Communication Technology Changing the World Competition
2016

Department of Computing
The Hong Kong Polytechnic University

Vehicle Enabled Big Data Platform

Yik Him Ho
Chi Ho Pun
Wing Chun Kung

August 15, 2016

1 Executive Summary

Smart cities employ technologies to improve the lives of citizens by bolstering the efficiencies of business transactions, healthcare, traffic or energy systems. We are building an intelligent big data platform by combining multiple technologies that involve a wide range of disciplines, including Internet of Things, mobile computing, cloud computing, big data and machine learning.

In this project, we aim to build a big data platform of environmental and traffic data for others to use. We utilise cars to collect data when they are moving on the road. By developing sensors and attaching in on vehicles, huge amount of data can be collected. This exposes and visualises the conditions of cities and enables cities to provide a far more accurate estimate of future conditions in different areas of the city, because we are collecting information from many points of the city, rather than from fixed locations that are not mobile at all.

We utilise big data and machine learning technologies to build predictive analytic models. Data collected will be put into an artificial intelligent training tools to recognise patterns and build predictive models.

We aim to make the data available to the public. We provide application programming interface (API) for others to make use of the data. Developer can build a smart alert mobile app to notify users when air quality at specific area is bad. City planner can analyse the traffic data to build a better city for all citizens. Observatory can make use of the environmental data to complement the weather forecast in a micro level. We hope that by providing access to the data and making it available for everyone, the platform will be the enabler to huge amount of innovative applications, enriching the live of all citizens in a smarter city.

2 General Description

Internet of Things (usually abbreviated to IoT) refers to a network for every day objects to communicate and exchange information automatically. Objects are vitally everything ranging from electronic devices, sensors, vehicles to furniture, clothing, etc. While IoT applications in home automation, fitness and healthcare field are popular nowadays, interests of IoT in other fields are emerging, such as vehicles.

We can see that vehicle is one of the important components in IoT ecosystem because vehicles generate huge amount of data when operating and vehicle network can be regarded as a subset of IoT network. It is believed that with the aid of information and communication technology (ICT), vehicles will be able to communicate among other vehicles (V2V), with the infrastructure (V2I), and ultimately with any other objects (V2X). This shares a similar concept of IoT, which objects will be able to communicate among other objects and with the infrastructure and everything (Internet of Everything, IoE).

Vehicle not only equips lots of sensors and collects huge amount of data for internal use (e.g., engine rpm, intake pressure, etc.), but also it has high potential to be a carrier to collect external data, such as temperature, humidity, air quality, etc. This information can be useful for prediction of environmental information in a local, district level or even street level.

In this project, we build a big data platform by capturing traffic and environmental data through vehicles. We make use of external sensors and attach it on vehicles. The system utilises sensor technology to collect data, mobile technology to transmit data, big data and machine learning technologies to provide a real-time analysis and prediction. Compared to traditional fixed data collection points, vehicles can collect far more data in different locations, and thus the coverage of data point will be much larger, so that high level of prediction accuracy can be achieved ideally.

We also take vehicular network (IEEE1609 and WAVE standard) into consideration when developing the system, so that it is compatible and ready to deploy onto a vehicle network when the infrastructure is ready.

Our application involves the use of ordinary cars on the roads to collect information about the conditions nearby. Utilising vehicles to collect data has advantages over other data collection methods. Data collection by monitoring stations requires large and expensive equipment built at a fixed point to collect data. The drawbacks are that it only take samples on that location which the results are not representative. As car is a highly mobile device, it is able to reach areas that are sometimes impossible to reach on foot within reasonable time. If we utilise cars to collect data, we will be able to obtain a large pool of data that can be used for predictive analytics. The more data we collect, the more accurate and representative the predictions will be.

Table 1: Comparisons on ways of data collection

Ways of Collection	**Station**	**Human**	**Vehicle**
Location	Fix	Mobile	Mobile
Coverage	Limited One point per station	Limited	Unlimited
Number of Collector	Small	Large	Very Large
Cost	High	Low	Low

Thanks to the advent of hardware solutions, these sensors can be purchased and assembled without paying exorbitant amounts of money, mere mortals will be able to afford the solution without breaking the bank. In fact, the cost can be as low as 50 USD. With the prevalence of smartphone, we can also expect that adopters of this application to install the smartphone or smart entertainment system in their cars.

Our application is flexible and adaptable, as we are gathering data from mobile devices. It will also be increasingly accurate over time, owing to big data technologies, data mining and machine learning techniques. These technologies are able to find characteristics, patterns, and correlations between data points. The data will be precise on an unprecedented level, because we are collecting information from many points of the city, rather than from fixed locations that are not mobile at all.

We expect that the big data platform is beneficial to all citizens. We would provide the data for others to use to meet their purposes, such as city planning, research, urban data analysis, etc. All citizens can make use of the applications to make better informed decisions. They will be able to know a far more accurate environmental condition of their current location.

3 Technical Solution and Project Details

There are three main components in our system, namely (1) Arduino sensors, (2) Smartphone, (3) Machine Learning tools. Figure1 shows the architecture of the system. The data will be collected by the sensors attach on vehicles, and the smartphone will temporarily store the data and send the data to the cloud when network is available. The on-board unit (OBU) and road side unit (RSU) are part of the vehicular network infrastructure. Ideally, data are sent to the cloud if vehicular network is available to avoid consuming the bandwidth of cellular network. However, if the vehicular network is not present, the data can also be sent through Wi-Fi and cellular. (In this case the access point, OBU and RSU can be omitted). When the data reach the cloud, it can be displayed in a real-time manner, and also be put in a data mining model. The result of the model will be made available to public to access, to analyse and predict the condition of the city.

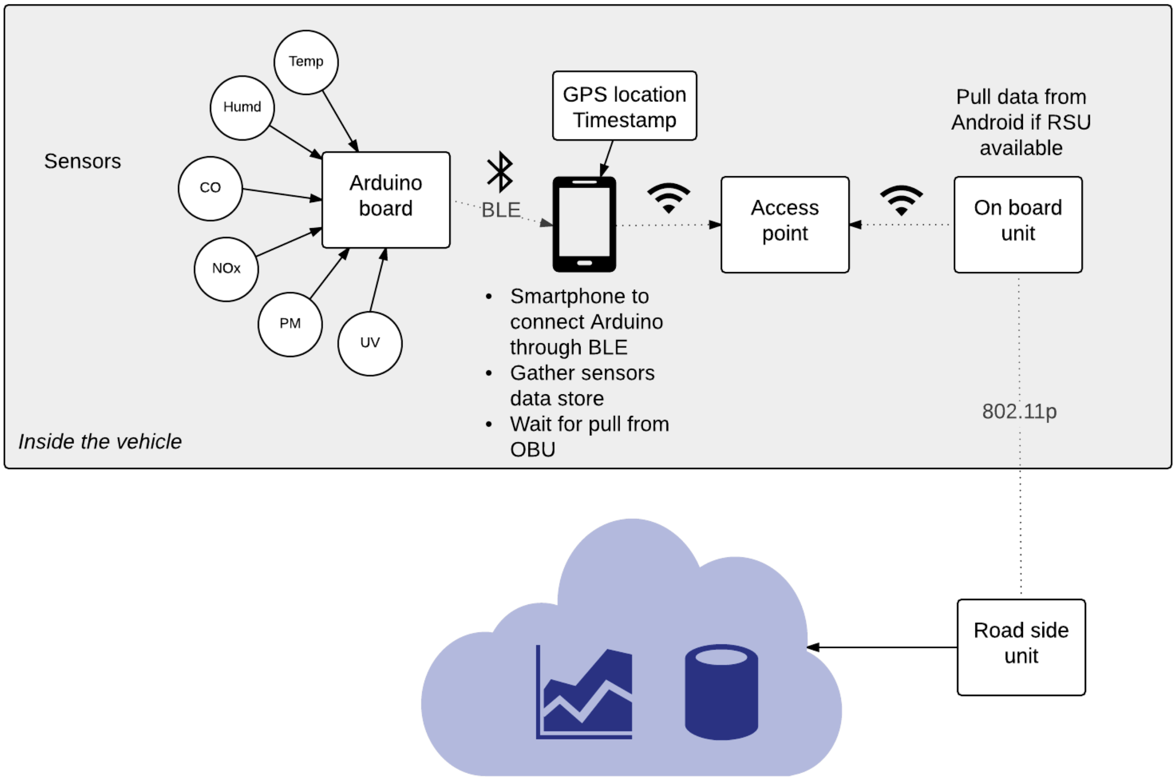

Figure 1: Architecture of environmental monitoring system based on V2X testbed

3.1 Sensors

Environmental information is collected continuously by Arduino sensors mounted on the car when the car is moving. Arduino is an open-source platform for developing hardware and software projects. The features are that it is easy-to-use, in the aspect of both hardware assembly and software coding, and the hardware cost is inexpensive. People tend to use this to develop hardware prototyping, proof-of-concept and DIY projects. Arduino provides high flexibility on assembly and combination of hardware. For example, developers can choose to attach a Wi-Fi shield or Ethernet shield on top of the main board to enable internet connectivity. Other hardware can be connected through Breadboard or other shields.

The board can be powered by USB, in which most of the vehicle has at least one comes from the cigarette charger. Beside the Arduino main board, there is also a Bluetooth Low Energy (BLE) shield for data transmission from Arduino to a smartphone. The sensors are shown in Figure2, 3 and 4. Currently, 5 sensors are used in the prototype, including

- Temperature
- Humidity
- Carbon monoxide

- Particulate matter
- Ultraviolet

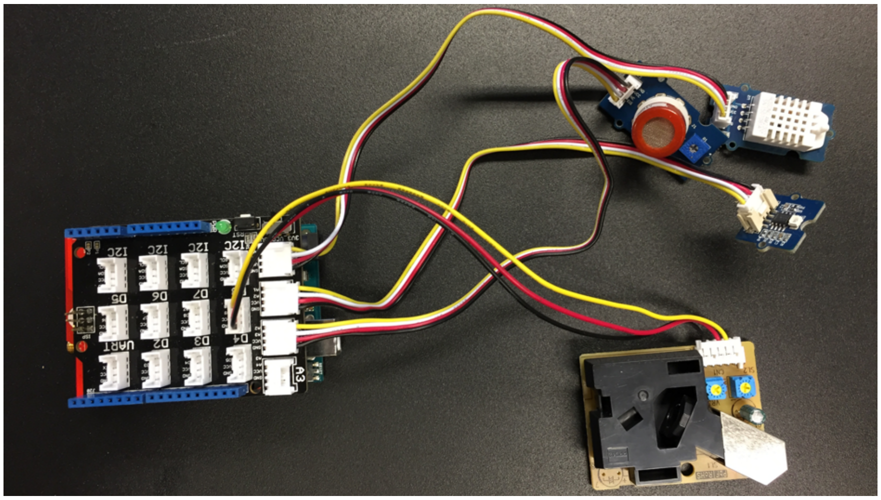

Figure 2: Arduino sensors

Figure 3: Sensors mounted on car

3.2 Smartphone

The role of the smartphone is to receive the data from the sensor and transmit to the cloud. To receive the sensor data, a smartphone with Bluetooth 4.0 is required. After connecting to the Arduino board, the app will listen for a new incoming data through BLE. Figure5 shows a screenshot

Figure 4: Sensors mounted on car

of the Android application receiving data. Because of the inter-compatibility of BLE, the connection between a smartphone and sensors is not limited to Android devices, but all Bluetooth 4.0 equipped devices. It requires no manufacturer-developed API in order to connect to the sensors, but BLE API which is available on all operation systems. After receiving a value, the phone will first save the data into local memory and wait for transmitting to the cloud.

The reason of utilising a smartphone to transmit data is that it allows additional processing for the raw data. For example, some of the data may require a conversion or calculation to obtain the true value (the gas sensor, for example). Also, it can temporary store the data on the phone if there is no connection to the cloud. Arduino board might not support these functions because of the limited computational resource.

The sensor data is transmitted to the cloud based on the Google Protocol Buffer structure. It is a data serialising method and interchange format which is similar to XML and JSON. Advantages of using Google Protocol Buffer include small data size and fast serialisation, because it encodes data into binary format. Another advantage is that once the data structure is defined, both front-end and back-end application can share the code implementation easily, without the problem of inconsistency.

Data is sent from the app to the cloud through TCP connection. In order to save the bandwidth

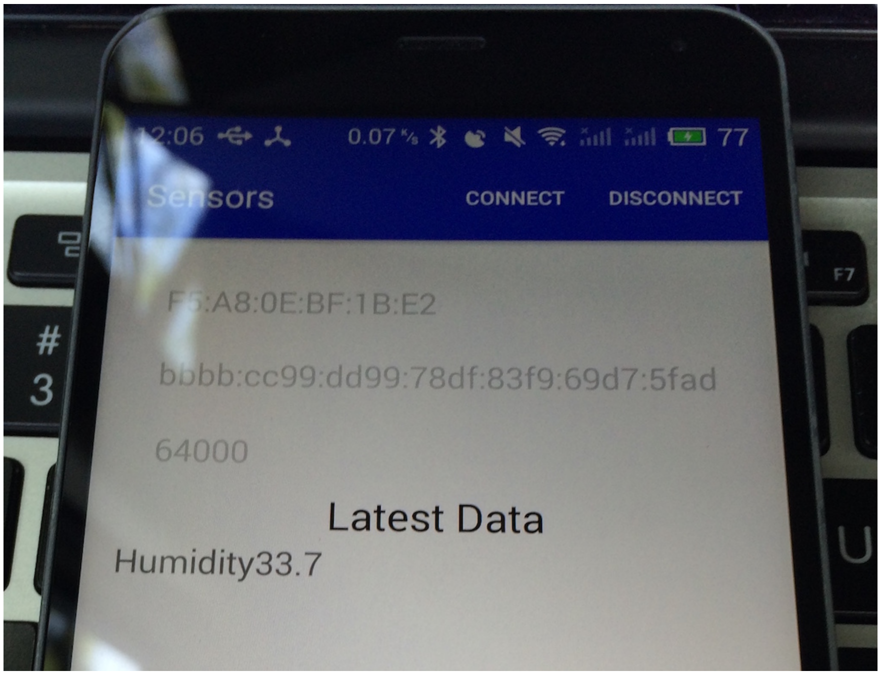

Figure 5: Android app receiving data

of Wi-Fi and cellular network, and because the data is not time-critical, data will be transmitted over the air through the vehicular network (i.e., IEEE 802.11p standard). However, as smartphone does not support 802.11p, the data will be sent through on-board unit (OBU) and road side unit (RSU). The smartphone and the OBU are connected to a on-board wireless access point, so that the smartphone can communicate to the OBU. The on-board access point will have internet connectivity if the OBU can receive RSU signal, because the RSU is always connected to the WAN through wireless mesh network. Once the car is driven into a RSU coverage area, connection can be established as the bridge between the smartphone and the cloud is formed, through the access point, OBU and RSU. If there is no RSU coverage or the connection is dropped, the smartphone will store the data and wait for the next available RSU coverage. Figure6 shows the IPV6 connection on vehicular network.

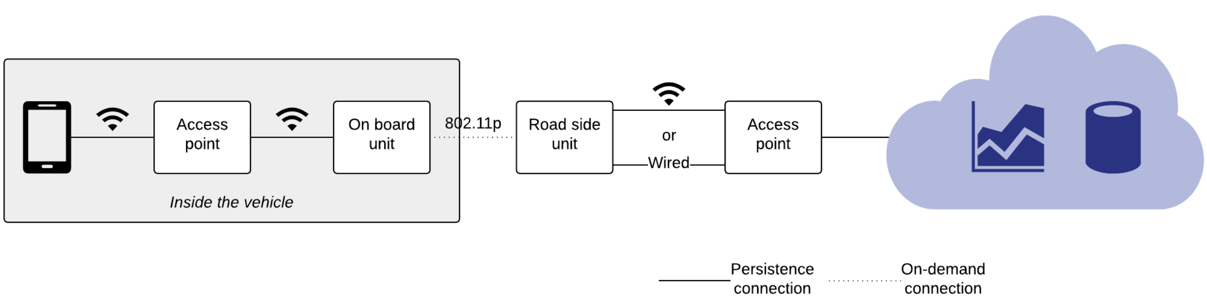

Figure 6: IPv6 Connection Bridge

3.3 Cloud

The cloud receives data from the smartphone. It stores the information into the database, then organises and prepares it for machine learning. Because of the flexibility of the cloud service provided by Microsoft Azure, the capacity of the web server and database can easily be scaled up to meet the increasing number of data.

3.3.1 Microsoft Azure Machine Learning

After collecting the sensor data from the vehicles, we can perform predictive analytics. We choose Microsoft Azure Machine Learning as an analytics tool for building predictive models. Machine learning is a process where the machine learns from existing data and find the patterns of history. There are three main machine learning categories for solving different problems, including classification, regression and clustering. Regression is a predictive model that predicts a continuous value based on independent variables, so it is suitable for predicting traffic speed and environmental data based on the latitude, longitude and time.

The data will be put in the Machine Learning studio and divide into two data sets, training data set and test data set. When the training data is input into the model, it finds the best pattern that fits the data. After the model is trained, it is used to make predictions based on the testing data. If the predicted value is close to the actual value of test data, the accuracy of the trained model is accepted. Figure7 shows an example of using Boosted Decision Tree Regression algorithm to build a model for temperature data. However, algorithm is not limited to this specific one, but different algorithms could be used for different data sets in order to create an accurate model. For example, neural network may be suitable for gas data, while clustering generates better result for UV. The choice of algorithm usually requires in-depth domain knowledge and analysis.

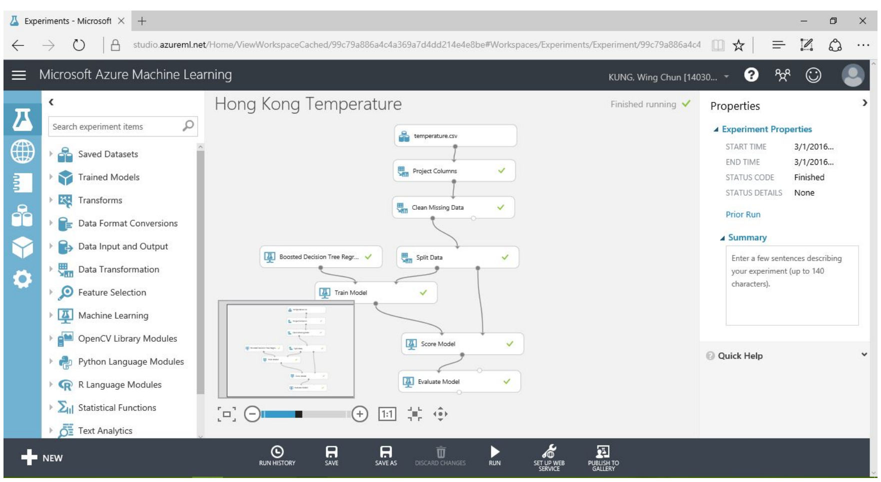

Figure 7: Azure Machine Learning

3.3.2 Front end applications

After an acceptable model is built, we convert the training experiment to a predictive experiment. It provides the best trained model and defines a web API service for forecasting the data in the future. Azure web API service allows users to send the input data (date, time, latitude and longitude) to the model and the model will produce the prediction results based on the model and input. Figure8 shows the architecture of the web service.

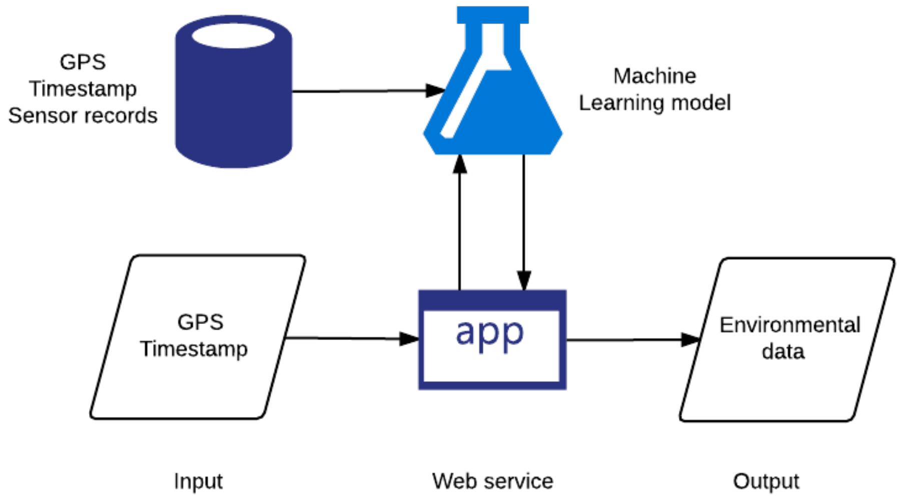

Figure 8: Machine learning structure

We could develop applications on the web or mobile devices that allow everyone to use the machine learning web service. People can make use of the data to build other applications, such as traffic speed analysis as shown in Figure10 and environmental data forcast as shown in Figure9. For example, users can simply choose date, time and location to predict the value when they decide to go out. Also, an mobile application can notify the users when and where the air quality will deteriorate so that they can avoid going to the areas of bad environmental condition (Figure11).

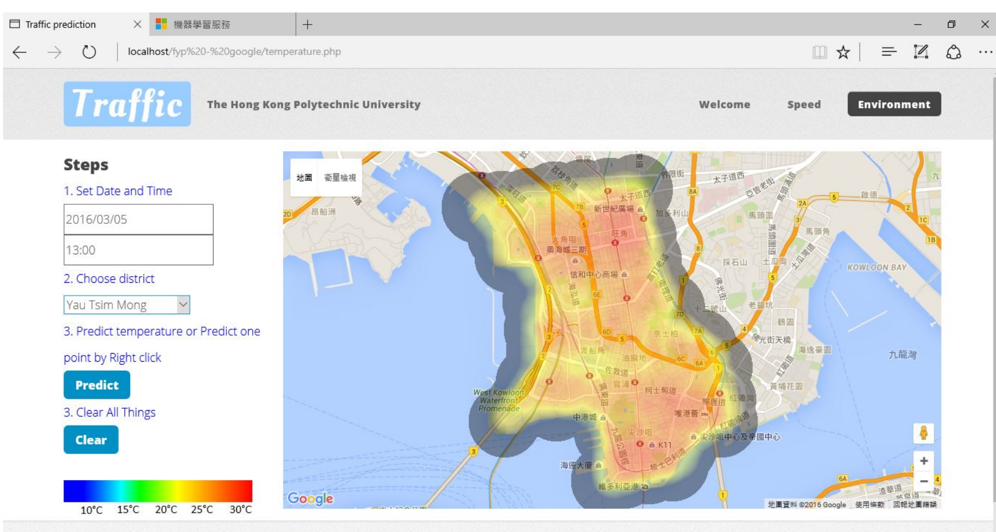

Figure 9: Temperature prediction

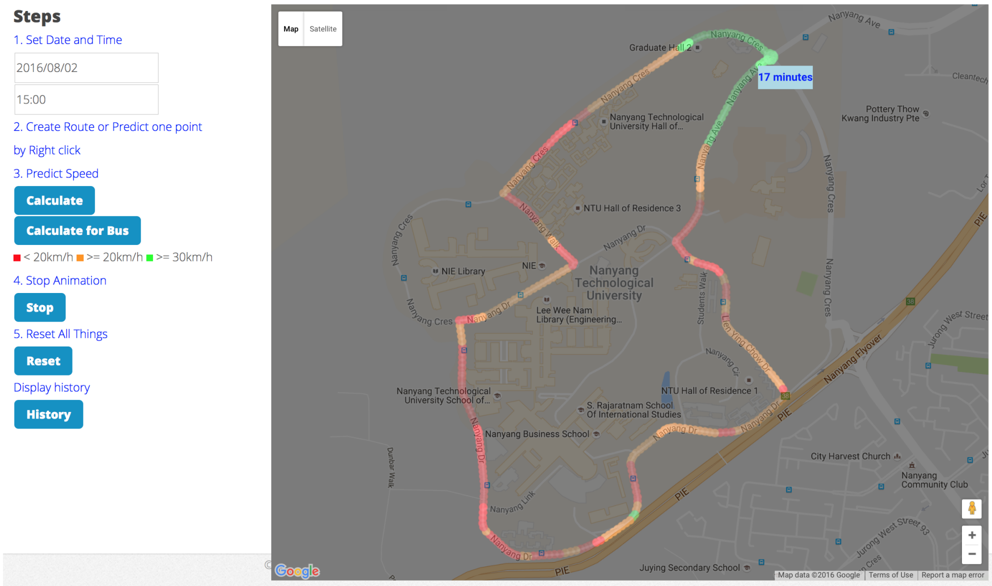

Figure 10: Traffic speed analysis

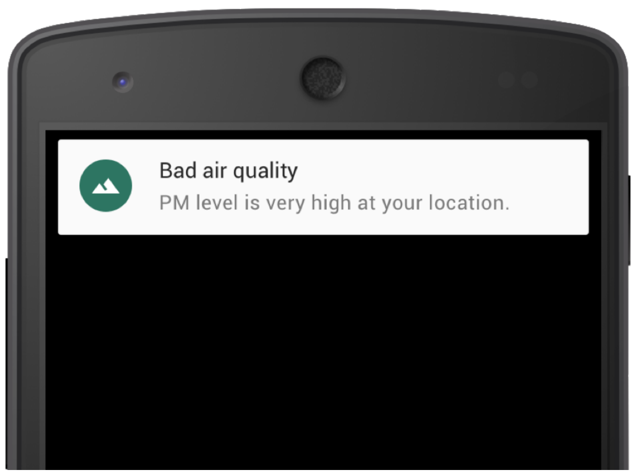

Figure 11: Notification of bad air quality

4 Social Impact on Humanity or Local Community

The system helps to collect the data and visualise these. It targets the general public as our consumers. All citizens can make use of our application to make better informed decisions. They will be able to know a far more accurate environmental condition of their current location.

We noticed that while traffic forecast technologies today offer decent accuracies, we think utilising real-time data generated by our system allow even higher accuracy for drivers. With this information, drivers will constantly be informed of the latest and most up-to-date information of the conditions on the road. This enables drivers to make better choices as to which route to use to reach their destination. We envision that traffic flow will be smoother as no longer will there be drivers crowding just some of the most popular roads. This will hence reduce traffic congestion, and therefore the trip time of drivers. With this also comes the possibility of increasing gas efficiency of the vehicle and reducing greenhouse gases.

Because the system can help to collect the data easily by the vehicles, it is much easier to survey and plan for the city in a more accurate and friendly way. By collecting huge amount of data and making it available to public, there will be increasing number of innovative smart city applications. The data is valuable to the public for city planning, research, analyse, policy decision making, etc. The micro-level data can also complement nowadays mature weather and environmental predictions provided by government or observatory. Citizens can decide what to do based on the forecast results. For example, someone who have lung diseases or heart diseases can avoid going to the location affected by the bad air quality. The decision of citizens can be more accurate because our application will comprehensively cover most of the roads. We believe that our platform will be the enabler for a smarter city, that bring benefits to all citizens.

5 Implementation Status, Testing and Trial

As shown in Figure2, 3, 4 and 5, a working prototype has already been developed. These figures show that the sensors work when attaching on the vehicle and send data back to the cloud. As this project is in collaboration with Nanyang Technological University (NTU) in Singapore, we have a chance to carry out on-side testing and modification at their Smart Mobility testbed on vehicular network. The prototype was also demonstrated at the Industry Partnership Showcase 2016 organised by NTU on 30 May 2016 on NTU campus.

The application also works using cellular or Wi-Fi. Testings and trials were also done in Hong Kong. Because vehicular network infrastructure is not available, the application relies on cellular network to transmit data to the cloud.

The machine learning models are also built based on the collected data and available on the internet. Figure9 and 10 are the demo web applications retrieving predicted results directly from the models.

6 Previous Submissions to Other Contests

This project has been submitted to Microsoft Imagine Cup competition 2016 in Hong Kong and received 3rd runner up during the National Final round.

7 Short video

A short video of this project can be found at: https://youtu.be/FQ6IGVm1hcU

8 Contact Information

Team Leader: Yik Him Ho
E-mail address: csyhho@comp.polyu.edu.hk
Phone: +852 97381567
Affiliation: Department of Computing, The Hong Kong Polytechnic University
Address: Dept. Comp, Rm QT418, 4/F, Q core, The Hong Kong Polytechnic University, Hung Hom, KLN, Hong Kong

Team Member: Chi Ho Pun
E-mail address: vincent.pun@connect.polyu.hk
Affiliation: Department of Computing, The Hong Kong Polytechnic University

Team Member: Wing Chun Kung
E-mail address: wing-chun.kung@connect.polyu.hk
Affiliation: Department of Computing, The Hong Kong Polytechnic University