Traj2User: exploiting embeddings for computing similarity of users mobile behavior

Andrea Esuli[1], Lucas May Petry[2], Chiara Renso[1], and Vania Bogorny[2]

[1]ISTI-CNR, Pisa, Italy
[2]Programa de Pós-Graduação em Ciência da Computação, Universidade Federal de Santa Catarina, Florianopolis, Brazil

Abstract

Semantic trajectories are high level representations of user movements where several aspects related to the movement context are represented as heterogeneous textual labels. With the objective of finding a meaningful similarity measure for semantically enriched trajectories, we propose Traj2User, a Word2Vec-inspired method for the generation of a vector representation of user movements as *user embeddings*.

Traj2User uses simple representations of trajectories and delegates the definition of the similarity model to the learning process of the network. Preliminary results show that Traj2User is able to generate effective user embeddings.

1 Introduction

The widespread use of GPS-equipped smartphones or positioning sensors applied to vehicles and animals, tend to produce a high number of trajectories, recording the spatio-temporal evolution of these objects. These raw trajectories can be enriched with semantic information to what is called *semantic trajectories* [15, 13], adding more meaning to the pure geometric movement facets.

In the era of Big Data, with the explosion of geolocated social media and other kinds of user generated data (e.g. Wikipedia, Flickr, etc), human mobility data can be significantly enriched with information that encompasses our daily life. Enriching information include weather conditions, the transportation means, the goal or the activity performed during the movement, the opinions and comments about people and places, the mood, being with a friend, just to name a few examples discussed in [6] and [2].

Being able to find similarities between trajectories enables several analysis methods like clustering or applications like recommendation systems. Several similarity measures have been proposed for both raw and semantic trajectories,
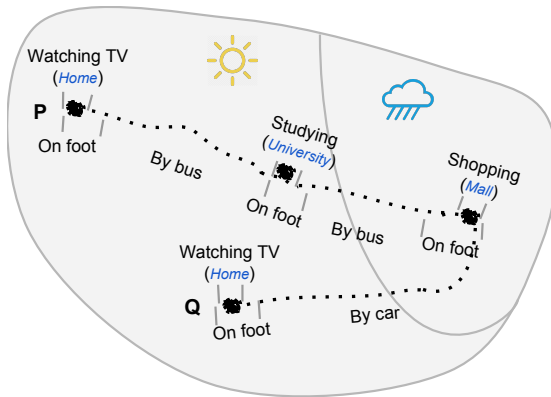
Figure 1: Example of two semantically rich trajectories P and Q.

such as EDR[4], LCSS [17], Cats [9], EDWp [14], UMS[7], MSM [8], and other approaches as presented in [1], [11] and [18]. However, these methods analyze a few trajectory attributes and are far from considering all the different semantic aspects that involve movement: previous works have mainly analyzed semantic trajectories over one single aspect at a time, such as stops and moves, or transportation means, or activities.

The great challenge here is how to integrate all such heterogeneous dimensions in a similarity measure dealing with space, time, and multiple semantics.

Figure 1 shows an example of two semantic trajectories P and Q from users u_1 and u_2. The question we want to answer is "*how similar are u_1 and u_2 given their semantic trajectories P and Q?*".

As can be observed, first of all, the size of trajectories P and Q is different. Both users are *shopping* at the same place (same spatial location) while *on foot* (same transportation means) and when the weather is rainy (same weather condition). On the other hand, both users at some time (time dimension) watch TV, but on a different spatial location. Indeed, while the user of trajectory P moves on foot and by bus, the user of trajectory Q moves on foot and by car.

Given all these different and heterogeneous data dimensions related to trajectories, where each dimension has its own similarity model, how can we compute the user similarity based on their trajectories P and Q considering all this information together?

In this paper we represent users by modeling the many semantic aspects that describe their movement habits. We take inspiration from *word embeddings* [12] methods that model the semantics of a word, and its similarity with other words, by observing the many contexts of use of the word in the language.

We therefore introduce here, for the first time, the new concept of *user embeddings* as a way to represent the semantically rich movements of a user. We also propose the TRAJ2USER model for measuring the semantic trajectory similarity of moving users. The main contribution of TRAJ2USER is that it does neither need any explicit definition of the similarity functions for the data dimensions nor any explicit modeling of the relations between data dimensions,

since these are implicitly learned from data.

Experiments show that the neural model of TRAJ2USER learns user embeddings that better capture user similarity than other embedding methods that are also inspired to language models. We believe that exploiting vectors to represent semantically complex movements is promising for the analysis of semantically enriched movement since several heterogeneous semantic aspects are uniformly modeled into a unique and compact form. One recent approach that considers Point of Interest (POI) and embeddings is the paper [5] where authors propose a method to jointly model the user preference and the sequence of POIs for predicting future visitors for a given POI.

2 Methodology

2.1 Basic definitions

We start our definitions with a state-of-art concept of raw trajectories, that has only space and time dimensions.

Definition (Raw trajectory) A raw trajectory is the sequence of timestamped locations of the traced moving object o in the form $< o, < x_1, y_1, t_1, \ldots x_n, y_n, t_n >>$, where each x_i, y_i represents the geographical coordinates and t_i the timestamp for each $i = 1, \ldots, n$.

In this paper we define a semantic trajectory based on trajectory segments (or movements).

Definition (Segmented Trajectory) A segmented trajectory of the moving object o is a pair $t = < o, s_1, ..., s_n >$ such that, for each $i = 1, ..., n$, s_i is a contiguous part of the trajectory split based on some criteria.

Examples of segmenting criteria are the stops and moves [15], the transportation means or the purpose of the trip [2].

Definition (Semantic Trajectory) A semantic trajectory of an object o is a pair $t = < o, \{< s_1, l_{11}, ..l_{1k} >, ..., < s_n, l_{n1}, ...l_{nk} >\} >$ such that, for each i=1,…,n and j=1, …, k l_{ij} is the j-th semantic label for the segment s_i.

A semantic trajectory, in the context of this work, is a segmented trajectory where each segment is enriched with a number of different labels that represent different semantic aspects of the trajectory segment. Considering again Figure 1, we see that the first trajectory has 5 segments while the second has 3 segments. The semantic representation of the first trajectory is $P = $ <u$_1$, <s$_1$, "on foot", "watching TV","sunny">, <s$_2$, "by bus, "going to University", "sunny">, <s$_3$, "on foot", "studying", "sunny">, <s$_4$, "by bus", "going shopping", "rainy">, <s$_5$, "on foot", "shopping", "rainy">>

2.2 Building the user embeddings

In this work we want to build vectorial representations in an embedding space, for each user, in a population of observed individuals. Each user is represented by a variable-sized set of semantically rich movements. We can consider each

3

movement as a context of the user's life. We observe that this situation is similar to the one in language modeling [16, 10], where one wants to model the semantic properties of a word and measure the semantic similarity between words by observing the many contexts in which the word appears. The intuition of this work is based on the observation that the similarity between two words (e.g., king, queen) can be inferred by observing that they frequently appear in similar contexts (e.g., near other words such as castle, crown, empire, throne...). Analogously, the similarity between two users can be inferred by observing that their semantic trajectories frequently have similar semantic values.

Following this parallel between words and users, we make here a first exploration of how some methods for the construction of word embeddings can be applied to the process of constructing *user embeddings*. In our approach we convert the label of each segment of a semantic trajectory into a vectorial representation (i.e., *a movement descriptor* as defined in Section 2.2.1). We then combine the various movement descriptors of a user into a single *user embedding*. We test three main models for the construction of user embeddings, which are described in Section 2.2.2.

2.2.1 Encoding of trajectories into vectors

The various labels of a semantic trajectory are heterogeneous by nature and therefore they may have very different forms. In this work we decided to build a *movement descriptor* by encoding all the values of the segment labels using a *one-hot encoding*. Given a label a with n^a possible different values, an actual value l^a for the label is encoded as a n^a-sized vector with a *one* in the position corresponding to the value l^a and $n^a - 1$ *zeros* for all the other possible values. The result of encoding a movement descriptor is thus a vector d of length $|d| = \sum_{a \in A} n^a$, where A is the set of labels.

We are aware that one-hot encoding does not explicitly model any complex relation between values, e.g., values on ordinal scales. Our approach follows the idea of leaving to the embedding generation method the burden of discovering the relations between the values of a label (and across labels).

2.2.2 Generating user embeddings

The set of vectors representing movements and associated to a user must be reduced into a single vector that represents the user embedding. A very simple approach is to sum all the vectors into a single vector, i.e., $e_i = \sum_{d_i \in D_i} d^i$, where D^i identifies all the descriptors associated to the user u_i (SUM method).

Stacking all the resulting sum vectors for all users produces a matrix M of size $|U| \cdot |d|$. This is similar to what is done when building a *word-context* matrix to create a language model, in which the position $T_{i,j}$ of a matrix T of size $|V| \cdot |V|$ stores the sums up of how many times the word $w_j \in V$ appears nearby (in the context of) the word $w_i \in V$, where V is a vocabulary.

Vectors deriving from raw sums can suffer skewness due to some values being much more frequent than others (e.g., weekday in our dataset). Such

4

very frequent values may dominate the directions of vectors and at the same time they may not be very discriminative.

Inspired by the approaches used in language modeling to address this issue, we tested a number of methods[1] to produce a weight-corrected matrix $\hat{M}$ of user embeddings starting from the *Sum* matrix[2] M:

Positive point-wise mutual information (PPMI) [3] measures the dependence between the two observed variables (a user and a specific attribute value) only on the occurrence of events, i.e., how much the probability of the two events to occur together differs from chance[3].

$$\hat{M}_{(i,j)} = PPMI(u_i, d_j) = max(\log_2(\frac{P(u_i, d_j)}{P(u_i)P(d_j)}), 0) \tag{1}$$

Softmax (SM) is often used to normalize a vector so that all its values are in the $[0, 1]$ interval and their sum is one, i.e., a categorical probability distribution.

$$\hat{M}_{(i,j)} = \sigma(M_{(i,j)}) = \frac{e^{M_{(i,j)}}}{\sum_{k=1}^{|d|} e^{M_{(i,k)}}} \tag{2}$$

None of the methods described so far actually explores the latent correlations between the values of the labels, within a label or among labels.

Methods based on matrix decomposition, such as singular value decomposition (SVD) may exploit these latent relations modeling them in a project space that is not constrained to the original encoding of values. Truncated SVD allows to generate shorter embeddings, possibly removing noise components from input data. We tested SVD in combination with the previously described methods (i.e., SVD-PPMI,

SVD-SM) with all the above listed methods, and using various reduction factors f, where the resulting user embedding length is $|d|/f$. In this paper we propose the TRAJ2USER method that is inspired to the WORD2VEC [12] method for the generation of word embeddings. The skip-gram variant of WORD2VEC learns word embeddings as a by-product of training a two-layer network on the task of predicting from a single word other words that may appear in its context.

In the TRAJ2USER network[4], the role of the input word is taken by a user id, and the context by a movement descriptor. Given a user u_i represented as a one-hot vector and one movement descriptors d_j^i, the first layer of the network selects the user embedding $e_i = W^T u_i$, i.e., the matrix W is the matrix of user embeddings. The second layer multiplies e_i to a second weight matrix W' and applies the sigmoid activation function[5] $S(x) = \frac{1}{1-e^{-x}}$ to predict the movement

---

[1]We also tested simple l^1 and l^2 normalization, with negative results.

[2]We also tested the application of such weighting methods to the distinct blocks of the matrix M that identify the values from a single label, again with negative results.

[3]A small dataset, like our, can generate negative values due to lack of a sufficient number of observations. For this reason negative values are clipped to zero.

[4]The WORD2VEC model uses many tricks, such as hierarchical softmax and negative sampling, to improve its performance. In this work we explore a "simple" model, leaving the evaluation of these methods to future work.

[5]We tested also the Softmax function, with negative results.

Label	Values	N.
Purpose	at home, at work, eat out, shopping (food), shopping (other), social events, study, entertainment (sport, theater, museum...), services (doctor, bank, hairdresser...), picking up/taking someone home, fueling, stop (changing transportation), not specified, incomplete tracking (dead battery, app crash)	14
Vehicle	car, bicycle, motorcycle, city public transport (but, metro), taxi, train, boat, on foot, not specified	9
Start hour	0-23	24
End hour	0-23	24
Duration	<5 min, 5-8 min, 8-12 min, 12-20 min, >20 min	5
Range	<1 km, 1 to 2 km, 2 to 4 km, 4 to 10 km, > 10 km	5
Weather	sunny, rain, fog, cloudy, not specified	5
Weekday	weekday, weekend	2
	Total count	88

Table 1: Attributes describing trajectories in the dataset.

descriptor $\tilde{d} = S(W'^T e_i)$. $\tilde{d}$ is compared to d_j^i so that backpropagation updates W' and W.

The length of user embeddings $|e_i|$ is a free parameter of the model. Differently from the previous methods, TRAJ2USER makes possible to set $|e_i| > |d|$, i.e., expanding the representation space. The idea supporting the use of a representation space bigger than the original descriptors, is that it may be able to capture and represent more relationships between the attributes.

3 Experiments

3.1 Dataset

The raw trajectories dataset has been collected by volunteers using GPS-enabled smartphones in the area of Pisa in the period from May 20, 2014 to September 30, 2014, in the context of the TagMyDay[6] experiment. Each user tracked his/her movements by using a GPS tracking *app* installed in their mobile phone. Each volunteer could freely decide which part of their daily movement to track by starting the application. Each trajectory has been uploaded into the experiment web site. Here, trajectory segments are automatically computed from the raw data identifying the different movements between two stops. Then, from the web interface the users could annotate these segments with semantic labels like the purpose of the trip, the means of transportation, the weather. Other information has been computed automatically from the raw trajectory like the

---

[6]We obtained this data under a non-disclosure agreement, therefore we cannot directly redistribute it. More information at http://kdd.isti.cnr.it/project/tagmyday

method	compression factor f				
	0.5	1	2	4	8
SUM	n/a	0.748	n/a	n/a	n/a
PPMI	n/a	0.708	n/a	n/a	n/a
SM	n/a	0.378	n/a	n/a	n/a
SVD-PPMI	n/a	0.708	0.706	0.666	0.605
SVD-SM	n/a	0.394	0.394	0.393	0.363
TRAJ2USER	0.858	0.858	0.844	0.803	0.774

Table 2: Comparison of the methods for the generation of user embeddings. Average MRR value across 1,000 test pairs.

temporal duration of the segment, the spatial length and the day of the week (weekday, weekend).

After the annotation task, a semantically enriched trajectory is identified by the attributes listed in Table 1.

The dataset contains traces of 157 users for a total of 10,880 segments. The distribution of the number of trajectories associated to the users follows an exponential decay with the most active user having produced 727 segments and a tail of 39 users with less than 5 segments. The encoding of a trajectory segment into a movement descriptor produces a vector of length $|d| = 88$ (see Table 1).

3.2 Experiments and results

We designed our experiment as a similarity search problem. Given two users u_a and u_b which are known to have very similar mobility habits, we rank all users in $U \setminus \{u_a\}$ by the similarity of their embedding with e_a, using the cosine similarity function, and observe the rank r_b^a of u_b. We repeat this on a large set of pairs of similar users P and measure the mean reciprocal rank $MRR(P) = \frac{1}{|P|} \sum_{(u_a, u_b) \in P} \frac{1}{r_b^a}$ across all pairs. The higher the MRR score the better, since it indicates that the embeddings capture the similarities among the (u_a, u_b) pairs.

Our dataset does not have an explicit evaluation of similarities among users. To solve this issue we created the pairs of similar users by randomly selecting a user u_a from U and then randomly distributing the movement descriptors of u_a between two 'virtual' users u_a' and u_a'', which actually define a test pair. Given a pair (u_a', u_a''), to capture the similarity between these two users we train user embeddings on the set of users $U \cup \{u_a', u_a''\} \setminus \{u_a\}$. We generated a test set of 1,000 pairs using this method. Each training of user embeddings consisted of 1,000 epochs, with shuffling after each epoch.

Results of experiments[7] (Table 2) show that TRAJ2USER outperforms the other tested method by a large margin, i.e., a 14.7% relative improvement over

---

[7]We implemented TRAJ2USER on PyTorch. Upon acceptance we will make the code available on GitHub.
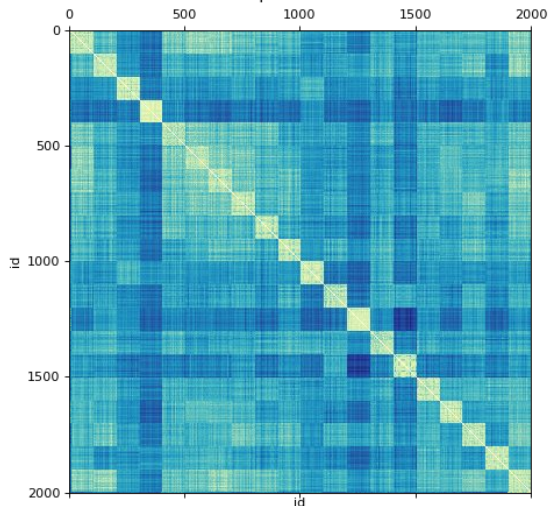
Figure 2: User-by-user similarity modeled by TRAJ2USER on a simulated population with 20 groups, each one of 10 users (the lighter the color the most similar the pair of users).

SUM, the second best performer[8]. An interesting negative result emerges from the comparison of SUM, with PPMI, SM and their SVD methods. These latter methods, that are typically applied with success on language modeling tasks perform poorly on our task. We measured that the distribution of frequency of values in the TagMyDay dataset follow a logarithmic distribution and not the typical Zipf distribution of words in text, yet it is hard to consider this difference as the cause of the drop in MRR. We leave the investigation of this aspect to future work. Softmax-based methods are by far the worst performers, indicating that forcing a probabilistic interpretation of the observed data is a wrong design choice. Shorter embeddings, either from truncated SVD or setting smaller size in TRAJ2USER, reduce the MRR. However, even the shortest TRAJ2USER embedding, with $f = 8$ ($|e_i| = 11$) outperforms SUM, indicating a graceful degradation of performance and the possibility of exploiting data compression. Larger embeddings performed as the original-length ones, indicating that the information contained in the relatively small dataset we were able to obtain was already fully modeled in the original-length embeddings.

We ran a further experiment to check if TRAJ2USER user embeddings are able to discover groups of similar users and to consistently model similarities across groups. We used the method for the creation of virtual users described in this section to create a population of 2,000 users composed of 20 groups of 100 virtual users, each one generated from a real user randomly sampled from the TagMyDay dataset. Figure 2 visualizes the cosine similarity among users

---

[8]The difference is statistically significant for a t-test on the reciprocal rank score across the 1,000 test pair with $p = 3.66 \cdot 10^{-5}$.

measured on TRAJ2USER user embeddings. Virtual users in the same group are adjacent. As shown, TRAJ2USER captures the (imposed) within-group similarity between users (the light blocks on the diagonal), and also consistently models the (casual) inter-group similarities.

4   Conclusions and Future Works

TRAJ2USER is an innovative way to generate effective user embeddings, starting from simple representations of semantic trajectories and delegating the definition of the similarity model to the learning process of the network. Although the TagMyDay dataset is limited to a few labels, the method is general enough to expand the amount and forms of semantic information (e.g. interactions of the user with social platforms like ratings and comments on Foursquare or TripAdvisor). For example, when the information comes in the form of a piece of text, it can be encoded into a semantic-rich vector using neural language models, e.g., paragraph vectors [10]. As future works we plan to make experiments on other datasets publicly available gathered from social media, and to extend the TRAJ2USER network with a recurrent component, so as to model multi-segment trajectories as a single entity.

Acknowledgments

References

[1] Donald J Berndt and James Clifford. Using dynamic time warping to find patterns in time series. In *KDD workshop*, volume 10, pages 359–370. AAAI Press, 1994.

[2] Vania Bogorny, Chiara Renso, Artur Ribeiro de Aquino, Fernando de Lucca Siqueira, and Luis Otavio Alvares. CONSTAnT – a conceptual data model for semantic trajectories of moving objects. *Transactions in GIS*, 18(1):66–68, 2014.

[3] John A. Bullinaria and Joseph P. Levy. Extracting semantic representations from word co-occurrence statistics: A computational study. *Behavior Research Methods*, 39(3):510–526, Aug 2007.

[4] Lei Chen, M. Tamer Ozsu, and Vincent Oria. Robust and fast similarity search for moving object trajectories. In *Proc. of the ACM Inter-*

national conference on Management of Data (SIGMOD), pages 491–502. ACM, 2005.

[5] Shanshan Feng, Gao Cong, Bo An, and Yeow Meng Chee. Poi2vec: Geographical latent representation for predicting future visitors. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA.*, pages 102–108, 2017.

[6] Carlos Andres Ferrero, Luis Otavio Alvares, and Vania Bogorny. Multiple aspect trajectory data analysis: Research challenges and opportunities. In *XVII Brazilian Symposium on Geoinformatics*, GEOINFO '16, pages 1–12, Campos do Jordao, SP, Brazil, 2016. GEOINFO.

[7] Andre Salvaro Furtado, Luis Otavio Campos Alvares, Nikos Pelekis, Yannis Theodoridis, and Vania Bogorny. Unveiling movement uncertainty for robust trajectory similarity analysis. *International Journal of Geographical Information Science*, 32(1):140–168, 2018.

[8] Andre Salvaro et al. FURTADO. Multidimensional similarity measuring for semantic trajectories. *Transactions in GIS*, 20(2):280–298, 2016.

[9] Chih-Chieh Hung, Wen-Chih Peng, and Wang-Chien Lee. Clustering and aggregating clues of trajectories for mining trajectory patterns and routes. *VLDB J.*, 24(2):169–192, 2015.

[10] Quoc V. Le and Tomas Mikolov. Distributed representations of sentences and documents. *CoRR*, abs/1405.4053, 2014.

[11] Hechen Liu and Markus Schneider. Similarity measurement of moving object trajectories. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on GeoStreaming (IWGS '12)*, pages 19–22, New York, NY, USA, 2012. ACM.

[12] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems 26*, pages 3111–3119. Curran Associates, Inc., 2013.

[13] Christine Parent, Stefano Spaccapietra, Chiara Renso, Gennady L. Andrienko, Natalia V. Andrienko, Vania Bogorny, Maria Luisa Damiani, Aris Gkoulalas-Divanis, José Antônio Fernandes de Macêdo, Nikos Pelekis, Yannis Theodoridis, and Zhixian Yan. Semantic trajectories modeling and analysis. *ACM Comput. Surv.*, 45(4):42, 2013.

[14] Sayan Ranu, Deepak P, Aditya D. Telang, Prasad Deshpande, and Sriram Raghavan. Indexing and matching trajectories under inconsistent sampling rates. In *31st IEEE International Conference on Data Engineering, ICDE 2015, Seoul, South Korea, April 13-17, 2015*, pages 999–1010, 2015.

[15] Stefano Spaccapietra, Christine Parent, Maria Luisa Damiani, Jose Antonio de Macedo, Fabio Porto, and Christelle Vangenot. A conceptual view on trajectories. *Data and Knowledge Engineering*, 65(1):126–146, 2008.

[16] Peter D Turney and Patrick Pantel. From frequency to meaning: Vector space models of semantics. *J. of artificial intelligence research*, 37:141–188, 2010.

[17] M. Vlachos, G. Kollios, and D. Gunopulos. Discovering similar multidimensional trajectories. In *Proceedings of the 18th International Conference on Data Engineering*, pages 673–684, San Jose, CA, USA, 2002. IEEE.

[18] Xiangye Xiao, Yu Zheng, Qiong Luo, and Xing Xie. Inferring social ties between users with human location history. *Journal of Ambient Intelligence and Humanized Computing*, 5(1):3–19, 2014.