Paper

Heuristic Analysis of Transport System Efficiency Based on Movement of Mobile Network Users

Grzegorz Sabak

Abstract— The paper describes results of introductory research focused on possibility to use location data available in a mobile network for the analysis of transport system status and efficiency. The details of a system capable of detecting abnormal traffic situation (accidents, heavy congestion) are described. This system (called VASTAR) uses a neural network to learn and store certain characteristic of the analyzed part of a road system. Based on a measured divergence from normal characteristic, a notification about non-typical situation is triggered. The results of a computational experiment using real-world location data and simulation of abnormal situation are provided. The proposed system can be a relatively low cost way to improve competitiveness of a mobile network operator by allowing him to offer new type of informational service. It could also aid municipal authorities by providing support for decisions regarding road traffic control and management and be used by emergency services as a monitoring an alarming tool for detecting abnormal road traffic situations when other means of observation are unavailable.

Keywords— congestion detection, decision support systems, neural networks, transport system.

1. Introduction

It is a common observation frequently found in newspapers that transport service is vital for both operations of the countries' economies and quality of life of the ordinary people. We all frequently experience either a pleasure of traveling along scenic highway or strong inconvenience of being stuck in a traffic jam or in a crowded bus traveling slowly through streets of a city centre.

Other domain of technology that became extremely important nowadays is mobile communications. Having mobile phone and being almost instantly accessible is assumed for all people living active professional or social life.

These two domains are similar in many ways. Firstly, their purpose is a change of place of availability of either goods (transport) or information (communication) with high efficiency. This typically means low cost and high speed. Secondly, problems encountered in research in these two areas are to a certain extent similar (e.g., routing, scheduling, congestion). Lastly, we all have an indispensable need of using them regularly, and we are actually doing that almost all the time.

These observations lead to the concept of analysis of transportation system state based on the behavior of users of mobile communication network. Since some of the required information is already available at relatively low cost in a mobile network, it may be a valid opportunity for business and government entities. This research is aimed to exploit this opportunity using information technology tools and techniques.

2. Location Services

Typically, among other value added services, the mobile network operator provides a set of services based on the customer's location. Customer's location, means position of a mobile station (MS) consisting of a user equipment and a SIM card. User equipment is usually one of the following: a mobile phone, a PC data card, or a dedicated GSM module installed in a vehicle. Business case of location services (LCS) assumes that information or service which would be strongly related to the current situation of the customer (his location in this case) is so valuable for him, that he is willing to pay a premium.

In general, there are four types of location services [1]:

- Commercial LCS – all services that are available to customers (e.g., SMS, MMS or WAP services offering content dependent on the location).
- Internal LCS – in which location information is used for internal purposes of a mobile network operator.
- Emergency LCS – enforced by law, provide location information in case of emergency situations.
- Lawful intercept LCS – also typically enforced by law, supports various legally sanctioned services.

Regardless the type, implementation of all LCS require presence of a gateway mobile location centre (GMLC) in the network. The GMLC plays key role as it actually enables different applications to access customer location information. Different interfaces can be used by LCS client to use location service, however mobile location protocol (MLP) is most widely adopted, being now considered sufficiently standarized.

In GSM networks, different positioning methods are available to choose from. Most of them are based on the features of a radio access network such as cell identification (cell ID), radio signal strength, and various time variables. They are not described in this paper, but detailed specifications are available, e.g., in [1], [2].

The cell ID method is especially important and shall be explained in more detail. This positioning method is most widely used and does not require significant investment in operator's access network or his core network. Cell ID parameter is used frequently in a mobile network. In this method, all cells in the network are mapped to geographical coordinates which point to a location considered to be the most likely position of the mobile station active in the cell.

3. Problem Formulation

3.1. Introductory Definitions

Let *an event* will be a pair $e = \langle t, P\rangle$, where t is a timestamp and P is a list of parameters (attributes) defining event type.

The sequence of events ordered according to ascending timestamp will be called *an event stream* and denoted as

$$\mathscr{E} = \{e_1, e_2 \ldots, e_k\}.$$

Given two event streams $\mathscr{E}_1$ and $\mathscr{E}_2$ an operation of *stream junction* $\oplus$ can be defined. This operation creates a new event stream that contains all events from both streams and preserves the order of timestamps.

Let an *event stream filter* will be a function $f : e \to \{0,1\}$. Each event mapped to value 1 will be considered as passed through the filter. All other events will be considered as blocked by the filter. Although filter is a function defined for sequence elements, for convenience we will denote it also as $f(\mathscr{E})$, which indicates that all sequence elements are processed by filter.

An operation $F(\mathscr{E}, f)$ removing all events for which $f(e) = 0$ from an event stream $\mathscr{E}$ will be called a *filtering operation*. The result of applying a filtering operation to a stream of events is also a stream of events.

In order to be able to model input data and define a problem let us define the following:

- $x, y \in R$: geographical coordinates; no particular coordinate system is assumed;
- AREA $\subset R^2$: area on which the problem is defined;
- MS: a set of all mobile stations in a mobile network;
- $p = \langle x, y\rangle \in$ AREA: location of a mobile station;
- $\hat{p} = \langle \hat{x}, \hat{y}, r\rangle$: mobile station location estimate defined as a circle (with a centre in $\langle \hat{x}, \hat{y}\rangle$ and radius r) in which the station is located.

Let *a location event* be defined as an event $l = \langle t, P\rangle$ with $P = \langle m, \hat{p}\rangle$, where $m \in$ MS is a mobile station which was located and $\hat{p} \in$ AREA is the location estimate of the mobile station.

A sequence of location events

$$\mathscr{L} = \{l_1, l_2 \ldots, l_k\}$$

will be called *a location event stream*.

3.2. Definition of Problems

Based on the definitions above, two problems can be defined. These problems will become a base for future considerations.

Trip time prediction. Given AREA, MS, and $\mathscr{L}$ at the moment t^* estimate time $\hat{t}$ that is required for mobile station $m \in$ MS to move from location p_1 to p_2 ($p_1, p_2 \in$ AREA) assuming that mobile station is minimizing trip time.

Congestion detection. Given AREA, MS, and $\mathscr{L}$ find the place and time when high congestion happened.

3.3. Further Definitions

For further considerations let *a move event* be an event $v = \langle t, P\rangle$, where $P = \langle m, t_b, \hat{p}_b, t_e, \hat{p}_e\rangle$, where: $m \in$ MS is a mobile station, t_b is a starting moment of the move, $\hat{p}_b = \langle \hat{x}_b, \hat{y}_b, r_b\rangle$ is a location estimate at the start of the move, t_e is an ending moment of the move, $\hat{p}_e = \langle \hat{x}_e, \hat{y}_e, r_e\rangle$ is a location at end of the move.

A sequence of movement events

$$\mathscr{V} = \{v_1, v_2 \ldots, v_k\}$$

will be called *a move event stream*.

For each move event the following variables can be calculated:

- movement duration $t_{dur} = t_e - t_b$,
- distance $d = \sqrt{(\hat{x}_e - \hat{x}_b)^2 + (\hat{y}_e - \hat{y}_b)^2}$.

Creation of the move event stream. The move event stream is created from a location event stream according the following procedure. For every $l_i = \langle t_i, m_i, \hat{p}_i\rangle \in \mathscr{L}$.

1. Let j be the highest number, where $m_j = m_i$ and $j < i$. The $\hat{p}_j$ is the last known location estimate of mobile station m_i. The t_j is the timestamp associated with the location event.
2. If j cannot be determined (which means that $\hat{p}_i$ is the first location event of mobile station m_i) continue with processing of $\mathscr{L}$.
3. If j can be determined create a movement event $v_k = \langle t_k, m_k, t_{b_k}, \hat{p}_{bk}, t_{e_k}, \hat{p}_{ek}\rangle$, where $t_k = t_i$, $m_k = m_i$, $t_{b_k} = t_j$, $\hat{p}_{bk} = \hat{p}_j$, $t_{e_k} = t_i$, and $\hat{p}_{ek} = \hat{p}_i$.

This procedure can be effectively implemented using, for example, a sorted list of last known location estimates.

4. The VASTAR Project

Analysis of the mobile network subscribers movement provides unique, nearly real-time information about the status of the transport system they use. To achieve this, a number of problems must be dealt with and solutions to them must be found.

Potential attractive applications of a system performing such analysis include.

- Provision of city or highway traffic data supporting urban traffic management and control (UTMC) process with no additional costs of traffic measurement units. Obviously, limited accuracy of such data would probably mean that traditional methods would have to be used in key areas.
- Support tool for monitoring city and highway traffic able to detect abnormal situations (e.g., congestions, accidents). This would be especially useful in areas without video surveillance.
- Business opportunity for mobile network operators to attract their subscribers to the website presenting up to date traffic information.
- Decision support system for local authorities optimizing investment in transport infrastructure.

This list defines objectives for a design of traffic decision supporting system. In Section 3, model of available location data was defined. In this section an architecture of such system is proposed. Let this information system be called VASTAR (name can be resolved to value added services for streets and roads) which for the sake of convenience can be shortened to VA∗. The VA∗'s high level logical architecture presented further does not limit its functionality to solving only one kind of problems but rather provides a framework which can support different types of analysis.

Elements constituting VA∗ are presented in Fig. 1. The functions of different modules are described in the following paragraphs, and an application of this architecture to the task of detecting abnormal road conditions is described later in Section 6.

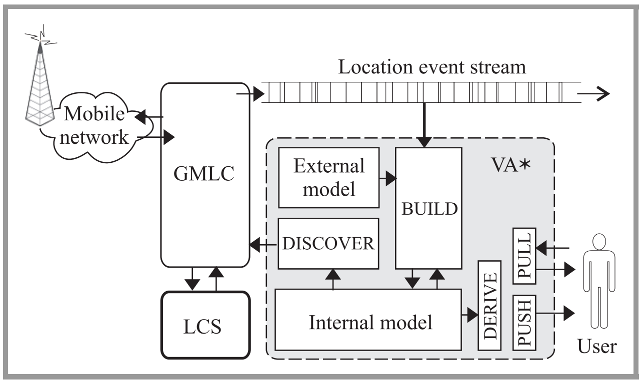

Fig. 1. Architecture of the VA∗ system.

External model. The purpose of the external model is to provide the VA∗ with information regarding transport system which is important to the application and known a priori to the user. This could include, for example, model of the transport infrastructure (roads, their types, nodes, etc.) or other key assumptions to the problem being solved.

Internal model. Represents VA∗'s knowledge about the analyzed transport system, including information about objects and their attributes.

BUILD module. This module is responsible for the analysis of incoming data stream, processing it, and acquiring information needed to perform tasks requested by users. This information is stored in the system by modifications of parameters of the internal model.

DERIVE module. The module performs analysis of information stored in the internal model in order to provide responses to the queries submitted by users via PULL/PUSH modules.

PUSH/PULL modules. These modules are responsible for processing queries from the users (PULL) and for notification about the events in the transport system (e.g., expected congestion) based on the requirements provided by the users (PUSH).

DISCOVER module. This functional element of the VA∗ shall have an ability to request a location of the specific mobile station. Its goal is to improve information quality of the location event stream from the point of view of the system's application.

5. Analysis of Real-World Data

In order to assess informational value of available data a basic statistical analysis was performed. A sample of the GMLC logs was used as an input data. The GMLC operates in the network of mobile network operator in Poland which uses cell ID as a location method. The consequence of limiting input data only to location streams processed by GMLC is that only a small fraction of all mobile stations are "observed". However, by doing this we avoid costs of collecting location data from network elements, which can be quite problematic. The sample used contained GMLC log entries from eight consecutive days from 25 Sept. to 2 Oct. 2008.

The key findings of analysis are as follows.

- The observed mobility of mobile stations, understood as a number of locations visited by a MS is different for working days (higher) and weekends (lower).
- The mobility of mobile stations is not equal throughout the day, but significantly lower between 10PM and 5AM.
- In a real world, location errors are encountered. For example, mapping of the cell ID to geographical coordinates can be incorrect.

- Some locations are visited more frequently then others (majority of them are actually near important transport routes, e.g., Warsaw-Katowice express road).

- It is visible that there is a process running in the network that is monitoring location of a set of mobile stations. This can be accounted to one of location services offered by the operator.

- Only a part of collected data is related to real movements of mobile stations. In many cases mobile stations are not changing their locations or a *virtual* movement between two or three locations is observed.

Virtual movement of mobile stations. In the land mobile networks it is very common that multiple cells overlap. This is a result of the requirement that holes in the coverage are eliminated. A situation when two cells cover the place where mobile station is located is shown in Fig. 2.

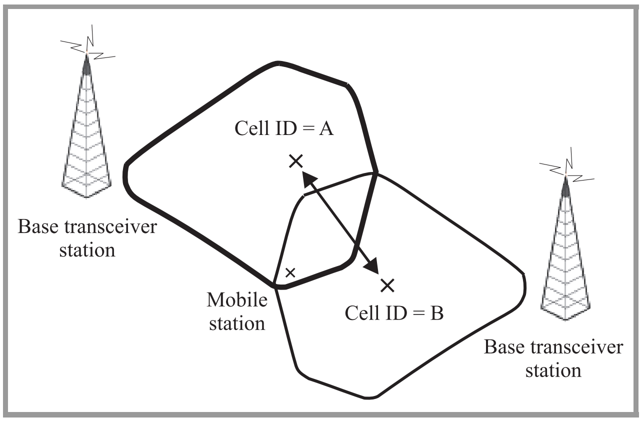

Fig. 2. Two overlapping sector cells.

In this case, despite the fact that real location of the mobile station remains the same, changing conditions (physical environment factors and mobile network status) cause mobile station to be reported as present in either cell A or cell B. The MS appears to move back and forth between centers of the cells A and B. Similar situation can be encountered when three or more cells cover location of the mobile station.

Summary. Analysis of available data shows that information about users of road system which can be used to evaluate status of this system is available in the GMLC logs. However, there are difficulties which need to be addressed.

- Lack of information about accuracy of the location data. When location data comes from the GMLC, radius of the circle representing position estimate is not available. Typically for cell ID method accuracy is in the 50 m – 1 km and 1 km – 35 km ranges in the urban and rural areas, respectively [3].

- Lack of information which MS are actual road system users. Location events found in GMLC logs do not necessarily refer to the positions of the mobile stations attached to the vehicles or people using transportation system. A way to distinguish between MS of his type and all other MS has to be proposed or a method that can effectively deal with this kind of "information noise" has to be used. "Information noise" is a data present in the event stream, but not useful for the system.

- The virtual movement of mobile stations. This introduces additional "information noise" and causes situation when static mobile station is seen by the VA$*$ as changing its position.

- High volumes of data to be processed. Extensive use of the location services in the mobile network may require processing millions of records, which may be a problem when a real-time operation of the system is required.

- Incorrect data due to errors in the positioning process. As in all real world systems, especially such complex and depending on the physics of nature as mobile network, incorrect position reports cannot be excluded. In some cases an error of location estimate can reach several hundred of kilometers.

The nature of these problems indicate that some kind of heurisitics method should be used to deal with them.

6. Accident Detection

Based on the high level architecture described in Section 4, a monitoring system capable of detecting abnormal traffic situations can be built.
Accidents or heavy congestions are generally situations when velocity of vehicles drops down dramatically. This means that in a given period of time the distances by which vehicles move are significantly smaller. It can be assumed that for every move event the distance d between estimated starting and finishing points depends on the following variables:

- starting point coordinates (x, y);
- direction of the move a;
- time of the day t_{day};
- duration of the move t_{dur}.

Let this dependency be called *a move length characteristics* and be denoted as $d(x, y, a, t_{day}, t_{dur})$. This characteristic is directly related to vehicles' average speed (the lower speed the lower value of characteristic). An assumption is made, that abnormal traffic conditions can be detected by monitoring changes of this characteristic.

In the accident detection application the VA∗ determines the move length characteristic of a road system defined by the given external model. Later on, during operation it monitors how much this characteristics for the current moment differs form the known one. A difference exceeding a pre-defined threshold may indicate abnormal conditions, e.g., accident or high congestion.

6.1. Internal and External Models of a Road System

In order to verify the proposed approach a computational experiment was conducted. In this experiment the analysis was limited to only a single road segment. In this case the road is represented as one line section. For the sake of further possible generalization of knowledge, the move events starting and finishing points are transformed to the new coordinate system. This coordinate system is defined by setting coordinates of the road segment in a way that it becomes a line section with ends at points $(0,0)$ and $(0,1)$. The effects of such transformation (for the road segment and for one of the move events) are shown in Fig. 3.

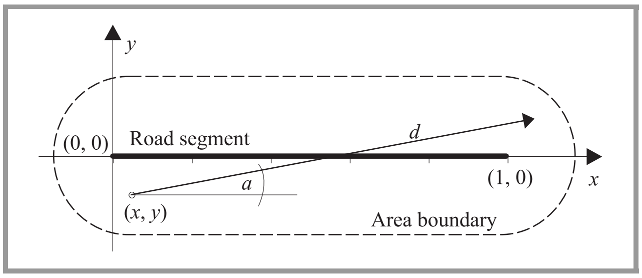

Fig. 3. Transformation of road segment.

In this application the VA∗ calculates move distance characteristics within an DERIVE module and stores it in the internal model. This characteristics is obtained from movement events created from location event stream available from the GMLC. In the internal model a neural network [4]–[6] (dual layer perceptron) is used to approximate this characteristic. The architecture of such network is shown in Fig. 4.

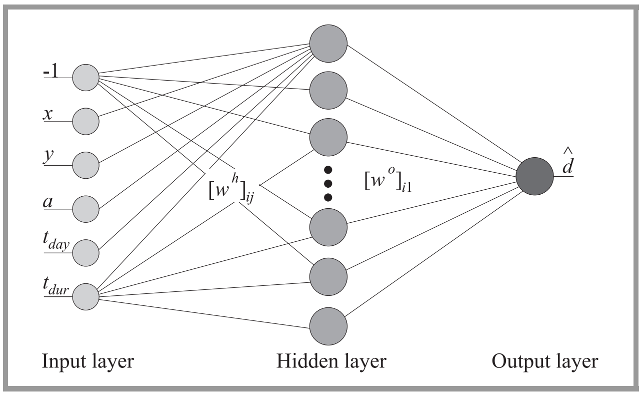

Fig. 4. Architecture of neural network.

The neural network used had six inputs nodes with the following input variables:

- constant value -1 (which makes activation thresholds in neurons unnecessary);
- x coordinate of the starting point;
- y coordinate of the starting point;
- a: angle between the move vector and $0x$ axis,
- t_{day}: moment of the day the move began;
- t_{dur}: duration of the move.

All nodes in a hidden layer use a sigmoid activation function. The output value of the ith neuron in a hidden layer of the perceptron is calculated according to the following formula:

$$o_i^h = g(w_{i1}^h \cdot -1 + w_{i2}^h \cdot x + w_{i3}^h \cdot y + w_{i4}^h \cdot a + w_{i5}^h \cdot t_{day} + w_{i6}^h \cdot t_{dur}),$$

where w_{ij}^h is a weight associated with jth input of a ith neuron, and $g()$ is a sigmoid activation function (logistic curve) defined by equation:

$$g(x) = \frac{1}{1+\exp^{-2\beta x}}.$$

Parameter β is called a slope factor. Plots of the sigmoid function for different values of β are shown in Fig. 5.

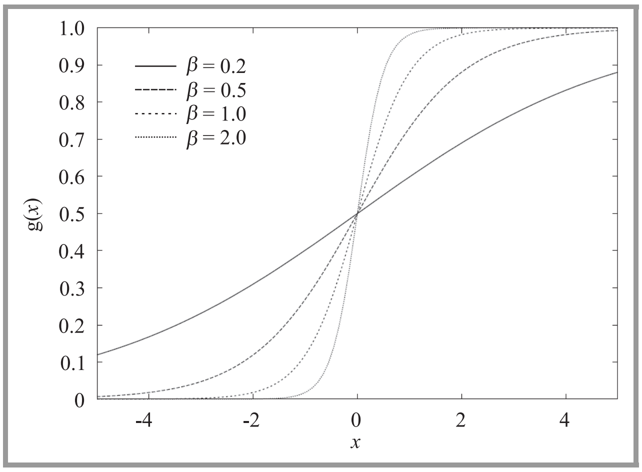

Fig. 5. Activation function.

There is only one node in an output layer. The output value of this node is calculated acording to the following equation:

$$\hat{d} = \sum_{i=1}^{n\_hidden} w_i^o \cdot o_i^h,$$

where n_hidden is a total number of neurons in the hidden layer, and w_i^o is a weight associated with ith input of the output neuron. For the output neuron activation function is not used. Because of that, the perceptron is able to learn function values not limited to any specific range, e.g., $(0,1)$ in case of a sigmoid function.
The output value of the network is the estimated value of a move length characteristics, denoted as $\hat{d}$.

6.2. Learning Process

From the VA∗'s architecture point of view, the development of the internal model is done by modification of the weights of a neural network. An important, but quite straightforward in implementation, function of the BUILD module is filtering out all move events that does not start within defined distance from the road segment (5 km in this case). During the learning, weights were adjusted according to an *error backpropagation* [4] algorithm with the step α.

When learning is completed the VA∗ knows, with some level of error, how mobile stations behave in the analyzed area. As new move events come in the stream a moving window average of error is calculated. When this value exceeds an average value of the error by a certain factor the notification is triggered that the traffic conditions on this part of the road differ from typical ones. The above describes principles of the DERIVE module operation.

6.3. Model of an Abnormal Situation

It was not possible to gather information regarding occurrence of real-world accidents which happened during period when data was collected, so an option to simulate such situation by modification of the movement stream was chosen.

Let t^A_{start} and t^A_{stop} be the start and end time of an abnormal situation period. Let $\gamma \in \langle 0, 1\rangle$ be a congestion intensity. For every move event $v_k = \langle t_k, m_k, \hat{p}_{b_k}, \hat{p}_{e_k}\rangle$.

- $t_k < t^A_{start}$ – move event is not changed, it is assumend that accident did not influence the move.
- $t^A_{start} \le t_k \le t^A_{stop}$, the move can be impacted by the accident. With the probability:

$$p_{block} = \gamma \frac{t_k - t^A_{start}}{t^A_{stop} - t^A_{start}}$$

  $\hat{p}_{e_k}$ is set to $\hat{p}_{b_k}$.
- $t^A_{stop} < t_k$ – move event remains unchanged.

The probability of impacting the move event is greater at the beginning of a congestion period and it decreases as congestion is coming to its end. For a given t_k, p_{block} is proportional to γ.

This model assumes that during the abnormal period mobile station movements are impacted in a following way: mobile station move is blocked with a probability depending on the relative time that elapsed from the beginning of the congestion (Fig. 6).

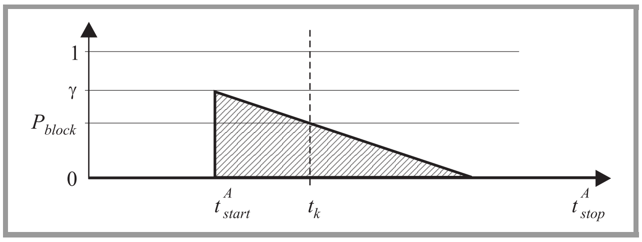

Fig. 6. Probability of blocking a move.

6.4. Computational Results

In this experiment the analysis of location events which took place during workdays between 25 Sept. and 2 Oct. 2008 was performed. The area covered was limited to the part of the Warsaw-Katowice express road between Tomaszów Mazowiecki and Rawa Mazowiecka. The summary of the main configuration parameters is presented in Table 1.

Table 1
Experiment configuration and parameters

Parameter	Value
Tomaszów node latitude	51°32'58.10" N
Tomaszów node longitude	19°59'16.54" E
Rawa node latitude	51°46'4.80" N
Rawa node longitude	20°15'24.22" E
Days used for learning	25 – 26 Sept., 29 Sept. – 2 Oct.
Max. dist. from road model	5 km
Simulated congestion period	25 Sept. 3PM – 4PM
Learning step α	0.2
Slope factor β	2.0
No. of neurons in hidden layer	20

The number of neurons in the hidden layer and other parameters of neural network learning process were determined by a series of experiments. The learning process in case of different parameter values is presented in Figs. 7, 8, and 9 (the results were averaged over 10 repetition of the learning process).

The conclusion of these experiments is that 20 neurons in the hidden layer, $\alpha = 0.2$, and $\beta = 2.0$ result in a good learning speed and accuracy.

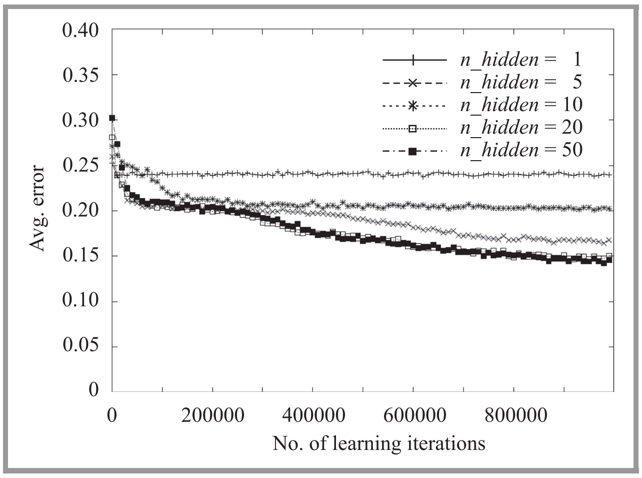

Fig. 7. Average error for different number of hidden neurons.

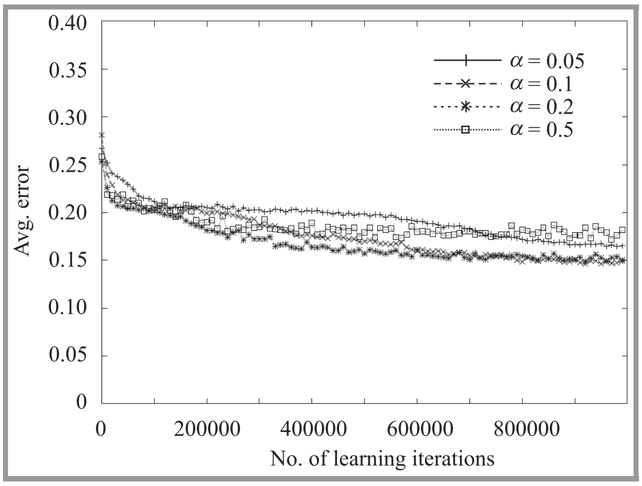

Fig. 8. Average error for different α.

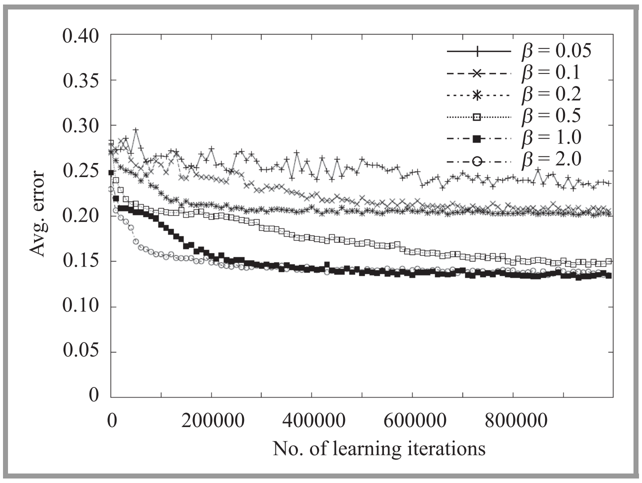

Fig. 9. Average error for different β.

In order to verify whether congestion can be detected by monitoring divergence between observed and estimated values of the move length characteristics, the following calculations were performed. For $\gamma \in \{0.25, 0.5, 0.75.1.0\}$ and different durations of congestion period (dur), a ratio of the average $\hat{d}$ estimation errors with congestion simulation and without simulation was calculated. This calculation was repeated 10 times for different, randomly chosen values of t^A_{start}. The average values are shown in Table 2.

For $\gamma = 1.0$ the error calculated for the move stream with congestion simulation is 22–30% higher in comparison to the average error in the case without congestion. As values of γ get smaller, the error ratio gets closer to 1 which means that this two cases became difficult (or even impossible) to be distinguish.

The conclusion of this computational experiment is that for high congestions it is possible to detect such situations by monitoring level of the average error of the move length characteristic estimation.

Table 2
Error ratio values

$dur[h]$	$\gamma = 1$	$\gamma = 0.75$	$\gamma = 0.5$	$\gamma = 0.25$
1	1.292630	1.250090	1.177957	1.128258
2	1.227444	1.151583	1.088391	1.026890
3	1.247943	1.188703	1.104728	1.059480
4	1.303445	1.252754	1.165497	1.097391
5	1.224714	1.156673	1.113232	1.019861

7. Conclusions and Further Work

In this paper a concept to use mobile stations location estimates to monitor status of a transport system was proposed and explained. A high level architecture of a decision support system (called VA∗) was proposed, its main functional modules outlined and their purpose described. This architecture represents an approach to analysis of input data (available as a sequence of location events) which can be used in different applications, depending on models and methods of analysis and reasoning used.

The analysis of a sample of a real-world data logged in a mobile network showed that such data can be used for this purpose. However, lack of location accuracy and substantial "information noise" present in the input data limit the number of methods which can be used.

The computational experiment, although limited to a small geographical area and using a simulation of the accident situation, showed that this approach can be effectively used for congestion detection. In this experiment a neural network was used as a function approximation tool. However, not unlike other heuristic methods some effort to tune parameters (in order) to achieve the best results is required.

Further research should include not only a neural network parameter tuning but also:

- verification whether other function approximation tools can give better results;
- defining reasoning about congestion detection from move length characteristics as a statistical hypothesis and verification of such hypothesis through statistical testing;
- what other move event stream characteristic can be proposed and how they change in case of accidents or congestion situations;
- analyzis how effective the usage of different strategies for DISCOVER module can be and how it can improve overall VA∗'s performance;
- verification whether data mining methods can be effectively used to deal with data available;
- how VA∗ can support objectives listed in Section 4 other than congestion detection.

A particular effort is planned to be made to determine what are the conditions influencing quality (e.g., speed, accuracy) of the results of VA∗'s operation.

References

[1] 3G TS 23.171 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Functional stage 2 description of location services in UMTS, Release 1999.

[2] W. Michalski, "Przegląd metod określania lokalizacji abonentów w ruchomych publicznych sieciach komórkowych GSM/UMTS z uwzględnieniem dokładności dostarczanej informacji, technicznych możliwości wdrożenia oraz czynników ekonomicznych i prawnych. Etap 1: Charakterystyka metod służących do określania lokalizacji abonentów w sieciach GSM i UMTS". Warsaw, Instytut Łączności, 2007 (in Polish).

[3] CGALIES Final Report, Report on implementation issues related to access to location information by emergency services (E112) in the European Union, 2002.

[4] W. S. McCulloh and W. Pitts, "A logical calculus of ideas immanent in nervous activity", *Bull. Math. Biophys.*, no. 5, pp. 115–133, 1943.

[5] F. Rosenblatt, *Principles of Neurodynamics*. Washington: Spartan, 1962.

[6] J. Hertz, A. Krogh, and R. G. Palmer, *Wstęp do teorii obliczeń neuronowych*. Warsaw, WNT, 1995 (transl. in Polish).

Grzegorz Sabak received the M.Sc. degree in computer science in 1998 from the Warsaw University of Technology, Poland. Currently he is working toward a Ph.D. degree. His research interests include application of heuristic methods to real-world problems, machine learning, and artificial intelligence. In his professional work he is responsible for design and management of value added services for one of the mobile network operators in Poland.

e-mail: grzegorz.sabak@home.pl
Faculty of Cybernetics
Military University of Technology
Gen. S. Kaliskiego st 2
00-908 Warsaw, Poland