

# A Rule-based Method for Discovering Trajectory Profiles

Lucas André de Alencar,  
Luis Otavio Alvares and Vania Bogorny  
Universidade Federal de Santa Catarina (UFSC)  
Florianópolis, Brasil

Chiara Renso  
ISTI-CNR  
Pisa, Italy

Alessandra Raffaeta  
DAIS - Università Ca'Foscari Venezia  
Venice, Italy

**Abstract**—The discovery of people profiles such as workers, students, families with kids, etc, is of interest for several application domains. For decades, such information has been extracted using census data, and more recently, from social networks, where people’s profile is clearly defined. A new type of data that has not been explored for discovering profiles, but which stores the real movement of people, are trajectories of moving objects. In this paper we propose a rule-based method to represent socio-demographic profiles, a moving object history model to summarize the daily movement of individuals, and define similarity functions for matching the profile model and the history model. We evaluate the method for single and multiple profile discovery.

## I. INTRODUCTION AND MOTIVATION

The knowledge about people living in a city or country has great value for the public administration as well as for enterprises. To know the population profile may help smart city planners, public transportation administrators, government services or companies to decide if and where to install a new store or to personalize an advertisement, for example. Most attempts to discover and measure the population profiles are through human surveys, and the most well known example is the socio-demographic census with diary activities, periodically done in almost all countries. However, the main drawbacks of the census data are that they: 1) are not up to date since they are usually collected every 5 - 10 years; 2) are expensive to collect, and cover only a small - although statistically significant - part of the population for a short period of time; 3) do not collect the actual movement of the individuals, but only the activities performed during the day and which are mentioned by the user during the interview.

We believe that nowadays we can infer much knowledge and the real behavior about people from their every day movement, about where people really go, when they go, and for how long. We are entering the era of big data, where the real movement behavior of a society can be extracted from its individuals everyday movement. In daily life, in general, we all follow a *routine*, going more or less to the same types of places everyday (e.g. work, gym, supermarket, restaurant, etc). The routine of one person during one week, one month or one year represents the *general pattern of movement* of this person. For instance, a typical routine of a *worker* is to go, in general, four or five times a week to work, while a *student* goes to school/university four or five times a week. On the contrary, an *Unemployed* may have a different routine, as not having a workplace. The routines followed by a similar group of people as the students, workers, or unemployed we call *profiles*.

With the increasing number of GPS trajectory datasets and the definition of semantic trajectories in GPS data [1],

it is possible to infer the real places visited by an object, the duration of the visit, and the frequency of the visits. Based on these visits, it is possible to obtain the routine of an object. An example of semantic trajectory is shown in Fig. 1, where the moving object visits four places (home, university, shopping mall and bar).



Fig. 1. Example of semantic trajectory A.

In the literature of moving object trajectories there are several works for extracting “general patterns” and that summarize the movement of objects, but no works have tried to look deeper into the data to infer more knowledge about the moving object. Only a few works address the discovery of user profiles, but from a different perspective and for different mobility data. For GPS trajectories, which is the focus of this paper, [2] defines as object profile the representative trajectory of a set of similar trips, for car pooling, [3] defines as profiles the users that visit similar places at similar times. [4] is the only work that proposes to infer socio-demographic profiles, but for social network data integrated to GPS trajectories, not only from pure GPS data. In summary, in these works a profile is considered as a set of features which characterize a type of user or a group of users, but not for socio-demographic inference.

In this paper we propose a different perspective. We assume that a description of a mobility behavior for specific socio-demographic categories of users is available and can be represented as “rules”. These rules can be defined by domain experts who describe which is a typical behavior of a specific category (workers, students, unemployed) in a certain application. Another possibility is to run data mining methods on census data or on GPS trajectories to identify groups of users with similar behavior, and label them with the socio-demographic category like “workers” or “students” [5]. Thus, given domain knowledge about how to describe a socio-demographic profile, we propose a *profile model* based on the rules that a moving object should fulfill to belong to a specific profile category. This model allows the user to specify, in a *simple way*, the types of profiles that are interesting for his/her application. How to match GPS trajectories to the profile model is the second focus of this paper, which proposes a moving object history model and a set of similarity functions that are capable to take into account the blurred aspect of

such profiles in two ways: (1) the temporal match is defined considering the overlapping portion between a profile model and the trajectories behavior; (2) the matching function assigns to the match a similarity degree between a profile model and the trajectory behavior. In other words, a trajectory may be matched to several profiles with different similarities, thus being able to discover multiple profiles.

Fig. 2 gives an overview of our proposal. Taking as input GPS trajectories, we first compute the trajectory history model. Then, we compute the similarity between the history model and the profile rules. The output is a set of trajectories labeled with one or more profile names.

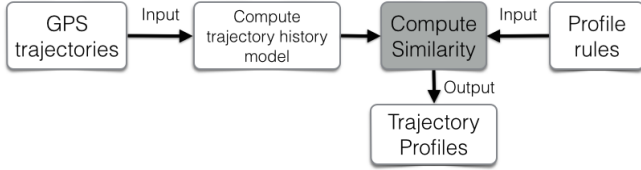


Fig. 2. Overview of the proposal.

The rest of this paper is organized as follows: Section II presents the related work; Section III introduces the basic and new concepts for this work; Section IV presents the algorithm T-Profiles for extracting socio-demographic profiles from trajectories; Section V presents the experimental evaluation of the method with real trajectories; and finally, Section VI presents the conclusion and future work.

## II. RELATED WORK

The inference of user profiles from GPS trajectory data, which is the focus of this paper, is very recent, and existing works for GPS trajectory mining have not addressed this problem to extract social information. For instance, [2] defines a set of representative trips performed by the object in his/her historical movement, and a profile is the spatio-temporal trajectory which is frequent in the object's movement history. Profiles in this work are computed for car pooling. Similarly, [6] defines object profile as a sequence of regions frequently visited by the object, and those with similar visits are clustered to infer communities of people. Both previous works focus on raw trajectories, where the object history is a set of space-time points, while we focus on semantic trajectories. [7] proposes a similarity measure that estimates the similarity among semantic trajectories, and the similarity is computed based on the matches of the sequences of categories of visited places between the trajectories. [3] proposes a similarity measure, considering not only the sequences of places but also the travel time to the place and the duration of the visit. Similar to previous works, [8] defines as user profile the mobility pattern of an object, computing the regions of interest (dense regions) and the duration of stays, but does not identify socio-demographic profiles.

In GSM data management, a user profile is defined as his/her mobility pattern [9], [10], [11], extracted from phone calls. Since telecommunication companies have normally a set of information about the user, it becomes trivial to infer user profiles from this type of data. In web logs and social networks, the inference of profiles has been an active area of research.

However, in these networks the user profile is available in the data, while GPS trajectories have only the position and time of the object.

Our work is different from all previous ones since we use a set of rules to describe the behavior of a profile. We also propose a *moving object history model*, which summarizes the individual user movement history in a way that it can be matched with the profile model. As a result, we give the similarity of a user with a given socio-demographic profile or multiple profiles.

## III. BASIC DEFINITIONS

Considering that we may infer the profile of people from places they visit, we make use of semantic GPS trajectories and stops [1]. A semantic trajectory  $A$  is a sequence of stops  $\langle stop_1, \dots, stop_i \rangle$  ordered in time, where each stop is associated to a POI type. Fig. 1 shows an example of semantic trajectory that has four stops  $\langle Home, University, ShoppingMall, Bar \rangle$ .

**Definition 1 (Stop).** Let  $POIType$  be a type of Point of Interest (POI),  $startTime$  and  $endTime$  be the start and end time that delimit the interval  $[startTime, endTime]$  in which a moving object  $oid$  stays at a POI of  $POIType$ . Then, a stop is a tuple  $(oid, POIType, startTime, endTime)$ .

In the following section we present the rule-based model (Section III-A), propose a history model (Section III-B), and define similarity measures for matching the rules and the history model (Section III-C).

### A. Profile Modeling

A profile is a set of features that represent a group of people with similar characteristics. These characteristics describe a profile/category. For example, the features *go to school, four or five times a week* describe a *student* profile. *Go to work, five times a week*, describe a *worker* profile. These examples of profiles are not mutually exclusive, since a worker can also be a student.

In order to extract socio-demographic profiles from trajectories we define a profile model with features that can be extracted and compared to moving object semantic trajectories. To make the model as simple as possible, we assume that four main features describe a socio-demographic profile: the *type of place* where people go (called  $POIType$ ), *when they go*, *how often* and for *how long* they stay there. With this set of features we define a *profile rule*. We denote with  $\mathcal{P}$  the set of profiles we want to investigate.

**Definition 2 (Profile rule for a  $POIType$ ).** Let  $POIType$  be a type of POI and  $p \in \mathcal{P}$  be a profile name. Then a profile rule  $r$  for a  $POIType$  and a profile  $p$  is a tuple of this kind:

$$r = (p, POIType, freq, \omega_f, timeU, weekPeriod, dayPeriod, duration, \omega_d),$$

where  $freq$  is the frequency that a  $POIType$  is visited in a time unit  $timeU$ , during certain periods of the day  $dayPeriod$ , and the period of the week ("weekday", "weekend" or "week")  $weekPeriod$ ,  $duration$  is an interval that describes the expected amount of time spent at  $POIType$  in the specified period of the day and week.  $\omega_f$  and  $\omega_d$  are the weights for

the attributes  $freq$  and  $duration$ , respectively, that should be in the interval  $[0,1]$  and their sum must be equal to 1.

An interesting part of this approach is that, to make the rules more expressive, we added a weight  $\omega$  to the attributes  $freq$  and  $duration$  to indicate the importance of the attribute to a specific rule.

A rule can express that a specific POIType should not be visited. For instance, a Retired should not have a Workplace. To support this type of profile we allow the definition of positive and negative rules, which are expressed through the attribute frequency. For positive rules the frequency attribute should be above zero ( $freq > 0$ ), and for negative rules  $freq = 0$ .

If any of the attributes  $weekPeriod$ ,  $dayPeriod$  or  $duration$  are not relevant, they can be set as Not Applicable (NA). The only exception is the attribute  $timeU$ , which can only assume NA when the profile rule is negative. Having defined the set of rules for a POIType we can define the profile model, given in Definition 3.

**Definition 3 (Profile model).** Let  $p \in \mathcal{P}$  be a profile name, a profile model for  $p$ , called  $\mathcal{R}^p$ , is a set of profile rules for POITypes associated with the profile name  $p$ .

### B. Moving Object History Modeling

The set of all stops of a moving object characterize the *movement history*. This history corresponds to the whole period that the object was tracked (e.g. one week, one month), i.e., the mobility diary. Definition 4 formalizes the object history extracted from semantic trajectories.

**Definition 4 (Object History).** An object history  $h = \langle stop_1, \dots, stop_n \rangle$  is the sequence of stops belonging to the same object such that

$$\forall i \in \{1, \dots, n-1\}, endTime_i \leq startTime_{i+1}$$

where  $endTime_i$  and  $startTime_i$  refer to the  $endTime$  and  $startTime$  of the  $i$ -th stop of the sequence, respectively.

From the object history, for each place (POIType) visited by an object in his/her trajectories, we compute the values in Definition 5 to summarize the trajectory information.

**Definition 5 (Moving Object History Model).** Let  $oid$  be a moving object identifier and  $h$  be its trajectory history. Then, a *moving object history model* for the object history  $h$ , called  $\mathcal{M}_h$ , is a set of tuples of this kind:

$$m = (oid, POIType, avgFreq, weekPeriod, dayPeriod, avgDuration)$$

where  $POIType$  is a type of POI,  $avgFreq$  is the average frequency that  $oid$  visits  $POIType$ ,  $weekPeriod$  specifies when this happens (weekdays, weekends or whole week),  $dayPeriod$  indicates the period of the day (morning, afternoon, evening, night) that  $oid$  visits  $POIType$ , and  $avgDuration$  is the average amount of time that the object spends at  $POIType$  at that weekPeriod and dayPeriod. All these values are extracted from the object history  $h$ .

Each tuple  $m \in \mathcal{M}_h$  represents the summary of a subset of stops from the object history  $h$  with the same  $POIType$

for a  $weekPeriod$  (weekday, weekend and whole week) and  $dayPeriod$ .

### C. Moving Object History Model and Profile Models Matching

As defined in section III-A, there can be two types of profile rules: positive and negative. For each type of rule the matching process is different. In Equation (1) we give the function that computes the similarity between a positive profile rule  $r$  and a tuple  $m$  of the moving object history model  $\mathcal{M}_h$ . It represents the sum of the similarities of frequency and average duration multiplied by their corresponding weight. The tuple  $m$  should have the same POIType, weekPeriod and dayPeriod of the ones in the profile rule  $r$ , in order to be analyzed.

$$sim_{pos}(m, r) = sim_f \cdot \omega_f + sim_d \cdot \omega_d \quad (1)$$

where  $POIType_m = POIType_r$ ,  $weekPeriod_m = weekPeriod_r$ , and  $dayPeriod_m = dayPeriod_r$ .

The similarity functions for frequency  $sim_f$  and duration  $sim_d$  are defined by functions that follow the same idea of a set membership function in fuzzy logic, and are detailed in section IV, that describes the algorithm.

The similarity for negative rules is defined by Equation (2), where if the  $POIType$  defined in the profile rule  $r$  is not present in the moving object history model  $\mathcal{M}_h$  it has  $sim = 1$ .

$$sim_{neg}(\mathcal{M}_h, r) = \begin{cases} 1 & \text{if } POIType_r \notin \mathcal{M}_h \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

The total similarity between a *moving object history model*  $\mathcal{M}_h$  and a profile name  $p \in \mathcal{P}$  is given by the function *MATCH* in Equation (3). In general words it is the sum of the similarities of the positive rules  $sim_{pos}$  and the sum of the similarities of the negative rules  $sim_{neg}$ , divided by the total number of rules of that profile name  $p$ .

$$MATCH(\mathcal{M}_h, p) = \frac{\sum sim_{pos}(m_i, r_j) + \sum sim_{neg}(\mathcal{M}_h, r_k)}{|\mathcal{R}^p|} \quad (3)$$

where  $\mathcal{R}^p$  is the set of rules of the profile name  $p$

In the following section we present the algorithm T-Profiles, to extract socio-demographic profiles from trajectories.

## IV. T-PROFILES: AN ALGORITHM FOR DISCOVERING TRAJECTORY PROFILES

Listing 1 shows the pseudo-code of the proposed algorithm to extract profiles from trajectory data, named T-Profiles. The algorithm receives as input a set of semantic trajectories  $T$ , a set of profile names  $\mathcal{P}$ , a set of profile models  $\mathcal{R}$ , and the minimal similarity degree  $\epsilon$  for an object to be considered similar to a profile name. The output is a set of moving objects labeled with a profile name  $p$  and the similarity degree between the object and the profile.

The first step is to compute the history model for each moving object in  $T$  (lines 12, 13). We summarize the trajectories

to the structure of Definition 5, for each visited place. Once computed, the moving object history model will be compared with all rules of each profile name  $p$  in the profile model  $\mathcal{R}^p$  (lines 14-38).

The similarity of positive rules (line 18) is computed according to Equation (1), with  $sim_f$  defined by Equation (4) and  $sim_d$  defined by Equation (5), presented in the following. The similarity of negative rules (line 24) is calculated according to Equation (2). These values are used to compute the matching between a moving object history model and a profile name (line 33).

If one or more of the negative rules of a profile name are not satisfied by the object history model, the similarity is set to zero, since the negative rules are mandatory (line 31). In case the total similarity is greater than the threshold  $\epsilon$ , the moving object identifier, the profile name and the similarity degree are added to the output set  $\psi$  of trajectory profiles (line 36). This step finishes the analysis of one profile name and the algorithm returns to line 14 to test the next profile name in  $\mathcal{P}$  with the current object history. Notice that the algorithm has the capability to return multiple profiles, i.e., a moving object can belong to several profiles in case the match is above  $\epsilon$ .

Listing 1. Pseudo-code of the algorithm T-Profiles

```

1  Input:  $T$  // set of semantic Trajectories
2   $\mathcal{P}$  // set of profile names
3   $\mathcal{R}$  // profile models  $\mathcal{R} = \cup_{p \in \mathcal{P}} \mathcal{R}^p$ 
4   $\epsilon$  // minimal similarity degree for
5  // an object belonging to a profile
6
7  Output:  $\psi$  // set of moving object profiles
8
9  Method:
10
11   $\psi = \{\}$  // empty set
12  for each moving object history  $h \in T$  do
13     $\mathcal{M}_h = \text{buildMovingObjectHistoryModel}(h)$ 
14    for each profile name  $p \in \mathcal{P}$  do
15      sumPos = 0
16      for each positive rule  $r \in \mathcal{R}^p$  do
17        for each  $m \in \mathcal{M}_h$  do
18          sumPos = sumPos +  $sim_{pos}(m, r)$ 
19        end for
20      end for
21      negativeRulesNotHold = False
22      sumNeg = 0
23      for each negative rule  $r \in \mathcal{R}^p$  do
24        aux =  $sim_{neg}(\mathcal{M}_h, r)$ 
25        sumNeg = sumNeg + aux
26        if aux = 0
27          negativeRulesNotHold = True
28        end if
29      end for
30      if negativeRulesNotHold
31        MATCH = 0.0
32      else
33        MATCH = (sumPos + sumNeg) /  $|\mathcal{R}^p|$ 
34      end if
35      if MATCH >  $\epsilon$ 
36         $\psi$ .add( $h.oid, p, \text{MATCH}$ )
37      end if
38    end for
39  end for
40  return  $\psi$ 

```

### Similarity functions

The frequency and the duration similarity can be implemented in different ways. After performing several experiments we implemented the following in T-Profiles: the similarity for frequency ( $sim_f$ ) is defined by Equation (4),

where  $avgFreq_m$  is the average frequency computed in the history model tuple  $m$ ,  $freq_r$  and  $timeU_r$  are respectively, the frequency and the time unit defined in the profile rule  $r$ . The function  $days(timeU_r)$  returns the number of days that the time unit represents (e.g. if  $timeU_r = week$ , then  $days(timeU_r)$  returns 7). When  $avgFreq_m = 0$ , means that the object did not visit the POIType, so  $sim_f = 0$ . When  $avgFreq_m$  is lower than the frequency defined in the profile rules (represented by  $\frac{freq_r}{days(timeU_r)}$ ), then  $sim_f$  increases linearly from 0 to 1. If  $avgFreq_m$  is greater than the frequency defined in the profile rules, then  $sim_f = 1$ .

$$sim_f = \begin{cases} 0 & \text{if } avgFreq_m = 0 \\ \frac{avgFreq_m \cdot days(timeU_r)}{freq_r} & \text{if } avgFreq_m < \frac{freq_r}{days(timeU_r)} \\ 1 & \text{if } avgFreq_m \geq \frac{freq_r}{days(timeU_r)} \end{cases} \quad (4)$$

The duration similarity ( $sim_d$ ) is defined by Equation (5), and is illustrated in Fig. 3 for  $duration_r$  defined as the interval [1:00, 2:00]. Fig. 3 shows that an  $avgDuration_m$  between 1 and 2 hours will have  $sim_d = 1$ . For an  $avgDuration_m$  between 0.5 hour and 1 hour the similarity increases linearly from 0 to 1, and for an  $avgDuration_m$  between 2 and 2.5 hours the similarity decreases linearly from 1 to 0.

$$sim_d = \begin{cases} 0 & \text{if } avgDuration_m < minGlobal \vee \\ & avgDuration_m > maxGlobal \\ 1 - \frac{minDur - avgDuration_m}{minDur - minGlobal} & \text{if } minGlobal < avgDuration_m \wedge \\ & avgDuration_m < minDur \\ 1 & \text{if } minDur \leq avgDuration_m \wedge \\ & avgDuration_m \leq maxDur \\ 1 + \frac{maxDur - avgDuration_m}{maxGlobal - maxDur} & \text{if } maxGlobal > avgDuration_m \wedge \\ & avgDuration_m > maxDur \end{cases} \quad (5)$$

where

$$duration_r = [minDur, maxDur] \\ minGlobal = minDur - (minDur * 0.5) \\ maxGlobal = maxDur + (minDur * 0.5)$$

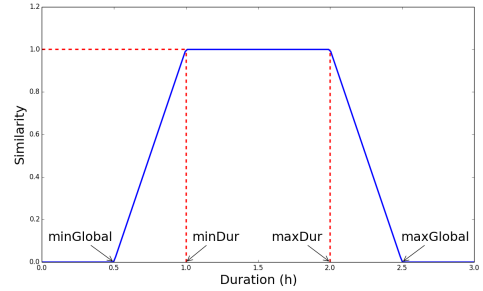


Fig. 3. Duration similarity function  $sim_d$  for  $duration_r = [1:00 - 2:00]$ .

## V. EXPERIMENTAL EVALUATION

We evaluate our proposal using two datasets, a trajectory set built from census data where we have the ground truth (V-A), and a GPS trajectory dataset of car trajectories collected in Florence, Italy (V-B).

### A. Census Trajectories

As it is still very difficult to obtain a dataset of semantic trajectories with a ground truth, we first evaluate the algorithm T-Profiles on a “trajectory” dataset generated from census data, where we have the ground truth. This dataset is a census of

activity diaries collected in Italy in 2008, having the socio-demographic profile of each individual that was interviewed. Each activity diary corresponds to the activities of one person during one day, and can be seen as the “semantic trajectories” of each individual, because they contain the place of activity (that corresponds to the POIType of the stops), the activities performed at the place, and the begin and end time of the activities. Examples are: sleeping at home from 10PM to 8AM, profile *retired*; working at a workplace from 10AM to 5PM, profile *worker*; studying at the university from 2PM to 6PM, profile *student*, etc. The most significant profiles in the database are: *worker*, *retired*, *unemployed*, *housewife with kids* and *student*.

As one day of activities is not enough to determine the profile of a person, we preprocessed the data grouping diaries that belong to the same socio-demographic profile, considering 14 days of activities. As a result, we obtained trajectories of 14 days long for 829 objects. The amount of objects for each profile is shown in the second column of Table II.

TABLE I. PROFILE RULES.

Profile Name $p$	POIType	freq ( $\omega_f$ )	timeUnit	weekPeriod	dayPeriod	duration ( $\omega_d$ )
Worker	Workplace	4 (0.8)	week	NA	NA	03:00 - 09:00 (0.2)
Student	School/Univ.	3 (0.5)	week	weekday	NA	03:00 - 06:00 (0.5)
Retired	Workplace	0 (1)	NA	NA	NA	NA
Retired	School/Univ.	0 (1)	NA	NA	NA	NA
Retired	Bar	1 (1)	week	NA	Morn, Aftn	NA
Unemployed	Workplace	0 (1)	NA	NA	NA	NA
Unemployed	School/Univ.	0 (1)	NA	NA	NA	NA
Unemployed	Bar	1 (1)	week	NA	Evening	NA
Unemployed	Restaurant	1 (1)	month	NA	NA	NA
Unemployed	Sport court	2 (1)	month	NA	NA	NA
Housewife Kids	School/Univ.	3 (0.5)	month	weekday	NA	00:20 - 01:00 (0.5)
Housewife Kids	Commercial estab.	3 (1)	week	weekday	NA	NA

Table I shows the rules considered in this experiment. A *worker* is identified by the POIType Workplace, that should be visited with a frequency of 4 times a week with duration between 3 and 9 hours. We define a broad range for duration to obtain all types of workers (full time and part time). Notice that we defined a higher weight for the frequency (0.8), because this attribute is more important than the duration.

The rule for the profile named *Student* expresses that this profile should visit a POIType related to educational institutions, such as schools or universities, for at least 3 times a week on weekdays, with a duration between 3 and 6 hours per day, to include full time and part time students. The weights for the attributes *freq* and *duration* are both 0.5.

For the profile *Retired*, we defined two negative rules related to workplace and educational places, to distinguish between workers and students, since it is expected that most retired do not have a workplace and do not go to school. However, these rules are not enough to distinguish a retired from an unemployed. Then, as they are supposed to go more often to bars or cafes, we create a rule with this kind of POIType, in the period of morning and afternoon, i.e., during the day.

The profile *Unemployed* may have similar behavior to the *Retired*, having no working place and not going to school to distinguish these profiles. To distinguish an unemployed from a retired we define three positive rules: POIType Bar visited during the evening, POIType Restaurant visited only once a month, and visits to sport places.

A housewife that has children can be identified if the person

visits educational places such as schools. But the difference from the profile student is the frequency and the duration. The profile does not need to go every day to take the child to the school, but should at least visit a POIType school sometimes to express that there is a relationship with educational place. Defining a rule forcing a housewife with kids to go very frequently to educational places would limit the discovery only of cases where the housewife takes the kids to school everyday.

Table II shows the results for similarity  $\epsilon$  of 60%, 70% and 80%. For similarity 70%, for instance, T-Profiles detected 478 workers out of 479, and 73 out of 74 students. For the profile Housewife Kids, 24 instances were discovered. The most difficult classification is to distinguish unemployed and retired, because their behavior is very similar, but still 158 retired from 224 were detected.

TABLE II. PROFILES FOR 60%, 70% AND 80% SIMILARITY.

Profiles	Total	$\epsilon = 0.6$	$\epsilon = 0.7$	$\epsilon = 0.8$
Worker	479	479	478	473
Housewife Kids	35	28	24	20
Unemployed	17	9	9	9
Retired	224	185	158	158
Student	74	74	73	72

Table III shows the precision and recall for each profile, considering the similarities for each profile name as well as the average for all objects. T-Profiles shows a very high average precision, about 97%. The recall is also high, between 88% and 93% with these values of  $\epsilon$ .

TABLE III. PRECISION AND RECALL

	$\epsilon = 0.6$		$\epsilon = 0.7$		$\epsilon = 0.8$	
	Precision	Recall	Precision	Recall	Precision	Recall
Worker	1.000	1.000	1.000	0.997	1.000	0.987
Housewife Kids	0.583	0.800	0.750	0.685	0.769	0.571
Unemployed	0.529	0.529	0.529	0.529	0.600	0.529
Retired	0.953	0.825	0.957	0.705	0.957	0.705
Student	1.000	1.000	1.000	0.986	1.000	0.972
Avg.	0.960	0.934	0.968	0.895	0.970	0.882
Avg. F1 measure	0.946		0.926		0.921	

In these results we considered only the highest similarity for each profile name, but we can also analyze all similarities that are above the threshold  $\epsilon$ , having multiple profiles. Table IV shows some examples of the output of T-Profiles. Each row corresponds to an object. The multiple profile column shows all profile names that have similarity above 80%. For the object 842, for instance, the similarity with Worker and Housewife Kids is above 90%, so this object is labeled as Worker and Housewife Kids.

TABLE IV. MULTIPLE PROFILES FOUND USING  $\epsilon = 0.8$

oid	Worker	Housewife Kids	Unemployed	Retired	Student	Multiple profile
16	0.000	0.174	0.800	0.980	0.000	Retired, Unemployed
131	0.000	0.261	0.800	0.980	0.000	Retired, Unemployed
842	1.000	0.949	0.000	0.000	0.179	Worker, Housewife Kids
600	1.000	0.897	0.000	0.000	0.100	Worker, Housewife Kids

## B. Florence Dataset

The Florence dataset is a sample of car trajectories collected by an insurance company during one month, but the average tracking period of one object was 10 days. From this dataset, we selected all trajectories with more than 10 stops, labeled with their POITypes and with at least 10 days history.

Here we show the flexibility of T-Profiles, where the user can choose any level of profile category analysis, from the more general to the more detailed. We are interested in Full Time Workers, Part Time Workers, Weekend Workers and Night Workers. Considering the rules defined in Table I, we extended the rules set with new profiles of workers. Table V shows the rules for workers, where the duration distinguishes Full Time and Part Time Workers; while the frequency, week period, and day period distinguish Weekend and Night Workers.

TABLE V. PROFILE RULES FOR WORKER PROFILES.

Profile Name $p$	POIType	freq ( $\omega_f$ )	timeUnit	weekPeriod	dayPeriod	duration ( $\omega_d$ )
Full Time Worker	Workplace	4 (0.5)	week	NA	NA	07:00 - 09:00 (0.5)
Part Time Worker	Workplace	4 (0.5)	week	NA	NA	03:00 - 05:00 (0.5)
Weekend Worker	Workplace	1 (1)	week	weekend	NA	NA
Night Worker	Workplace	3 (1)	week	NA	Evening, Night	NA

Table VI shows the result for  $\epsilon = 0.8$ , where T-Profiles labeled 36 Full Time Workers, 16 Part Time Workers, 31 Weekend Workers, 21 Night Workers, 4 Students, and 2 Retired.

TABLE VI. PROFILES FOR 70%, 80% AND 90% SIMILARITY.

Profiles	$\epsilon = 0.7$	$\epsilon = 0.8$	$\epsilon = 0.9$
Full Time Worker	56	36	32
Part Time Worker	26	16	9
Weekend Worker	37	31	31
Night Worker	24	21	15
Student	4	4	4
Retired	2	2	0
Unemployed	1	0	0
Housewife Kids	0	0	0

Table VII shows some examples of the output of T-Profiles, with some single and multiple profiles. Notice that objects 893757 and 820817 were labeled with multiple profiles. For oid 893757 the similarity degree was 100% with Part Time and Weekend Worker, while object 820817 had similarity of 100% with Full Time and Night Worker. Objects 85000 and 288807 had similarity above 80% with the profile category Full Time and Night Worker, respectively. Another example is the object with oid 757727, that had similarity 100% with the profile Student.

## VI. CONCLUSION AND FUTURE WORK

In this paper we made a first attempt to go deeper in the analysis of moving object trajectories, analyzing every individual object mobility history in order to discover the socio-

TABLE VII. EXAMPLES OF THE OUTPUT OF T-PROFILES FOR  $\epsilon = 0.8$

oid	Housewife Kids	Unemployed	Retired	Student	Full Time Worker	Part Time Worker	Weekend Worker	Night Worker	Profile
893757	0.000	0.000	0.000	0.000	0.603	1.000	1.000	0.000	Part Time Worker, Weekend Worker
820817	0.000	0.000	0.000	0.000	1.000	0.500	0.000	1.000	Full Time Worker, Night Worker
85000	0.326	0.000	0.000	0.106	0.977	0.477	0.000	0.000	Full Time Worker
288807	0.000	0.000	0.000	0.000	0.743	0.318	0.636	0.848	Night Worker
757727	0.250	0.000	0.000	1.000	0.000	0.000	0.000	0.000	Student
1255063	0.000	0.750	0.852	0.000	0.000	0.000	0.000	0.000	Retired

demographic status of each individual. While the discovery of socio-demographic profiles is very trivial in social networks, GSM calls and weblog data, where far more information is available, the discovery of socio-demographic profiles from GPS trajectories is a challenge. In this paper we proposed a profile model, as a set of very simple rules that the user can express to discover any type of profile. We also introduced a moving object history model that summarizes the historical traces of moving objects, that is independent of a specific profile model. Finally, we proposed a matching process that provides the similarity between a given profile name and a moving object based on his/her trajectory summary. As future work, we will go deeper in the analysis to discover more complex profiles such as gender, marital status, income, etc.

## ACKNOWLEDGMENT

This work was supported by EU project FP7- PEOPLE SEEK (N.295179 <http://www.seek-project.eu>) and the Brazilian agencies CAPES and CNPQ.

## REFERENCES

- [1] L. O. Alvares, V. Bogorny, B. Kuijpers, J. A. F. de Macedo, B. Moelans, and A. Vaisman, "A model for enriching trajectories with semantic geographical information," in *Proceedings of the 15th annual ACM international symposium on Advances in geographic information systems*. ACM, 2007.
- [2] R. Trasarti, F. Pinelli, M. Nanni, and F. Giannotti, "Mining mobility user profiles for car pooling," in *Proceedings of the 17th ACM SIGKDD*, ser. KDD '11. ACM, 2011, pp. 1190–1198.
- [3] X. Xiao, Y. Zheng, Q. Luo, and X. Xie, "Finding similar users using category-based location history," in *Proceedings of the 18th SIGSPATIAL*. ACM, 2010, pp. 442–445.
- [4] V. W. Zheng, Y. Zheng, and Q. Yang, "Joint learning user's activities and profiles from GPS data," in *Proceedings of the 2009 LBSN*, ser. LBSN '09. New York, NY, USA: ACM, 2009, pp. 17–20.
- [5] S. Jiang, J. Ferreira, and M. C. González, "Clustering daily patterns of human activities in the city," *Data Mining and Knowledge Discovery*, vol. 25, no. 3, pp. 478–510, 2012.
- [6] C.-C. Hung, C.-W. Chang, and W.-C. Peng, "Mining trajectory profiles for discovering user communities," in *Proceedings of the 2009 LBSN*, ser. LBSN '09. New York, NY, USA: ACM, 2009, pp. 1–8.
- [7] J. J.-C. Ying, E. H.-C. Lu, W.-C. Lee, T.-C. Weng, and V. S. Tseng, "Mining user similarity from semantic trajectories," in *Proceedings of the 2nd ACM SIGSPATIAL LBSN*, ser. LBSN '10. New York, NY, USA: ACM, 2010, pp. 19–26.
- [8] X. Chen, J. Pang, and R. Xue, "Constructing and comparing user mobility profiles," *ACM Transactions on the Web (TWEB)*, vol. 8, no. 4, p. 21, 2014.
- [9] M. A. Bayir, M. Demirbas, and N. Eagle, "Mobility profiler: A framework for discovering mobility profiles of cell phone users," *Pervasive and Mobile Computing*, vol. 6, no. 4, pp. 435 – 454, 2010.
- [10] B. Furlotti, L. Gabrielli, C. Renso, and S. Rinzivillo, "Analysis of GSM calls data for understanding user mobility behavior," in *2013 IEEE International Conference on Big Data*, Santa Clara, California, 2013, pp. 550–555.
- [11] M. Dash, H. L. Nguyen, C. Hong, G. E. Yap, M. N. Nguyen, X. Li, S. Krishnaswamy, J. Decraene, S. Antonatos, Y. Wang, D. T. Anh, and A. Shi-Nash, "Home and work place prediction for urban planning using mobile network data," in *IEEE 15th MDM*, vol. 2, July 2014, pp. 37–42.