

DEEP LEARNING THROUGH ENVIRONMENTAL DETECTION

Junyi Lu¹, Yu Sun¹ and Fangyan Zhang²

¹ Department of Computer Science, California State Polytechnic University,
Pomona, CA, 91768

²ASML, San Jose, CA, 95131

ABSTRACT

Many of the urban areas are highly polluted. They are aware of the damage caused by pollution but lack efficient and economical solutions to address it. The purpose of this project is to design a portable pollution sensor that is able to communicate with an online database and allow users to access data through the internet. The algorithm of machine learning is able to create data models to predict future pollution level with existed data values.

KEYWORDS

Deep Learning, Environmental detection, Machine Learning, Wireless Network

1. INTRODUCTION

Pollution monitoring has been a great concern for ecologist. Without the knowledge of when, where, and what certain pollution occurred, scientists are hard to locate and find a correlation between the pollution levels. To resolve this problem, many laboratory researchers have collaborated with engineers and technician to search for solutions in technology. Environmental protection organization around the globe focuses on data collection and field research to monitor and reduce industrial water and air pollution in areas. But the organizations lack devices that could help volunteers to assess pollution without professional guidance. Furthermore, all of the real-time data should be sent to the online database and website to achieve wireless [8] transmission. Online users should also be able to locate the marker on the map and generate graphs [10] for the pollution level of the chosen area. This will provide an opportunity for anyone to access environmental factors [11] in a certain area online. The pollution data collected can be further processed to reach laboratory research or data analysis standard. The value can be processed through machine learning [7] algorithms. Machine learning [7] is a method of data analysis that automatically model correlated data points. Machine learning systems can learn from data, identify patterns and make decisions with minimal human interaction. Linear regression is a linear approach to modeling the relationship between a response and one or more variables. A Support Vector Machine (SVM) provides labeled training data, and it is an algorithm outputs a plane which categorizes new examples. Moreover, polynomial regression is a form of regression analysis in which the relationship between the independent variable and the dependent variable is modeled through degree polynomial.

2. MOTIVATIONS AND CHALLENGES

Many of the tradition pollution monitors focus on the quality of the data gathered from the machine. However, the new generation of pollution monitors should lean towards the remote [9]

data transfer and analysis. This allows the new generation to cover a wider range of remote [9] areas and explore new data correlations. Comparing to the old model of pollution detection, the new generation should be able to function fully automatic and transfer web data. In the old model, all of the pollution detectors require an outer power supply. However, the new model should use other power options to avoid power supply requirements.

The solar panels need to provide enough energy to the detector can run by its own. The energy from the solar panels cannot exceed the amount required. Thus, an algorithm is required to put the controller to sleep for a short time. Furthermore, all the data collected from the controller cannot be processed without a location value. A GPS booster was implemented to ensure the detector is able to read GPS signal quick and accurate.

3. REQUIREMENT

The detector shall be able to detect temperature from the outside environment. (including Light level, UV light, Humidity, CO₂, Noise pollution, Air pressure, Temperature, GPS, and TVOC) The detector shall attach to a circuit board which contains all of the sensors listed above. The detector shall be able to run independently by solar energy converted by attaching solar panels. The detector shall be able to transfer the values read by the sensor to Firebase [6] while connecting to the internet. The detector shall be able to show the marker location of the values on the Website. The website shall be able to display specific time and graphs [10] of the values upload by the detector. The AWS server shall be able to activate and run a python script to simulate a machine learning [7] model.

The website shall be able to refresh the machine learning [7] prediction value after the user enters the value of the time they want to predict. The first prototype of the detector allows the user to view and access data online provided by the firebase [6]. Later, a machine learning [7] function was able to create a multi-dimensional model that takes time, location, values as the input to simulate output. Users can approach the prediction in two different ways.

- Giving location and time to predict the value
- Able to predict different location but requires a larger database for higher accuracy
- Giving time to predict the value on the existing location
- Able to predict accurate value with limited information.

4. SOLUTIONS



Figure1: Components of detection system

The Raspberry Pi 3 is a micro-size computer with the implementation of a keyboard, mouse, display, power supply, SD card and installed Linux Distribution. Raspberry Pi Model 3 is used as sensors controller. The controller is protected by 3D printed case from the exposure in the environment and solar panels are attached to drive the controller without charging outlet. The sensors are on the back of the controller where they are connected with a circuit board. The touchscreen on the front of the controller allows the user to view data locally. A python script was coded in the controller such that it is able to transfer the sensors reading to firebase when it is connected to the internet. The website then receives the values from the firebase and displays it on the website.

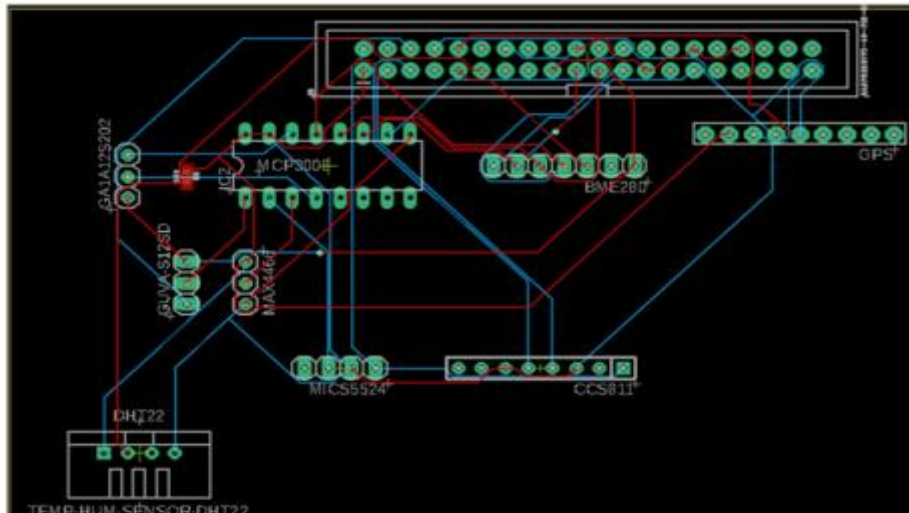


Figure 2. PCB for sensors

The website is composed of a map with locations' markers, data table, and graphs. Users are instructed to select a marker on the map and then select a time stamp to view the values in the selected time and location. The users can also select a prediction time on the website to view the prediction of all values. The prediction was simulated by data models on the AWS server. The server will obtain firebase real-time data and process a machine learning algorithm when it is activated. Raspberry Pi Model 3 has the capability of connecting to the internet. It also has a sustainable memory capacity to store local sensor readings. The clock, the amount of time to generate an electric pulse, compare to other micro-computers is much faster, as the project requires the controller to communicate with multiple frequently.

The sensors in the project use the Adafruit breakout PCBs.

- CCS811: Co2 and Tvoc
- GUVAS12SD: Analog UV Light Sensor
- DHT22: Temperature and Humidity Sensor
- MiCS5524: CO, Alcohol and VOC Gas Sensor
- MAX4466: Electret Microphone Amplifier
- Adafruit GPS: 66 channel w/10 Hz updates
- BMP085: Temperature Humidity Pressure Sensor
- GA1A12S202: Log-scale Analog Light Sensor

Google Firebase [6] is an online database that manages real-time data and helps users to store and obtain information efficiently. Amazon Web Service [12] provides a platform for

demanding cloud computing to individuals. It allows the users to have a virtual cluster of the computer to process data.

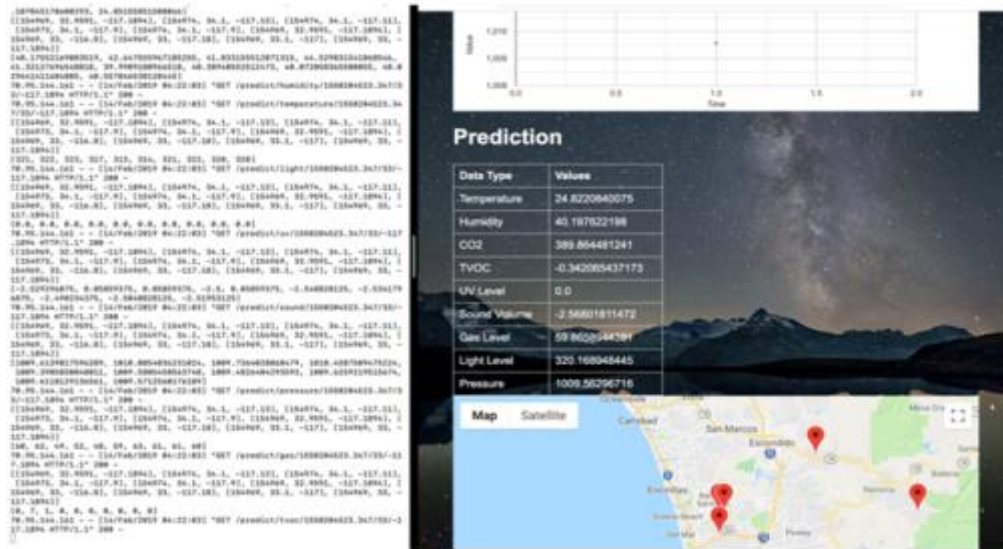


Figure 3. Website and Online regression

Figure 3 shows the machine learning [7] process and the website that displays the data. The website shows a map with markers with existing data points, which can be selected by any online user. When hovering over a marker, the related graphs and table of each value (CO2, Temperature, etc.) will be shown. The website also provides a drop-down menu for users to select different timestamps. After providing the number of days to predict in the textbox, the system will send related data to Amazon Web Server [12] to process machine learning [7] algorithms. Machine learning [7] algorithms used three regressions to predict possible data points. Linear regression [3], polynomial, and SVM are used to predict the data value.



Figure 4. Firebase Data Format

```

    },
    $.ajax({
    {
      type : "GET",
      url : "http://52.53.163.3:5500/predict/temperature/" + timestamp + "/" + cLat + "/" + cLon,
      data : {
      },
      success : function(result) {
        console.log("Get predicted value: " + result)
        $("#Temperature_Predict_Label").text(result);
      },
      error: function (jqXHR, exception) {
        console.log("Failed to get the prediction.")
      }
    }
  });

```

Figure 5. Example of Getting processed Machine Learning Value from AWS

5. RELATED WORK

The device included many accessible outside resources to achieve data reading, transferring, and analyzing purposes. Amazon webservice [12] provide the user the admin portal to connect and manage application and data. AWS setup cloud service and connect it through unique IP address. However, the data transferred is not encrypted which means any anonymous user can access data while it is transmitting through the channel. Moreover, AWS cannot be adjusted which indicates user do not have direct control of the channel. User required additional security service in order the data transmitting process can be completely trust without information leak.

6. CONCLUSIONS

The accuracy of the data value obtained from the sensors was tested by comparing the temperature and humidity with the weather report and mercury thermometer. The prediction from the machine learning achieved a 2-degree error with the actual result. However, with more data points, it can achieve higher accuracy. The solar panels need to provide enough energy to the detector can run by its own. The energy from the solar panels cannot exceed the amount required. Thus, an algorithm is required to put the controller to sleep for a short time. Furthermore, all the data collected from the controller cannot without a location value. A GPS booster was implemented to ensure the detector is able to read GPS signal quick and accurate. With test and evidence, the machine learning algorithm yields a considerable result. The detector can upload data required by a machine learning algorithm in the AWS server. Meanwhile, the server can simulate computer model base on the values existed in the server. The project achieved to detect pollution remotely. The project can be most effective in urban areas, where highly pullulated cities may cause exceeding pollution in the environment. Online users can locate the marker on the map and generate graphs for the pollution level of the chosen area. The device can be placed in areas through private or government organizations. The goal is to create an independent machine which does not require an outer power supply and human intervention.

REFERENCES

- [1] "Raspberry Pi 3 Model B." Rotate Display 90°? - Raspberry Pi Forums, Adafruit, www.raspberrypi.org/products/raspberry-pi-3-model-b/.
- [2] Gizem, Aksahya & Ayese, Ozcan (2009) *Coomunications & Networks*, Network Books, ABC Publishers.
- [3] Linear Regression." Wikipedia, Wikimedia Foundation, 2 Mar. 2019, en.wikipedia.org/wiki/Linear_regression

- [4] Tony Dicola. “Adafruit Ultimate GPS and wiring” www.learn.adafruit.com 2012. Adafruit <https://learn.adafruit.com/adafruit-ultimate-gps/circuitpython-parsing>
- [5] Dean Miller. “Raspberry Pi Wiring & Test” www.learn.adafruit.com 2017. Adafruit <https://learn.adafruit.com/adafruit-ccs811-air-quality-sensor/raspberry-pi-wiring-test>
- [6] “Firebase.” Wikipedia, Wikimedia Foundation, 20 Feb. 2019, en.wikipedia.org/wiki/Firebase
- [7] Bishop, Christopher M. *Pattern Recognition And Machine Learning*. Springer-Verlag New York, 2016.
- [8] Pillai, Pradeep G, and Saravanan. P. “A Study on Air Pollution Monitoring System Using Wireless Sensor Network.” *International Journal of Trend in Scientific Research and Development*, Volume-2, no. Issue-2, 2018, pp. 343–347., doi:10.31142/ijtsrd8374.
- [9] Choi, Sung-Woo, et al. “Method for Transmitting Data in Heterogeneous Network.” 2014 International Conference on Information and Communication Technology Convergence (ICTC), 2014, doi:10.1109/ictc.2014.6983245.
- [10] Jain, Rishu K., et al. “Big Data Big Cities: Graph Signals of Urban Air Pollution [Exploratory SP].” *IEEE Signal Processing Magazine*, vol. 31, no. 5, 2014, pp. 130–136., doi:10.1109/msp.2014.2330357.
- [11] Cecchi, L., et al. “Climate, Urban Air Pollution, and Respiratory Allergy.” *Climate Vulnerability*, 2013, pp. 105–113., doi:10.1016/b978-0-12-384703-4.00116-7.
- [12] Jackson, K. R., Ramakrishnan, L., Muriki, K., Canon, S., Cholia, S., Shalf, J., . . . Wright, N. J. (2010). *Performance Analysis of High-Performance Computing Applications on the Amazon Web Services Cloud*. 2010 IEEE Second International Conference on Cloud Computing Technology and Science. doi:10.1109/cloudcom.2010.69