Mobile road pothole classification and reporting with data quality estimates

Conference Paper · February 2018

DOI: 10.1109/MOBISECSERV.2018.8311437

L. Reznik
Rochester Institute of Technology

Igor Khokhlov
Rochester Institute of Technology

Mobile Road Pothole Classification and Reporting with Data Quality Estimates

Ayush Vora, Leon Reznik, Igor Khokhlov
Department of Computer Science,
Rochester Institute of Technology
Rochester, New York 14623–5603
Email: arv1924@rit.edu, lr@cs.rit.edu, ixk8996@rit.edu

***Abstract*—Harnessing the power of mobile computing platforms has opened up newer possibilities of gathering and classifying data by leveraging the use of crowd-sourcing. In the current generation that is being dominated by the mobile computing platform namely smartphones, crowd-sourcing can be achieved in a relatively hassle-free yet effective means of collecting data from a large set of ordinary users. Tapping into this source pool, however, has a drawback of unspecified data quality and security and hence the lack of trust in the data collected. This paper proposes an approach that aims to realize the most influential factors related to destructiveness of potholes that are encountered on roadways, while supplementing the data with quality estimates derived from the completeness of the data and security and quality factors of the input device itself that is involved in the data collection process, thereby addressing the drawback of data trust. The mobile application design is described. The application use cases are presented and discussed.**

1. Introduction

The problem of roadway conditions has persisted to this date. Poor maintenance can result in numerous potholes and lead to harm that may range from vehicular damages to the loss of human lives. Since this is an important task for the whole society, its significance makes it a perfect application field for an emerging citizen science. Modern technology provides ordinary Joes with data collection and communication tools such as smartphones that may facilitate collaboration with the government and the industry in addressing public needs.

Using smartphones for analyzing general infrastructure problems started recently but still is mostly within its infancy. A number of solutions exist that involve a real-time road pothole detection using accelerometers embedded in smartphone devices. The existing applications [1] attempt to collect and analyze the embedded accelerometers data in order to detect potholes in real time, rather than using external devices. This research follows up [2] but aims at a solution that utilizes more advanced heuristic real-time event detection using limited hardware and software resources. Some of the advantages of the system [1] are that potholes can be detected in a real-time without a user-intervention. Also, it could detect larger potholes at medium vehicular speeds with a fairly decent true-positive rate. However, the system requires a high accelerometer sensor sampling rate that results in a high battery consumption. It does not provide a detailed description of a pothole and lacks the contextual information that can put the significance of the pothole within a different perspective.

Pothole Detection and Warning System using Wireless Sensor Networks [3] uses dedicated accelerometers placed within the vehicle, embedded communication subsystem and external Wi-Fi Access Points, to detect and warn vehicular commuters about impending potholes. It proposes to use a Mobile Node (MN) located within the vehicle that does the job of sensing the pothole, transmitting certain pothole data to a Wi-Fi Access Point (AP) and receiving localized information on a pothole from an AP. However, installing dedicated sensors and setting up communication devices is an expensive process. It does not classify potholes for repair scheduling. It also lacks contextual information that can put the significance of the pothole within a different perspective.

Existing smartphone apps [4] like Commonwealth Connect and iReport, also harness crowd-sourcing. They address the problem of reporting potholes to authorities too. However, they fail to provide a detailed description or any contextual information of the pothole.

Our approach aims at realizing some general and universal policies related to classifying potholes via crowd-sourcing. One of the tasks is to determine the most appropriate influencing factors that would be able to signify the destructiveness of a pothole, because most approaches did not consider these conditions, while they can be used to assign a priority to potholes.

There is also a need to provide vehicle commuters or concerned citizens with a quick and efficient way to report pothole data related to factors that influence its priority and at the same time extract some contextual information, if possible, from users so as to assist the effort of scheduling the repair of the potholes. Moreover, this information should also be shared with commuters to serve as warnings.

It is not always possible to include all the information that is necessary to describe a pothole. Thus, we need to estimate the accuracy of the data pertaining to a pothole, which could be generated using some factors which are derived from the completeness of the data and certain factors of the input device itself that is involved in the data collection process. However, accuracy is just one important factor influencing the overall quality of data. Another one is the source data security. While this paper concentrates mainly on general design and influential factors for data accuracy, evaluation procedures for data security and trust can be integrated with the current system and are presented in our other publications [5], [6]. Both factor groups need to be taken into account in evaluating the trust degree, which a decision maker may have in the provided data. This confidence level will determine the prioritization of the pothole repair and road maintenance procedures as the data is sent, where it can be analyzed and used for scheduling pothole repair.

These data quality and security requirements are used to ensure that proposed solution does not drain a significant amount of smartphone resources like battery life etc. Only data of certain level of quality and security will be collected and communicated. This condition will require a smartphone user, who is willing to collaborate in the road maintenance program and provide data, to support and maintain a certain level of the device security by downloading new system versions, patches, but avoiding dangerous operations and downloads.

2. Mobile Application General Design

Harnessing mobile computing platforms becomes apparent at this juncture as we design and implement an Android smartphone application. It is also essential to establish a user-friendly user interface (UI), which assists the user to define the pothole characteristics and requests the user to provide some contextual information using visual cues and a survey-type activityWe recognize the processing capabilities of smartphones and propose to design and implement an intelligent system, which involves balancing the significance of various factors that influence the priority of the pothole and calculating the priority while taking into consideration the contextual information provided by the user. To calculate the accuracy estimates of the data collected, we propose to compute the completeness ratio (i.e amount of information that is received through the user input / amount of total information requested).

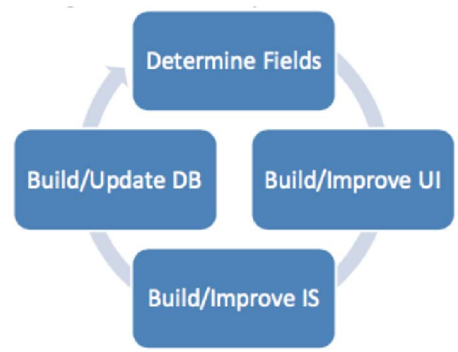


Figure 1. Iterative Development

We develop a system prototype to show the significance of the contextual information, which is used to supplement the data related to physical characteristics of potholes and also demonstrate a solution, which provides estimates on quality of the data being collected. We store and record the data in a developed database. The system can be extended to record meta data related to device's sensor identification and characteristics which would be useful for getting a security and data quality evaluation and trust score. To be able to continuously improve the system, we suggest to keep its development iterative in nature i.e, the smartphone application's implementation is made extensible to incorporate newer factors and context influencing the potholes priority, which is based on its destructiveness power. We refer to the proposed solution for each new iteration. Figure 1 shows an iterative development model for the proposed solution.

3. Data Models and System Analysis

3.1. Pothole Data Model for influential factors

By analyzing road maintenance policy and protocol documents from various counties, common factors have been established. The data model considers the following factors. *Type of Street* would specify whether the pothole existed on an arterial road, commercial road or an alleyway. *Weather Conditions* would specify the current weather conditions which could be used to further assess the damage rate of a pothole as time progressed. *Depth of pothole* is a numerical estimation of the depth of the pothole by the user. *Traffic Volume* would specify the current volume of traffic. *Area of pothole* is a numerical estimation of the area of the pothole by the user. *Average Speed* is the general speed limit of the roadway in question. *Safety Zone Rating* would specify whether the zone where the pothole exists in a spacial zone like school, etc. *Type of Junction* would specify whether the pothole exists is on an intersection or a t-junction or neither. *PPRRW* is the position of the pothole relative to the road width. *Road*

Material would specify the asphalt, cement, etc. *Existing Repair Conditions* would specify whether there was any kind of temporary or permanent repair work that was undertaken previously. *Type of pothole* would be extracted through visual cues provided by the user to identify which type of pothole it resembled based on the progress of the damage that road had endured [7]. This factor would specify whether it is a traditional pothole, a developing pothole, a tectonic pothole, a cliffhanger pothole or an inverted pothole. *GPS location* would specify the (x,y) gps co-ordinates accurate upto 10m. *Image* the image bitmap of the pothole for reference and analysis. *Priority* would specify the priority that is assigned to the system by the simple rule-based intelligence system. These factors are not exhaustive and serve as a basis for future prototypes and data models.

3.2. Data Quality and Security Model

We have to account for varying quality and security in collected data. *Camera Quality* would specify the quality of the camera that is used to capture an image of the pothole. *Data Completeness* would specify a user trust measure based on the number of questions answered by the user to the number of answers requested by the smartphone application. *Security Characteristics* is a measure that is based on [5], where it is proposed to employ a hierarchical framework to determine trust considering the analysis of installed applications, embedded mobile security features and sensor privacy and [6] where we explore a Security and data quality engine and use the accumulative SDQ indicator which characterizes the integral quality and security for a specific device model's sensor data based on its timeliness, consistency and accuracy. If a particular device does not already exist in its database it can volunteer to send periodic diagnostic sensor data to gradually develop an accurate security score.

3.3. Mobile Smartphone Application

3.3.1. User Interface. Since this application is based on crowd-sourcing, our approach aims at facilitating user's participation. As shown in Figure 2a, capturing the image and enabling locations services is encouraged and is the first suggestion made to the user, since this is one of the most important information for an official who might use this data in the database to decide on scheduling repair of potholes. Figure 2b illustrates capturing the pothole dimensions. The area, depth are captured using an Android Number picker, the limits are set to values that are practically observed and keeping in mind the average width of roads in the United States. In Figure 2c, we see another physical characteristic that needed to be captured, which was the type of pothole. Visual cues are used to get this information by using a custom UI component that uses a "SelectOneImage" approach, to help users choose which type is most applicable to the pothole they want to report. This information may help officials identify what approach they could follow in performing repair. Figure 2d and 2e explain the next section of the activities, that comprises of a survey for gathering contextual information that helps the intelligent system to assess the priority of the pothole being reported. This activity and the models that back it are kept generic so as to incorporate newer contextual factors. The survey activity is unique. It follows a fixed choice question style survey and uses a generic framework, wherein new questions and new choices can be easily defined and automatically incorporated into the UI. The choices are fixed, so as to avoid inaccurate information being recorded.

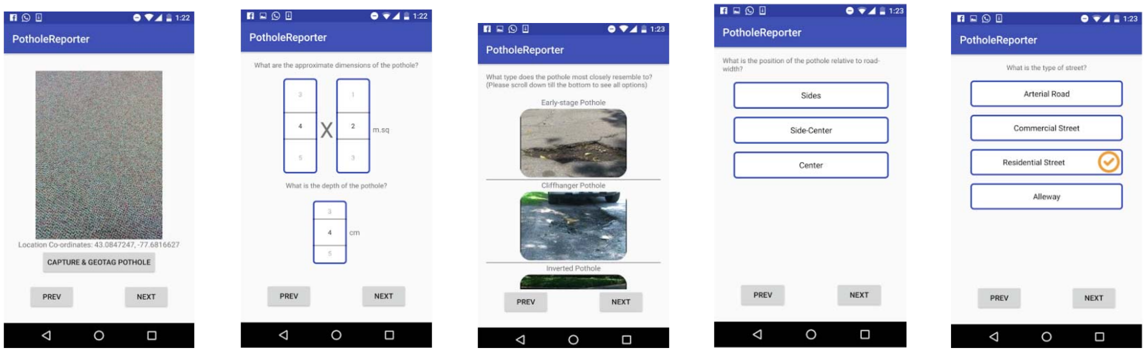
3.3.2. Outcomes. Once the survey activity is complete, the pothole data rule engine assesses the priority of the pothole being reported and assigns a priority score. The pothole data quality rule engine assesses the quality of the input and assigns a quality score. All the physical characteristics, contextual information that influence the priority of the pothole along with the pothole priority score as well as the data quality score are incorporated into a separate database that maintains this data. Officials can use this database for scheduling repair or analyzing the reported data.

The app is divided into different types of activities. The architecture allows different aspects of pothole related data to be extracted through different types of UI elements because the nature of the information may differ, while at the same time be necessary for arriving at accurate results. The app does not force the user to input all the information, respecting different levels of participation. However, they are encouraged to do so.

3.4. Intelligent System

3.4.1. Pothole Priority Calculation. In this section we discuss the approach followed to develop a simple forward-chained, rule-based intelligent system. Priority is scored as a percentage to determine a suitable priority class. All fields are given a fixed weight. All values of each field assigned a numeric value between (0-1). Figure 3 shows a forward chained rule-based hierarchical tree and evaluates each prioritized field in order it determines if the next fields weight needs to be adjusted based on its current value. If a weight is adjusted then weights of all fields down the hierarchy are re-adjusted accordingly, so that maximum priority sum does not change. It is important to note that, some fields which are partially dependent on the previous fields require a weight adjustment, while some others do not.

3.4.2. Data Quality Calculation. Figure 4 shows the data quality rule engine. Data quality is determined in



(a) Picture & Location (b) Dimensions (c) Pothole Type (d) Survey A (e) Survey B

Figure 2. Prototype User Interface.

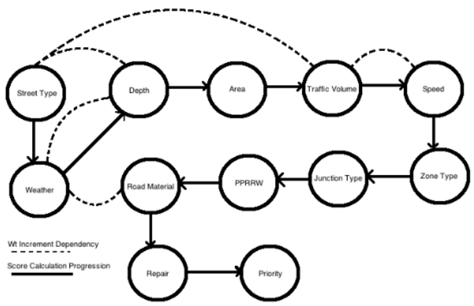


Figure 3. Priority Calculation Engine

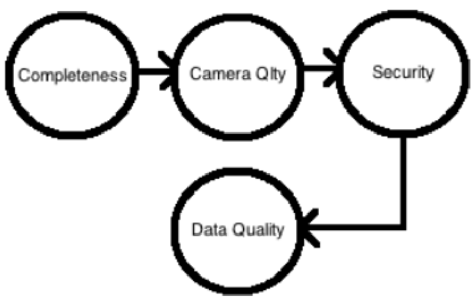


Figure 4. Data Quality Calculation Engine

a similar way as priority, taking into consideration the data completeness, camera quality and security in that order. No dependencies between the factors seem to be obvious hence weight adjustments are not considered in this model. Quality degree is scored as a percentage. All factors are assigned fixed weights. Completeness rating is calculated based on the ratio between number of questions answered by the user to the number of answers requested by the app. Camera quality rating determined through existing Android API's. Security rating is determined by analyzing the smartphone security characteristics. After determining the rating for each factor it is multiplied with each factor's weight. The final weighted sum gives the calculated data quality.

3.5. Purpose and Implementation of the Database connection

Once all the information is acquired, it is incorporated into a database so that it can be used for scheduling repair of potholes and even for further analysis. The intention is to analyze the data and detect patterns such as regions, which are extremely prone to specific types of potholes, or specific sized potholes, or how the weather conditions affect potholes in a specific region. After adequate amount of data gets populated in the database, it can exponentially increase the ability to answer many interesting questions and ultimately help in the effort of prognosis and actions on a pothole related maintenance. An Android-PHP-MySQL platform is used for the implementation. The Android smartphone application converts all the data related to the pothole into a JSON object and sends it to the Apache Web Server where a PHP file receives the JSON object and inserts it into the MySQL database.

4. Experiments and Results

A number of different pothole characteristic sets have been tested and analyzed. To better contrast the results, four instances are presented in Table 1 and discussed further. For case 1, we calculated the priority score as 98.3%. The priority class assigned to this instance evaluates to Extremely High, indicating that this is a one of the most dramatic potholes in the region. It must be immediately looked at and either requires a temporary repair (depending on weather conditions) or a permanent repair (if the weather permits and if temporary repair may have already been performed). We see that if weather conditions are snowy and the type of the street, where the pothole was observed is an Arterial street, where a considerably size pothole has appeared, with extremely high volumes of traffic

TABLE 1. POTHOLE CASES TABLE

Factors	Instance1	Instance2	Instance3	Instance4
Weather	Snowy	Sunny	Sunny	Rainy
Street Type	Arterial	Commercial	Alleyway	Commercial
Depth	16 cm	12 cm	4 cm	8 cm
Area	20 sq.ft	12 sq.ft	1 sq.ft	12 sq.ft
Traffic Vol.	Extremely High	Medium	Extremely Low	Medium
Avg. Speed	Extremely High	High	Extremely Low	High
Safety Zone	School Zone	None	None	School Zone
Junction	Intersection	T-Junction	None	Intersection
PPRRW	Center	Side-Center	Sides	Side-Center
Material	Asphalt	Asphalt	Concrete	Asphalt
Repair	Permanent	Temporary	Temporary	Permanent
Priority	Extremely High	Medium	Extremely Low	High

and extremely high speed traffic, is almost certainly a very high priority pothole and this can be verified empirically.

For instance 2, we calculated the priority score as 64.31%, The priority class assigned to this instance evaluates to Medium, indicating that although this pothole may be a nuisance, it is not likely to cause an enormous amount of damage. Pothole must be repaired soon to provide commuters a good experience. We see that if the general weather conditions are sunny and clear, the type of the street is commercial, with a fairly deep and big pothole, with medium volume of traffic and high speed traffic, indicates a fairly common pothole scenario and prioritized to be a pothole of medium significance and this can be verified empirically.

For instance 3, we observed that the priority score was 24.23%. The priority class assigned to this instance evaluates to Extremely Low, indicating that this pothole is not an immediate danger. It may be within an Alleyway or parking space. It must be repaired after all important potholes have been attended to. We see that characteristics such as sunny and clear weather conditions, the type of street as an alleyway or parking space, with a very small pothole, with extremely low volume of traffic and extremely low speed traffic, indicate the least significant kind of potholes. This is also verified empirically.

For instance 4, we calculated the priority score as 73.59%. The priority class assigned to this instance evaluates to High, indicating that this pothole may be a dangerous pothole owing either to its size or to its contextual significance. It should be repaired as soon as possible and should not be neglected as it has the potential to cause some serious damage. This is the instance where we can clearly see the contextual information at work in raising the significance of a pothole on characteristics other than physical. We see that the difference between Instance 2 and Instance 4 is the weather being rainy, safety zone type being a school zone, and the road being an intersection, which raises its priority score and evaluates to a High priority pothole even though Instance 4 reflects a smaller pothole with regards to its physical dimensions like depth.

5. Conclusion

The developed application facilitates an ordinary citizen's involvement into data collection and processing for improving road pothole detection. The research in this paper addresses some of the drawbacks that may be present in real time pothole detection and warning systems by establishing cost-effective approaches to classifying and reporting pothole data by introducing deeper contextual information related to potholes as well as providing information regarding the source data quality and security that can be used to prioritize road maintenance procedures. One of the goals of this research was to establish more context and factors that significantly influence the destructiveness of potholes and harness the pervasiveness and processing power of mobile computing platforms, where it is entirely viable to develop a system that would detect potholes in real time using accelerometer sensors as in [3] while consuming less power and also establish a warning system through wireless sensor networks as in [1], that would use these contextual factors. A simple prototype with an effective user interface was developed on the popular Android mobile operating system platform, to collect information related to potholes. Since, the prototype requires a high-level of user interaction which can be burdensome it suggests, as a future study, to explore approaches to automate the data collection process by various methods, while not draining a significant amount of battery power from the device. The research presented in this paper demonstrates the significance of contextual information that could be supplemented with physical characteristics of potholes. By using a simple intelligent system that harnesses this information, the classification of potholes is shown to be very close to what would be practically observed in a real world environment. The research results also demonstrate the significance of supplementing data with meta-data, which indicate the level of the data quality and security and establish a confidence and trust in the collected data that would drive important decisions. There is no universally accepted guide or policy existing to prioritize potholes. This research sets up the procedure that is based on the context and goals as well as the quality and trust in collected data. This research should serve as a guide for future researchers to develop and integrate various systems to ensure safer and pleasant roadways.

References

[1] A. Mednis, G. Strazdins, R. Zviedris, G. Kanonirs, and L. Selavo, “Real time pothole detection using android smartphones with accelerometers,” in *2011 International Conference on Distributed Computing in Sensor Systems and Workshops (DCOSS)*, June 2011, pp. 1–6.

[2] J. Eriksson, L. Girod, B. Hull, R. Newton, S. Madden, and H. Balakrishnan, “The pothole patrol: Using a mobile sensor network for road surface monitoring,” in *Proceedings of the 6th International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys ’08. New York, NY, USA: ACM, 2008, pp. 29–39. [Online]. Available: http://doi.acm.org/10.1145/1378600.1378605

[3] S. S. Rode, S. Vijay, P. Goyal, P. Kulkarni, and K. Arya, “Pothole detection and warning system: Infrastructure support and system design,” in *2009 International Conference on Electronic Computer Technology*, Feb 2009, pp. 286–290.

[4] I. SeeClickFix. (2008) Commenwealth connect. [Online]. Available: http://commonwealthconnect.io/

[5] R. Weiss, L. Reznik, Y. Zhuang, A. Hoffman, D. Pollard, A. Rafetseder, T. Li, and J. Cappos, “Trust evaluation in mobile devices: An empirical study,” in *2015 IEEE Trustcom/BigDataSE/ISPA*, vol. 1, Aug 2015, pp. 25–32.

[6] I. Khokhlov and L. Reznik, “Data security evaluation for mobile android devices,” in *2017 20th Conference of Open Innovations Association (FRUCT)*, April 2017, pp. 154–160.

[7] M. Barnhill. (2008) Know your: Potholes a guide to the street critters we love to hate. [Online]. Available: http://blog.nola.com/potholepatrol/2008/07/