

Analysis of GSM calls data for understanding user mobility behavior

Barbara Furletti*, Lorenzo Gabrielli*, Chiara Renso*, Salvatore Rinzivillo*

*KDDLlab, ISTI-CNR, Pisa, Italy

Email: {barbara.furletti,lorenzo.gabrielli,chiara.renso,salvatore.rinzivillo}@isti.cnr.it

Abstract—This information about our GSM calls is stored by the TelCo operator in large volumes and with strict privacy constraints making it challenging the analysis of these fingerprints for inferring mobility behavior. This paper proposes a strategy for mobility behavior identification based on aggregated calling profiles of mobile phone users. This compact representation of the user call profiles is the input of the mining algorithm for automatically classifying various kinds of mobility behavior. A further advantage of having defined the call profiles is that the analysis phase is based on summarized privacy-preserving representation of the original data. We show how these call profiles permit to design a two step process - implemented into a system - based on a bootstrap phase and a running phase for classifying users into behavior categories. We evaluated the system in two case studies where individuals are classified into *residents*, *commuters* and *visitors*. We conclude the paper with a discussion which emphasizes the role of the call profiles for the design of a new collaboration model between data provider and data analyst.

I. INTRODUCTION

The cellular phones based on the GSM protocol are probably the most popular device we carry almost everywhere during our life. Since GSM is based on the communication to an antenna covering a local area, the active connection - e.g. a call - to a certain antenna represents a spatio-temporal position information for the user. This information, as collected by the telecom provider, thus provides a spatio-temporal fingerprint of the users moving in a GSM covered area. For example, a user calling only in the weekdays and during the working hours in a given area can be considered as a *commuter* since we observe his presence in the monitored area only during working temporal slots. The observation above raises three considerations: the first one is that GSM data embeds a great deal of information about the user habits and behavior; the second one is that GSM data, when collected from the provider side, comes in large volume thus posing new challenges, not only from the collection and storage point of view, but also from the analysis and mining side; thirdly, the analysis of this large volume of personal data poses several privacy issues.

This paper is a proposal for facing these three issues with an approach for mining large volumes of GSM call data with the purpose of identifying the user behavior from his/her call habits. An important aspect of our method is that we propose a strategy for behavior identification, based on the calling profiles of phone users. These call profiles represent an aggregate pre-processed data on which to run mining algorithm for discovering various kinds of user behavior.

The advantage of having defined the call profiles is that the analysis phase is no more based on the original GSM raw (big and privacy sensitive) data, but on an aggregated privacy-preserving summary of the original data. We show how these call profiles permit to design a two step process based on a bootstrap phase and a running phase for classifying users into behavior categories based on their call habits.

More in detail, we present a user behavior inference system where an improved inductive learning phase and automatic classification replaces the two-phase learning steps adopted in [1]. Furthermore, the structure of the analysis process into a bootstrap and a running phase leads to another interesting result: the call profiles represent a quantifiable model to measure the amount of information transferred from the TelCo operator to the data analyst. The call profiles therefore can be seen as a common interchange model allowing both parties to implement a collaboration protocol based on the production and analysis of well-defined pieces of information. This brings two important advantages for the two parties: the volume of data transferred using pre-aggregated data like the call profiles is considerably lower, and the privacy is better preserved since no original raw data is moved to the data analyst. We present a large experimentation conducted in two Italian cities from calls data provided by an Italian mobile phone operator. In these experiments we classified users call habits into three mobility behavior categories: *Residents*, *Commuters*, and *Visitors*. This may enable several innovative applications like *mobility observatories* where monitoring systems periodically receive call profiles from the TelCo operator and then infer the mobility behavior users.

The paper is organized as follows: after discussing some related works in Section II, Section III describes the analysis process to discover user behaviors from GSM data. In Section IV the method is experimented in two case studies, while Section V discusses some important observations about the advantages of this method both for a new business model for the TelCo operator and for solving the privacy issues. Finally, section VI draws the conclusions and future works.

II. RELATED WORKS

A previous work of the authors of the present paper is [1] where an approach for profiling users based on GSM calls is presented. The novel contribution of the present paper compared to the previous one is manifold. First of all, the previous paper approach is only experimented in the city of

Pisa, while the present paper adds another city. Secondly, the software we have developed here has been improved with an automatic classification of call profiles into user categories. Thirdly, we envisage a new cooperation protocol between the TelCo operator and the data analyst where the data to be transferred for analysis are the pre-aggregated call profiles instead of the raw GSM data.

Among the literature on the analysis of GSM data we can recall a famous experiments on analysing GSM data for studying people movement have been run on Rome [2] and Graz [3]. They use GSM data to realize a real-time urban monitoring systems. The final objective is to realize a wide range of services for the city such as traffic monitoring and tourists movement analysis. A different approach comes from Schlaich et al. [4] where the authors exploit the GSM handover data - the aggregated number of users flowing between cells - to perform the reconstruction of vehicles trajectories to study the route-choice-behavior of car drivers.

Another use of GSM data is the identification of interesting users places as in [5], where the authors propose a method for the identification of meaningful places relative to mobile telephone users, such as home and work points. They use GSM data (both calls and handovers) collected by the phone operator. They distinguish between personal anchor points like home, work and other person-related places as the locations each user visits regularly, as for example a gym.

In Pereira et al. [6], the authors exploit cellular phone signaling data¹, focusing on the prediction of travel demand for special events. Since mobility data are dependent on mobile phone usage, they proposed to integrate the GSM dataset with external data to increase the quantity and the quality of the data. Quercia et al. [7] uses GSM data for recommending social events to city dwellers. They combine the locations estimated by mobile phone data of users in the Greater Boston area and the list of social events in the same area. Trajectories are extracted from GSM calls and combined with events from the web. By crossing the events and the stops, they identify a set of potential users participating to events.

Mobile phone records are analysed also in [8] where the authors propose a visual analytics framework to explore spatio-temporal data by means of SOM (Self-Organizing Map) analysis. They propose a method to cluster the dataset by either of the two dimension and evaluate the resulting aggregation on the other one. Although they show the potentialities of using SOM for analysing mobile phone records, they do not focus on identifying user profiles.

Differently from the above approaches, the aspect we want to study in this paper is to characterize the user behavior based on the call habits of the tracked users.

III. THE SYSTEM

The availability of huge quantity of mobile phone data stimulates more and more challenging questions that go

¹These data consist of location estimations which are generated each time when a mobile device is connected to the cellular network for calls, messages and Internet connections.

beyond the observation of people movements through the territories. What is interesting is to infer some “mobility behaviors” starting from the footprints left by the users during their mobile phone calls. Knowing where they make a call and when, how many times they call from a certain location are aspects implicitly stored in the GSM data and can be a valid indicator to draw user profiles.

In this section we introduce an inductive analysis method based on Self Organizing Maps [9], to infer user behaviors. These behaviors can identify people categories characterizing the mobility of the monitored individuals. In this paper we focus on three specific examples of these mobility behaviors distinguishing people who are resident in the monitored area from the commuters and the visitors. According to the instructions given by the domain experts, we can formally define these categories as follows:

- A person is *Resident* in an area A when his/her home is inside the A . Therefore the mobility tends to be from and towards his/her home.
- A person is a *Commuter* between an area B and an area A if his/her home is in B while the work/school place is in A . Therefore the daily mobility of this person is mainly between B and A .
- A person is a *Visitor* in an area A if his/her home and work/school places are outside A , and the presence inside the area is limited to a certain period of time that can allow him/her to perform some activities in A .

The GSM data used in our analysis method are of two kinds: spatio-temporal position of the calls - the Call Data Records (CDRs) [10] - and the corresponding spatial coverage of the GSM cells.

The user categories can be obtained from the data by the analysis process shown in Figure 1.

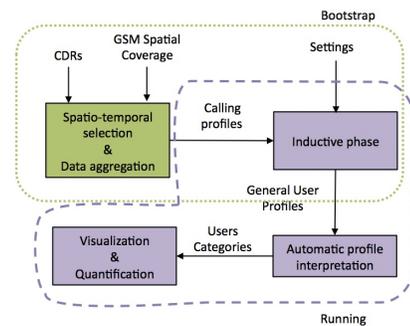


Fig. 1. The analysis process to infer the user categories from GSM data

To an abstract level, the process foresees two phases: a **bootstrap** (shown at the top of the figure) and a **running phase** (shown at the bottom). The bootstrap phase is aimed at finding a temporal profile of the calls for the users, and is based on the analysis of the raw GSM data. The running phase starts with the temporal calls profiles and performs inferences to extract the user categories. The distinction between bootstrap

and running phase is particularly important since we can decouple the analysis of GSM raw data - a large dataset of limited use due to privacy and storage issues - with the calls profiles needed to infer the user categories. In other words, the bootstrap phase can be done only once to establish the calls profiles needed to infer the user categories. Once the format of these aggregated data has been chosen and tested in the bootstrap phase, the running phase iterates the process, but starting from these temporal calling profiles instead of the GSM raw data. The importance of having established the calls temporal profiles is that in the running phase the data provider limit the data supply to these aggregated profiles, thus avoiding the transfer to the analyst of raw data. The running phase can therefore be started with new call profiles every time we need to refresh the dataset for a continuing monitoring of the area.

During the bootstrap phase we have a first computation step called *Spatio-temporal selection and aggregation*. Here, the CDRs are first analyzed and some statistics are calculated to learn about the data and get some aggregate information. These statistics are useful for the choice of parameters and thresholds in the following steps.

Each CDR represents an activity of a user with references to space (i.e. the cells where the call was performed) and time (i.e. when the user called). From this information, it is possible to reconstruct the call habits of each user. The spatial aspect is used when we select the area to be analysed, e.g. where the call has been issued. Depending on the application, the area selection may help in defining the different user profiles.

Using only the temporal information, we can derive a *Temporal Calling Profile* (TCP). Formally, given a user u and a time granularity (i.e. minute, hour, day, week, etc), a *Temporal Calling Profile* is a sequence $TCP_u = [a_1, a_2, \dots, a_n]$, where a_i is the frequency of calls done by u in time unit i . From the TCP of a user it is possible to infer his/her typical patterns of calls. According to the temporal granularity chosen, it is also possible to project a TCP on a time period cycle. For example, given a TCP based on day granularity, it is possible to project it on a weekly interval. Figure 2, shows an example of the extraction of a Weekly TP. The number of calls of each day of the month (in the example February 2012) of two users are shown in the top part: *User 1* (a) and *User 2* (b). The Weekly TCPs are obtained by projecting their respective TCPs on a single week, as shown in parts (c) and (d). The users calls over the same day through the 4 weeks are summed up in the representative weekly TCPs.

A TCP does not take into account the space dimension, i.e. where the user performed the calls. The spatial information of the CDRs is represented by the cell identifier used by the user to transmit. When referring to relevant places on the territory it is useful grouping different adjacent cells in one group. For example, a stadium is a large area that usually is covered by several cells: the group of this cells may be referenced as *stadium*. In this work, we aggregate adjacent cells by covering the urban area of a city. In general, given a partition of the cellular network in group of cells each representing an area, for each group A (i.e. for each area covered by

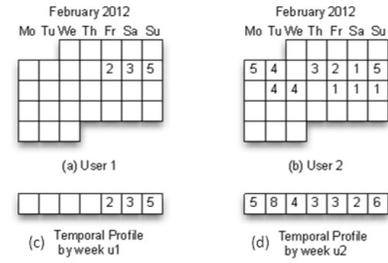


Fig. 2. Example of temporal profiles computed starting from the GSM calls.

G), we extract a *Space-Constrained Temporal Calling Profile* (SCTCP). Formally, given an area A as a group of cells, a user u and a time granularity, a *Space-Constrained Temporal Calling Profile* is a TCP extracted only from the call activities performed by u in the cells in A .

Both TCP and SCTCP use a single time granularity. Indeed, it can be useful to distinguish different time intervals within the same time unit. For example, given a TCP by day, it is interesting for an analyst to split the daily activities by time of the day, i.e. during the working hours or during the night. In this case, we can extend the two profiles above by considering a multilevel time granularity. Given a user u , and two time granularities TG_1 and TG_2 (where TG_2 is contained in TG_1), a *Multilevel Temporal Calling Profile* (MLTCP) is a matrix $TCP = \{P_{ij}\}$, where the entry P_{ij} represents the number of calls performed by u in the time unit i during the time interval j . As before, it is possible to define a space constrained MLTCP by considering only the activities in a specific area A . When TG_1 is *day* and TG_2 is *hour*, a MLTCP of a month contains 31 columns and 24 rows: the entry $P_{4,13}$ gives the number of calls of the user on 4th of the month between 13 and 14.

For the objective of our analysis, the information of a MLTCP can still be aggregated by losing some redundant details. In particular, for the objective of this work, we are interested only on the presence of a user in a given area during a time interval, not on his/her actual volume of calls. Thus, we can consider only a predicate to state if user u was present in the observation area during a specific time interval. From the previous example, the entry $P_{4,13}$ will yield *true* if the user performed at least one call in that time interval, *false* otherwise.

We can still aggregate the MLTCP profile, based on days and hour of day, by aggregating both temporal granularities with conceptual intervals. We call this specific aggregation *Multi-Dimensional Call Profile* (MDCP). A MDCP is a profile that captures the user's presence during the whole period of observation, splitted into specific temporal slots. In particular, two operations are performed on the two time granularities: 1) Aggregation of the days in weekday and weekend, and 2) definition of time intervals within the day, representing the interesting time windows during the day. For instance, we have used the following intervals:

- $t1 = [00:00:00 - 07:59:59]$, Early in the morning when

- people are usually at home;
- $t2 = [08:00:00 - 18:59:59]$, Mid day when people are out for work/school or other activities;
- $t3 = [19:00:00 - 23:59:59]$, Late in the evening and night when people are typically back home.

The choice of the daily time intervals is not trivial and it strongly depends on the application and on the categories of users we want to identify. For instance, to distinguish between the two categories of Residents and Commuters, we need to concentrate the analysis on a discriminating factor. An example is to compare the presence of people in the area of observation when they are at home and outside home (work, education, leisure) joint with the typical time windows used for home and work activities. Therefore, when a user shows a temporal profile where (1) during a *home* time windows he/she is always outside the monitoring area, and (2) during the *work* time slot and in the weekdays he/she is always inside the monitoring area, then we can conclude that this user is potentially a commuter. On the contrary, a resident is expected to show a call profile where his/her presence is inside the monitoring area at both the time slots (home and work) and during both weekdays and weekends.

From this example we can see that we need to analyse several dimensions at the same time: the daily temporal slots dimension, the day-of-the-week dimension, the spatial dimension. Therefore a Multi-Dimensional call profile represents all these aspects in one structure, as shown in Figure 3. The figure exemplifies the construction process: starting from the calls along the days, the presences over the time intervals $t1, t2, t3$ are computed, then these presences are aggregated over the weekdays and the weekend summing up all of them. The result is a sort of compact representation of the global call profile for a given user.

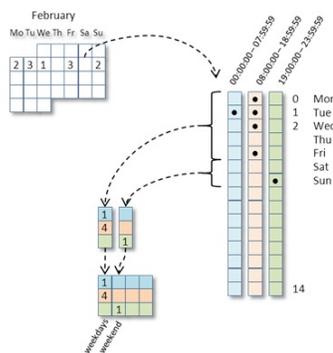


Fig. 3. Example of Multi-Dimensional Call Profile computed from the GSM data and representing the user's presence in the monitored area.

The *Inductive step* takes as input the call profiles (which can be simply Temporal, or Spatio-constrained Temporal, Multi-Level or Multi-Dimensional (depending of the kind of user behavior we want to define) and, according to their relative similarities, group them into homogeneous groups of *similar* call profiles. Here we use an implementation of the Self Organizing Map (SOM) [9]. A SOM is a neural network based

on unsupervised learning that produces a one/two-dimensional representation of the input space using a neighborhood function to preserve the topological properties of the input space. The SOM forms a sort of semantic map where similar samples are mapped close together and dissimilar apart are mapped far away. This produces a set of nodes, where each node represents a group of users with similar profile. We call each node the *General User Profile*.

The automatic classification of the General User Profiles in categories like Residents, Commuters and Visitors is done by using a K-Nearest neighbors algorithm where the training examples are a set of prototypes, defined by the domain experts. In the classification phase, each General User Profile is assigned to the label corresponding to the most similar prototype. This phase returns a first attempt of automatic classification that can be later modified by the expert by using the system interface. The quantification step concludes the analysis evaluating the percentage of users for each category and computing a general statistics. The visualization of the results allows the analyst to discover new profiles that were not considered in the extraction of the profiles. For example, we were able to distinguish in Pisa the nightly visitors from the daily visitors.

This inference process has been implemented in a running system written in Java and including a state of the art SOM algorithm [9]. The data are stored in a PostgreSQL relational database with geo-spatial extension (PostGis).

IV. CASE STUDY

The method has been tested with several datasets coming from different geographical areas [1], [11]. In this paper we focus on the analysis on two Italian cities: Pisa (in the Center of Italy) and Cosenza (in the South). The final objective is the deployment of a permanent monitoring system for people moving in the cities. The two GSM datasets are provided by one of the Italian mobile phone operators and consist of one month of CDRs of users having an Italian phone contract (no roaming users are included in the datasets)².

We are interested in understanding the behavior of people “using” the city and their locations. Therefore, for the experiments we selected only the calls performed in the urban areas, thus discarding the calls performed outside the selected area. Furthermore, we discard the *People in Transit*, thus individuals who only transit through the urban area without performing relevant activities. We identify *People in Transit* as the users who make their calls only inside a (short) temporal interval of k hours in whole observation period. The value of k is estimated based on the minimum time needed to cross the urban area without performing any significant activity. For both Pisa and Cosenza, k is set to 1 hour according to the indication given by the experts in the domain. For both experiments we use the same temporal constraints and the

²Data have been provided by the Mobile Phone Operator as part of a long-term cooperation with severe data and privacy constraints. As a consequence the data cannot be publicly released and we can only disclose the research results.

same settings for the SOM algorithm. In particular: (i) for each user we build a temporal profile that captures the weekly presence as described in Section III; (ii) for the SOM, we set the dimension of the output map to 10x10. These values have been empirically identified as the good trade-off to guarantee a sufficient number of General Profiles to analyze.

a) *Profiling people in Cosenza*: The whole dataset contains around 25 millions of CDRs collected from October 14th to November 15th, 2012 in the Italian region Calabria. The calls are of about 352.600 users. After applying the spatial filter, the users calling only in the urban area are about 104.600. For each of these users we computed the Multidimensional profiles and then we applied the SOM algorithm to group these profiles into General profiles. The quantification step returns the statistics depicted in Figure 4 (left), that shows the percentages of people for each category.

b) *Profiling people in Pisa*: The dataset contains around 7.8 million of CDRs collected from January 9th to February 8th 2012 in the Province of Pisa. The calls are relative to about 232.200 users. After the spatial filter, the users making calls in the urban area of Pisa are about 103.000. Similarly to the Cosenza case, we first computed the Multi-Dimensional profiles (MDTP), then apply the inductive step (SOM algorithm) to obtain the corresponding similarity map. It is worth noticing that among the General Profiles some interesting new cases can be identified. For example, we can identify new general profiles belonging to two new categories: *Night Visitors* and *Occasional Visitors*. The Night Visitors are people calling only during the night and can be interpreted as persons who visit Pisa only for the “nightlife”. We know that Pisa offers several entertainment and it attracts young people for its nightclubs and pubs, while more mature persons may be attracted by shows and cultural events. Occasional Visitors are, instead, people who visit Pisa only few occasional times during the month. The two statistics in Figure 4 are a good indicator to describe a sort of “dynamics” of the two cities. The categories of people that are present in a city can be in fact considered an indicators on how the presences may change over time. Indeed, according to the definitions given above, the Commuters are not present during the weekends and during the nights while People in Transit cross the city but they don’t stop for performing some activity.

Looking at the statistics, we can see that Pisa counts many more People in Transit compared to Cosenza. This is because the urban area of Pisa is crossed by 2 high traffic roads (an important Municipal road and a motorway), a railway with an important rail station and an Airport. Particularly the Airport generates high flows of people arriving and departing from the city and this data has been confirmed by their managers since the Airport generates a monthly average of 100.000 arrivals during the winter. This trend is also confirmed by the Pisa Tourist Office observing the fact that many incoming visitors actually use the city as a “hub” to reach Florence or other nearby touristic attractions. By studying the density of the People in Transit we have another confirmation of this behavior. The percentage of Residents and Commuters are

similar for both the towns, while the percentage of Unknown (individuals unclassified due to the few phone calls) is higher in Cosenza than in Pisa.

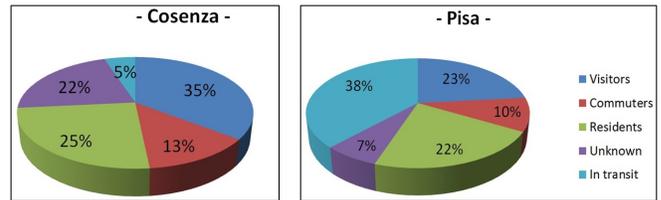


Fig. 4. Results of the quantification for Cosenza and Pisa.

These experiments show some examples of user behavior inference using the system we have proposed. These are real case studies where we use the inference process to establish a kind of monitoring of the city: the call profiles are fed by the mobile operator and the “running phase” of the system computes the user categories and their statistics. When this process is iterated periodically (on a weekly or monthly basis) it realizes the core component of a more complex infrastructure like a *urban mobility observatory*. Naturally, this can be enabled only if we can ensure (1) that the data to be transferred is reduced to an aggregated summary and (2) the data to be transferred for the analysis preserves the privacy of the monitored individuals.

V. BUSINESS MODEL & PRIVACY

The proposed system provides a quantification of social behaviors on a territory that are extremely valuable for local governments. At city level, it became crucial to evaluate the incidence of distinct groups of user that exploit the resources of the city, like water, energy, wastes, etc. A real breakthrough of this profiling activity may come from the foundation of a continuous observatory capable of sensing these population changes along the year or, more in general, in the long term history. One of the main obstacles for the implementation of a continuous monitoring is, usually, the availability of GSM data. In fact, these data are usually protected by laws on individual privacy preservation and it may become difficult for the TelCo operator to easily distribute such information. Moreover, TelCo operators consider these data as an asset of the company and they tend to adopt a conservative approach towards their disclosure.

The analytical phase of our system assumes that the only data available are represented as a set of individual temporal profile. This assumption is important for two main aspects: 1) the temporal profiles represent a quantifiable model to measure the amount of information transferred from the TelCo operator to the data analyst; 2) the agreement on a common interchange model structure allows both parties to implement a protocol of collaboration based on the production and analysis of well-defined pieces of information.

When it comes to find an economic agreement with a data consumer, it is difficult to give objective criteria for estimating

the cost of the data. To clarify this point, let us consider the basic granularity of GSM data, i.e. the CDR format. A single CDR log contains the information about a single event on the cellular network. Per se, this information may be of little utility, and it depends on other activities performed by the same user. Thus, the pricing criteria may be influenced by this observation and the unit cost of each CDR log depends on the behavior of the corresponding user. Our approach try to overcome this problem, by aggregating the behavior of a use in a single model. The system we propose serves as connection between the data provider and the data analyst, providing an objective quantifiable set of models that can be used by both parties also for an economic agreement. For example, the data analyst may correspond to the TelCo operator a monthly fee based on the number of distinct temporal profiles received. Note that such agreement may not be possible if considering the basic CDR logs.

On the basis of this common protocol of data exchange, the TelCo operator is protected against leakage of important corporate information, since the transmitted data is in an aggregated form and it provides the minimum set of information needed to enable the sociometer service. Moreover, the TelCo is free to rearrange internally the organization of the information and it is constrained only on the requirements of the models to be produced. In the bootstrap phase, both parties have to agree on a set of procedures to produce the aggregates, i.e. both define the methods and technologies needed to extract the temporal profiles from CDR data. Once this phase is concluded, the extraction method may be automate in order to produce a set of profiles periodically, for example on a monthly basis. One of the most relevant advantages of this approach is the possibility of measuring and handle the privacy risks involved in the distribution of individual habits. Since the Multi-Dimensional temporal profile is a spatio-temporal generalization of individual CDR logs, this transformation brings in the system a protection against user re-identification. We consider, in particular, a linking attack where a malicious party knows part of the activities of a user, and his objective is to look in the set of the MDT profiles to infer new information about other activities of the same user. We set up a series of experiments to measure the probability of re-identification on the GSM datasets we used for the experiments. We generated a series of basic knowledge configurations for the attacker, i.e. a combination of observations for a user, ranging from few observations to the 99% of the user's activities. From this empirical evaluation we found out that the highest probability of re-identification is 0,21% for the 5% of the users in the dataset, i.e. each combination of observation we have generated selects at least 3,6k candidates among the available MDT profiles. This indeed is a valuable protection to ensure that the individual information is not at risk.

VI. CONCLUSION AND FUTURE WORKS

This paper proposes a process to analyse the big GSM calls data collected by the mobile operators to understand the mobility behavior of people. This is done by aggregating

and summarizing the calls records, then applying data mining algorithms to automatically label the different calling profiles into user behaviors. The method has been implemented into a running system and experimented in two case studies where data has been provided by an Italian mobile operators for two cities. We also discussed a strategic feature of the method: we base our analysis on the aggregation of the raw GSM data into calls profiles, thus giving both a new cooperation model between TelCo operator and data analyst, and, at the same time, providing a good trade off between analysis outcomes and privacy preservation.

Possible improvements include replace the SOM with a more sophisticated clustering algorithm and to design a more efficient strategy for the automatic classification of the General profiles to reduce the percentage of "Unknown". More in general we intend to test the new cooperation protocol with the local municipality and the TelCo operator to better understand which is the economical impact of this new paradigm in the business model of the operator.

ACKNOWLEDGMENT

This work has been partially supported by EU projects DATASIM N. 270833 and LIFT N. 255951.

REFERENCES

- [1] B. Furletti, L. Gabrielli, C. Renso, and S. Rinzivillo, "Identifying users profiles from mobile calls habits," in *In the Proc. of UrbComp'12*, 2012.
- [2] F. Calabrese, M. Colonna, P. Lovisolo, D. Parata, and C. Ratti, "Real-time urban monitoring using cell phones: A case study in rome," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, pp. 141–151, 2011.
- [3] C. Ratti, A. Sevtsuk, S. Huang, and R. Pailer, "Mobile landscapes: Graz in real time," 2005, MIT Senseable City Lab.
- [4] J. Schlaich, T. Otterstätter, and M. Friedrich, "Generating trajectories from mobile phone data," in *Proceedings of the 89th Annual Meeting Compendium of Papers, Transportation Research Board of the National Academies*, 2010.
- [5] R. Ahas, S. Silm, S. Järv, and E. Saluveer, "Using mobile positioning data to model locations meaningful to users of mobile phones," *Journal of Urban Technology*, vol. 17, no. 1, 2010.
- [6] F. C. Pereira, L. Liu, and F. Calabrese, "Profiling transport demand for planned special events: Prediction of public home distributions," 2010, available online at www.scienceDirect.com.
- [7] D. Quercia, N. Lathia, F. Calabrese, G. D. Lorenzo, and J. Crowcroft, "Recommending social events from mobile phone location data," in *International Conference on Data Mining, ICDM*, 2010, pp. 971–976.
- [8] G. Andrienko, N. Andrienko, P. Bak, S. Bremm, D. Keim, T. von Landesberger, C. Poelitz, and T. Schreck, "A framework for using self-organising maps to analyse spatio-temporal patterns, exemplified by analysis of mobile phone usage," *Journal of Location Based Services*, vol. 4, no. 3–4, 2010.
- [9] T. Kohonen, *Self-Organizing Maps*, ser. Springer Series in Information Sciences. Springer, 2001. [Online]. Available: <http://books.google.it/books?id=e4igHzyfO78C>
- [10] Wikipedia, "Call data record," http://en.wikipedia.org/wiki/Call_detail_record.
- [11] M. Nanni, R. Trasarti, B. Furletti, L. Gabrielli, P. V. D. Mede, J. D. Bruijn, E. D. Romph, and G. Bruil, "Mp4-a project: Mobility planning for africa," 2013, mobile Phone Data for Development. Analysis of mobile phone datasets for development of Ivory Coast.
- [12] F. Giannotti, M. Nanni, D. Pedreschi, F. Pinelli, C. Renso, S. Rinzivillo, and R. Trasarti, "Unveiling the complexity of human mobility by querying and mining massive trajectory data," *VLDB J.*, vol. 20, no. 5, pp. 695–719, 2011.