Poster: Cloud-Based Data Fusion in GreenIoT for Smart Cities

Xiuming Liu[1], Abdul Rahman Khankan[1], Mohamad Alsioufi [1], Zhitao He[1,2], and Edith C.H. Ngai[1]

[1]Department of Information Technology, Uppsala University

[2]Networked Embedded Systems Group, SICS Swedish ICT

Abstract

Applications based on internet-of-things (IoT) for smart cities attract intensive attentions recently. One of the important elements in a IoT platform is processing the sensor data. The motivation of abstracting useful information from large scale and heterogeneous sensory data drive us to design and implement a cloud-based sensor data fusion system, which bridges the sensor networks and external applications. A system architecture is illustrated and tested with two use cases in this poster.

1 Introduction

In the ongoing process of global urbanization, urban air pollution is one of the major challenges for major cities around the world. According to the reports from World Health Organization, air pollution has very wide and direct impacts on the health of citizens and generates significant economic cost [1]. Major air pollutants in the urban area include nitrogen-dioxide (NO_2), ozone (O_3), and particulate matters (e.g., PM_{10} and $PM_{2.5}$), which are mainly generated by transportations and other industrial activities [2]. For air pollution data measurement, analysis, and visualization, a smart city must provide an information system based on sensor networks.

GreenIoT project [3] aims to develop an internet-of-things (IoT) testbed consisting of a network of low cost sensors and a cloud-based air quality data analysis service for smart cities. The GreenIoT project builds an ecosystem for smart cities with three layers: 1) Embedded sensors which are low cost and portable; 2) Cloud-based data fusion services, including storage, analysis, and visualisation; 3) Applications powered by the data from IoT and cloud. Those three layers are connected via machine-to-machine (M2M) communications (e.g., MQTT + 6LoWPAN) and mobile cellular networks.

International Conference on Embedded Wireless
Systems and Networks (EWSN) 2017
20–22 February, Uppsala, Sweden

ISBN: 978-0-9949886-1-4

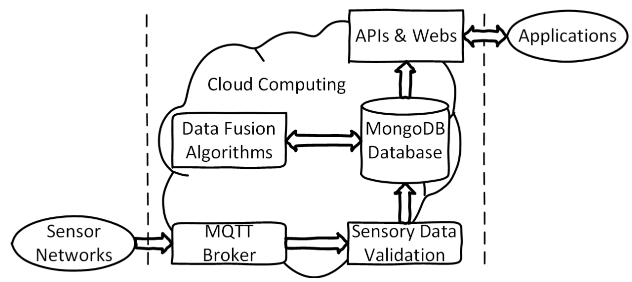

Figure 1: System architecture of the cloud-based services.

In this poster, we present the cloud-based data fusion services for smart city applications. The analysis of data are based on statistical modelling of sensory data. Based on the learned dynamic model, we are able to provide services such as air quality forecast and spatial-temporal interpolation.

2 System Design and Implementation

The cloud-based data analytic service is one of the central components in the GreenIoT testbed, not only because it bridges sensor networks and applications which might be developed and operated by different organisations, but also because of discovering knowledge from noisy sensory data.

2.1 System Architecture

The cloud-based system consists of the following subcomponents: communication interfaces, a sensory data validation mechanism, a document-oriented database, and a collection of data fusion algorithms (Figure 1).

Communication Interfaces

The M2M communication protocol MQTT is applied for the interface between sensors and the cloud. After the registration of device, the data will be streamed to the cloud in a publish-subscribe fashion. The sampling rate of the sensor network can be dynamic and heterogeneous. For the communication between cloud and applications, a set of web application programming interfaces (APIs) are implemented. An external application may query data from the cloud via sending HTTP requests which contains a JSON array of GPS locations and time stamps.

Data Validation

In order to provide high quality data fusion services, it is important to prevent malformed sensory data being stored and used. Firstly we check a sensor's maintenance history.

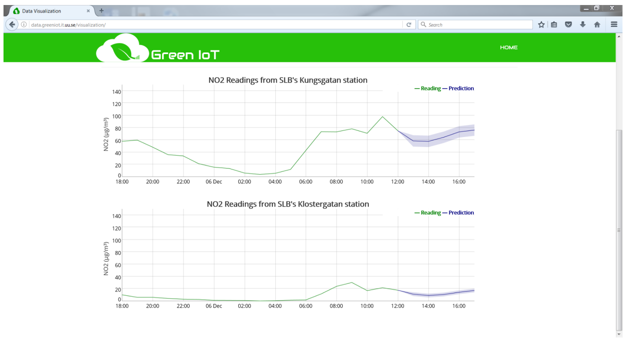

Figure 2: The web page visualizes the air pollution level forecast results for the next 5 hours.

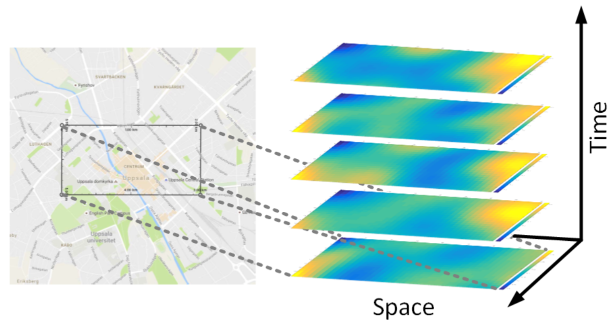

Figure 3: The spatial-temporal air quality data in Uppsala.

If the sensor has not been maintained properly (e.g., calibration), then readings received from this sensor will be discard until it is repaired. Secondly we check the value of a reading and will discarded it if the value is obviously incorrect (e.g., readings out of possible data range).

Data Storage

A data storage system based on MongoDB is implemented for the system. The database stores all data (including both raw sensory data and data fusion results) as binary JSON documents.

Data Fusion

The central task of the data fusion service is to process sensory data and provide useful information. There are several steps in the data fusion algorithm. First of all, the algorithm learns and updates the statistical model of the air pollution data from high-precision monitoring stations. Secondly, it filters received measurements from low-cost gas sensors to mitigate random errors. Thirdly, the algorithm performs a spatial-temporal interpolation in order to provide a continues air pollution map for the service area based on a limited number of sensors. Finally, it provides useful insights from the measurements, for instance a forecast of air pollution level.

2.2 Forecasting the Time Series Data

The measurements published from the sensor networks are a set of time series data. We use the state-space model to represent the dynamics of air qualities and the sensing mechanism:

$$\mathbf{x}_t = f(\mathbf{x}_{1:t-1}) + \mathbf{u}_t, \text{ and } \mathbf{y}_t = g(\mathbf{x}_t) + \mathbf{e}_t. \tag{1}$$

In the first equation, $\mathbf{x}_t$ are the true value of the concentration of pollutants, $\mathbf{u}_t$ are random system inputs, and $f(.)$ is the dynamic function. In the second equation, $\mathbf{y}_t$ are the measurements from sensors, $\mathbf{e}_t$ are the random errors, and $g(.)$ models the sensors' characteristics.

Various tools can be applied for modelling the dynamic process of air quality, such as auto-regressive (AR) model, Gaussian process (GP) [4] and so on. We apply the GP model to handle the stochastic process of air quality which contains strong periodic patterns. To model the dynamic function with GP, let

$$\mathbf{x}_t \sim \mathcal{GP}(m(t), k(t,t')) \tag{2}$$

where the mean function $m(t)$ and the kernel function $k(t,t')$ are learned and updated based on both historical measurements and latest readings. Assume that $g(.)$ is linear

$$\mathbf{y}_t = \mathbf{H}\mathbf{x}_t + \mathbf{e}_t, \tag{3}$$

and the sensing matrix $\mathbf{H}$ is known for calibrated sensors. Based on the GP model, we perform filtering and prediction for the air quality data (Figure 2).

2.3 Handling the Spatial-Temporal Data

The inputs from a network of sensors are a set of spatial-temporal data. In Figure 3, we take the central area of Uppsala as an example. The concentration level of air pollutants in a geographic area is projected into a grid system, which corresponds to the $\mathbf{x}_t$ in Eq. (2). A snapshot of the readings from the sensor network is $\mathbf{y}_t$ in Eq. (3). The number of sensors is usually much smaller than the size of the service area ($\text{size}(\mathbf{y}_t) \ll \text{size}(\mathbf{x}_t)$). Therefore, a spatial-temporal interpolation is required to provide a continuous air pollution map. To do so, we extend the kernel function $k(.)$ in Eq. (2) to both the time and space domains. Moreover, when the sensors are mobile, the sensing matrix becomes a time-variant matrix $\mathbf{H}_t$. The value of $\mathbf{H}_t$ will be determined by both the sensors' properties and locations at time stamp t.

3 Conclusions and Future Work

An information system for smart cities should not only have the ability to gather data from IoT but also to abstract useful information from raw sensory data and visualize the knowledge in an intuitive way. A part of the GreenIoT project's goal is to explore research questions raised in handling spatial-temporal and heterogeneous sensory data in smart cities.

For future works, the project aims to integrate data from other dimensions such as road systems, energy systems, and social media networks to further improve the quality of services for IoT in smart cities.

4 References

[1] "Economic cost of the health impact of air pollution in europe: Clean air, health and wealth," World Health Organization (WHO), 2015.

[2] "Air quality in europe - 2015 report," European Environment Agency (EEA) of European Union, 2015.

[3] B. Ahlgren, M. Hidell, and E. Ngai, "Internet-of-things for smart cities: Interoperability and open data," *IEEE Internet Computing*, 2016.

[4] C. Rasmussen and C. Williams, *Gaussian Processes for Machine Learning*. University Press Group Limited, 2006.