Towards an Affective Semantic Trajectory Generator (ASTG)

Antonios Karatzoglou
Karlsruhe Institute of Technology and Robert Bosch, Corporate Sector Research and Advance Engineering
Germany
antonios.karatzoglou@kit.edu
antonios.karatzoglou@de.bosch.com

Markus Szarvas
Karlsruhe Institute of Technology
Germany
markus.szarvas@student.kit.edu

Michael Beigl
Karlsruhe Institute of Technology
Pervasive Computing Systems
Germany
michael.beigl@kit.edu

Abstract*—Trajectory modelling, trajectory analysis and trajectory prediction have become very important tools in the hands of mobile service providers, whether in respect to resource management (e.g., mobile network management), or to building intelligent, context-aware mobile applications. Most of the existing modelling approaches are highly data-driven. For this reason, the need of large, high-quality datasets has become enormous in the recent years. It is very costly and time-consuming to collect real-world data such as human trajectory data. Moreover, new privacy laws and restrictions make it even more harder. Thus, data turned into a bottleneck for algorithm developers of all kinds. Synthetic data generators provide a solution for this problem. There exists a variety of approaches for producing synthetic trajectories and many extra features have been investigated such as the transportation mode, the proximity to friends and the activity to name but a few. However, none of them has explored the use of psychological features, such as the personality and the emotional state of the users. In this work, we try to give insight into the impact of the aforementioned features on the generation process of (semantic) location trajectories. For this purpose, we designed a novel multi-agent synthetic trajectory generator that takes, among others, these features explicitly into account. We refer to it as *Affective Semantic Trajectory Generator (ASTG)*. In order to evaluate our approach and the use of personality and emotions, we compared the produced trajectories with the outcome of two large-scale studies (> 25.000 participants each) conducted in Germany and Chicago, USA in 2008. It can be shown that dynamic data, such as emotions, can lead to a better performance, a fact that makes ASTG particularly interesting for further investigation.

***Index Terms*—mobile users, context awareness, urban movement patterns, trajectory generator, semantic trajectories, personality traits, emotional state**

I. Introduction

The interest in neural networks and artificial intelligence has grown exponentially over the past few years. The need for big and high-quality data is as high as never before since they play a major role in the models' performance. However, getting a big amount of high-quality data is not easy. In many cases, it can be very expensive to collect or buy a good dataset. Furthermore, recent privacy laws and restrictions make the situation even more difficult, especially in the field of Location Based Services (LBS) and mobility or trajectory modelling and analysis. A solution for this problem can be provided by synthetic trajectory generators, which have the goal to produce realistic human trajectories that are statistically as close as possible to real human trajectories.

There exist many different trajectory generators. These can be categorized into two main classes. The ones that use GPS data as basis, and the ones that use log-diaries or surveys. In both cases, a big variety of algorithmic approaches has been suggested by now. These vary depending on their core model, as well as on the number and type of training features that the respective model takes into consideration. Although many different training features have been explored by now, to our knowledge, none of the existing trajectory generators considers the personality and the emotional state of the users as input for their models.

The presented work aims at exploring the impact of the two aforementioned psychological features (among other, e.g., location, activity, companionship, transportation mode, weather, etc.) on the performance of a trajectory generator in terms of output accuracy and representativity. We believe that considering the users' emotions and personality would award the generator with a certain degree of emotional intelligence and thus make it more accurate in generating human location sequences. For this purpose we designed and implemented a probabilistic multi-agent survey-based trajectory generator, which we refer to as *Affective Semantic Trajectory Generator (ASTG)*. Goal of ASTG is to produce synthetic semantic trajectories, that is, temporal sequences of both location (type) and activity (see Section III). We evaluate our approach, and thus indirectly the influence of emotions and personality, by comparing the ASTG's results with the initial survey data, as well as with the statistical outcome of two large-scale mobility studies conducted in Germany (MiD2008[1]) and Chicago, USA [1] in 2008 with more than 25.000 and 30.000 participants respectively. It can be shown that the ASTG and personal dynamic data such as emotions can lead to a higher accuracy and representativity.

This work is structured as follows. Section II gives a brief overview about some of the basic related work in this field.

[1]http://www.mobilitaet-in-deutschland.de/mid2008-publikationen.html

Then, in Section III we provide insight into the basic concepts described in our work. Section IV and V describe in detail the user survey and the architecture of our approach, ASTG, while in Section VI we present and discuss our evaluation results. Finally, Section VII summarizes our work and contains some suggestions about potential future work.

II. Related Work

Spatio-temporal trajectory generators can be categorized into two main classes, the *network-based* and the *network-free* generators. Network-based generators assume that moving objects follow a certain network and take this fact explicitly into consideration, while network-free generators don't, like in [2] and [3]. Objects moving along a network underly a set of fundamental restrictions, such as the number of nodes and interconnections, the connection's maximum speed and capacity and how moving objects affect other moving objects with regard to travelling time and chosen route. The same kind of restrictions experience humans in an urban environment when moving from one location to another.

Brinkhoff implemented in 2002 as one of the first a network-based trajectory generator [4]. His work serves as basis for most of the network-based generators, whereby the term *network* usually refers to the road and/or the railway network. With Brinkfoff's generator, one is able to determine a set of both network and moving object characteristics and boundary conditions. These include definitions such as the class and the maximum possible speed of an edge in the network or an object. The weather, traffic jams and similar events can also be defined with respect to their impact on the maximum speed and capacity of a certain edge or group of edges between nodes in an area. Both the starting and the destination nodes are computed by using a 1- or a 2-dimensional distribution function such as the uniform or the Gaussian distribution, whereas the destination node depends additionally on the starting node and the length of the route (which can be manually defined). The lifetime and the number of the moving objects is a time-dependent function. For instance during the rush hour, there are much more moving objects in the streets than late at night. Furthermore, moving objects cease to exist after arriving at their goal destination or after reaching a maximum time threshold. The generated trajectories are computed based on the premise that most moving objects tend to choose the fastest way to their destination. For this reason, Brinkhoff uses the A* search algorithm.

As in Brinkhoff's work, the core principle of most trajectory generators underlies a certain framework model that defines a set of boundary conditions, e.g., the schedule of an office worker during the week, short driving times of cars in the city in contrast to longer ones made by trucks in the highways, etc. For instance BerlinMOD [5], a further network-based generator, utilizes both a rule-based and a probabilistic component in its trip creation algorithm. Duntgen et al. use a set of rules to define certain situation-specific temporal and spatial behaviours, e.g., home leaving and arrival times during the weekdays and the weekends for working people, etc. In general, BerlinMOD shows many similarities with Brinkhoff's work, but in contrast to Brinkhoff, their algorithm can be used for long-term observations of moving objects. Giannotti et al.'s work [6] is based on the GSTD generator [3], a general, network-free spatio-temporal generator, which among others offers the users the option of determining the agility degree, that is, how often a moving object changes its direction, as well as the option of defining the urban infrastructure. Their work places particular emphasis on cellular networks and extends GSTD by taking GSM antenna covering areas additionally into account. With MNTG [7], Mokbel et al. provide an extendable wrapper for existing trajectory generators. Their framework comes shipped with the Brinkhoff and the BerlinMOD generator, as well as with the U.S. Tiger and the OpenStreetMap road networks, but it can be easily extended to support other generators and road networks as well.

Pelekis et al. introduce in [8] *Hermoupolis*, a pattern-aware network-based trajectory generator. Hermoupolis takes as input an existing set of trajectory patterns together with information about the road network and a set of within lying POIs[2]. Unlike the aforementioned research, Pelekis et al. follow the trend of semantic trajectory analysis [9] and provide a way to include semantic meta-information in the trip modelling and generation process, e.g. information about the activity performed at a certain POI, etc. In [10], Pelekis et al. extend their previous work by enlarging upon the semantic description of the taken routes and define so called semantic episodes. Semantic trajectories (see Section III), gained in the last years increasingly in significance, whether in respect of human trajectory analysis [11], as well as of destination prediction [12]–[14]

Pelekis et al.'s work matches another group of generators, that relies increasingly on further enriching semantically their core models, aiming at improving the representativity of the generated data. However, the particular generators are not solely location generators, but focus rather on modelling and generating both travel patterns and activity sequences. The TASHA generator of Miller et al. is such a travel and activity generator [15]. TASHA's heuristic and rule-based algorithm generates a sequence of travel episodes, which describe the daily routine of moving objects, that is the corresponding times, places and activities. In contrast to the aforementioned work, Miller et al.'s work is based on a trip diary dataset. This is particularly important in cases where no activity data are available, like for many urban areas of interest. ST-ACTS is a further spatio-temporal generator that similar to TASHA takes the activities explicitly into consideration [16].

In a recent work [17], Doudali et al. present an interesting framework that uses as input and generates GPS traces as well as text in form of check-ins as found in Location-based Social Networks like Twitter and Facebook. Their algorithm follows a set of probabilistic distributions and supports similar boundary conditions like the former generators, such as the maximum distance and the maximum duration of a daily route.

[2]Point Of Interest

Finally, a certain number of recent works explore the use of artificial neural networks and deep learning for generating spatio-temporal trajectories and sequential data in general following the hype of the latter ([18] and [19] respectively).

As we saw above, trajectory generators have explored a big variety of heterogenous data as their input by now, such as GPS recordings and travel diaries, transportation network and urban infrastructure, POI bases and common sense knowledge rule sets (e.g., daily workers' schedules, etc.), to name but a few. However, none of them investigated the impact that psychological features like the personality and the emotional state of the mobile objects might have on their output data quality and representativity. And this despite the fact that many studies suggest that both personality and emotions play a significant role in a human's future actions. Ajzen et al. show in their work that the personality could be used to help understand certain attitudes and predict future social behaviour [20]. Some years later, in [21], Ajzen et al. introduce the *theory of planned behaviour* (TPB). Apart from the person's character, TPB takes subjective and social norms, as well as the behavioural power and control into consideration. The behavioural control refers to what degree a person is willing to act in a certain way even if the conditions are not favourable. Hunecke et al. point the significance of sociodemographic and external factors like the infrastructure in their work as well [22]. For instance a person can only travel by bus if there is a bus near to get on, despite fulfilling the other conditions. At the same time, the choice of the transportation means itself can be partly attributed to the weather [23], [24]. TPB has been used successfully to predict and explain people's behaviour in various fields such as in the marketing and the healthcare domain. Rhodes et al.'s study confirms the relationship between certain personality traits and exercise behaviour, as well as the performance of an extended theory of planned behaviour model [25]. Kim et al. use the users' personality as input in their model to provide predictions about the locations to be visited next [26], [27].

In [28] it could be shown that emotions influence the choice of food and the overall eating behaviour, such as, for instance, that bad mood often leads to eating fast food, which in turn can affect the people's movement patterns. In addition, Forgas describes in his work the overall significance that emotional state has when it comes to taking decisions [29], a fact that can be also carried over to our mobility pattern case.

Thus, including the personality and the emotional state into a synthetic trajectory generator seems to be more than reasonable. In the presented work, we explore whether and to what extend the aforementioned psychological features may lead to a better semantic trajectory generator in terms of representativity of the outputted trajectories and matching degree with the results of the German national statistics institute, as well as with the outcome of a large-scale field study in Chicago (see Section VI).

III. BASICS

This section gives insight into the basic concepts described in this work.

A. Semantic Trajectories

The concept *semantic trajectories* first appeared in the works of Spaccapietra [30] and Alvarez [11]. They show that knowing the semantics within the available trajectories can be very beneficial since it can contribute to a better understanding and thus a more extensive and accurate trajectory analysis. A typical semantic trajectory reduces a GPS track into a small sequence of significant locations, like the one shown in Fig. 1. Eq. 1 and Eq. 2 define a GPS and a semantic trajectory respectively.

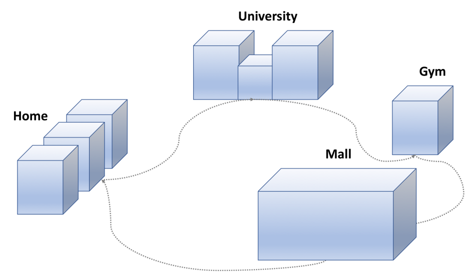

Fig. 1. Typical daily semantic trajectory.

$$(long_1, lat_1, t_1), (long_2, lat_2, t_2), ..., (long_n, lat_n, t_n) \quad (1)$$

$$(semLoc_1, t_1), (semLoc_2, t_2), ..., (semLoc_m, t_m) \quad (2)$$

with $long_i$ the longitude, lat_i the latitude, t_i the timestamp and $m < n$, since each semantic location comprises a number of GPS points within a given radius and temporal threshold. An element of a semantic trajectory may comprise more than just the location type depending on the semantic enrichment level, such as the user's activity and the overall purpose of visit. Our generator operates with exact these type of trajectories.

B. The Big Five Factor Model

The *Big Five Factor Model* is a broadly used psychological model to describe a person's personality [31]. Its structure is based on 5 orthogonal factors, which describe 5 main personality characteristics (*personality traits*). The five factors covered by the Big Five model are:

- Openness to experience
- Conscientiousness
- Extraversion
- Agreeableness
- Neuroticism

Each of the above personality traits can be more or less pronounced depending on the character of the person. The respective extend can be evaluated and mapped onto a numeric scale, e.g., from 1 to 5. This makes it particularly suited for feeding it as input into a generator model.

C. Markov Model

A *Markov model* (or *Markov Chain*) describes a memoryless stochastic process. The term *memoryless* refers to the Markov property, which states that predictions about future states of a model show similar results, whether these are based on just a short history part or on the entire history available. If the Markov model takes just the previous state into consideration, then we are talking about a 1. Order Markov model. This can be expressed through the conditional probability in Eq. 3:

$$p(s^{k+1}|s^1, s^2, ..., s^k) = p(s^{k+1}|s^k), \tag{3}$$

with p being the conditional probability and s^k a random state at time step k. In our work, we use the Markov model for generating temporal and semantic location distributions.

IV. User Survey

Our generator is mainly a log diary-based generator. In order to collect the respective necessary diary data we conducted an online user survey. For this purpose and to ensure that the sample will be representative we used the online platform Clickworker[3]. We surveyed a total of 100 people with regard to:

- Demographic information
- Personality
- 7-day mobility and activity diary
- General relations between location types, activities, time of day, companionship and emotional and mental state

The collected demographic information comprise among others the age, the sex, the occupation, the household size and income, the education level and the ownership of a car or motorbike. In order to determine the participants' personality we used the 10-item personality test of Rammstedt et al. [32]. Core of our survey was the 7-day long mobility and activity diary. The survey participants were interviewed with regard to their daily movement and activity patterns over a period of one week, whereby the users were given, beside free text boxes, a predefined set of location and activity types to chose from (see Table I). Furthermore, they were invited to provide additional information about their emotional and mental state during the day, as well as about the weather condition, their companion (if any) and their transportation means. Finally, we asked people to bring the given set of main location categories into connection with certain emotions and activities, as well as with the information whether they visit these places usually alone (e.g., gym) or not (e.g., cinema) and when. The considered main location and activity categories can be found in Table I. In addition, Table I lists also the surveyed emotional states.

After collecting the surveys, we preprocessed the data by clustering single locations and activities into the above mentioned predefined categories for the free text box cases in the survey. In addition, we filtered out all inconsistent, empty or erroneous entries.

[3]https://www.clickworker.de

TABLE I
LIST OF MAIN LOCATION AND ACTIVITY TYPES AND EMOTIONAL STATES

Location types	***Activity types***	***Emotional states***
Residence	earn money	energetic
Business & Services	shopping	excited
School or University	sports	happy
Culture & Entertainment	family	angry
Nightlife	education	bored
Free time & Nature	entertainment	frustrated
Travel & Transportation	relax	neutral
Food		sad
Event		stirred up
		hungry
		sleepy
		stressed
		ill

V. Affective Semantic Trajectory Generator (ASTG)

The generator we introduce in this paper is a survey-based semantic trajectory generator that takes both emotional and personality characteristics explicitly into account. For this reason, we refer to it as *Affective Semantic Trajectory Generator (ASTG)*. Our generator produces a sequence of semantic locations that comprises both the locations and the performed activities. This is especially useful since many locations are multi-purpose locations, at which a certain number of different activities may take place (e.g., home-relax and home-work, or mall-meet friends, mall-shopping, mall-having a haircut, etc.). Pappalardo et al.'s probabilistic approach, DISTRAS [33], served partly as basis for our generator.

The ASTG semantic trajectory generation process consists of 5 steps illustrated in Fig. 2. First, the survey data is being preprocessed. This step includes primarily filtering out inconsistent, illogical and empty entries. Additionally, single location and activity entries are clustered into a set of main categories as described in Section III and shown in Table I. Lastly, the filtered survey data, which contain partially free text, are formatted in the form of JSON objects for further processing in the following steps.

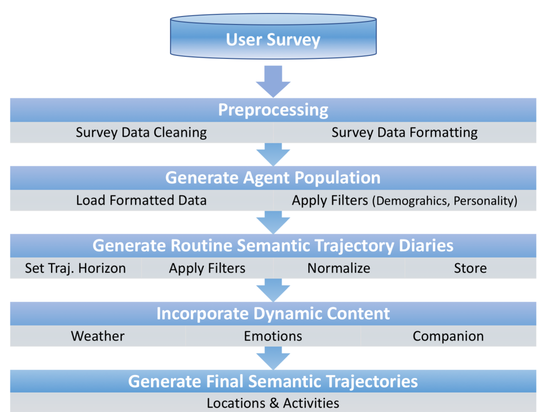

Fig. 2. ASTG process layout diagram.

The next step involves the generation of the agent population. The number of the agents and the agents themselves as well as their characteristics can be either configured manually, or generated automatically based on the corresponding distributions derived from the survey data, that is, based on the demographic and the personality distributions lying within. The manual configuration can be easily achieved by setting and applying one or more *filters* for the generation of just a single or more than one agent groups. So for instance it is possible to produce a group of agents that are 26-35 years old, full-time employees, extrovert and live alone. The resulting agent groups can be illustrated with a graph as shown in Fig. 3. In the graph, each agent is a node and an edge links agents

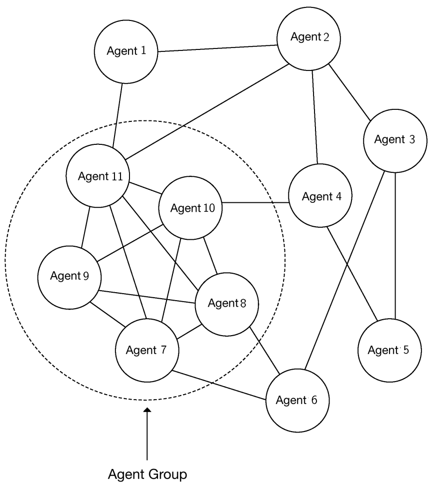

Fig. 3. Agent groups in a graph.

that share the same (more or less) features. In this way, agent groups (e.g., have a car, age between 26-35 and work full time) are represented by highly interconnected and dense parts.

After generating the agents, ASTG produces a set of multi-dimensional semantic diaries based on the afore-generated agents and the corresponding distributions derived from the available 7-day long log diary data. In particular, ASTG creates for each agent a multiple set of diaries, one for each dimension of interest and all their combinations (e.g., semantic location, activity, activity duration, emotions (with regard to location), weather, etc.). For this purpose, it iterates over the formatted survey results and writes the desired information into the diary objects, which in our case are Python Dictionaries. Each diary represents a temporal sequence of which each element contains the feature of interest (e.g., location, activity, ...) and a list of transition probabilities to each other element in the respective diary. Fig. 4 shows an example of a semantic location diary. The activity diary looks similar, only that the IDs would in this case refer to activities instead of locations. The single dimension diaries are then combined to produce the final set of semantic trajectory diaries. The combination process relies on the Naive Bayes Theorem, where all feature dimensions are regarded as independent from each other. Therefore, the final overall transition probability values are

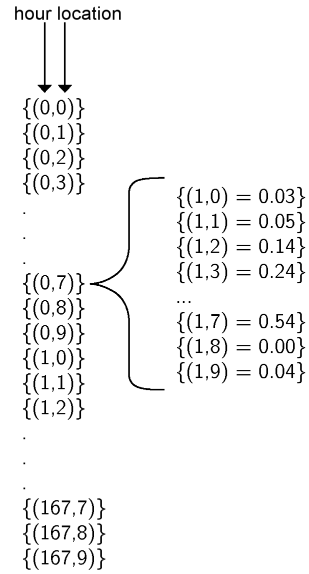

Fig. 4. Example of a 1-week long location diary (7x24=168 hour slots). Each element contains here the time, a location, and a list of next location transition probabilities.

a result of the product of the corresponding single ones. As in the previous step, the final semantic trajectory diaries can be configured by applying filters in terms of which dimension should be regarded and to what extend. For example, one can create a semantic trajectory diary that contains the probabilities for the next Location depending 80% on the time, 15% on the emotions and 5% on the current Location. The length of the generated trajectories (*trajectory horizon*) can also be set manually. In this work, we have set the trajectory horizon and the diary's temporal resolution to 1 week and 1 hour accordingly. This step is mainly responsible for generating basic regular movement behaviours based on the survey data.

In the following step, ASTG focuses on the irregular behaviour potential by taking dynamic data, such as the emotional state, the transportation means, the existence of companion and the weather, explicitly into consideration. Same as in the steps before, all above features can be either configured manually, or generated automatically based on the respective distributions found in the survey data or completely randomly. For instance, one could pass on the weather forecast for the next couple of days to the ASTG manually and get a set of synthetic trajectories for the coming time. Importing dynamic data into the generator leads to certain feedback effects when it comes to producing or choosing the next location of the semantic trajectory. This can be easily illustrated in the case of emotions, which influence both the current location as well as the people's choice of the location to be visited next.

Finally, in the last step, ASTG uses the resulting semantic trajectory diaries to generate the final semantic trajectories. By default, the generated trajectories set the initial location and time as "home" and "Monday 00:00am" respectively. Afterwards, the next location is selected based on the distribution in the available diaries. Then, it uses the same method for choosing the next location based on the result before and so on until the desired trajectory length is reached. At each iteration, ASTG starts by crawling the full-dimensional diaries. But a sparse survey dataset would lead to a set of sparse, lower-dimensional semantic trajectory diaries (e.g., some emotions in relation to certain location types may be missing, etc.). This makes the determination of the distribution of certain features of interest highly difficult or even impossible, which in turn would lead to transition probabilities of 0%. If this is the case and the selection of a next location is impossible, ASTG tries to use a lower-dimensional semantic trajectory diary, which contains one feature less. This is done until it doesn't fail to return a next location. After a next location is found, a matching activity will be chosen depending on both the produced (next) location and the old one. Then, after a matching activity is selected, ASTG determines the (temporal) length of the activity. A new activity will be chosen when either the old activity is expired or there is a new location. The described procedure runs as many times as the number of agents that was defined at the very beginning. The final output of our ASTG generator can be either completely random, or based on the within the survey underlying distributions, or target-oriented by applying a series of filters, e.g. a 1-week long set of trajectories for an extrovert happy man, between 18 and 25, who lives in a big city alone and works full-time in a stormy week.

VI. Evaluation

We evaluated our approach and hence the impact of personality and emotions on the generation of representative semantic trajectories in a number of ways with respect to the following aspects:

- Purpose of trip (e.g., shopping, entertainment, etc.)
- Number of trips per person
- Trip length per day
- Distribution of locations
- Distribution of activities
- Average stay time at locations
- Average activity duration
- Average daily patterns during the week
- Average daily patterns during the weekend

We compared the ASTG generated set of semantic trajectories from 1000 agents over a period of 1 week (trajectory horizon) with the statistical analysis data from the MiD2008[4] study conducted in Germany by the German Ministry of Transport, Building and Urban Affairs in 2008 as well as with the outcome of an activity-based travel survey with over 30.000 participants conducted in Chicago also in the period of 2008 [1]. The MiD2008 survey includes the data of 25.000 households across Germany. In addition, similar to other diary-based related works, we compare the generators outcome with the initial surveyed dataset to verify its correctness.

[4]http://www.mobilitaet-in-deutschland.de/mid2008-publikationen.html

During our evaluation, we tested the impact of all possible feature combinations, that is, we compared the result based on every possible semantic trajectory diary, beginning with the 1-dimensional diaries, where only one feature was considered (e.g., location, activity, ...) and ending with the diaries with all features taken into account. Due to space limitations, the results and plots presented in this section refer mainly to the best configuration. Generally, it could be shown that 1-dimensional diaries lead to the worst overall results. Solely the time as feature showed some reasonable results with regard to the daily trip length. The generators based on 2-dimensional diaries yield partly better results, with the generator that takes both the current location and the time into consideration performing best in terms of trip length and average stay time. It is interesting to note here, that some of the 2-dimensional diaries led to worse results than some of the 1-dimensional ones. This could be attributed to the sparse survey dataset that served as a basis for the generator and does not cover every possible combination. For instance, adding the feature of transportation mode (a piece of information that appears rarely in the survey) has an extremely negative effect on the generated trajectories. The overall best results were achieved by the generator that takes *(current location, time, emotions, weather and companionship)* into consideration. It achieves an average trip length of 3.18 locations compared to 3.10 found in the survey. The results below refer primarily to this particular generator.

Fig. 5 and Fig. 6 show the location distribution and the respective stay times of the generated trajectories by the best ASTG variant in comparison to the survey values. We can see

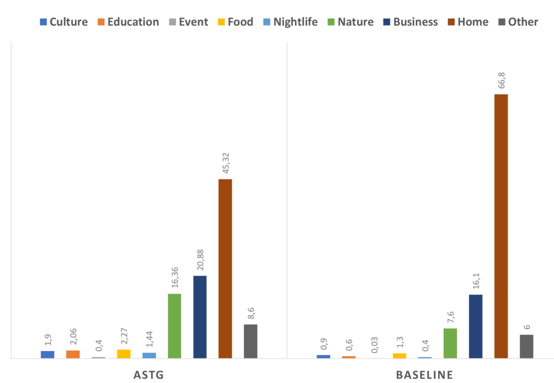

Fig. 5. Location distribution over a period of 1 week for the case of the 5-dim ATSG (1000 agents, (location, time, emotions, weather, companionship)) and the survey values.

that both distributions strongly resemble one another. Solely in the case of the *Travel & Transport* locations is a noticeable difference in terms of stay duration to identify. Fig. 7 compares the location distribution of ASTG with the one of the MiD2008 survey with respect to the basic location types of the latter. The generated patterns of ASTG matches the trend of the MiD2008 outcome very well. However, we can see a slight discrepancy when it comes to how often working locations are visited.

In Fig. 8 and Fig. 9, we compare the location and stay time

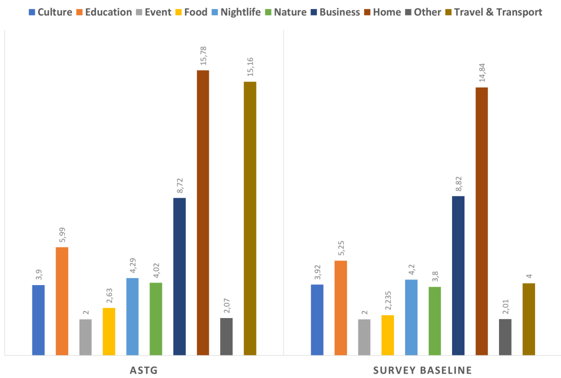

Fig. 6. Location stay time distribution over a period of 1 week for the case of the 5-dim ATSG (1000 agents, (location, time, emotions, weather, companionship)) and the survey values.

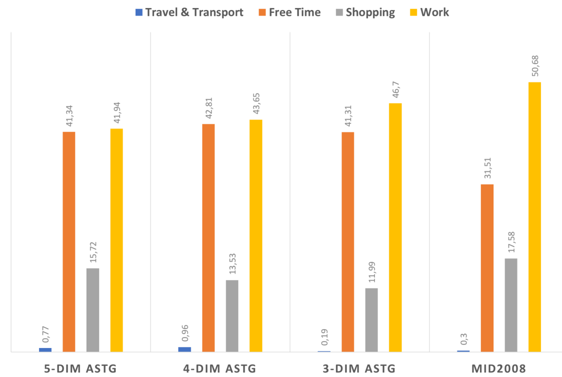

Fig. 8. Location distribution over a period of 1 week for the case of 5-dim ATSG (1000 agents, (location, time, emotions, weather, companionship)), 4-dim ATSG (1000 agents, (location, time, emotions, weather)), 3-dim ATSG (1000 agents, (location, emotions, companionship)) and the MiD2008 survey.

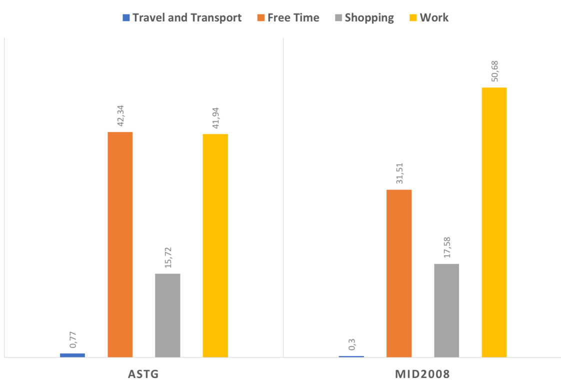

Fig. 7. Location distribution over a period of 1 week for the case of the 5-dim ATSG (1000 agents, (location, time, emotions, weather, companionship)) and the MiD2008 survey.

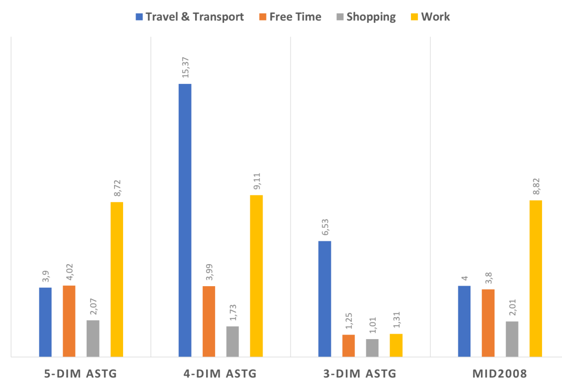

Fig. 9. Location stay time distribution over a period of 1 week for the case of 5-dim ATSG (1000 agents, (location, time, emotions, weather, companionship)), 4-dim ATSG (1000 agents, (location, time, emotions, weather)), 3-dim ATSG (1000 agents, (location, emotions, companionship)) and the MiD2008 survey.

distribution of our (5-dimensional) ASTG and the MiD2008 survey with the ones of the best 3- and 4-dimensional ASTG versions. The number of dimensions here refers to the number of features taken into consideration by each generator as we show in Section V. The best 4-dimensional generator displayed in the respective figures considers *(current location, time, weather and emotions)*, while the best 3-dimensional generator uses a (location, emotions and companionship)-diary as basis. We can see that with regard to location distribution all generators perform well. The produced patterns are very similar to the ones found in the MiD2008 survey. However, the same doesn't hold in terms of stay times, despite the fact that the 4-dimensional generator takes *time* also into account. The 5-dimensional ASTG does a much better job in producing stay times in contrast to the the lower dimensional generators. This shows how valuable heterogenous information can be when it comes to (re)producing semantic trajectories. Furthermore, this provides a strong indication that both weather and emotions play a significant role when it comes to estimating temporal behaviour, a fact that underpins our initial expectations.

Fig. 10 and Fig. 11 displays the activity distribution, as well as the corresponding activity durations. While ASTG performs almost perfectly in the first case, the generation of location sequences, with respect to activity it seems to have it slightly more difficult when it comes to generating temporal information such as stay times and activity durations. Because of the fact that different activities may take place on the same location and vice versa, as observed in the survey, activity sequences are much harder to reproduce by a generator that takes the current location into account, as ASTG. This might partly explain the respective irregularities in the figure.

In Fig. 12, 13 and 14, we can see the average daily activity pattern produced by our ASTG compared with the initial survey values and the MiD2008 survey outcome. As shown in these figures, we investigated additionally the performance of a primarily full-time working agent population to a primarily half-time working agent population. We can see that the starting points of the synthetically produced patterns in both cases (full-time and half-time) are slightly shifted to the right. This matches to the also shifted initial survey pattern explaining in this way the shifting effect of the former two patterns (compared to the MiD2008 pattern). All in all, all daily patterns show strong similarities, with the activity *working* to perform at best. The worst result belongs to the *shopping*

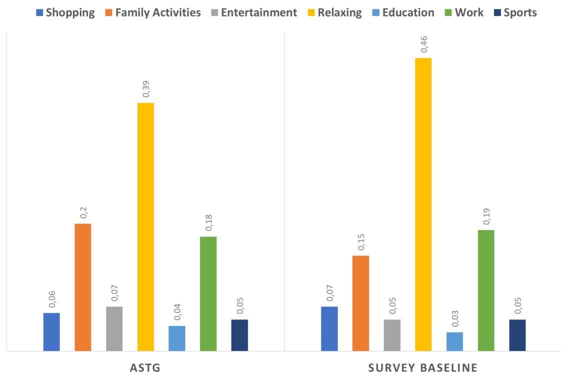

Fig. 10. Activity distribution over a period of 1 week for the case of the 5-dim ATSG (1000 agents, (location, time, emotions, weather, companionship)) and the MiD2008 survey.

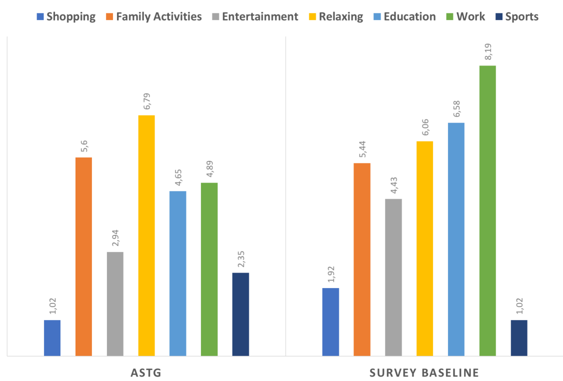

Fig. 11. Activity duration distribution over a period of 1 week for the case of 5-dim ATSG (1000 agents, (location, time, emotions, weather, companionship)) and the MiD2008 survey.

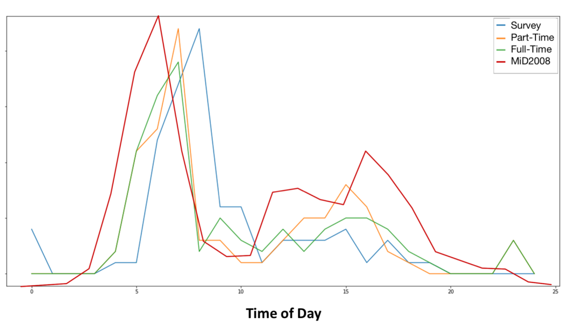

Fig. 12. *Average daily activity* = *Work*-pattern for the case of the 5-dim ATSG (1000 agents, (location, time, emotions, weather, companionship)), the initial survey values and the MiD2008 survey.

activity. Although the starting points fit almost perfectly, the first half of the generated daily patterns is very irregular and far away from the MiD2008 baseline. The second part is much better though.

Fig. 16 shows the average daily activity patterns produced by 1000 agents by ASTG over a period of one week. If we compare it with the results of the 30.000 participants Chicago survey of 2008 [1] shown in Fig. 15, we see that they are almost identical. Fig. 17 and 18 show the average

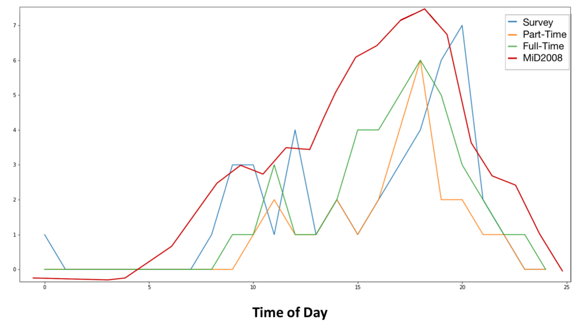

Fig. 13. *Average daily activity* = *Free time*-pattern for the case of the 5-dim ATSG (1000 agents, (location, time, emotions, weather, companionship)), the initial survey values and the MiD2008 survey.

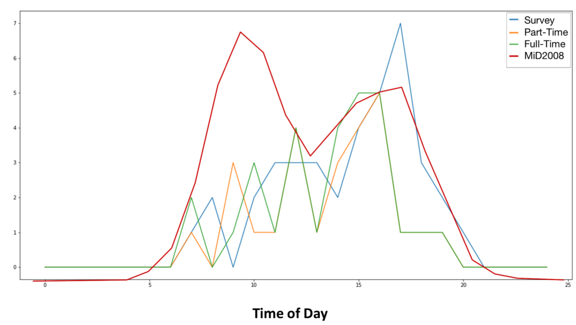

Fig. 14. *Average daily activity* = *Shopping*-pattern for the case of the 5-dim ATSG (1000 agents, (location, time, emotions, weather, companionship)), the initial survey values and the MiD2008 survey.

daily patterns during the weekends. Here, the differences are larger, although the basic patterns are similar. Especially the *home* and the *free time* activities show a strong resemblance, which is to expect during the weekends. The fact that the rest generated activity patterns don't reflect the Chicago survey data could be explained by the sparsity of our initial survey dataset. Additionally, a certain degree of mismatch might be attributed to cultural differences, since our generator is based on data collected in Germany, in contrast to the results from the Chicago study. For instance, if we take a look at Fig. 18, our agent population prefers to stay at home more often than in Chicago (Fig. 17), where people seem to be more active and tend to participate more often in outdoor activities.

Lastly, we also compared several agent groups with each other based on their characteristics and how these affect the generator. The agent group with the characteristics *(full-time occupation, owns car, age 26-35, personality trait: openness)* stood out and showed the overall best results. This can be explained mainly by the fact that a full-time occupation constitutes a very strong indicator for regularity in movement patterns. At the same time, openness could be interpreted here as a disposition to often going out and thus can also be related to a higher occurrence of certain trajectory patterns. The rest of the groups showed roughly similar results.

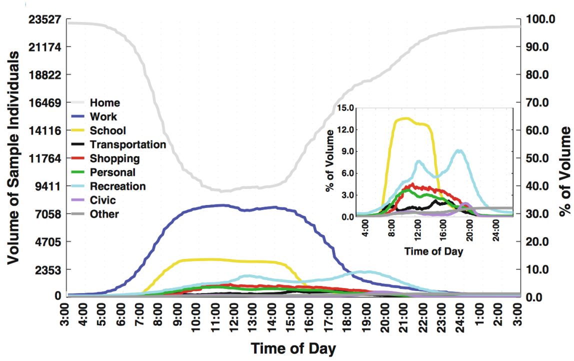

Fig. 15. Average daily activity patterns during the week (2008 Chicago survey, [1]). The inset figure is a zoom in of the middle of the picture.

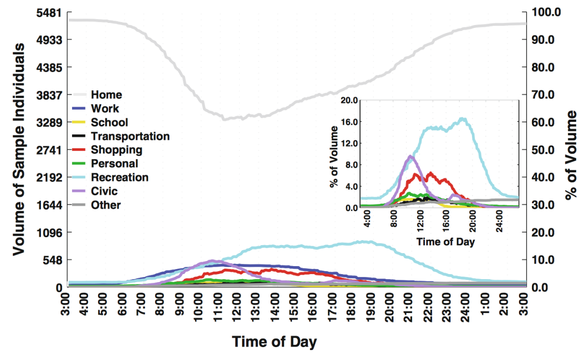

Fig. 17. Average daily activity patterns during the weekend (2008 Chicago survey, [1]). The inset figure is a zoom in of the middle of the picture.

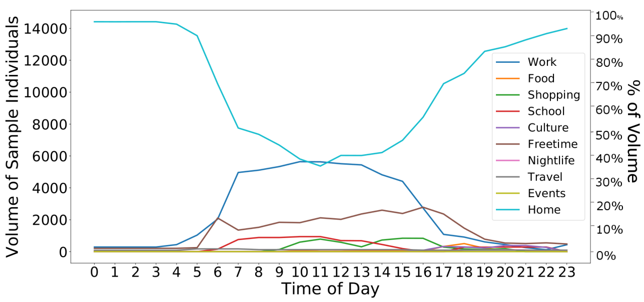

Fig. 16. Average daily patterns during the week for the case of the 5-dim ATSG (1000 agents, (location, time, emotions, weather, companionship)).

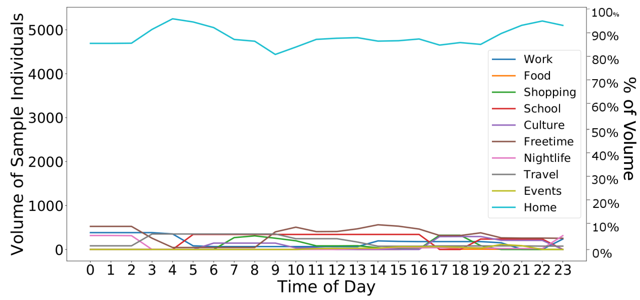

Fig. 18. ASTG process layout diagram.

VII. Conclusion

Goal of the presented work is to explore the impact of emotions and personality on the generation of synthetic semantic trajectories. For this purpose, we designed and implemented a multi-agent log diary-based generator that takes these information explicitly into account. We refer to it as *Affective Semantic Trajectory Generator (ASTG)*.

Similar to related work, we evaluated our approach with respect to a number of aspects such as the daily trip length, the location and activity distribution, the stay time distribution, the average daily location patterns, to name but a few. We compared our results with the respective outcome of two large-scale mobility surveys conducted in Germany and Chicago in 2008 with over 25.000 participants each. All in all, we can say that incorporating dynamic data into the semantic trajectory process can lead to a better performance. Especially the consideration of emotions seems to have a significant impact on the produced trajectories and their representativity. We could also observe the fact that the more features we considered, the better for the generator, with time being the most important of all.

However, we could also identify some limitations in the form of discrepancies between the generated and the real trajectory data. These can be mainly attributed to the small size of the initial 7-week long survey dataset that served as basis for our generator. A bigger dataset would probably lead to better and more accurate results.

For the future, we plan to investigate other modelling techniques, such as deep neural networks and dynamic Bayes networks. Moreover, in a next step, we aim at working with GPS data and at automating the semantic enrichment process.

Acknowledgment

The authors would like to thank Imanuel Richter for assisting in the design, conduction and evaluation of our online user survey.

References

[1] 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.

[2] J.-M. Saglio and J. Moreira, "Oporto: A realistic scenario generator for moving objects," *GeoInformatica*, vol. 5, no. 1, pp. 71–93, 2001.

[3] D. Pfoser and Y. Theodoridis, "Generating semantics-based trajectories of moving objects," *Computers, Environment and Urban Systems*, vol. 27, no. 3, pp. 243–263, 2003.

[4] T. Brinkhoff, "A framework for generating network-based moving objects," *GeoInformatica*, vol. 6, no. 2, pp. 153–180, Jun 2002. [Online]. Available: https://doi.org/10.1023/A:1015231126594

[5] C. Düntgen, T. Behr, and R. H. Güting, "Berlinmod: a benchmark for moving object databases," *The VLDB Journal?The International Journal on Very Large Data Bases*, vol. 18, no. 6, pp. 1335–1368, 2009.

[6] F. Giannotti, A. Mazzoni, S. Puntoni, and C. Renso, "Synthetic generation of cellular network positioning data," in *GIS*, 2005.

[7] M. F. Mokbel, L. Alarabi, J. Bao, A. Eldawy, A. Magdy, M. Sarwat, E. Waytas, and S. Yackel, “Mntg: An extensible web-based traffic generator,” in *Advances in Spatial and Temporal Databases*, M. A. Nascimento, T. Sellis, R. Cheng, J. Sander, Y. Zheng, H.-P. Kriegel, M. Renz, and C. Sengstock, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 38–55.

[8] N. Pelekis, C. Ntrigkogias, P. Tampakis, S. Sideridis, and Y. Theodoridis, “Hermoupolis: a trajectory generator for simulating generalized mobility patterns,” in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2013, pp. 659–662.

[9] C. Parent, S. Spaccapietra, C. Renso, G. Andrienko, N. Andrienko, V. Bogorny, M. L. Damiani, A. Gkoulalas-Divanis, J. Macedo, N. Pelekis *et al.*, “Semantic trajectories modeling and analysis,” *ACM Computing Surveys (CSUR)*, vol. 45, no. 4, p. 42, 2013.

[10] N. Pelekis, S. Sideridis, P. Tampakis, and Y. Theodoridis, “Hermoupolis: a semantic trajectory generator in the data science era,” *SIGSPATIAL Special*, vol. 7, no. 1, pp. 19–26, 2015.

[11] 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, p. 22.

[12] A. Karatzoglou, S. C. Lamp, and M. Beigl, “Matrix factorization on semantic trajectories for predicting future semantic locations,” in *Wireless and Mobile Computing, Networking and Communications (WiMob),*. IEEE, 2017, pp. 1–7.

[13] A. Karatzoglou, H. Sentürk, A. Jablonski, and M. Beigl, “Applying artificial neural networks on two-layer semantic trajectories for predicting the next semantic location,” in *International Conference on Artificial Neural Networks*. Springer, 2017, pp. 233–241.

[14] A. Karatzoglou, D. Köhler, and M. Beigl, “Purpose-of-visit-driven semantic similarity analysis on semantic trajectories for enhancing the future location prediction,” in *International Conference on Pervasive Computing and Communications Workshops (PerCom)*. IEEE, 2018.

[15] M. R. Eric Miller, “Prototype model of household activity-travel scheduling,” in *TASHA*, 2002. [Online]. Available: https://doi.org/10.3141/1831-13

[16] G. Gidofalvi and T. B. Pedersen, “St–acts: a spatio-temporal activity simulator,” in *Proceedings of the 14th annual ACM international symposium on Advances in geographic information systems*. ACM, 2006, pp. 155–162.

[17] T. D. Doudali, I. Konstantinou, and N. Koziris, “Spaten: A spatio-temporal and textual big data generator,” in *2017 IEEE International Conference on Big Data (Big Data)*, Dec 2017, pp. 3416–3421.

[18] D. Y. Kim and H. Y. Song, “Method of predicting human mobility patterns using deep learning,” *Neurocomputing*, vol. 280, pp. 56 – 64, 2018, applications of Neural Modeling in the new era for data and IT. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925231217317691

[19] J. M. Kanter and K. Veeramachaneni, “Deep feature synthesis: Towards automating data science endeavors,” in *2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, Oct 2015, pp. 1–10.

[20] I. Ajzen and M. Fishbein, “Understanding attitudes and predicting social behaviour,” 1980.

[21] I. Ajzen, “The theory of planned behavior,” *Organizational Behavior and Human Decision Processes*, vol. 50, no. 2, pp. 179 – 211, 1991, theories of Cognitive Self-Regulation. [Online]. Available: http://www.sciencedirect.com/science/article/pii/074959789190020T

[22] M. Hunecke, S. Haustein, S. Grischkat, and S. Böhler, “Psychological, sociodemographic, and infrastructural factors as determinants of ecological impact caused by mobility behavior1,” *Journal of Environmental Psychology*, vol. 27, no. 4, pp. 277–292, 2007.

[23] B. S. Flynn, G. S. Dana, J. Sears, and L. Aultman-Hall, “Weather factor impacts on commuting to work by bicycle,” *Preventive medicine*, vol. 54, no. 2, pp. 122–124, 2012.

[24] S. Saneinejad, M. J. Roorda, and C. Kennedy, “Modelling the impact of weather conditions on active transportation travel behaviour,” *Transportation research part D: transport and environment*, vol. 17, no. 2, pp. 129–137, 2012.

[25] R. E. Rhodes and K. S. Courneya, “Relationships between personality, an extended theory of planned behaviour model and exercise behaviour,” *British Journal of Health Psychology*, vol. 8, no. 1, pp. 19–36, 2003.

[26] S. Y. Kim, H. J. Koo, and H. Y. Song, “A study on influence of human personality to location selection,” *Journal of Ambient Intelligence and Humanized Computing*, vol. 7, no. 2, pp. 267–285, 2016.

[27] ——, “A study on estimation of human personality from location visiting preference,” *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–14, 2017.

[28] M. Macht and G. Simons, “Emotions and eating in everyday life,” *Appetite*, vol. 35, no. 1, pp. 65 – 71, 2000. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0195666300903258

[29] J. Forgas, “Mood effects on decision making strateges,” vol. 41, pp. 197 – 214, 02 2011.

[30] S. Spaccapietra, C. Parent, M. L. Damiani, J. A. de Macedo, F. Porto, and C. Vangenot, “A conceptual view on trajectories,” *Data & Knowledge Engineering*, vol. 65, no. 1, pp. 126–146, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0169023X07002078

[31] L. R. Goldberg, “An alternative description of personalit: the big-five factor structure.” *Journal of personality and social psychology*, vol. 59, no. 6, p. 1216, 1990.

[32] B. Rammstedt and O. P. John, “Measuring personality in one minute or less: A 10-item short version of the big five inventory in english and german,” *Journal of research in Personality*, vol. 41, no. 1, pp. 203–212, 2007.

[33] L. Pappalardo and F. Simini, “Data-driven generation of spatio-temporal routines in human mobility,” *Data Mining and Knowledge Discovery*, pp. 1–43, 2017.