Sensors talk and humans sense

towards a reciprocal collective awareness smart city framework

Athena Vakali, Lefteris Angelis, Maria Giatsoglou
Department of Informatics
Aristotle University
Thessaloniki, Greece
{avakali,lef,mgiatsog}@csd.auth.gr

Abstract-*Smart city infrastructures provide unique opportunities for innovative applications developing and testing. Sensor city installations offer the ground for experimenting with user-oriented services, which at the same time can test and improve the infrastructure itself. The proposed work summarizes principles and methodology for and experiment, entitled SEN2SOC which will bridge sensor measurements and social networks interactions via natural language generation for supporting smart city services. SEN2SOC aims at exploiting the SmartSantander infrastructure in a sensor to social reciprocal fashion such that the sensor measurements will be and communicated to the public (citizens, authorities, etc), while social networks users activities in relevance to sensors social postings will be analyzed and summarized both to verify sensors reporting and to develop collective aware applications.*

Keywords-smart city,sensors data management, social networks mining, collective aware applications

I. *INTRODUCTION*

The Internet of today enables users to access an always increasing amount of information ubiquitously, anytime and from any device. In the emerging Internet of Things (IoT) scene, sensors act as information storehouses of their own, capable of collecting and transmitting real-time data. Utilizing such data to their surrounding environment (people, places and things) has major impact in the smart city short-term and long-term activities. Urban environments offer unique opportunities for developing and testing new applications and platforms in line with the vision of the Internet of Things. European IoT platforms have already begun emerging over the last few years inline with the future Internet momentum. Large smart city infrastructures have now been set up worldwide (e.g. Songdo *http://www.songdo.com/*, South Korea) and in Europe the SmartSantander *http://www.smartsantander.eu/* infrastructure offer a fertile ground for experimentation. The growing need and interest in smart city innovation was highlighted by the Commission in its report "Internet of Things in 2020: A roadmap for the future", in which it identified key topics such as "Smart living" as part of what it termed a "mastered continuum of people, computers and things". There is a growing number of innovative social and people-centric application areas, including social networking, smart metering, smart data collection, city information models and so on. It is also clear that with the growth of Web 2.0 and social media, a wide sharing of information and know-how is held and such social networking activities can be properly harvested for the creation of smarter urban living environments. Utilizing sensor data production and social interactions can lead to beneficial applications addressing both individual and authority needs.

Current IoT and Ios applications offer opportunities to test services and infrastructures, and they offer merely quantitative solutions since they typically collect and store data and information from technical devices and sensors [1]. Networked sensors are widely used in the context of smart cities in order to collect accurate measurements of physical parameters and phenomena that have an impact of the citizens' life. With appropriate analysis methods, such kind of data can be leveraged to identify the occurrence of special phenomena (such as traffic congestions or atmospheric pollution at a particular area/time period), and take short-term (e.g inform/alert citizens and especially vulnerable social groups) or long-term actions (e.g. understanding the progress of the phenomena and trying to address them in order to improve the city's conditions). So far, smart sensor networks provide information about some parameters of a given phenomenon without having the knowledge of its effects on humans and the way that the citizens sense it. Thus, there is clearly the need for extending the existing sensor networks by "human-networks" that will "sense" and provide information for the same or similar facts [14].

This paper is about an ongoing experiment which is motivated by the fact that it is now time to go beyond quantitative outcomes, since people are primarily interested in qualitative human-oriented solutions. The proposed experiment framework (SEN2SOC)[1] is designed and will be implemented for a smart city setting, namely under the SmartSantander infrastructure. SEN2SOC considers reciprocities in quantitative sensor-generated data along with qualitative human-generated data exposed in social networking platforms. With such an approach storytelling and "listening" by networked objects (sensors) is enhanced and vetted by human storytelling, thereby getting closer to realizing collective awareness tasks and collaborations. Such a twofold experiment embeds the next major scientific, research and technical challenges :

[1] *This work is supported by the EU Smart Santander project under its 2[nd] call for experiments under the experiment "SEN2SOC : bridging SENsor measurements and SOCial networks interactions via natural language generation for supporting smart city services", implemented by Aristotle University, 2013.*

- *Data flows and demanding data processing* since sensors' large-scale flows or streams of data need recording and measurement in a continuous manner. Real-time processing requirements are demanding since users recommendations and instructions/ suggestions must be based on the most recent values of sensors' measurements. Heterogeneous data collected from both sensors (traditional and human) and people's activities (within SEN2SOC applications and third-party location-based Web2.0 applications) must be integrated to result in more accurate collective awareness applications. Various parameters (e.g. temperature, humidity, traffic, etc) need to be captured in a sufficient and at the same time their interpretation in natural language should be carefully implemented;

- *Improving and extending the existing capabilities* of the SmartSantander infrastructure by evaluation and validation of the proposed services which will capture both the users' opinions and infrastructure's performance. SEN2SOC platform will be tested to verify the robustness and suitability of the SmartSantander for supporting the SEN2SOC services, as well as to evaluate the efficiency of the SEN2SOC server-side modules. The platform testing will include measuring the SS responsiveness rates based on scaling number of requests for real time sensor data (focusing on particular areas), as well as the overall response rates of the different SEN2SOC services (including the network latency and the algorithms' time complexity).

- *Qualitative and public privacy requirements*, since SEN2SOC applications must support quality of services under both a user-friendly mobile and web interfaces. Since information is pushed to and pulled from users (citizens, tourists, authorities) particular concern is placed on the issues of anonymity and privacy preservation. In terms of users navigation monitoring a collective activity capturing is suggested such that there is no supervision of the individual's navigation and preferences but only group's activities. The SEN2SOC members will impose anonymity and privacy preservation in all relevant to data collection and user information management tasks. This will be implemented by embedding privacy guaranteed practices while exploiting conclusions raised in recent relevant efforts which enable privacy-aware useful services [4].

The rest of the paper is structured as follows. Section II reviews the current state regarding Internet of Things in smart environments and natural language processing approaches that will leverage sensor to human interactions. Section III presents the methodology to be followed for sensors talking and human sensing whereas in Section IV a motivational scenario highlights the experiment's impact and the paper concludes with emphasis on impact and collective awareness nature.

II. RELATED WORK

Providing effective scalable applications and services which will exploit existing smart city infrastructures is still at an initial stage due to their recent emergence and applicability. Monitoring of such new IoT infrastructures can exploit data that capture conditions (environmental and other) and can lead to improved information analysis, which is vital for significantly enhancing collective awareness and policy making. The SEN2SOC dual nature imposes the need to exploit current approaches in emerging and evolving scientific fields, such as the IoT in smart contexts, awareness and recommendations on location-aware frameworks as well as mobile applications and services. The proposed framework is inline with the principles set in [14], where sensor data production is utilized for social networks exposure.

Sensors of all kinds are key sources of numerical and non-textual data, embedding knowledge that has to be extracted. Statistical and data mining methods are used to manipulate the data; however there is always the need for humans to receive the raw data or the results of their processing in a language that can easily understand. That is why a relatively recent research area is evolving, aiming to the communication and interaction between the sensor data and the humans through Natural Language Generation systems. Natural Language Generation (NLG) systems generate texts in human language from non-linguistic data. A number of data-to-text systems have been developing for more than a decade now and there are already several applications of data-to-text systems and the research is ongoing and challenging as the available data become more and more complex (*http://www.nlg-wiki.org/systems/*).

Important research work on NLG and their applications in decision support has been done by researchers at Aberdeen University in a series of papers spanning more than a decade ([12],[16],[11],[5]). Applications involve the production of textual summaries based on different disciplines (such as data from gas turbines, medical applications, etc). Moreover, applications utilize NLG for geographical and spatio-temporal information for geographic descriptions, route directions and wheather forecasts ([14],[8]). In [10] the problem of generation of narrative summaries of daily events, based on sensor data from various sources is tackled. Specifically, in this paper a 'clustering' of data into events facilitates choices on the basis of user 's interests. Research on NLG is directed also towards more theoretical and formal issues. In [7] a uniform mobile terminal software framework is presented for providing systematic methods for acquiring and processing useful context information from a user's surroundings, through multiple sources and sensors, and giving it to applications. Research has been conducted also on technical issues and performance of NLG systems and specifically of sentence planners [3] and [2].

In SEN2SOC the goal of the proposed natural language processing is to convert sensor and other numerical data to sentences that can be understood by humans. The text produced by such an NLG approach will be in the form of a short summary of limited text (i.e. a posting).

III. SEN2SOC FRAMEWORK : SENSORS TO SOCIAL NETWORKS ENCOUNTERS

SEN2SOC is based on the idea of bridging sensor measurements and social networks interactions via natural language generation for supporting smart city services. As explained above it is motivated by the fact that utilizing smart interconnected objects (sensors) along with social networking activity contributes in collective awareness and intelligent urban landscapes. The experiment is in accordance to the Future Internet smarter cities vision, and its particular IoT and Internet of Services (IoS) priorities. The SEN2SOC experiment will support a dual side platform which will utilize sensors data production and social networks interactions and it will increase collective awareness for various audiences and communities.

SEN2SOC has designed and has under development a dual side platform, which will exploit SmartSantander infrastructure in a twofold manner: i) utilize the sensor side measurements which will be interpreted in natural language postings uploaded in popular social networks and ii) follow the social networks side activities around sensors postings and capture "human sensing" and crowd observations. Different all sensor types and categories will be utilized covering different topics of relevance to city audiences such as traffic conditions, pollution, temperature, humidity etc. With respect to the users' services provision, SEN2SOC will exploit IoT (sensors) data and also user information provided either directly (users' explicitly stated preferences) or indirectly (analysing users activity on social networks) to support activities of specific targeted audiences under the SmartSantander infrastructure.

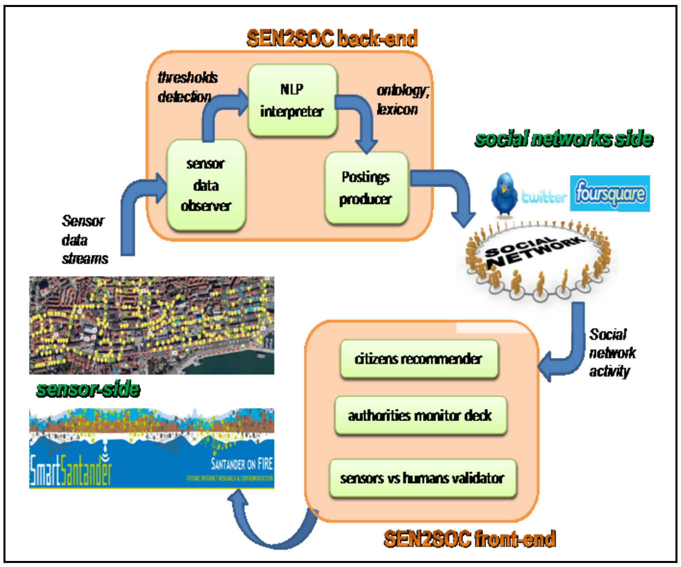

Figure 1. *SEN2SOC framework outline*

An overview of the SEN2SOC framework is depicted in Figure 1 which emphasizes the fact that there is a need to tune the quantitative measures from the infrastructure level (i.e. sensors installations) with the qualitative users' activities at the client level (i.e. users with access on popular geolocated social networks). SEN2SOC platform functionality is based on:

- *a back-end module* for sensor data streams observations and postings generation when sensors thresholds are reached;
- *a sensor to social interpretation mechanism* which will get sensor data streams and will produce textual feeds;
- *a front-end module* which will address individual, authorities and also platform testing requirements and will be updated via the social networking activity

The methodology of the experiment involves particular tasks which will be realized by state of the art algorithms and techniques which will allow appropriate sensor and social networks data bridging and integration.

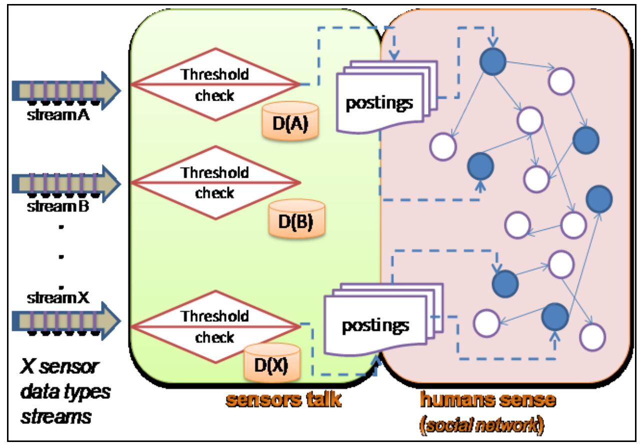

Figure 2. *Sensors talk and humans sense flow*

A. *SENSORS TALK via bursts detection and alerting*

Sensors produce data streams in accordance to their scope and in a constant manner. Such data streams are of purely numeric character with values for the monitoring condition (humidity, noise, etc) and with the time stamp of the recorded value. As depicted in Figure 2 ("sensors talk" left side) for a given sensor setting of X sensor types, data streams are produced which are monitored in order to deliver postings once specific threshold limits are violated or once value ranges are determined. The sensor burst detection methodology will rely on detecting patterns in the sensors values deviating activities and for this purpose, an analytic model will be used to define the threshold limits for each sensor type.

A specific sensor stream is represented by a set $SenStr_X$ which refers to a particular sensor type X (i.e. humidity, noise, light, etc) which at a particular time period of T time units produces N specific measurements in the range of the X sensor type values (as specified by the set V(X) for each sensor type). Therefore such data stream for each sensor is expressed by :

$$SenStr_X = \{(v_i, t_i) / v_i \in V(X), t_i \in [0,T], i \in N\}.$$

Data collected from multiple sensors involve measurements of each sensors set's monitoring conditions (humidity, temperature, noise, traffic, parking, etc) and a suitable storage scheme is designed, with appropriate scalable indexing schemes to cope with the ever-increasing sizes and unpredictable rates of sensor data streams. Then, the "sensors talk" can be formulated as next.

Problem Definition I : Given a set of timestamped sensor values $SenStr_X$ detect data deviations from specified threshold values and identify bursts such that natural language postings are generated.

ALGORITHM I : SENSORS TALK ALGORITHM

```
REQUIRE : T, N values of SenStr_X , thresholds_X
-----------------------------------------------
(v_i,t_i) ε SenStr_X
burst_indicator ← 0
time_step←TS

for j=1 to T step TS
    for i=1 to N
        if v_i NOT IN thresholds_X
                then burst_indicator ++
        if monitor (burst_indicator)=TRUE then
                posting_generate(X, t_i,social_net)
```

Algorithm I summarizes the proposed algorithm such that simple steps are followed in order to capture out of thresholds uprising and momentum. These captures are crucial for triggering and initiating alerting which will publicize deviating from the norm city conditions. The *posting_generate* process will generate simple sentences which will be posted on a social network such as Twitter (with 140 characters limitation) and automatically be assigned to specific dedicated sensor accounts created by the experiment team. This sensor to natural language interpretation will be carried out via a simplistic mapping of sensor value ranges to specific sentences (postings) which will be compliant with each of the X sensor types scope. More specifically, the natural language posting will be enabled by the use of the particular sensor type dictionary D(X) such that for each sensor type X the corresponding terminology is properly identified. This dictionary will then be used to syntax the specific sentences which will be the actual postings in the social network.

B. HUMANS SENSE via social network interactions

Since sensor postings and signaling will be produced automatically, a social network of its own will emerge with both sensors and humans exchanging sensing experiences. As depicted in Figure 2 ("humans sense" right side) the social network involves both sensors (dark) and human (blank) nodes which interact on the basis of "sensors talking". For example, once a posting is published at a sensor account for humidity, citizens can comment and also add preferences and opinions which can be populated in the social network itself. Then this specific social network activities can be monitored to capture humans sensing and to validate sensors postings for a specified set of postings P, a set of U_S (sensor) and U_H (human) social network users and their comment set C.

Problem Definition II : In a specified social network with a given a set of postings P originating either by U_S sensor or by U_H humans accounts, specific analysis is required to result in groups of both sensors and humans with respect to their similarities in social network activity.

This problem is resolved by using an appropriate similarity measure and a simple clustering technique which will result in clusters of (both sensor and human) users. The following similarity value will capture common users activity for a total set of U users ($U = U_S \cup U_H$) which expose their comments $c_{user,p}$ (in the form of terms, ratings etc) in response to a particular posting p :

$$Sim(u_i, u_j) = \frac{\sum_{p \in P} (u_i, c_{i,p}, p) \cap (u_j, c_{j,p}, p)}{\max(\sum_{p \in P} (u_i, c_{i,p}, p), \sum_{p \in P} (u_j, c_{j,p}, p))}$$
$$\forall (u_i, c_{i,p}, p), (u_j, c_{j,p}, p), \text{where } u_i, u_j \in U$$

Based on this similarity a clustering algorithm (such as k-means) will detect k groups of users (i.e. clusters) which exhibit similar opinions.

The front-end part can then utilize these clusters to offer services which will embed collective awareness (such as recommendations, sensor-aware route directions etc). These SEN2SOC services (indicative ones are outlined at the front-end in Figure 1). Such services can and will leverage users-relevant activity metrics (postings per time unit, overall number of likes and/or views, etc) to offer recommendations for citizens, status statistics for authorities, as well as sensor to human validations to IoT researchers. As indicative scenario for such services is given in next Section IV.

IV. TOWARDS IMPLEMENTING SEN2SOC IN SMARTSANTANDER

SEN2SOC is at its initial stage and it has proceeded in identifying the modules, the methodology to be followed and the audiences targeted. In terms of the technologies used it will support data collection from various sensors and from geo-located social networks and the refinement of these data sets will result in an integrated data repository. At the same time the social data collection will be held with techniques which adapt to the publicly available Application Programming Interfaces (APIs) of location-based Web 2.0 services (such as Foursquare and twitter API). For data collection and storage NoSQL solutions will be utilized to support social media data streams heterogeneity and evolving data requirements. Next subsections provide a motivational scenario and SEN2SOC impact highlights.

A. Motivational scenario:a SEN2SOC day in Santander

It was 9am in the morning when the Spanish minister of culture suddenly announced that in 2 hours he'll arrive to Santander in an unofficial visit (since he was visiting a nearby location) and he expressed his wish to communicate mostly with the University students. The City Council had to respond effectively both for organizing his mobility around the city (from his arrival at the Railway Station to the City Council and then to all of his visits) and for publicizing his talk to the University of Cantabria students. For this purpose, City Council officers access SEN2SOC Web platform (authorities monitor deck) and they get all of the minister's routes nearby sensors and human postings in terms of the city's (traffic, noise, atmospheric, etc) status for the last 6 hours. City council officer noticed that there was a red alert posting coming out of the Calle de Alfonso sensors in terms of the traffic and noise levels and this was also verified by the corresponding (human, sensors) social network postings, so they immediately deviated his transport route via a better quality sensors neighbourhood which had no alert postings in the last two hours.

It was noon time when Carlos, who is a junior student for the degree in History at the University of Cantabria (Santander, Spain) enrolled at the "Early Contemporary Spanish History" course, was at the University campus when at his class facebook account the posting for the Minister's talk arrived. He noticed that this talk was scheduled to be held at the Biblioteca y Casa-Museo de Menéndez Pelayo at 2pm. Along with his classmates decided to go to the talk, but since Carlos suffers from acute chronic asthma attacks he had to look for the atmospheric conditions in downtown Santander in order not to risk his health conditions. Therefore, he accessed SEN2SOC mobile application (citizens recommender)to get summaries of sensors and humans postings in relevance to humidity, temperature levels and real feeling which greatly influence his health. He noticed that they was a one hour earlier posting ("Estoy en la plaza, hace demasiado ruido, me voy. Es my molesto.") from sensors side located at Plaza de Toros which was verified by about 10 more users in twitter, whereas there were no postings alerts originating from sensors and/or humans social network in an alternate route he picked from his origin (University) to his destination (Biblioteca. Since there were no reports of measurements and of real feeling over the allowed limits for his health condition he decided to join his classmates and attend the talk.

B. *SEN2SOC Expected impact*

SEN2SOC will define new ways of capturing city data from sensors and utilizing collective activities by: i) making recommendations to users, in order to improve their navigation within the urban environment, ii) allowing city authorities to have access to patterns and knowledge mined from city data through user friendly visualization interfaces. SEN2SOC will increase collective awareness for seamlessly communicating sensor data to services oriented to citizens such that city administration and e-governance is improved. SEN2SOC will provide new technology for city authorities for understanding of issues citizens encounter in their real life and of sensor data and comments that citizens provide to initiate handling of these issues. SEN2SOC will provide new technology for coupling such contextualization of issue reporting with data coming from multi-sensors, to draw policies and plans aiming at giving citizen a better understanding of governmental actions and governmental actors increased insight via citizens' collective intelligence.

The SEN2SOC techniques will ensure increased levels of environmental and mobility awareness for decision makers (i.e. city authorities) and shorter feedback cycles among citizens and authorities. Via a web-based application the city authorities will have access to data from sensors and users acting as "sensors" showing the current state in each area of the city with respect to environmental, mobility and users parameters (e.g. user likes and comments). Apart from raw data, city authorities will also have access to historical data and to data mining applications that allow them to analyze data from different perspectives and summarize it into useful information (e.g. mobility problems that persist in specific areas versus those occurring episodically, maybe due to an event). This will result in appropriate policy planning which considers all the location dependent variables (such as noise, temperature and CO levels) and aims to improve living conditions in areas detected as "problematic" which both exhibit over-threshold sensor measurements and constant citizens reports.

REFERENCES

[1] L. Atzoria, A. Ierab, G. Morabitoc: "The Internet of Things: A survey", Computer Networks, Volume 54, Issue 15, 28 October 2010, Pages 2787-2805

[2] N. Chambers. Real-Time Stochastic Language Generation for Dialogue Systems. European Workshop for Natural Language Generation, Aberdeen, Scotland. pp 40-48. 2005

[3] J. Chen, S. Bangalorey, O. Rambow, and M. A. Walkery.Towards automatic generation of natural language generation systems. COLING 19th international conference on Computational linguistics, 2002.

[4] D. Freni, C. Ruiz Vicente, S. Mascetti, C. Bettini, C.S. Jensen. "Preserving location and absence privacy in geo-social networks". Proceedings of the 19th ACM International Conference on Information and Knowledge Management CIKM, pp. 309-318, 2010

[5] D. Guinard, V. Trifa, E. Wilde: "A resource oriented architecture for the Web of Things", Internet of Things (IOT) Conference, Dec. 2010.

[6] Hunter J, et al. Automatic generation of natural language nursing shift summaries in neonatal intensive care: BT-Nurse. Artif Intell Med (2012), Artificial Intelligence in Medicine, In Press, Corrected Proof, Available online 12 October 2012

[7] Korpipaa, Panu, Jani Mantyjarvi, Juha Kela, Heikki Keranen, and Esko-Juhani Malm. "Managing context information in mobile devices." Pervasive Computing, IEEE 2, no. 3 (2003): 42-51.

[8] Martin Molina and Javier Sanchez-Soriano, Generating text descriptions for geographically distributed sensors, IADIS International Conference Applied Computing 2011.

[9] L.D. Prete, L. Capra: differS: A Mobile Recommender Service. In Proceedings of the 2010 Eleventh International Conference on Mobile Data Management (MDM '10). IEEE Computer Society, Washington, 21-26, 2010.

[10] Josheph Reddington, Nava Tintarev. Automatically Generating Stories from Sensor Data. Proceedings of the 16th international conference on Intelligent user interfaces. 2011, Pages 407-410.

[11] E. Reiter, An architecture for data-to-text systems, in: Proceedings of the 11th European Workshop on Natural Language Generation (ENLG-07), Schloss-Dagstuhl, Germany, 2007, pp. 97–104.

[12] Roy D.K. and E. Reiter. Connecting language to the world. Artifcial Intelligence, 167(1-2):1,12, 2005.

[13] T. Sandholm and H. Ung: Real-time Location-aware Collaborative Filtering of Web Content. In Proceedings of the 2011 Workshop on Context-awareness in Retrieval and Recommendation (CaRR '11). ACM, New York, 14-18.

[14] L. Srivastava and A. Vakali : "Towards a Narrative-Aware Design Framework For Smart Urban Environments", EU Future Internet Assembly (FIA) Book 2012, Springer, 2012

[15] Ross Turner , Somayajulu Sripada , Ehud Reiter , Ian P. Davy, Using spatial reference frames to generate grounded textual summaries of georeferenced data, Proceedings of the Fifth International Natural Language Generation Conference, June 12-14, 2008, Salt Fork, Ohio.

[16] Yu J, Reiter E, Hunter J, Sripada S. SUMTIME-TURBINE: a knowledge-based system to communicate time series data in the gas turbine domain. IEA/AIE-2003. Springer; 2003. p. 379—84.