Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 121 (2017) 415–422

Procedia Computer Science

www.elsevier.com/locate/procedia

CENTERIS - International Conference on ENTERprise Information Systems / ProjMAN - International Conference on Project MANagement / HCist - International Conference on Health and Social Care Information Systems and Technologies, CENTERIS / ProjMAN / HCist 2017, 8-10 November 2017, Barcelona, Spain

Anomaly Detection in Roads with a Data Mining Approach

Nuno Silva[a*], João Soares[a], Vaibhav Shah[a], Maribel Yasmina Santos[a], Helena Rodrigues[a]

[a]*ALGORITMI Reasearch Centre, Universidade Do Minho, Azurém Campus, 4804-5333 Guimarães Portugal*

Abstract

Road condition has an important role in our daily live. Anomalies in road surface can cause accidents, mechanical failure, stress and discomfort in drivers and passengers. Governments spend millions each year in roads maintenance for maintaining roads in good condition. But extensive maintenance work can lead to traffic jams, causing frustration in road users. In way to avoid problems caused by road anomalies, we propose a system that can detect road anomalies using smartphone sensors. The approach is based in data-mining algorithms to mitigate the problem of hardware diversity. In this work we used scikit-learn, a python module, and Weka, as tools for data-mining. All cleaning data process was made using python language. The final results show that it is possible detect road anomalies using only a smartphone.

Peer-review under responsibility of the scientific committee of the CENTERIS - International Conference on ENTERprise Information Systems / ProjMAN - International Conference on Project MANagement / HCist - International Conference on Health and Social Care Information Systems and Technologies.

Keywords: Road Anomalies, Data Mining, Data Analitics

1. Introduction

Road condition has an important role in our daily live. Anomalies in road pavement can cause vehicle mechanical failure or even accidents. Also, roads in bad conditions can affect driver and passenger comfort and increase stress.

* Corresponding author. Tel.: +351 967161737.
E-mail address: b7204@algoritmi.uminho.pt

1877-0509
Peer-review under responsibility of the scientific committee of the CENTERIS - International Conference on ENTERprise Information Systems / ProjMAN - International Conference on Project MANagement / HCist - International Conference on Health and Social Care Information Systems and Technologies.
10.1016/j.procs.2017.11.056

Finally, governments spend millions of euros in road maintenance, which can cause traffic jams, increase driver stress, fuel consumption and cause delays in our daily live.

A road condition/anomaly tracking system can help governments planning and scheduling road maintenance in a way that decreases the impact in people daily live, while, an anomaly warning system for drivers can increase anomaly awareness, which may decrease the number accidents, costs in car maintenance and driver stress. Also, when planning a trip, information about roads condition can increase comfort both in driver and passengers[1,2].

However, detecting and surveying road condition/anomalies requires expensive and specially designed equipment and vehicles that cost considerable amounts of money, while also requiring specialized workers to operate them[3]. On the other hand, most drivers have smartphones, which are equipped with several different sensing capabilities, namely inertial and location sensors[4]. Also, smartphones are part of our everyday lives, with most people owning, at least, one device. Thus, these devices could be used as road monitoring equipment, "transforming" every driver/car into specialized road-monitoring technicians/vehicles.

In this work, we describe a system that, based on data collected by smartphone sensors, automatically detects and classifies road anomalies using data mining algorithms. Specifically, we describe the methodology used for the data mining process, and discuss some problems and challenges resulting from this process.

However, the use of smartphone devices imposes several problems and limitations when compared to the more expensive and specialized road-monitoring equipment. Specifically, smartphone devices use less expensive sensors that typically offer less sensitivity and precision when compared to specialized equipment. This is especially true for inertial and GPS sensors. Additionally, the huge variety of smartphone and sensor makes and models result in a data heterogeneity problems, since different sensors typically measure different values for the same physical action[3].

This heterogeneity is not only related to smartphones and sensors. The automotive park suffers from the same problems, with many different car makes and models, distinct suspension types and tires. All this has direct influence and impact on the collected data, thus existing solutions that detect and categorize road anomalies based on pre-defined thresholds are not suited for real world scenarios[5,6,7]. Additionally, GPS sensors of smartphones typically have an accuracy error ranging from a couple to tens of meters[8,9], meaning that the same physical location can have distinct coordinates (latitude and longitude) in different acquisitions. Furthermore, driving speed has a considerable impact in vehicle vibration, creating a higher vibration amplitude in accelerometer data[5,10].

2. Related Work

When we talk about road condition monitoring systems, we typically talk about data centered systems. Data is typically collected by devices and analyzed to better understand the problem in hand, and to acquire knowledge about it. However, raw data is difficult to analyze and hard to understand. Data needs to be transformed in order to be more comprehensive, increasing its usability in data mining algorithms.

In literature, we see many different approaches to solve these problems. In the data acquisition stage, labeling is one of the hardest to solve. Seraj et al.[11] uses video cameras to analyze road condition and a subtitle editor to annotate anomalies. Speed is another problem to take into consideration, as GPS signals can be erratic. Also, routes need to be well defined as to avoid high traffic and pedestrian volumes, not only due to safety reasons, but also to prevent data quality from being compromised by these interferences. Finally, diversity of vehicles, drivers and recording/sensing devices are important measures that need to be taken for preventing algorithm overfitting.

In the data cleaning and transformation stage, one needs to clean and process erratic data, and know which features must be extracted from data. When collecting data from accelerometer sensors, a considerable amount of noise is collected due to events like, slamming doors or vibration caused by motor running. To remove this noise Vittorio et al.[1] and Mukherjee et al.[12] use a low pass filter. Douangphachanh et al.[13] use a high pass filter, based on the Android Developer Reference, to remove gravity from all accelerometer axis, combined with a low pass filter. For erratic data, Chen et al.[14] deals with it by deleting it or using interpolation to fill that gaps. For extracting features, Nomura et al.[4] calculates vertical acceleration variances, while Mukherjee et al.[12] uses acceleration means and Bello-Salau et al.[15] uses standard deviation of accelerometer data, and presents a new measure called "z-difference square", that uses the square of z-axis magnitude(Z0- Z1) in order to get only positive values. Rajamohan et al.[16] use standard deviation of all three accelerometer axis. Others use more complex algorithms to do signal analysis. Seraj et al.[11] uses Stationary Wavelet Transformation (SWT) at four levels of decomposition to analyze data from gyroscope sensors Joubert et al.[17] use Kalman filters, and Mukherjee et al.[12] use Tturn, Tmin, Tbank and Twin filters to help reducing the number