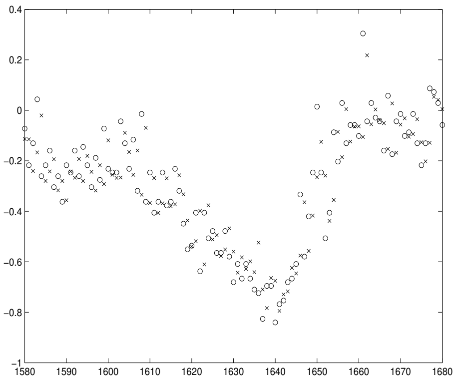
Conditional channel

On the left picture, the second channel is predicted without any knowledge of the first one. As the model is static and thus without any information about the historical values of the signals, the best prediction is mean value of the second channel. On the right picture, we admitted knowledge of the actual value of the first channel into the prediction of the second one. This non-causal prediction has limited practical use, but the mutual correlation of channels leading to the considerable improvement in the prediction indicates possibility to construct a good causal predictor. Mixture of first order models was estimated for getting such a predictor. By model order the depth of history of modelled variable necessary for modelling its actual item is meant.

Predictions of channel 2 with first order model



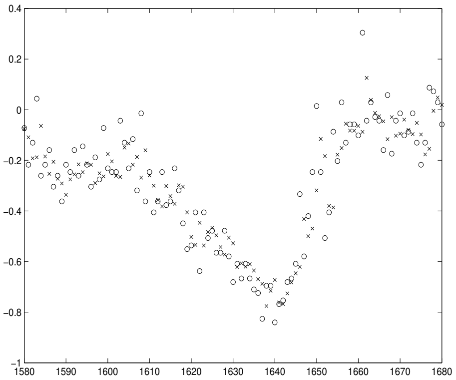
Marginal channel



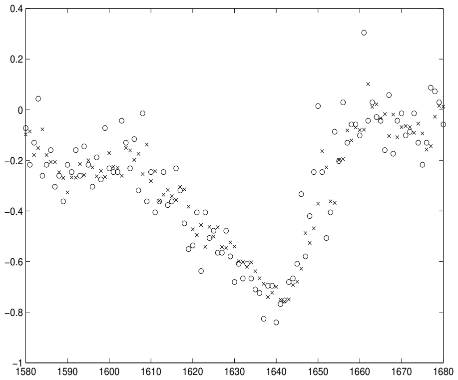
Conditional channel

Here, first order mixture model is used for estimation. The left picture presents prediction of the second channel without knowledge of the actual value of the first channel, the right one with it. The difference between both the pictures is negligible. The quality of prediction is approximately the same as in the previous "irregular" case with static model and knowledge of the actual value of neighbouring channel. This example underlines the importance of higher order models for description of some data that are traditionally referred to and treated as static.

Predictions of channel 2 with higher order models



Second order conditional prediction



5th order conditional prediction

These pictures follow the previous idea of using higher order models. They show, that further increasing of the model order does not bring expected results. The first order seems to be fully sufficient.

For precise comparison of the last six pictures, here are the ratios of standard deviation of prediction error and that of data channel. They are 1.00, 0.53, 0.55, 0.55, 0.51 and 0.50, respectively.

10 Conclusions

General recursive, potentially adaptive, estimation of mixture models was presented and applied to detection of clusters in two-dimensional traffic data. The data are intensity and density of traffic flow. Their configuration can be taken as an evidence of the current state of traffic. Static dependence of these variables has parabolic shape. It varies between roads and may change with time. The presented algorithm estimates simply the relevant shape. Its individual application allows to create specific building blocks of complex models. It also provides reliable estimate of the traffic state of this element.

The performed experiments show that the main purpose, detection of data clusters, is reached satisfactorily. It is important that the algorithm works not only on artificial examples in which underlying assumptions are met but also on real data where the simplifying assumptions are violated. There, universal approximation property [3] of normal mixtures plays probably the decisive role. The basic mixture of static models was shown to provide reasonable traffic predictions when non-causal models exploiting correlation of multi-variate records were used. The results with first order models were, however, shown much better and moreover they provide naturally causal predictors.

The experiments confirmed that the initial estimation conditions can considerably influence quality of results. The automatic initialization proposed in [9] that solves also problem of structure estimation was found quite helpful for the studied problem. At the same time, the prior information is abundant in the case of transportation data and its wide utilization is desirable and useful.

This paper represents promising but only initial step in using mixtures for modeling, estimation, prediction and control of complex traffic systems. A lot remains to be done. At present, for instance:

- The initialization procedure has been substantially improved and its thorough testing is performed.
- The assumption on constant component weights, justified from approximation viewpoint but illogical from interpretation perspective, is relaxed. Evolution of activity of components is modelled by Markov chain and, importantly, theoretically justified extension of the presented estimation has been designed.
- The straightforward extension of the presented solution to several lanes and crossroads is addressed.

It is worth of stressing that the limited solution for a single traffic flow can be directly used for better modelling and predicitions.

References

[1] G. C. Bianco, L. and P. Reverberi. Optimal location of traffic counting points for transport network control. In *IFAC Transportation Systems, Chania*, pages 890–895. Grece, 1997.

[2] G. Bowen and R. Bretherton. Latest developments in scoot - version 3.1. In *International Conference on Road Traffic Monitoring and Control. Inst. of Electrical Engineers*. London, 1996.

[3] S. Haykin. *Neural networks: A comprehensive foundation*. Macmillan College Publishing Company, New York, 1994.

[4] D. Helbing. Traffic data and their implications for consistent traffic flow modelling. In *Transportation Systems, Chania*, pages 809–814. Greece, 1997.

[5] F. M. Hounsell, N. and S. Ishtiaq. Journey time prediction for bus priority at traffic signals. In *IFAC Transportation Systems, Chania*, pages 911–916. Greece, 1997.

[6] N. Hounsell and J. Landles. Bus priority at traffic signals in london. progress, performance and opportunities. In *Second World Congress on Intelligent Transport Systems '95*. Yokohama, 1994.

[7] S. I. Hounsell, N. and F. McLeod. Short term forcasting of urban traffic congestion. In *6th World Conference on Transport Research, Lyon*. France, 1992.

[8] M. Kárný, J. Kadlec, and E. L. Sutanto. Quasi-Bayes estimation applied to normal mixture. In J. Rojíček, M. Valečková, M. Kárný, and K. Warwick, editors, *Preprints of the 3rd European IEEE Workshop on Computer-Intensive Methods in Control and Data Processing*, pages 77–82, Praha, September 1998. ÚTIA AV ČR.

[9] M. Kárný, P. Nedoma, I. Nagy, and M. Valečková. Initial description of multi-modal dynamic models. In V. Kůrková, R. Neruda, M. Kárný, and N. C. Steele, editors, *Artificial Neural Nets and Genetic Algorithms. Proceedings*, pages 398–401, Wien, April 2001. Springer.

[10] W. Kronjäger and P. Konhäuser. Applied traffic flow simulation. In *IFAC Transportation Systems, Chania*, pages 805–808. Greece, 1997.

[11] R. Kulhavý and M. B. Zarrop. On general concept of forgetting. *International Journal of Control*, 58(4):905–924, 1993.

[12] M. Papageorgiou. *Applications of Automatic Control Concepts to Traffic Flow Modelling and Control.*, volume 50. Lecture Notes in Control and Information Sciences, Springer-Verlag, 1983.

[13] V. Peterka. Bayesian approach to system identification. In P. Eykhoff, editor, *Trends and Progress in System Identification*, pages 239–304. Pergamon Press, Oxford, 1981.

[14] D. Titterington, A. Smith, and U. Makov. *Statistical Analysis of Finite Mixtures.* John Wiley & Sons, Chichester, New York, Brisbane, Toronto, Singapore, 1985. ISBN 0 471 90763 4.

[15] J. Viegas and L. Baichuan. Traffic control system with intermittent bus lines. In *IFAC Transportation Systems, Chania*, pages 905–910. Greece, 1997.