Applied Computer Science, vol. 12, no. 4, pp. 18–27

Submitted: 2016-11-30
Revised: 2016-12-12
Accepted: 2016-12-15

potholes, accelerometer, road quality, cloud computing

Marcin BADUROWICZ[*], Tomasz CIEPLAK[**],
Jerzy MONTUSIEWICZ[***]

ON-THE-FLY COMMUNITY-DRIVEN MOBILE ACCELEROMETER DATA ANALYSIS SYSTEM FOR ROAD QUALITY ASSESSMENT

Abstract

In this paper the authors are discussing a community-driven system for reporting via smartphones road acceleration data, processed on-the-fly in the cloud computing system for finding possible road artefacts as well as assessing overall road quality on the driver-friendly RRUI scale. The proposed system uses smartphones mounted in a car with little to no calibration or initial setup. By performing a fast analysis in the cloud, data are made immediately available for other users. The system continuously sends to end users' devices data about road quality issues "in exchange" for acceleration information.

1. INTRODUCTION

Road conditions affect road users in different ways: fuel efficiency, vehicle and driver fatigue, and user comfort. Published data about the state of roads may be used to optimize traffic, but also in professional transport, for example chassis configuration for trucks (Lundstrom, 2009). The negative impact

[*] Institute of Computer Science, Faculty of Electrical Engineering and Computer Science, Lublin University of Technology, Nadbystrzycka Str. 36D, 20-618, Lublin, Poland, m.badurowicz@pollub.pl
[**] Management Faculty, Lublin University of Technology, Nadbystrzycka Str. 38A, 20-618 Lublin, Poland, t.cieplak@pollub.pl
[***] Institute of Computer Science, Faculty of Electrical Engineering and Computer Science, Lublin University of Technology, Nadbystrzycka Str. 36D, 20-618, Lublin, Poland, j.montusiewicz@pollub.pl

of different types of road artefacts – potholes, speed bumps and similar is also important (Badurowicz & Montusiewicz, 2015). Information about road condition is also a crucial factor in decisions as to which roads should undergo maintenance in the next round and how they degrade with time, important for sparsely populated regions.

There are currently discussed standards on reporting road surface profiles – in the Polish law system there is an instruction provided (Duda & Adamczyk, 1999) for assessment of the technical quality of the road, which must be provided by the road operator for preparing strategic plans of improving road quality, while assessment is provided by a group of experts using paper protocols, rather than automatically, while better methods using current technology were proposed by Gajewski (2013). This kind of methods, while useful for road operators, experts and governmental units, are however not so useful for average road users, as they require additional knowledge to be properly analyzed, and as this kind of data is usually not publicly distributed to road users on demand. It is also impossible for expert groups to continuously monitor road quality on a vast area of the country.

The authors propose the use of smartphones mounted in road users' cars for basic road quality monitoring. Smartphones are much more prevalent than any kind of specialized equipment, they have a variety of built-in sensors and communication units. Based on this, the authors propose a crowdsourcing project using personal smartphones to record accelerometer data for road quality assessment.

2. DESIGN AND MOTIVATIONS

The general idea is based on the road artefact detection concept the authors proposed in earlier works (Badurowicz & Montusiewicz, 2015). In this paper, the authors concentrate on streaming analysis of data coming from road users' devices.

The key concept of the proposed solution is to use road users' smartphones mounted in their cars in a way that the device can record acceleration data of the car. This means the device has to be mounted in a stable hold, and it is the only requirement: device orientation however is not important, as the authors are using an orientation-agnostic way of acceleration data recording by means of a rotation matrix calculated from the array of the device's sensors. This means that little to no calibration is required for the end user to use the system. The main sensor device used is the accelerometer, which reacts to shakes from the road surface, going through all the car's parts. Taking into account the results of all the previously performed experiments, which used different cars from different manufacturers and in a different overall state, there is always a direct connection between data recorded by the accelerometer and road profile (Badurowicz & Montusiewicz, 2015; Astartita et al., 2014).

The current solution is acquiring data on acceleration in all three axes, acceleration data after rotation matrix transformation, current location, speed, course, time (synchronized with Network Time Protocol server) and GPS (Global Positioning System) accuracy. Data from many devices, which is actually data from multiple cars, are processed in the cloud-computing system in streaming analysis fashion. To achieve communication between the devices and the computing system, the MQTT protocol is used.

The authors are performing different kinds of analysis on the data received by the system, where one of the key factors of road quality estimation is to examine the quantity and intensity of road artefacts - various kinds of potholes, road surface damages and similar, as well as the overall state of the road, which is information taken directly from the statistical analysis of the collected signals.

As similar techniques for using accelerometers for International Roughness Index calculation were previously used in (Du & Lin & Wu & Jiang, 2014) and (Hanson et al., 2014), the authors believe that end users, to whom the system will be offered, need more user-friendly ways of calculating road quality than professionals, and a simple scale allowing for quick understanding while driving. The concept of a social system for road defect detection was previously suggested by Aksamit and Szmechta (2011), however in the present authors' solution there is a significant difference – data are not only being analyzed in stream analytics fashion in the cloud computing system, but the resulting information is also being sent back to the device. Also, the whole concept of road condition monitoring was also discussed earlier (Chen at el, 2011), however in the authors' proposal the sending and receiving device is just the simple smartphone, with everything integrated in the software, not relying on external sensors.

3. PRINCIPLES OF ON-THE-FLY DATA ANALYSIS

The data obtained from the mobile devices produce a continuous stream. Hence, the amount of information grows over time. Not every event produced by the mobile sensors is important but each one must be analyzed. The number of mobile devices included in the system is not limited, so the amount of events passing through the system is difficult to predict. Making such assumptions, the overall system can be described with two models: as Big Data and as the Internet of Things.

In general, there are two main models for big data computing. The first one is big data batch computing (BDBC), where the flagship solution is Hadoop. The second model is called big data stream computing (BDSC) (Sun & Zhang & Zheng & Li, 2015). In the case of BDSC, the straight-through computing model has implementations in such systems as: Apache Storm, S4, Akka and IBM InfoSphere Streams. In contrast to batch computing, where the data are first stored and then computed, stream computing is sufficient for many real-time application scenarios, where the result must be updated every time the data change. In the case of the real-time applications where the input stream incurs multistaged computing, stream computing ensures low latency to produce an output stream (Hummer & Satzger & Dustdar, 2013).

A continuous data stream is an infinite sequence of data sets, and parallel streams produce more than one stream at the same time. A big data input stream has the characteristics of high speed, real time, and large volume for applications, like the system described here.

In an endless stream of data, the data flow can be partitioned into subgroups, called windows. Depending on the solution, a number of types of windows can be distinguished. Each window has a start condition and an end condition. The basic type of query window is the growing window which binds data values to be available over the entire stream. In contrast, the single item window is where each single event is considered separately. The third type is the tumbling window, where a new window is created if there is currently no other open window. The tumbling window fills with data, performs whatever operator is specified, and then moves to an entirely new window of data. It tumbles forward trough the data, end over end. The next type of window is the sliding window, where it fills with its first subgrouping of data, performs whatever it is an element of, then continues with these same data, adding them incrementally, or sliding, as new data are received. The last type of window is a landmark window query, where once a window is opened, it remains open indefinitely. In contrast to the growing window which binds some values over the entire stream, the landmark window considers different portions of the stream over time. The stream data can be promoted into windows by defining a number of conditions, such as row count, elapsed time or marker in the stream.

4. ROAD QUALITY ESTIMATION FROM ACCELERATION DATA

To evaluate the quality of the road travelled an accelerometer was used, located inside a conventional mobile device, i.e. a smartphone. This examination of the quality of the road leads indirectly to determining the comfort of driving along this road.

The application for the smartphone collects the data indicating acceleration in the global coordinate system associated with the current sample time and the current location from the GPS system. In the experiment we used constant sampling frequency of 10~Hz. This value was chosen on the basis of the ana-lysis of the speed of the vehicle on the roads of various types and after conducting preliminary studies. Moving on the road of bad quality was possible at speeds not exceeding 30 km/h, and on seeing the different types of road damage, e.g. potholes, the driver had to perform a braking maneuver and, if possible, a partial bypass.

In such situations, the vehicle speed was reduced to about 10-20 km/h. On the application of sampling it was possible to record waypoints distant from each other about 28 to 85 cm. While driving along the city's ring road, the car reached the speed of about 90 km/h – the obtained sample corresponded so sections of the road at 250 cm. Preliminary studies have confirmed that on roads in very good condition there is no need to collect information at a higher sampling rate, since there are no artefacts that should be noted.

The developed road test methodology included the following activities:

- choice of appropriate sections of the road for the research,
- road of very good quality (Lublin ring road),
- a gravel road (Snopków village near Lublin),
- paved road - so-called cobblestones (Lublin, Agronomiczna Street).
- a test drive along the route, verification of its state,
- selection of software parameters for the registration of the road's condition,
- sending the data to the computer cloud.

For the purpose of getting information of interest to end users, the authors decided to perform road artefacts (potholes, speed bumps, road surface quality changes) detection from a data stream, by using the streaming computing method. The artefacts are "deviant" points in the Z-axis acceleration data stream. To find those points a design pattern called "outlier detection" was applied (Ballard et al., 2012). The computing solution was built by applying the IBM Infosphere Streams system. The aim of the solution is to forward only those tuples (data points) that are considered outliers. The outliers are identified by computing the distance of a data value from a mean. The mean is calculated over a sliding window the size of 25 tuples.

The algorithm considers a data point as an outlier when the Z-axis acceleration value is higher or lower than the sum of the average value of Z-axis acceleration and the product of standard deviation of the Z-axis acceleration value multiplied by sigma. The value of sigma was established experimentally and equals 2.25 in the case of the present research. The average value and standard deviation of the Z-axis acceleration were calculated by using functions provided by the Aggregate operator (see Fig. 1-B). The Aggregate operator is a part of stream the diagram depicted in Figure 1.

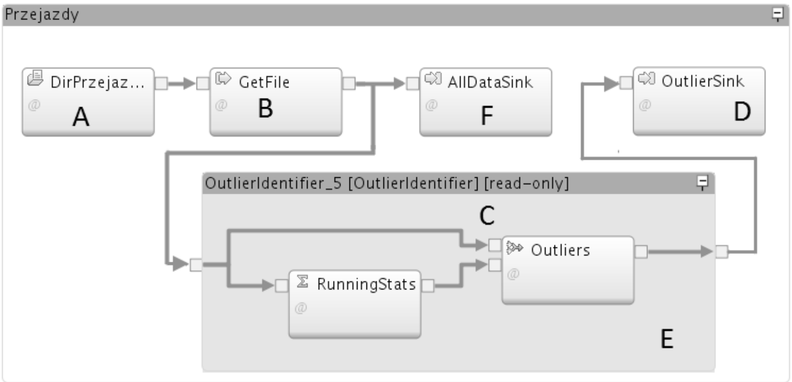
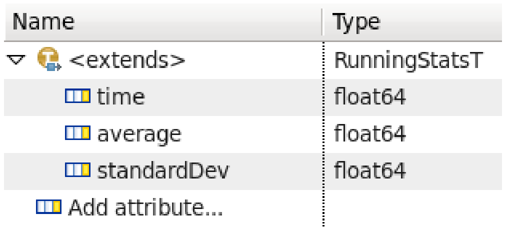


Fig. 1. Stream graph representing the system for outlier detection (own study)

The ingest of the stream consumes all the data from the input port – the input port of the whole system is built upon the MQTT source operator (see Fig. 1-A). It connects with the MQTT broker and receives information produced by the device placed in the car. Then data flows into the OutlierIdentifier operator (see Fig. 1-E) as well as into the AllDataSink operator (see Fig. 1-F). The latter is used just for debugging purposes, to accumulate all the data received from the MQTT source.

In the OulierIdentifier operator the stream is divided into two substreams. The first one goes into the input port of the Aggregate operator where the mean and standard deviation are computed. The outcome of the Aggregate operator produces a substream with tuples described by the schema depicted in Figure 2.



Name	Type
<extends>	RunningStatsT
time	float64
average	float64
standardDev	float64
Add attribute...	

Fig. 2. Data scheme of the outcome of Aggregate operator (own study)

To get the outliers, both streams – all the data and substream received from the Aggregate operator – must be merged into one, containing only outlier tuples. To achieve this, the Join operator was used (see Fig. 1-C). Join operator contains the isOutlier function (see Listing 1). The isOutlier function computes the Z-axis acceleration distance from the mean. If it returns true, the data is an outlier and is forwarded to the output. Then outliers are placed in the OutlierSink (see Fig. 1-D).

5. EXPERIMENTAL RESULTS

All data received from mobile devices were analyzed but only three types of roads are going to be shown in this part of the article. To clarify the explanation in Figure 3 only four seconds of the test drive are shown. On the X-axis the time in unix timestamp format is presented, the Y-axis depicts the acceleration. Standard deviation was computed in two ways: the first one was obtained over all data of the research sample – the sample contains 5008 measuring points. In this case the standard deviations equals 0.2003. The second way of standard deviation calculation was done during streaming analysis. It is not a constant in time, but depends on calculation in a window and changes acceleration value along the Z-axis. Every new tuple coming into the window removes the oldest one and at that time a new value of the mean and standard deviation are produced.

In Figure 3 the data points received from the research are marked as grey stars, constant deviation is marked as a Y bar error and standard deviation computed over the window is marked as a black line with triangle marks. The output of the streaming analysis – outlier data – is marked as black dots. When the road quality decreases the slope of the window, the calculated standard deviation line increases.

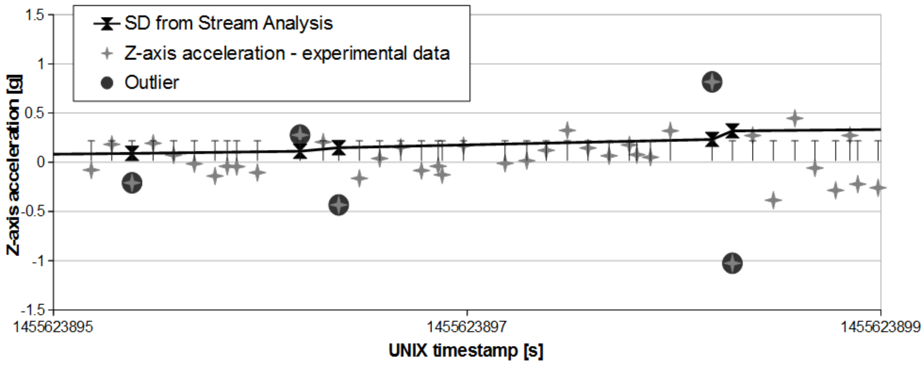


Fig. 3. Measurements of data quality represented by Z-axis acceleration vs. time for a gravel road (own study)

On the following three figures outliners are going to be shown for three different types of roads. To simplify the observation only outliers are depicted on the graph. The time duration for each graph equals 100 s. Thereby, the number of measurements points equals 1000.

In the case of very good road quality (see Fig. 4) only a small number (14 points) of outliers was found. The acceleration value over the Z-axis is close to zero and the value of the outliers does not even exceed 0.5g.

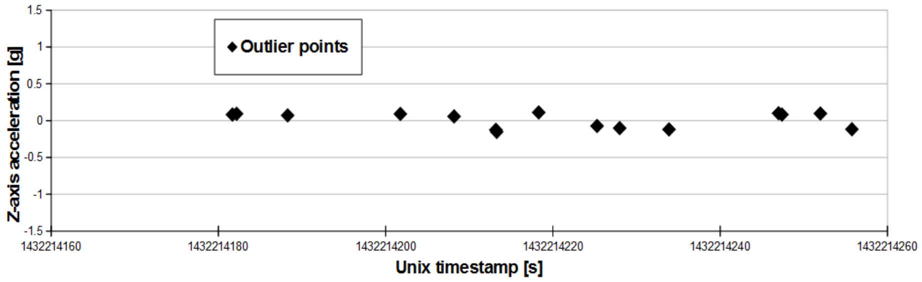


Fig. 4. Outlier points for a road of very good quality (own study)

With deterioration in the quality of the road surface in the case of a paved road – so-called cobblestones (see Fig. 5), the number of outliers increases and reaches 31 points. At the same time the value of Z-axis acceleration increases, and in some points surpasses the value of 0.5g.

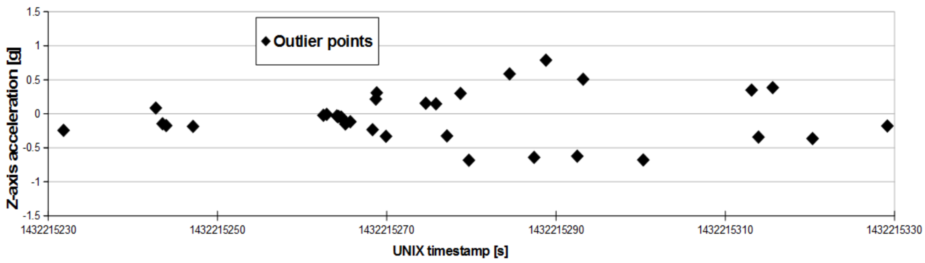


Fig. 5. Outlier points for a paved road – so-called cobblestones road (own study)

In Figure 6 the worst road surface examined is shown. Compared to the previous two samples, the gravel road presented here is characterized by a high number relative to the paved road, of the outlier points (49 of them). With a higher number of distinguished outlier points, their value of Z-axis acceleration exceeds the value of 1.0 g.

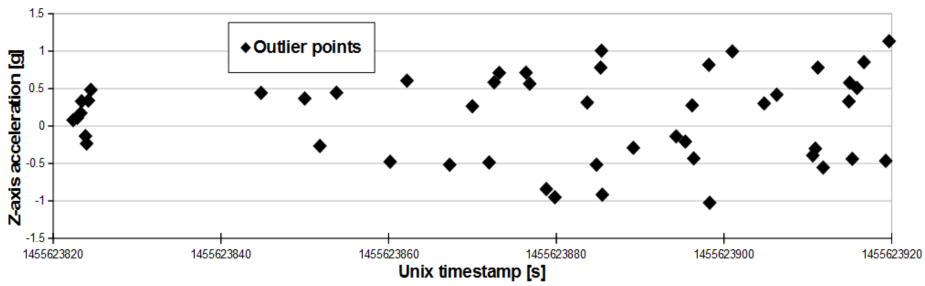


Fig. 6. Outlier points for gravel road (own study)

5.1. The assessment scale

While the position of individual road artefacts is important for the driver, the overall state of the road is also a must. Therefore, a road classification may be proposed, similar to the one presented by Turkay and Akcay (2015) – the same road classes from A-H will be used, where A is a high quality road and H is the lowest quality (terribly poor) road, but the value will be based on standard deviation of Z-axis acceleration. Both the A and H class roads were tested experimentally, and the authors will use the A road standard deviation on the Z axis as the baseline for any other kind of road, using the formula:

$$RRUI = (\sigma_Z - A) \cdot 10^2 \qquad (1)$$

The RRUI, or Road Relative Unevenness Index, is calculated as the difference between the standard deviation of acceleration in vertical axis from the road fragment being assessed (σ_Z) and the average standard deviation of an A-class road (Lublin ring road) (A), multiplied by 100 to achieve more readability. Based on streaming analysis principles, the authors asses not whole roads, but their fragments which are in one sliding window of the cloud computing analysis.

The gravel and paved roads acceleration data were statistically analyzed in the second stream analytics system, this time built by using the Tumbling Window technique on Microsoft Azure platform, again with the window size of 25 tuples. The values calculated from the first road are presented in the table 1, below:

Tab. 1. Calculated statistic values from Z-axis and RRUI for a road fragment (source: own study)

Fragment	Average acceleration	Standard deviation	RRUI value
1	-0.02529	0.062724	1.748547
2	-0.01818	0.057316	1.207696
3	-0.01836	0.045599	0.036032
4	-0.01305	0.058717	1.34784
5	-0.02467	0.050069	0.483021
6	-0.01938	0.191661	14.64223
7	-0.02452	0.166082	12.08434

The calculated RRUI is 14.64 at its highest. The same analysis was performed for the gravel road and the RRUI was up to the value of 45.03.

The proposed scale may be directly transformed into A-H alphabetic scale, allows end users to quickly assess road fragments directly ahead of them, when the feedback information is being sent to their devices. The outlier detection finding hazardous road artefacts, on the other hand, warns users about dangerous places and advises them to slow down.

6. CONCULUSIONS AND FUTURE WORK

The authors performed streaming analytics, on-the-fly and without human intervention, of road acceleration data for finding different types of road artefacts and for assessment of road quality. This information is being sent to end-user devices who are participating in the social system, so in exchange for the road acceleration data they are being warned about incoming road quality issues.

The proposed RRUI scale is not the defining index of the road quality – there is a need for preparing a multi-factor road assessment, using the overall road quality from the RRUI scale, but also density and factor of the road artefacts found, and the creation of this kind of scale, useful for end users, will be the authors' target in the future.

REFERENCES

Aksamit, P., Szmechta, M. (2011). Distributed, mobile, social system for road surface defects detection. In: *Computational Intelligence and Intelligent Informatics (ISCIII)* (pp. 37-40).

Astarita, V., Vaiana, R., Iueleand, T., Maria, V.C., Giofre, V.P., Masi, F.D. (2014). Automated sensing system for monitoring of road surface quality by mobile devices. *Procedia - Social and Behavioral Sciences 111*, 242-251.

Badurowicz, M., Montusiewicz, J. (2015) Identifying road artefacts with mobile devices. *Information and Software Technologies, vol. 538*, Springer

Ballard, C., et al. (2012). *IBM InfoSphere Streams: Assembling Continuous Insight in the Information Revolution*. IBM Redbooks.

Chen, K., Lu, M., Fan, X., Wei, M., Wu, J. (2011). Road condition monitoring using on-board three-axis accelerometer and gps sensor. In: *Communications and Networking in China (CHINACOM)* (pp. 1032-1037).

Du, Y., Liu, C., Wu, D., Jiang, S. (2014). Measurement of international roughness index by using z-axis accelerometers and gps. *Mathematical Problems in Engineering 2014*

Duda, J., Adamczyk, M. (1999). *Instrukcja oceny stanu nawierzchni dróg publicznych w miastach*. Generalna Dyrekcja Dróg Publicznych

Gajewski, W. (2013). Nowoczesne metody diagnostyki stanu, czyli profilaktyka w drogownictwie. *Magazyn Autostrady Nr 10,* pp. 46-51.

Hanson, T., Cameron, C., Hildebrand, E. (2014). Evaluation of low-cost consumer-level mobile phone technology for measuring international roughness index (iri) values. *Canadian Journal of Civil Engineering 41(9)* (pp. 819-827)

Hummer, W., Satzger, B., Dustdar, S. (2013). Elastic stream processing in the cloud. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 3(5)* (pp. 333-345)

Lundström, J. (2009). Road roughness etimation using available vehicle sensors

Sun, D., Zhang, G., Zheng, W., Li, K. (2015). Key technologies for big data stream computing

Turkay, S., Akcay, H. (2015). Spectral modeling of longitudinal road profiles. In: *2015 IEEE 28th Canadian Conference on Electrical and Computer Engineering (CCECE)* (pp. 477-482)