Esopo: sEnsors and SOcial POllution measurements

Vittoria Cozza
IIT CNR, Pisa, Italy
vittoria.cozza@iit.cnr.it

Ilaria Guagliardi
University of Calabria, Italy
ilaria.guagliardi@unical.it

Michelangelo Rubino
Expert System, Modena, Italy
michelangelo.rubino@email.com

Raffaele Cozza
Department of Computer Science
University of Milan, Italy
raffaele.cozza@studenti.unimi.it

Alessandra Martello
University of Calabria, Italy
a.martello@mat.unical.it

Marco Picelli
OverIT S.p.A.,
Fiume Veneto, Pordenone, Italy
marco.picelli@overit.it

Eustrat Zhupa
Uist *St. Paul the Apostle*,
Ohrid, Macedonia
eustrat.zhupa@uist.edu.mk

Abstract

In the following we present the idea of a smart sensor distributed platform where users collect pollution measurements by simply placing a small smart device out of their office or home window, a device that interacts with their smartphones. They provide time-geolocalized information that, trough an app, will be made available to the community that will have the chance to control the pollution level and eventually share it on the most popular social networks along with the related user's opinions and feedbacks. The big data coming from sensors and social networks will be analysed, in combination with local setting area data, in order to have a thorough view of the place the people live in and enhance our environmental conscience. The design of such a project, named Esopo, implies and requires technologies capable of providing data privacy, as it deals also with storing sensitive data, and efficiency, being the corresponding output prone to becoming unusable if not produced in real-time.

1 Introduction

The interest for the environmental issues is increasing continuously in many countries. The rapid development of urban areas has changed the physical, chemical, and biological composition of living environment (Guagliardi et al., 2012). As a consequence, millions of people living in and around urban areas are exposed to an unnatural and unhealthy environment. The increasing awareness that air, water, soil pollution induces a wide variety of adverse physiologic effects to humans, makes people more alert about environmental conditions. In particular, the effects of increasing amount of air pollutants, such as airborne particulate matter, nitrogen dioxide, sulfur dioxide, carbon monoxide on human health have been intensely studied in recent years, leading to unequivocal conclusions: high levels of pollutants are linked to increased rates of allergies in the less serious cases, to high risk of neoplasias in the worst ones. The problems connected to this subject pushed the World Health Organization (WHO, 2015) to define pollution guidelines and thresholds for each pollutants: e.g., the particulate matter annual mean is set to 10 mg/m3 (PM2.5) and 20 mg/m3 (PM10). This attention to environmental pollution can be supported by setting up a platform based on sensors detecting pollutants in the air. It has to be said that there are already systems used to control quality of air, yet the environmental sensors they use are limited in number and distributed in the main roads, consequently they cannot cover the whole area of interest.

We propose a system for collecting pollution data through portable sensors positioned in many points of a specific area and requiring no user intervention (passive detection) and for analysing this data. The monitoring platform should be composed of sensors that can detect enviromental measures such as Carbon Dioxide, Carbon Monoxide, Oxidizing and Reducing Gases and particle sensor. A group of people should be equipped with such sensors to detect pollution in a metropolitan area. Obviously, sensors need to be in contact with the air.

As regards the flow of data collected with sensors, measurements will be sent via bluetooth to people's personal devices and managed by the data logger mobile application, then will be transferred to remote servers and stored permanently. This requires the sensors to be connected to the platform through smartphones and tablets and, even more important, the data about the GPS position (latitude and longitude) to be sent as well as time data (timestamp). Pollution measurements, space and time are the three dimensions, detected anonymously, that will allow us to create a more detailed pollution map to monitor quarters, main and secondary streets of the cities. Pollution data about the current air conditions will be available in realtime to everyone by downloading the mobile application Air quality.

This is the second part of the project, what can be called as active detection. Indeed, every user can also share measurements on the most popular social networks, for instance by posting the level of PM10 in a particular area and adding a photo or a comment in Twitter as in Figure 1. This way,

Figure 1: Pollution measurement shared on Twitter.

everyone will be able to get information about the quality of air or to compare their perceptions with what is shown by the Air quality app. The project is therefore aimed at developing an application able to map data related to the environmental pollution and sharing it via social networks with the possibility of adding personal sensations.

Actually, many geographical areas are covered by pollution detectors, they are analysed according to pollutant thresholds that change on the basis of the law. Yet enviromental sensors have a sparse distribution on the territory. Such a system will provide the users with realtime information about the current air pollution levels within a denser map. As a plus, Esopo will analyse their opinions when they share air information on the most common social network, adding personal comments.

Geolocalized and time-tagged data along with user's opinions, will be used to generate a big pollution dataset.

The collected data will be accessed through multidimensional indexing structures (based on space, time and text). For deeper analyses it will be merged with other data regarding the area (data fusion) (Guo and Hassard, 2008; Guo et al., 2010).

2 Social and Smart Sensor networks

Nowadays, the interest for environmental issues is not restricted to few people, but it is growing, as well as the demand for pollution data in the Healthcare and the Security sectors. Actually, almost everyone owns a personal smart device, be it a smartphone or a tablet. The pollutants are usually concentrated in specific spots, e.g., the main city roads, as they are typically emitted by motor vehicles, but just few people know that their concentration can be different based on the time, usually more concentrated in the morning and less in the evening and night. We aim at designing a wireless sensor network to collect "affordable" information. When we use the adjective "affordable" we refer to open data, available for a free access. Citizens can monitor directly the quality of air combining the environmental sensors with their own mobile sensors. On the market, there is a number of sensor devices to be implemented in such kind of network and useful to control pollution 24/7. By law, the distribution of pollution control units must be revised periodically and, during this period, many things may change. Typically, just few control units can map all the pollutants. A smart distributed sensor network would cover a larger territory, also detecting more pollutants. This solution would integrate the control units network with low costs for citizens and private companies. Once implemented, it will provide people with valuable information related to their life and this knowledge may eventually motivate them to advocate for a change.

We propose to build a sensor network collecting chemical, physical, and biological measurements, monitoring the environment with the minimum of user intervention: enhancing a number of user's smartphones with ad hoc sensors.

2.1 Sensors devices

We believe that a so called "social" wireless network of sensors should satisfy some general requirements, such as:

1. Sensors should be able to monitor the air quality around a user, especially nitrogen dioxide (NO2), sulphur dioxide (SO2), carbon monoxide (CO) and the so-called PM10.

2. The monitoring process should need very limited user intervention.

3. Sensors should be extensible, namely should be possible to connect, through an external interface, other peripherals, in order to extend the sensors capabilities for future purpose.

4. Sensors must connect in a wireless way to some portable smart devices, in order to gather data and send it to the back end infrastructure. A smartphone is believed to be the most suitable device for this task.

5. Sensors must be commercial off-the-shelf (cots).

Moreover, it is important to safeguard user privacy. To this aim, we define the following software requirements:

- All the possible API to access the sensors capabilities must be open source, so that it would be possibile to inspect them looking for threats to the privacy of the user.

- All applications developed must be released under open source licenses, so that each user - or central authority - could inspect them and be assured that there is no threat for the privacy of the user.

Given these requirements, we find out that there are just a few sensors suitable for our purpose (at the time of our study) namely: Air.Air sensors[1], CitiSense sensors[2], Sensordrone sensors[3] and M-Dust sensors, an innovative low-cost smart pm sensor[4]. Unfortunately, none of the previous options are able to satisfy all the requirements. For options 1) and 2), we were unable to find out whether these devices were really cots or just a proof of concept. Moreover, we were unable to understand exactly what kind of air parameter they are able to monitor and the API available to communicate with the sensors. Option 3) instead seems very interesting. Sensordrone (Sensordrone, 2015), in Figure 2, is a portable & wearable multisensor connectable to portable smart devices and can be turned into a multi-function environmental monitor: a carbon monoxide leak detector, a non-contact thermometer, a lux meter, a weather station.

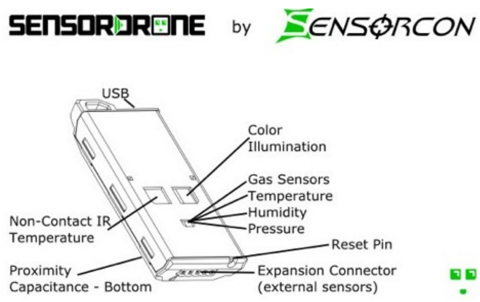

Figure 2: Sensordrone

From the information available on the website, it is able to communicate with Android and Apple devices through Bluethooth technology and it offers a well-defined, open source, API. Moreover, it is able to monitor most of the air quality parameters we need, but not the PM-10. On the contrary, M-Dust (Must, 2015; Soldo et al., 2012) can monitor PM10 and PM2.5 parameters, but it is not capable of directly communicating with a device such as a smartphone.

The idea then is to use the Sensordrone, extended with the particle sensor MDust for controlling PMs. This is possible since Sensordrone has also an expansion port were we can connect, as an expansion device, M-Dust. In interfacing Sensodrone with M-Dust, we must take into ac-

[1]https://www.kickstarter.com/projects/1886143677/airair-portable-air-quality-detector

[2]http://ucsdnews.ucsd.edu/pressrelease/small_portable_sensors_allow_users_to_monitor_exposure_to_pollution_on_thei

[3]http://sensorcon.com/sensordrone-1/

[4]http://www.particle-sensor.com/

count a technical inconvenient though, as the expansion port of Sensordrone accepts, as input, a signal ranging from 0 to 3 V, while M-Dust outputs a signal ranging from 0 to 4 V. The different input/output requirements, however, can be compensated using a voltage divider. However, at this stage, we are unable to exactly pinpoint the effect of the loss of resolution in doing this adaption.

2.2 Data flow

All the data measured with the environmental sensors, being it on demand or periodic, is saved in a local database on a personal Sensordrone device, that can be easily sent as a csv file to our data logger application, installed on the user smartphone. Each hour the data logger downloads a new csv file from Sensordrone and merges data acquired by the pollution sensor, the timestamp and other information coming from the smartphone, such as the GPS data tracking the position of the user.

Detected data is sent to a repository. Indeed the continuous data flow, composed of measurements, timestamps and user location, is sent and stored to the big data repository. At the same time, the user can visualize the current measurements, as well as the data of the past, by using an air quality application on his/her smart device. The repository will also collect text written by the users while they share and comment sensor data on the most popular social networks. The chosen repository will support georeferenced data, a hybrid multidimensional index to speed up soft real-time and offline analyses. The data can be retrieved with a popularity spatio-temporal-keyword search engine, as proposed in (Cozza et al., 2013). The system will also provide a polarity detection module to understand whether comments express a positive or negative sensation about the air and an opinion mining module to extract relevant information from unstructured social comments.

2.3 Privacy concerns

Constraints about profiling and trustworthiness will need to be matched. As regards profiling, different user profiles will be investigated to understand to what extent confidentiality and multiple view and map customisations can be satisfied, namely not only the information to show but also how to show it. Particular attention will be given on storing, processing and sharing that data referring to more than one subject (so called "multi-subject personal data") (Gnesi et al., 2014). We envisage a support architecture based on privacy policies, through which users can edit their privacy preferences, appropriately enforced at the time of the actual data processing (Casassa-Mont et al., 2015).

The second aspect to consider is the Data trustworthiness: an in-depth analysis of data trustworthiness is required to identify and test a model able to exploit geotagged data and to get the highest level of reliability. Sensor data cannot be linked to any user and its physical location.

Privacy concerns arise beyond data content and focus on context information such as the location of a sensor initiating data communication. The problem of data unintentionally shared when using and producing georeferenced information is formulated and discussed in (Friginal et al., 2013), (Cortez et al., 2015).Table 1 summarizes the data categories involved in Esopo: volunteered entered by the users, observed and inferred. Data can be inferred by information intelligence analysis: data fusion can enhance informativeness of data coming from sensors, Opinion Mining and Polarity Detection or any Social Network Analysis can be performed on these data shared on social networks through the air Quality app.

Volunteered data	**Observed data**	**Inferred data**
Pollution measurements	Online activity (time, location from GPS), Pollution level exposures	Habits and lifestyle
Interaction with airQuality app: sharing, comment, likes	Online activity (time, location from GPS)	Opinions on air quality, relationship: friends, followers, mentions

Table 1: Social and sensor data

For a state-of-the-art survey of existing privacy-preserving techniques in WSNs readers should refer to (Li et al, 2009).

2.4 Applications

Due to the variety and the volume of data retrieved and analyzed we can offer several applications.

This big data can be "consumed" in real-time, as it is "produced", and the user can get the current level of pollution. This way, we can inform users about the air pollution at a specific moment of the day and can get reports, statistics and charts to evaluate the pollution level in a particular area, over the last weeks or months. This will be possible through a modern user friendly interface application. Anytime the users have the chance to check on personal devices the information about pollution (consuming information) and are able to share measurements, they will send comments and images (producing new information) to the most popular social networks. It is already available an air quality mobile application for Sensordrone providing the end users with a view about the quality of air for a specific area and a second application that allows them to take measurements with the Sensordrone and to post them to Facebook, Twitter, Google+.

Big data can be "consumed" offline as well, this means that we can combine information about the environment: pollutants, humidity, oxidizing and reducing gases with social information and other information about an area and therefore produce statistics and predictive analyses. Furthermore, we think that if we carry out experiments and analyses in a metropolitan area of Italy, let's say Milan, where there are already air pollution stations, then it would be significant to combine this data with that coming from the wireless sensor network to enrich the information provided to end users and have a wider view of the environmental conditions in different locations of that metropolitan area, even those ones not monitored by Arpa (Arpa, 2015).

All the data acquired so far has a relevant added value for others applications too, as it is collected not only pollution data, but also data about the people daily movements for instance: the starting point of each journey and where it ended, the most common paths (let's call it mobility data); data about the busiest areas; data about public places where people connect to the internet (libraries, council houses, schools...). In a nutshell, a large amount of data that could be further analysed.

3 Related Work

In the main geographical areas, air quality data currently available from government agencies does not provide enough detailed measurements within particular neighbourhoods, then several projects have focused on increasing the spatial resolution of air pollution data using ubiquitous sensor networks. These works did raise the spatial granularity compared with data from fixed air pollution monitoring sites. In (Devarakonda et al., 2015) the authors present a vehicular-based mobile approach for measuring fine-grained air quality in real-time. They provide users with a small sensor that they should bring on their veichles or in public transportation to collect realtime information.

In (Hu et al., 2014) the authors combine air pollution and human energy expenditure data to give individuals real-time personal air pollution exposure estimates. They monitor pollution in an area and at the same time analyse users' life behaviours, specially they apply multiple data mining techniques to find out associations among activity modes, locations and the inhaled pollution.

The authors understand the relevance of automatically analysing how pollution level are perceived by people and combining air pollution exposure with personal health.

In (Leonardi et al., 2014) the authors propose SecondNose, an air quality mobile crowdsensing service, aimed at collecting environmental data to monitor some air pollution indicators to foster participants reflection on their overall exposure to pollutants. At the time of the work, SecondNose aggregates more than 30k data points daily from 80 citizens in Trento, northern Italy. Esopo has many features in common with SecondNose, in addition it encourages users to share pollution measurements on the social networks and, consequently, the combined analysis of sensor data and social data.

4 Conclusions

In this work we have described the project idea of a smart sensor network that stores environmental information from sensors and eventually collects social network comments about it. Sensor data can be shared through users that will have a real-time snapshot of the enviromental pollution in a defined area at a specific time and that may want to add their personal sensations. Sensors will share minute-by-minute air quality measurements that could provide a better understanding of risks related to potentially harmful exposure in the area and eventually identify patterns for any given day, week, month or year.

Furthemore, the analysis of sensor data combined with people's sentiment on social networks related to it, permits a semantic analysis of collected measuremens through sensations perceived by people.

5 Acknowledgments

Work partly supported by the Registro.it project My Information Bubble (MIB). The authors thank Paolo Palana and Antonio Gabriele for sharing their technical knowledge on sensors and giving them support.

References

Arpa, Agenzia regionale per la protezione dell'ambiente. Homepage: http:/ita.arpalombardia.it/ITA/index.asp

Casassa-Mont, Marco and Matteucci, Ilaria and Petrocchi, Marinella and Sbodio, MarcoLuca: Towards safer information sharing in the cloud International Journal of Information Security, volume 14, number 4, pages: 319-334. issn 1615-5262, Springer Berlin Heidelberg.

M Nunez del Prado Cortez, J Friginal, Geo-Location Inference Attacks: From Modelling to Privacy Risk Assessment. Tenth European Dependable Computing Conference (EDCC), New Castle, UK., May 2014.

V. Cozza, A. Messina, D. Montesi, L. Arietta, M. Magnani: Spatio-temporal keyword queries in social networks. In Springer, editor, B. Catania, G. Guerrini, and J. Pokorn(Eds.): ADBIS 2013, volume 8133 of LNCS, pages 70-83, 2013.

Srinivas Devarakonda, Parveen Sevusu, Hongzhang Liu, Ruilin Liu, Liviu Iftode, and Badri Nath. 2013. Real-time air quality monitoring through mobile sensing in metropolitan areas. In Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing (UrbComp '13).

J. Friginal, J. Guiochet, M.-O. Killijian: Towards a Privacy Risk Assessment Methodology for Location-Based Systems. Mobiquitous2013.

Stefania Gnesi and Ilaria Matteucci and Corrado Moiso and Paolo Mori and Marinella Petrocchi and Michele Vescovi: My Data, Your Data, Our Data: Managing Privacy Preferences in Multiple Subjects Personal Data Proceedings of Privacy Technologies and Policy - Second Annual Privacy Forum, APF 2014, Athens, Greece.

I. Guagliardi, D. Cicchella, R. De Rosa, 2012. A geostatistical approach to assess concentration and spatial distribution of heavy metals in urban soils. Water, Air & Soil Pollution 223: 5983-5998.

Yajie Ma, Yike Guo, Moustafa Ghanem: Distributed pattern recognition for air quality analysis in sensor network system. IADIS International Journal on Computer Science and Information Systems Vol. 5, No.1, 2010, pp. 87-100ISSN:1646-3692

Yike Guo, John Hassard: Air pollution Monitoring and Mining based on Sensor Grid in London Sensors (Basel). Jun 2008; 8(6): 3601-3623.

Ke Hu, Timothy Davison, Ashfaqur Rahman, and Vijay Sivaraman. 2014. Air Pollution Exposure Estimation and Finding Association with Human Activity using Wearable Sensor Network. In Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis (MLSDA'14).

Chiara Leonardi , Andrea Cappellotto , Michele Caraviello , Bruno Lepri , Fabrizio Antonelli, SecondNose: an air quality mobile crowdsensing system, Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational, October 26-30, 2014, Helsinki, Finland

Na Li and Nan Zhang and Sajal K. Das and Bhavani Thuraisingham: Privacy preservation in wireless sensor networks: A state-of-the-art survey. Ad Hoc Networks 7 (2009), p. 15011514, 2009.

D. Soldo, A. Quarto, V. Di Lecce: M-DUST: an Innovative Low-Cost Smart PM Sensor. International Instrumentation and Measurement Technology Conference. Measurement Technology Conference: Industrial Track - Protecting the Environment. Graz (Austria) - April 14th, 2012, pp. 1823-1828.

Sensordrone project. Homepage: http://sensorcon.com/sensordrone/

MDust homepage: http://www.particle-sensor.com/

World Health Organization (WHO), 2015. Homepage: http://www.who.int/en/