Article citation information:

Monika N. BUGDOL¹, Marcin D. BUGDOL², Marcin GRZEGORZEK³, Andrzej W. MITAS⁴

ROAD TRAFFIC ESTIMATION USING BLUETOOTH SENSORS

Summary. The Bluetooth standard is a low-cost, very popular communication protocol offering a wide range of applications in many fields. In this paper, a novel system for road traffic estimation using Bluetooth sensors has been presented. The system consists of three main modules: filtration, statistical analysis of historical, and traffic estimation and prediction. The filtration module is responsible for the classification of road users and detecting measurements that should be removed. Traffic estimation has been performed on the basis of the data collected by Bluetooth measuring devices and information on external conditions (e.g., temperature), all of which have been gathered in the city of Bielsko-Biała (Poland). The obtained results are very promising. The smallest average relative error between the number of cars estimated by the model and the actual traffic was less than 10%.

Keywords: Bluetooth; ITS; traffic estimation

¹ Faculty of Biomedical Engineering, The Silesian University of Technology, Roosevelta 40, 41-800 Zabrze, Poland. E-mail: monika.bugdol@polsl.pl.
² Faculty of Biomedical Engineering, The Silesian University of Technology, Roosevelta 40, 41-800 Zabrze, Poland. E-mail: marcin.bugdol@polsl.pl.
³ Pattern Recognition Group, University of Siegen, Hoelderlinstr. 3, D-57076 Siegen, Germany. E-mail: marcin.grzegorzek@uni-siegen.de.
⁴ Faculty of Biomedical Engineering, The Silesian University of Technology, Roosevelta 40, 41-800 Zabrze, Poland. E-mail: andrzej.mitas@polsl.pl.
1. INTRODUCTION

Every year the intensity of road traffic grows constantly, with this trend unlikely to change in the near future. People spend more and more time unproductively in vehicles, which directly result in economic losses of the part of the individual and the entire economy. The search for an alternative route is very important, not only from the driver’s point of view but also for managing the traffic in this area.

The existing systems for notifying an individual driver about traffic jams, which remain the responsibility of traffic management centres, involve variable message signs installed at selected routes. This information, although extremely precise, is very often delayed with respect to the point in time at which the decision about changing the route should be made. Traffic data are collected, based on sensors mounted at appropriate points along the road network (e.g., at intersections). For this purpose, the most commonly used sensors are inductive loops, video detection, acoustic detection and radio detection [1].

Drivers who are equipped with cellular phones have access to many different navigation systems, which can provide up-to-date traffic information on a given road stretch. This allows for the most time-efficient route to be planned in advance. A major drawback of such systems is the fact that the data used to determine the traffic situation are collected from users of these applications. Therefore, the estimation accuracy is strongly dependent on the number of service users, meaning that information can be very inaccurate if there are too few users. Furthermore, some users are concerned about their privacy being invaded by IT corporations, given that information about their location is sent to location-based services, which operate mostly in continuous mode recording every position change. The most significant problem regarding all mobile applications, however, is the lack of confirmed information about their accuracy. Moreover, the results are presented in the form of colours, which are not defined in a quantitative way and may represent different speeds on different types of roads.

There are many studies on the estimation of vehicle velocity, location and travel route on the basis of different mobile phone information. The possibility of using data from mobile phones in local IT systems has been described in [2]. Location data have been gathered and tests carried out under simulation conditions. The measurements presented in [3] have been performed on a rather short section of the road (2.5 km), while the mobile phones in question must have been equipped with dedicated software. In [4, 5], it has been proven that information from the cellular network can be used to detect traffic congestion, while, in [6], an undefined wireless vehicle-to-vehicle communication protocol is used to estimate traffic density. Traffic estimation on the basis of GPS data and speed profile is presented in [7]. In [8], the same problem is resolved by employing GSM data gathered in a large database. Noisy data filtering and the estimation of vehicle localization have also been presented. The detection of dangerous road events using smartphone sensors has been described in [9].

An important disadvantage of these systems is that they use mobile networks to collect data and thus require operators’ approval. Much more flexible are solutions that are based on generally available communication protocols, such as Bluetooth or Wi-Fi, available in each new mobile device.

In [10], the suitability of Bluetooth technology is examined in terms of the provision of IT services. This study tested parameters, such as maximum connection range, delays in connection set-up and the influence of vehicle speed on communication quality. The results prove that this technology can be employed in traffic monitoring. In [11, 18], the employment of Bluetooth in vehicle-to-vehicle communication was studied in order to increase road safety. In [12], the system for car tracking using Bluetooth devices is limited to vehicles that
travel at 70 km/h at most, which is insufficient for traffic monitoring on high-speed roads. In [13], an analysis of the influence of different Bluetooth data aggregation methods on the accuracy of travel time estimation is presented. A complex system using Bluetooth data for traffic monitoring in a large city is presented in [14], which proves that this kind of technology can be successfully employed within IT services.

In this work, a novel system for road traffic estimation has been developed. It consists of several modules, while its main task is to provide information to drivers about the road traffic situation, which is evaluated on the basis of the number of detected Bluetooth devices, as well as weather conditions and time parameters.

2. METHODOLOGY

The whole system consists of a filtration module, data archiving, statistical analysis of historical data, and traffic estimation. Its task is to provide short- and long-term forecasts of road traffic using Bluetooth sensors, which have been developed in the course of another part of the project.

2.1. Filtration

Bluetooth sensors are distributed according to the user’s needs. Each of them gathers signals from all Bluetooth-enabled devices within its range. The first step is the filtration of the acquired data in order to find measurements that could corrupt the subsequent calculations.

The following road users have been identified:
- Objects that are still or moving at low speed (e.g., pedestrians or cyclists)
- Cars or trucks
- Buses

The residence time of the device within the wireless user detector (WUD) served as the basis for the classification.

Data filtering should take place at the local station to reduce the amount of transmitted information to the central repository and thus shorten the processing time to determine the parameters of traffic on the road network. In addition, a smaller amount of transferred data decreases the minimum technical characteristics of certain elements of the infrastructure, which directly affects the price of the system.

The following indications are used in formulas and diagrams:
- \( \text{ArrVeh} \) – array containing the last 20 residence times of vehicles within the Bluetooth sensor
- \( \text{ArrVehArch} \) – array containing aggregated archival residence times of vehicles within the Bluetooth sensor
- \( \text{ArrNoVehArch} \) – array containing aggregated archival number of occurrences of Bluetooth devices that have been filtered
- \( E(t) \) – value of the moving average of the residence time of vehicles within the Bluetooth sensor
- \( \alpha \) – coefficient of the moving average
- \( tVehBt \) – time period in which the Bluetooth device has been within the sensor range
Archived data are stored in the database table *archival_data*, containing average residence times for a given sensor. In each record in this array, information is stored on the type of day and time given to the nearest 5 min for which these averages have been calculated. In addition, the number of Bluetooth devices registered in the time period, which are classified as belonging to pedestrians, cyclists or other persons whose position does not change significantly over time (e.g., in buildings), is saved and stored in a data archive.

Current data are stored in the *ArrVeh* array, which has a “last in, last out” (LILO) structure. The length *N* of array *ArrVeh* can be modified depending on the individual needs at each location. Taking into account the results of the tests, however, it should be presumed that such modifications will certainly be unnecessary or, in some cases, performed only a few times.

The *ArrVehArch* array contains historical data, including *tVehBt*, which refers to the times of residence of vehicles with active Bluetooth devices in the range of the WUD sensor, aggregated every 5 min. First, these data are obtained during the observation period, but this table should be updated periodically.

The *ArrNoVehArch* array can be optionally completed after the observation period. In this case, it includes the aggregated number of devices, which, at a given time, were recorded by the sensor, but did not belong to moving vehicles.

The filtering algorithm for Bluetooth devices is as follows:

1. At the moment when a new Bluetooth device is no longer detected by the sensor, the time frame in which it had been in its range is saved to variable *tVehBt*.
2. It is then checked whether the *ArrVeh* array is completed. If not, *tVehBt* should be added to the queue. This step is carried out only in the preliminary phase of the system work.
3. Value *E(t)* is equated to 0:
   \[ E(t) = 0. \]  
   Such a situation takes place only once, after inserting the first *N* records into the *ArrVeh* array. Therefore, like step 2, this is only required in the first phase of the system work. If the value *E(t)* equals 0, it should then be set as the arithmetic mean of the first *N* records:
   \[ E(t) = \frac{1}{N} \sum_{i=1}^{N} ArrVeh(i), \]  
   before going to step 1.
4. The condition is checked:
   \[ tVehBt > k_1 \cdot E(t). \]  
   If this is fulfilled, the analysed device belongs to a pedestrian, cyclist or a person in a stationary vehicle. The record should be inserted into *ArrNoVehArch*, before going back to step 1.
5. The condition is tested:
   \[ tVehBt > k_2 \cdot E(t). \]  
   If it is fulfilled, the record must be classified as incorrect and deleted, before going to step 1.
6. If conditions 4 and 5 have not been met, the last entry from the *ArrVeh* array should then be removed, the remaining records should be shifted by one position and a new *tVehBt* should be added in the first place.
On the basis of data from the ArrVeh array, the moving average $E(t)$ is computed. It is a weighted average with exponential weights, the value of which is calculated according to the following formula:

$$E(t) = \alpha \cdot C(t) + (1 - \alpha) \cdot E(t - 1),$$

where:
- $\alpha$ – coefficient of the moving average
- $C(t)$ – value of the element with index $t$
- $E(t - 1)$ – weighted average from $t - 1$ periods

Coefficient $\alpha$ of the moving average is calculated using the following formula:

$$\alpha = \frac{2}{N+1},$$

where:
- $N$ – array length

Bus detection should be performed.

### 2.1.1. Bus detection

Bus detection should proceed as follows:

1. For each analysed time $t_{VehBt}$, it should be checked whether, in less than $t_{BUS}$ seconds, at least $k_3$ MAC addresses of Bluetooth devices occurred. If this condition is not fulfilled, then $t_{VehBt}$ does not belong to a device inside a bus; otherwise, proceed to step 2.

2. It should be checked whether all MAC addresses from step 1 have stopped being visible in a time interval shorter than $t_{BUS}$ seconds. If that condition is fulfilled, all these MAC addresses belong to devices travelling in a bus; otherwise, $t_{VehBt}$ does not belong to a device in a bus.

### 2.1.2. Data update

Data update should be performed daily using only those $t_{VehBt}$ values that have been registered when there was no traffic congestion. For each aggregation time period, the mean value of the time when the vehicle is within the WUD range ($E_{arc}$) should be updated according the formula:

$$E_{arc,new} = 0.1 \cdot E_{current} + 0.9 \cdot E_{arc}.$$  

### 2.2. Models

Models should be individually constructed for each sensor. Their correction can take place any time, but it is not recommended to carry this out more frequently than the database update. The resulting linear models take the following form:

$$\hat{y} = a \cdot v,$$
2.2.1. Adding a new variable

The procedure should begin with the addition of a new variable containing information about the temporal distance from the peak hours: dist\_rush\_h. In order to compute it, the peak hours for a given location must first be determined. The average number of vehicles for each time interval between 6 am and 10 am must be calculated; the one for which the maximum is reached must be indicated at the morning rush hour mrh. A similar procedure is performed when evaluating the afternoon rush hour arh, but the averages are calculated for points in time between 2 pm and 6 pm. Next, a new variable should be added in each record in the database:

\[ \text{dist\_rush\_h} = \min[|h - \text{mrh}|, |h - \text{arh}|], \]

where:
\( h \) – hour in the given record

2.2.2. Variables selection

The selection of variables is performed independently using two algorithms: Hellwig’s method [15] and stepwise forward selection [16]. If both methods give the same results, there is no need to assess the quality of fit of the model to empirical data [17].

Let \( k \) denote the number of potential exogenous variables. After adding dist\_rush\_h, there are \( m \) variables:

\[ m = k + 1, \]

Since, in contrast to Hellwig’s method, the stepwise forward selection is a popular tool, only the latter will be described below. In Hellwig’s method, all possible subsets of potential variables are analysed, with the best combination selected on the basis of integral capacity indicators \( H \). The higher the value of the integral capacity indicator for a given variable combination, the better is this combination in terms of the built model.

First, the number \( L \) of all possible combinations is computed:

\[ L = 2^m - 1, \]

All possible variable combinations must be determined (from \( m \) one-element combinations to one \( m \)-element combination). Next, the individual information capacity \( h \) must be evaluated:

\[ h_{kj} = \frac{r^2_{kj}}{\sum_{i=1}^{m_k} |r_{ij}|}, \quad j = 1, 2, ..., m_k \]

where:
\( k \) – combination number
\( m_k \) – number of variables in the \( k \)-th combination
\( j \) – variable number in the given combination
The following step is to compute the integral capacity for the given combination:

\[ H_k = \sum_{j=1}^{m_k} r_{kj}, \quad h = 1, 2, \ldots, 2^m - 1, \quad H_k \in [0; 1] \] (13)

The best variable combination \( k_{max} \) is chosen as follows:

\[ k_{max} = \arg \max_k H_k \] (14)

### 2.2.3. Building the model

For both sets of variables, selected with the methods mentioned above, two models are built, that is, with and without the constant term (maximum of four models; the number can decrease if two methods give the same results). The values of the model parameters \( a_i \) are estimated using the least-squares method:

\[ a = (x^T x)^{-1} x^T y \] (15)

where:
- \( a \) – parameters matrix
- \( x \) – exogenous variables
- \( y \) – reference data

The model for which the highest adjusted coefficient of determination \( \tilde{R}^2 \) has been obtained should be chosen:

\[ \tilde{R}^2 = 1 - \frac{n \cdot (1 - R^2)}{n - m - 1} \] (16)

\[ R^2 = \frac{\sum_{t=1}^{n} (\hat{y}_t - \bar{y})^2}{\sum_{t=1}^{n} (y_t - \bar{y})^2} \cdot (1 - R^2) \] (17)

where:
- \( \bar{y} \) – mean value of \( y \)
- \( y_t \) – actual number of vehicles at moment \( t \)
- \( \hat{y}_t \) – theoretical number of vehicles at moment \( t \) (from the model)
- \( n \) – number of measurements (records)
- \( w \) – number of exogenous variables

### 2.2.4. Predicting the number of cars

The estimation of the number of vehicles is conducted every 5 min on the basis of historical data, aggregated in 5-min intervals, as well as the current data collected from the sensors and subjected to filtration.

From the reference data, the average percentage of vehicles with an active Bluetooth transmitter for the moment \( t_c \) and for the moment \( t \pm 5 \) should be read. These values will be denoted as \( b_{(t-5)}, b_t, b_{(t+5)} \) (these are already averaged values for the given hour). Next, the arithmetic mean \( \bar{b} \) should be computed:

\[ \bar{b} = \frac{b_{(t-5)} + b_t + b_{(t+5)}}{3} \] (18)

Using the model, the estimated number of cars \( \hat{s} \) for given atmospheric conditions and active Bluetooth devices is computed. Finally, the number of cars \( \bar{x} \) is calculated as:
\[\bar{x} = \frac{(n_{BT} + y)}{2}, \quad (19)\]

where:
\(n_{BT}\) – number of vehicles with active Bluetooth devices observed in the previous 5 min

3. RESULTS

Bluetooth and weather sensors were distributed in 40 places in Bielsko-Biala, Poland. After a preliminary observation period of two months, four locations were chosen for reference data acquisition. These data were collected in a time period of one month. The stored aggregated information included the number of Bluetooth devices, date, starting hour of the aggregation interval, average air temperature, average road temperature, average humidity, average air pressure, average salinity, freezing point, dew point, water film, average precipitation, and the number of Bluetooth devices registered by the sensor (after filtration).

For filtration purposes, parameters were set on the basis of the authors’ observations as follows:

\[N = 20, \quad k_1 = 2, \quad k_2 = \frac{1}{2}, \quad k_3 = 5, \quad t_{BUS} = 5, \quad (20)\]

Yet, these can be modified for each new location if data obtained during the observation period indicate the need to do so.

Outlier elimination was performed using the Dixon Q test. As a result, four records were deleted from the database, since incomplete data were not considered.

The prepared reference data, combined with the according input data, were randomly divided into the training set and the test set in proportions of 80/20, 85/15 and 90/10. Modelling was then performed on each training set, with the resulting models evaluated on appropriate test sets. For every sensor and proportion, this procedure was repeated 10 times, after which the results were averaged. Table 1 presents the obtained relative errors for the selected sensors.

<table>
<thead>
<tr>
<th>Training set percentage</th>
<th>Test set percentage</th>
<th>Relative error [%]</th>
<th>Sensor</th>
<th>Sensor</th>
<th>Sensor</th>
<th>Sensor</th>
</tr>
</thead>
<tbody>
<tr>
<td>80%</td>
<td>20%</td>
<td>16.4%</td>
<td>14.8%</td>
<td>15.4%</td>
<td>15.2%</td>
<td></td>
</tr>
<tr>
<td>85%</td>
<td>15%</td>
<td>13.1%</td>
<td>12.2%</td>
<td>13.2%</td>
<td>10.6%</td>
<td></td>
</tr>
<tr>
<td>90%</td>
<td>10%</td>
<td>8.6%</td>
<td>8.2%</td>
<td>7.7%</td>
<td>8.6%</td>
<td></td>
</tr>
</tbody>
</table>

For all analysed sensors, the results exhibit a similar characteristic. The relative error decreased from about 15-16% to about 8-8.5%, while the share of the training set changed from 80% to 90%. When the test set included 15% of all available data, the relative error was approximately 12-13%.
The strongest influence on the relative error involved aggregation periods where there was a small number of cars, given that, in this case, even a small absolute error could have been the source of a large relative error.

4. CONCLUSIONS

In this paper, a novel road traffic estimation system using Bluetooth sensors has been presented. Data collected by Bluetooth sensors, supplemented with data describing the external conditions, serve as a basis for the assessment of the number of cars passing a sensor location in a given time period.

The obtained results are very promising. The smallest average relative errors (less than 10%) occurred when 90% of the collected data served as the training set, which leads to the conclusion that the length of the observation period has a significant impact on the accuracy of the prognosis. If there is a possibility to repeat the observation period, then the model should be retrained and recalculated in order to provide the best possible estimations.

Acknowledgements

The system described has been developed as a part of the Multimodal System of Road Traffic Monitoring project, founded by the National Centre for Research and Development in Poland, in the context of the INNOTECH-HiTech programme.

References


Received 11.05.2017; accepted in revised form 15.08.2017

Scientific Journal of Silesian University of Technology. Series Transport is licensed under a Creative Commons Attribution 4.0 International License