Geophysical Research Abstracts
Vol. 19, EGU2017-16389, 2017
EGU General Assembly 2017
© Author(s) 2017. CC Attribution 3.0 License.

Matrix Factorisation-based Calibration For Air Quality Crowd-sensing

Clement Dorffer (1), Matthieu Puigt (1), Gilles Delmaire (1), Gilles Roussel (1), Romain Rouvoy (2), and Isabelle Sagnier (3)

(1) LISIC, ULCO, F-62228, Calais, (2) Spirals project-team, Inria Lille Nord Europe, F-59650 Villeneuve d'Ascq, (3) ATMO Hauts de France, F-59044, Lille

Internet of Things (IoT) is extending internet to physical objects and places. The internet-enabled objects are thus able to communicate with each other and with their users. One main interest of IoT is the ease of production of huge masses of data (Big Data) using distributed networks of connected objects, thus making possible a fine-grained yet accurate analysis of physical phenomena.

Mobile crowdsensing is a way to collect data using IoT. It basically consists of acquiring geolocalized data from the sensors (from or connected to the mobile devices, e.g., smartphones) of a crowd of volunteers. The sensed data are then collectively shared using wireless connection—such as GSM or WiFi—and stored on a dedicated server to be processed. One major application of mobile crowdsensing is environment monitoring. Indeed, with the proliferation of miniaturized yet sensitive sensors on one hand and, on the other hand, of low-cost microcontrollers/single-card PCs, it is easy to extend the sensing abilities of smartphones. Alongside the conventional, regulated, bulky and expensive instruments used in authoritative air quality stations, it is then possible to create a large-scale mobile sensor network providing insightful information about air quality. In particular, the finer spatial sampling rate due to such a dense network should allow air quality models to take into account local effects such as street canyons.

However, one key issue with low-cost air quality sensors is the lack of trust in the sensed data. In most crowdsensing scenarios, the sensors (i) cannot be calibrated in a laboratory before or during their deployment and (ii) might be sparsely or continuously faulty (thus providing outliers in the data). Such issues should be automatically handled from the sensor readings. Indeed, due to the masses of generated data, solving the above issues cannot be performed by experts but requires specific data processing techniques.

In this work, we assume that some mobile sensors share some information using the APISENSE® crowdsensing platform and we aim to calibrate the sensor responses from the data directly. For that purpose, we express the sensor readings as a low-rank matrix with missing entries and we revisit self-calibration as a Matrix Factorization (MF) problem. In our proposed framework, one factor matrix contains the calibration parameters while the other is structured by the calibration model and contains some values of the sensed phenomenon. The MF calibration approach also uses the precise measurements from ATMO—the French public institution—to drive the calibration of the mobile sensors. MF calibration can be improved using, e.g., the mean calibration parameters provided by the sensor manufacturers, or using sparse priors or a model of the physical phenomenon.

All our approaches are shown to provide a better calibration accuracy than matrix-completion-based and robust-regression-based methods, even in difficult scenarios involving a lot of missing data and/or very few accurate references. When combined with a dictionary of air quality patterns, our experiments suggest that MF is not only able to perform sensor network calibration but also to provide detailed maps of air quality.