

Urban and Social Sensing for Sustainable Mobility in Smart Cities

Giuseppe Anastasi, Michela Antonelli, Alessio Bechini, Simone Brienza, Eleonora D'Andrea,
Domenico De Guglielmo, Pietro Ducange, Beatrice Lazzerini, Francesco Marcelloni, Armando Segatori
Dipartimento di Ingegneria dell'Informazione
Università di Pisa
Largo Lucio Lazzarino 1, 56126 Pisa, Italy
Email: {g.anastasi, m1.antonelli, a.bechini, p.ducange, b.lazzerini, f.marcelloni}@iet.unipi.it

Abstract—In this contribution, we describe the SMARTY project. The project is funded by the Tuscany Region and aims to develop innovative services for sustainable transport and mobility in smart cities. These services are based on data collected by environmental and social sensors: such data are pre-processed and analysed by data mining techniques for determining useful information such as state of the traffic flow and parking lots, special events, demonstrations and accidents. All this information is used by SMARTY to suggest optimal routes to the users, taking also the multi-modality into account.

Index Terms—Data Mining, Smart Cities, Social Sensing, Urban Sensing

I. INTRODUCTION

One of the main concerns addressed by the European program HORIZON 2020 [1] is the development of innovative solutions for sustainable transport and mobility, with particular emphasis on cities. In fact, nowadays, cities are growing more and more: it is estimated that more than 60% of the European population lives in cities. This enormous concentration of people in small areas has highlighted some important issues, such as inadequate transport systems, inadequate city services, and increasing pollution. Cars are the most popular transport systems and their large use causes increased emissions, traffic jams, waste of time for drivers and economic losses. The use of the Information and Communication Technologies (ICTs) can be of the utmost importance for promoting new sustainable transport systems with low environmental impact, and thus improving the lives of citizens [2].

In this paper, we describe the work we are carrying out in the framework of SMARTY (SMART Transport for sustainable city), a research project for sustainable mobility promoted and funded by Tuscany Region (Italy) [3]. In particular, we focus on the urban and social sensing features, which are included in the SMARTY software platform.

The SMARTY platform can be actually considered as an Intelligent Transportation System (ITS) [4]. An ITS is an infrastructure which integrates ICTs with transport networks, vehicles and users. By sharing information, an ITS allows improving the management and the safety of transport networks. Further, by using an ITS, it is possible to reduce the impact on the environment, e.g., by providing real-time information about traffic congestion and regulation, and by

planning efficient (e.g., shortest, fast driving, less polluting) routes.

The purpose of SMARTY is to develop an ICT platform which provides tools and services for innovative mobility services easily accessible to the users. This objective is achieved by promoting the use of flexible transport systems (e.g., bike, car sharing, bus/car-pooling) integrated with traditional transport systems and by providing a number of related services (e.g., electronic payment, parking information availability, etc.). The aim is to encourage sustainable mobility solutions by reducing the use of infrastructural and environmental resources.

SMARTY will collect several kinds of information, such as traffic flow, location of users, weather reports, pollution levels, delays of transport services, parking availability. All this information will be processed to keep the users posted and to provide them with optimal routes in real-time taking into account the traffic situation, the availability and the possible delays of means of transport (both scheduled and flexible), the availability of a parking lot, the level of pollution, possible closed roads due to accidents or works at the pavement, traffic jam, etc.. Further, SMARTY will offer to the users services available through mobile devices to pay parking/transport tickets, to book parking lots, etc.

In collecting information for managing a sustainable mobility, urban sensing strategies play a fundamental role by using low-cost sensors deployed across the city. Further, SMARTY allows single users to share the information of personal sensors (personal weather station, gas sensors and so on) by connecting them to the platform by means of domestic WIFI- or UMTS-based Internet connections. Thus, the sensing capability can be improved without increasing the costs of the infrastructure.

In addition to environmental sensors, another innovative aspect of SMARTY will be the possibility of extracting knowledge from “social sensors”. These sensors are related to social interactions [5] such as those in social networks. In fact, people intensely use social media (e.g., Twitter, Facebook) and usually write posts referring to events happening around them. Hence, SMARTY can be used as a social sensing platform, in order to detect, through several kinds of sensors (including smartphones, with posts as sensor information) indications regarding the life of people in a city (e.g., real-time traffic

monitoring, prediction of events, outline of social behavior, feedbacks about the satisfaction of a service). Further, by creating virtual communities, the use of social networks allows improving the usability of some services, especially those based on mutual trust between users, such as bus/car-pooling.

Given the huge amount of data gathered by the sensors of the SMARTY platform, specific techniques are required to process these data and extract useful knowledge. To this aim, data mining techniques [6] such as classification, regression, clustering and frequent pattern analysis, suitable for “big data” will be integrated into the SMARTY platform.

The paper is organized as follows. In Section II, we give a brief overview of the SMARTY platform. Sections III, IV and V discuss, respectively, the urban sensing, the social sensing, and the data pre-processing and mining modules. In Section VI, we describe the current state of the implementation of the described modules. Finally, Section VII provides concluding remarks.

II. SMARTY OVERVIEW

In this section, we briefly give an overview of the platform that will be developed in the framework of the SMARTY Project. SMARTY offers a number of integrated services to users, such as route planning, real-time traffic monitoring, parking reservation, car and bike sharing, car pooling. As shown in Fig. 1, SMARTY needs to interact with several sources of information.

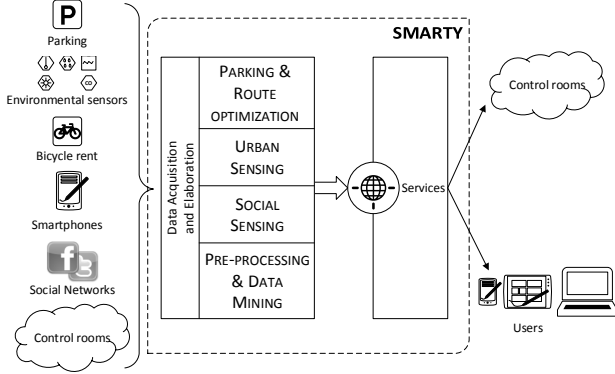


Fig. 1. Overview of the SMARTY Platform.

Parking occupancy and traffic congestion levels are continuously reported to the SMARTY platform by city traffic control rooms. Physical sensors deployed in the urban area (CO₂ sensors, temperature sensors, etc.) provide SMARTY with up-to date information about environmental conditions. Further, SMARTY interacts with smartphones and hand-held devices to obtain information regarding user positions, bike sharing service utilization, service feedbacks, transport means and parking reservations, and service payments. In addition, user posts and messages in social networks (Facebook, Twitter, etc.) are continuously monitored to extract useful knowledge, such as the user’s satisfaction for a certain service, or to predict events in the city.

As shown in Fig. 1, SMARTY is composed of different modules, aimed at supporting the services offered by the platform. In this contribution, we focus on Social Sensing, Urban Sensing, and Pre-processing and Data Mining modules.

The *Urban Sensing* module performs the acquisition of data from the environmental sensors deployed in the urban area. User social interactions through social networks (e.g., Facebook, Twitter, MySpace) are monitored and analyzed by the *Social Sensing* module. Finally, the *Pre-processing and Data Mining* module is responsible for the operations of *cleaning, integration, selection* and *transformation* of the heterogeneous acquired data and for performing the extraction of *knowledge* from the big amount of pre-processed data. The SMARTY platform has been designed as an Event Driven Infrastructure, built on a Service Oriented Architecture and based on the FUSE ESB integration platform [7].

III. THE URBAN SENSING MODULE

Currently, the monitoring of environmental parameters in urban areas, for instance the detection of the level of pollutants in a specific area, is done through appropriate and quite expensive stations positioned in predetermined points of the city, e.g., at traffic lights. This method allows acquiring precise measurements, but punctual and limited to the physical parameter of interest.

With the aim of covering the entire urban area, in SMARTY we plan to adopt a very large number of less precise but also less expensive sensors. Thus, we can obtain a real-time detailed map of the environmental parameters of interest, such as temperature, humidity, level of CO₂, along the overall urban area. In order to reduce the installation costs, the sensor nodes will be powered by batteries. This solution also makes the deployment of the nodes more flexible, because the nodes can be easily moved from one point to another of the city.

As shown in Fig. 2, data will be collected through an opportunistic approach [8], [9]. In this case, we use a short-range, energy-efficient wireless technology (i.e., IEEE 802.15.4/Zig-Bee) and the data gathering is performed by mobile collector nodes mounted on buses.

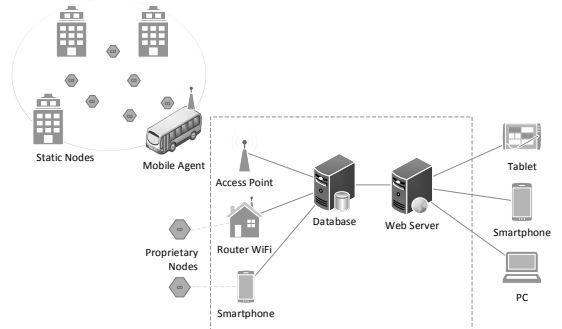


Fig. 2. Urban Sensing.

In addition, SMARTY provides a *participatory sensing* mechanism. As it can be observed from Fig. 2, data can be obtained also by *proprietary* nodes, i.e., low-cost sensors that

users decide to install and manage on their own. Specifically, data can be sent to a collection server directly by sensor nodes through the home connection (e.g., Wi-Fi). As an alternative, data can be collected using a smartphone and transmitted later to SMARTY. Data collected by using the urban sensing module can be exploited for implementing services such as selection of non-polluted paths.

IV. THE SOCIAL SENSING MODULE

The social sensing module introduces an additional novelty in the SMARTY framework, namely the use of online social networks as source of information to detect events such as accidents, demonstrations, or traffic congestion.

In the last years, online social networking sites have gained significant popularity, becoming the most visited web sites on the Internet. The social networks allow people to create an identity and let them share it in order to build a community. Users who join a social network publish their profile, share their comments and their links with other users. The resulting social network is then a basis for maintaining social relationships, for finding users with similar interests, and for locating content and knowledge [10].

SMARTY uses this shared knowledge to analyze the social life of the city in order to highlight social phenomena that can generate traffic or parking problems, such as accidents, demonstrations, concerts and other particular events.

One of the most common questions people ask each other is “where are you?” followed by “what are you doing?”. By means of the Status Update Messages (SUMs), provided by almost all the social network websites, users can follow other users real-time and can know where other people are and what they are doing.

In the framework of the SMARTY project, the aim of the social sensing module is to monitor and to analyze these large amount of SUMs in order to detect specific events. These SUMs will come from the profiles of registered users who will agree to share their information with the Smarty platform. Further, other SUMs may be downloaded from public profiles or pages which are related to a specific subject. Smarty will also provide social pages, built with Facebook and/or Twitter, for its services and events. By using these pages, users will be able to obtain information and to write comments and feedbacks.

In order to analyze the information coming from the social networks, the social sensing module exploits text mining techniques [11]. These techniques are applied to unstructured textual information to extract useful information. In Fig. 3 we show the process of text mining applied to an SUM. This process has been already implemented as a service that the social sensing module is able to provide.

The first step of the text mining process is the fetch of the SUM. Then, before performing any operation, the SUM is tokenized. This is a fundamental step that converts each stream of characters into a stream of processing units called tokens. Each token is then converted to a standard form by the process of stemming, that reduces different grammatical word forms



Fig. 3. Text Mining stages applied to an SUM.

to its root form called stems. After executing the stemming process, the token undergoes the next step of Part of Speech (POS) tagging. In any natural language, words are organized into grammatical classes or parts of speech, thus the POS tagging deals with the classification of the tokens into predefined classes. Finally, the resulting tokens are grouped into units called phrases, that will undergo to the step of sequence analysis. This last step extracts meaningful features that will be used for analyzing the SUM with data mining techniques, such as the classification among predefined categories or classes.

As an example, we consider two classes, one class for the SUMs related to an accident, and another class for all other SUMs. By monitoring the social network updates, we may classify each SUM into the two aforementioned classes. If we find a significant number of SUMs originated nearby a specific location and belonging to the class related to accidents, we may infer the event *accident* in that specific location and predicting, depending on the specific time of the day, possible high traffic. Then, SMARTY may suggest alternative paths to the users. Similarly, we can detect events related to any other social phenomenon and provide real time warnings to the users.

V. PRE-PROCESSING AND DATA MINING MODULE

The SMARTY platform will acquire a large amount of data: traces of users obtained through the GPS receiver integrated into their smart phones, states of parking spaces in several parking lots in the city, measures from the urban and social sensors, data from municipal, provincial and regional databases, etc. Before extracting useful knowledge from these data, we need to pre-process and filter them. For example, data collected by sensors may be incomplete or noisy. In fact, we may have missing measurements, due to problems in the sensor or in the transmission. Further, some outliers may be collected.

Several approaches have been proposed in the literature for filtering and preparing data [6]. The data preprocessing module of the SMARTY platform will integrate a number of algorithms: i) to filter noise and eliminate inconsistent data (data cleaning), ii) to combine and make data from heterogeneous sources consistent (data integration), iii) to find and select relevant data (data selection) and iv) to prepare the data for the algorithms of data mining (data transformation).

In the last years, many algorithms have been proposed to extract knowledge from the data [6]: these algorithms

are identified with the term “data mining” algorithms and include methodologies for classification, regression, clustering, frequent patterns discovery, outliers detection. By using data mining techniques it is possible to achieve a sort of “Knowledge Acceleration”, which is one of the strategic objectives of the Tuscany Region. The knowledge provided by the data mining module will be useful not only for users and the administrators of the platform, but also for the local governments. For example, the analysis of the traces left by the GPS will allow us to study how people move in the city and, thus, indirectly analyze how traffic flows. Furthermore, it may be possible to carry out a real time analysis of anomalous traffic situations by comparing the normal flow of traffic, stored in a database, with the real-time collected traffic data. Finally, we may be able to identify areas that are particularly crowded where billboards may be placed. As stated above, the detection of special events, such as traffic jams and demonstrations, may be also carried out by applying data mining algorithms to the social data, i.e. to user posts, tweets and comments extracted from the social networks.

As regards the parking management, the collected information can be used for real-time monitoring of parking lots. In particular, it may be possible to indicate to users the availability of parking lots or to identify possible frauds. Indeed, the availability of accurate information on the use of the individual parking lots allows SMARTY to identify in real time whether a parking lot is occupied without having paid the correct fee.

The data mining module will integrate state-of-the-art algorithms, paying a special attention to multi-objective evolutionary fuzzy systems [12]. These systems, while ensuring good accuracy in classification and regression problems, have the great advantage of being interpretable, i.e. of being able to clearly explain, by using linguistic rules, how the inputs are elaborated to produce an output. Finally, in order to manage big data, distributed data mining algorithms, based on the cloud computing paradigm, will be integrated into the Smarty platform [13].

VI. THE CURRENT STATE OF THE PROJECT

The research and development activities of the SMARTY project started in October 2012. As regards the urban sensing module, we have already implemented the mobile agent discovery mechanism, the communication protocol between static nodes and the access point. Currently, we are carrying out the first on-field experimentations.

As regards the social sensing module, we have already implemented some services such as i) the automatic downloading of tweets, i.e., SUMs from Twitter, ii) the preliminary text cleaning, which removes useless text such as hash tags and user information, and iii) the complete scheme shown in Fig. 3. Finally, a number of state-of-the-art data filtering and data mining algorithms have been implemented and wrapped into services available to the other modules of the platform.

Preliminary promising results have been obtained in the classification of tweets between SUMs related to traffic con-

gestion events and SUMs with no connection with these events. We built a dataset consisting of 500 tweets for each class. By using the C4.5 classifier and performing a ten-fold cross-validation we obtained a classification rate of 93.73%.

VII. CONCLUSIONS

In this paper, we have briefly described the SMARTY project, which aims to develop a platform which will feature innovative mobility tools and services for a smart city.

We have mainly focused our description on the social and urban sensing modules. The urban sensing module deals with environmental data gathered by employing common long-range wireless sensor networks and mobile data collectors.

The social sensing module aims to exploit data extracted from social networks, such as posts, tweets and comments, so as to infer information on some specific events.

Finally, we have also described the pre-processing and data mining module. This module pre-elaborates and filters the heterogeneous gathered data and then extracts useful knowledge from them.

ACKNOWLEDGEMENTS

The SMARTY project is funded by “Programma Operativo Regionale (POR) 2007-2013” - objective “Competitività regionale e occupazione” of the Tuscany Region. The authors would like to thank all the partners of the project: Almaviva spa, Bassilichi spa, KKT srl, Softec spa and Dipartimento di Sistemi e Informatica of the University of Florence.

REFERENCES

- [1] “Horizon 2020 - the framework programme for research and innovation,” 2011. [Online]. Available: <http://ec.europa.eu/research/horizon2020>
- [2] A. Caragliu, C. Del Bo, and P. Nijkamp, *Smart cities in Europe*. Vrije Universiteit, Faculty of Economics and Business Administration, 2009.
- [3] “The SMARTY project,” 2013. [Online]. Available: <http://www.smarty.toscana.it/>
- [4] J. Zhang, F.-Y. Wang, K. Wang, W.-H. Lin, X. Xu, and C. Chen, “Data-driven intelligent transportation systems: A survey,” *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 4, pp. 1624–1639, 2011.
- [5] T. Sakaki, Y. Matsuo, T. Yanagihara, N. Chandrasiri, and K. Nawa, “Real-time event extraction for driving information from social sensors,” in *Proceedings of IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems*, 2012, pp. 221–226.
- [6] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, 2011.
- [7] “Fuse open source community,” 2013. [Online]. Available: <http://fusesource.com/>
- [8] M. Di Francesco, S. K. Das, and G. Anastasi, “Data collection in wireless sensor networks with mobile elements: A survey,” *ACM Trans. Sen. Netw.*, vol. 8, no. 1, pp. 7:1–7:31, 2011.
- [9] K. Shah, M. D. Francesco, G. Anastasi, and M. Kumar, “A framework for resource-aware data accumulation in sparse wireless sensor networks,” *Computer Communications*, vol. 34, no. 17, 2011.
- [10] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, “Measurement and analysis of online social networks,” in *Proceedings of the 7th ACM SIGCOMM Conference on Internet measurement*, 2007, pp. 29–42.
- [11] S. Weiss, N. Indurkha, T. Zhang, and F. Damerau, *Text Mining: Predictive Methods for Analyzing Unstructured Information*. Springer Verlag, 2004.
- [12] P. Ducange and F. Marcelloni, “Multi-objective evolutionary fuzzy systems,” ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2011, vol. 6857, pp. 83–90.
- [13] S. Owen, R. Anil, T. Dunning, and E. Friedman, *Mahout in Action*. Manning Publications, 2011.