Annotating Trajectories by Fusing them with Social Media Users' Posts

Ricardo Gil Belther Nabo[1], Renato Fileto[1], Mirco Nanni[2], Chiara Renso[2]

[1]Dept. of Statistics and Informatics (INE), Santa Catarina Federal University (UFSC), University Campus of Trindade, 88040-900, Florianópolis-SC, BRAZIL

[2]Institute of Information Science and Technologies "Alessandro Faedo" (ISTI), National Research Council (CNR), Via G. Moruzzi, 1, S. Cataldo, 56124, Pisa, ITALY

Abstract. *The widespread use of mobile devices allows gathering large amounts of moving objects' trajectories. However, just trajectories are often not enough to enable movements understanding. On the other hand, users' posts in social media can be regarded as sparse and freely annotated movement traces, which can also be collected via mobile devices. This paper proposes a method for automatically fusing trajectories with social media users' posts based on their spatiotemporal compatibility. The results are trajectories annotated with posts contents, that may help to explain movement goals, and relations with places and events, among other information. The proposed method has been implemented and evaluated in experiments with real GPS trajectories and tweets.*

1. Introduction

The popularization of mobile devices equipped with positioning technologies (e.g. GPS navigators, smartphones, tablets) has increased considerable in the recent years. This growth has led to the gathering of large volumes of raw trajectories, i.e., time ordered sequences of spatiotemporal positions of moving objects holding mobile devices. The trajectories collected by using modern devices can have quite accurate spatiotemporal coordinates (e.g., 3 to 30 meters), which are collected in such a rate (e.g., at each second) that allows capturing many movement details. However, such purely spatiotemporal data lacks information (e.g., in the form of textual data) to help understand the movements, such as related places of interest, events, and goals.

Several works have been proposed for trajectories data processing and mining [Spaccapietra et al. 2008, Alvares et al. 2007, Parent et al. 2013, Pelekis and Theodoridis 2014], but spatiotemporal coordinates are not enough to explain movements [Yan et al. 2013, Fileto et al. 2013], making trajectories annotation crucial to realize their information analysis potential. Thus, many solutions have been proposed for trajectories annotation. Nevertheless, these methods have limitations on the characteristics of the annotations produced and/or rely on human labor. The former problem limits the use of the generated annotations. The latter makes the methods unsuitable for daily use with large quantities of trajectories, because annotating is a laborious task, that can easily become tedious for people. On the other hand, the sparse spatiotemporal data available in social media (e.g. Twiter, Facebook), have textual information (e.g., hashtags, comments) that can help describe and analyze trajectories. These data can be regarded as sparse and freely annotated movement traces [Azmandian et al. 2012], and are also frequently collected via mobile devices. Thus, they can be used to annotate trajectories and analyse movements.

This paper proposes a method to annotate raw trajectories by fusing them with social media users data (e.g. tweets, posts in Facebook). It uses as the matching criteria the proximity of each trajectory (or trajectory segment) with sequences of social media users' posts. Our strategy is analogous to some of those that are common practice in many engineering areas to fuse data originated from different kinds of devices [Castanedo et al. 2010]. The resulting trajectories annotated with textual contents of social media posts (hashtags, comments, etc.) can feed semantic enrichment, analysis, and mining processes for extracting useful information of large quantities of spatiotemporal data. In fact, some collections of sparse, geo-referenced, and timestamped textual contents of social media users' posts are already being proved useful to explain goals, places, and events related to movements of people [Fileto et al. 2013, May and Fileto 2014]. The method proposed in this paper aims to ally the virtues of vast collections of trajectories (usually accurate, detailed) and social media posts (having rich textual contents). It has been implemented in a prototype, and evaluated in experiments with real GPS trajectories and tweets, both with geographic coordinates inside the city of Fortaleza Brazil.

The remaining of this paper is organized as follows. Section 2 discusses related works. Section 3 defines some key concepts for understanding the proposal. Section 4 presents the proposed method. Section 5 describes some experiments and their results. Finally, Section 6 concludes the paper, and provides a glimpse of our future work.

2. Related Works

One of the most accepted conceptual models for structuring trajectories is proposed by [Spaccapietra et al. 2008]. According to this model trajectories can be segmented in stops and moves, which are specializations of episodes (Definition 3). It introduces the possibility of interpreting a subsequence of spatiotemporal points (stop or move) as an aggregation with distinguishable characteristics. It abstracts irrelevant details, and allows annotations to be associated with stops and moves instead of trajectory points.

The method proposed by [Alvares et al. 2007] aims to semantically enrich trajectories by mining stops in places of interest (POI) of a given collection, and moves between these stops. Their method allows the efficient calculation of stops and moves, to build the conceptual representation trajectories proposed in [Spaccapietra et al. 2008]. This method, originally called SMoT (Stops and Moves of Trajectories), can also be called IB-SMoT (Intersection-Based SMoT). CB-SMoT (Cluster-Based SMoT) [Xiu-li and Wei-xiang 2009] is another method to mine stops and moves. It aggregates spatiotemporal points of raw trajectories in subsequences that present similar characteristics (e.g., around the same speed). CB-SMoT can identify stops by clustering adjacent positions in which the moving object is stationary or moves slowly, regardless of where they occur. Notwithstanding, IB-SMoT and CB-SMoT produce limited annotations. Both label segments as stops or moves, and IB-SMoT associates each stop with the respective POI of the given collection.

The annotation platform proposed in [Yan et al. 2013] progressively transforms the raw trajectories into semantic trajectories. The trajectory segments are annotated with concepts such as as home and work, or POIs. These annotations are based on predetermined hot spots, and produced by trajectory mining algorithms. They derive from behavior found in trajectories and/or external data.

The DayTag annotation system [Rinzivillo et al. 2013] helps an individual to reconstruct her/his travel diary from the GPS trajectories collected by using a smartphone. The user uploads his trajectories and interacts with the system to visualize and annotate trajectories. It generates diaries a posteriori, instead of annotating trajectories during real time on mobile devices, as done in works like [Doulamis et al. 2012, Broll et al. 2012]. These tools sometimes they infer basic information, such as the kind of transportation means (motorized vehicle) or nearby places, by using spatiotemporal analysis or checking the places that are close to the user's location in available databases. However, these semi-automatic annotation tools still demand a lot of user effort to confirm what is inferred, and mainly to provide additional information for annotations (e.g., places and events of interest, goals). Our work, on the other hand, proposes a totally automatic method that can be applied to huge data volumes, without demanding additional user effort.

The movement mining algorithms presented in [Azmandian et al. 2012] use as inputs another source of movement data: sequences of social posts. It examines the movement patterns of Twitter users and cluster moving objects according to their spatiotemporal these patterns. The results of this work show that it is possible to infer part of the underlying transportation network from Tweets alone, and uncover interesting differences between the behaviors exhibited by users across cities. [Gabrielli et al. 2013] exploits mobility data mining techniques along with social network analysis methods to aggregate similar trajectories, and point out hot spots of activities, and flows of people that vary over time, according to the number of tweets sent from each place. They apply and validate the proposed trajectory mining approaches to a large set of trajectories built from geo-positioned tweets gathered in Barcelona during the Mobile World Congress 2012.

Other works dealing with social media posts to peoples activities and behavior as they move in the geographic space include [Kisilevich et al. 2010, Cheng et al. 2011, Yin et al. 2011, Zigkolis et al. 2011, Wakamiya et al. 2012]. However, none of these proposals use the textual contents of social media posts to annotate trajectories or to help explain movements.

3. Basic Definitions

This section first presents definitions related to trajectories, and social media users' trails (i.e., sequences of social media user's posts). Then, it describes the problem of fusing them to produce trajectories annotated with the textual contents of the posts. These subjects are fundamental to understand the rest of the paper.

3.1. Trajectories

Raw trajectories are temporally ordered sequences of spatiotemporal positions occupied by a moving object. In this work, we consider trajectories of small objects (e.g., people, vehicles), whose positions are represented by spatiotemporal points.

Definition 1. (**Spatiotemporal Point**). Position represented by the quadruple: $p(p\_id, x, y, t)$, where:

- p_id is a point identifier;

- (x, y) is a pair of geographic coordinates; and
- t is a time instant.

A mobile device that collects locations samples in the form of spatiotemporal points within a certain time interval generates a raw trajectory.

Definition 2. (**Raw Trajectory**). Temporally ordered sequence of spatiotemporal positions visited by a moving object, represented by the triple: $RawTraj(mo\_id, t\_id, p\_seq)$, where:

- mo_id is the mobile object identifier;
- t_id is the trajectory identifier; and
- p_seq is a temporally ordered sequence of spatiotemporal points $(p_1, \ldots, p_n)$, with each p_i $(1 \leq i \leq n)$ of the form stated in Definition 1.

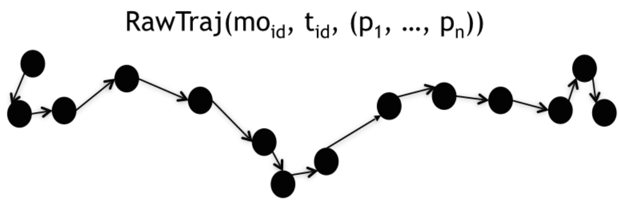

Figure 1. Raw Trajectory

A raw trajectory can be segmented in subsequences of spatiotemporal points satisfying certain conditions. These subsequences are called episodes, as defined in the following.

Definition 3. (**Episode**). Maximal subsequence of spatiotemporal points of a raw trajectory that satisfies a given predicate. An episode is represented by the quadruple: $episode(t\_id, e\_id, e\_type, p\_subseq)$, where:

- t_id is the trajectory identifier;
- e_id is the episode identifier;
- e_type is the episode type (e.g. $Stop$,$Move$); and
- p_subseq is a maximal subsequence of spatiotemporal points $(p_i, \ldots, p_j)$ from a raw trajectory $RawTraj(mo\_id, t\_id, (p_1, \ldots, p_n))$ that satisfies the predicate $(p_i, \ldots, p_j) \implies \{true, false\}$ $(1 \leq i \leq j \leq n)$.

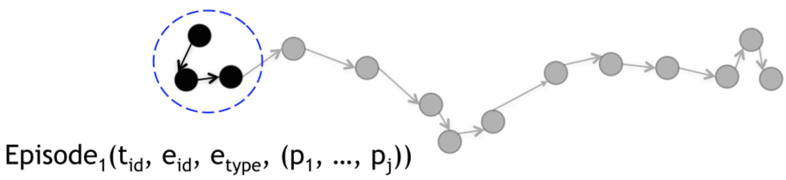

Figure 2. Episode

Temporally ordered episodes of a trajectory, constitute another representation of the movement, called a structured trajectory, as defined in the following.

Definition 4. (**Structured Trajectory**). Temporally ordered sequence of non nested episodes. Each element of the sequence is represented by a pair: **StrTraj(**$stid, Ei$**)**, where:

- $STid$ is the structured trajectory identifier; and
- Ei is an episode.

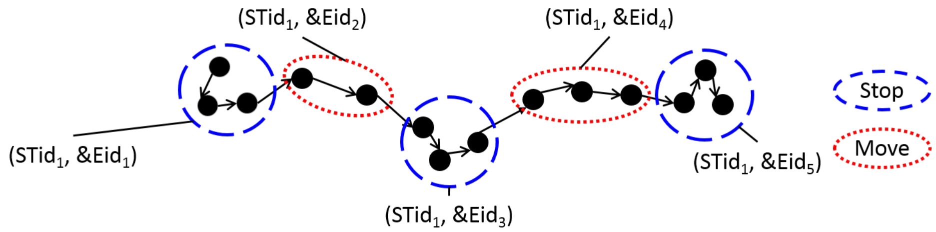

Figure 3. Structured Trajectory

3.2. Movement Data collected on Social Media

A social media footprint is the record of an interaction between an user and a social media (eg, Twitter, Facebook, Foursquare). When the user posts an associated information (eg, space-time position, photo) is recorded in the respective media (eg, Twitter, Facebook) and accessible via specific each media API. A temporally ordered sequence of footprints is a trail.

Definition 5. (**Social Media Footprint**). Social media system record of an iteration performed by a user, represented by the quintuple: **SMF**$(MOid, SMFid, Smid, P, c)$, where:

- $MOid$ is the mobile object identifier;
- $SMFid$ is the footprint identifier;
- $SMid$ is the social media identifier (e.g., Twitter, Facebook);
- P is a reference to a spatiotemporal point (Definition 1);
- c are the contents of the footprints (e.g., tags, pictures, texts).

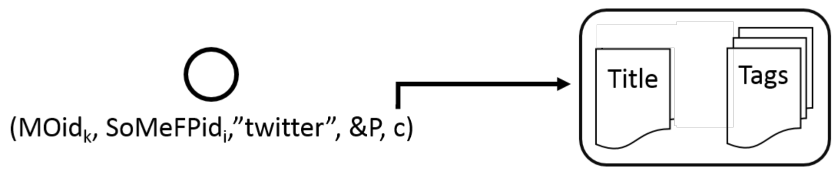

Figure 4. Social Media Footprint

Definition 6. (**Social Media Trail**). Temporally ordered social media footprints sequence, generated by the same user. Each element of this sequence is represented by the pair: **SMT**$(SMTid, SMF)$, where:

- $SMTid$ is the social media trail identifier; and
- SMF is a reference to a social media footprint (Definition 5).

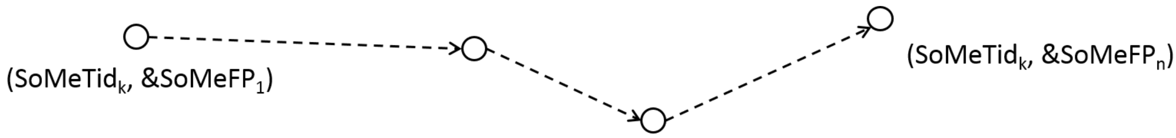

Figure 5. Social Media Trail

Both raw trajectories and social media trails refer to sequences of spatiotemporal positions. However, trajectories usually have better spatiotemporal accuracy than trails. Raw trajectory points are usually sampled at short and fixed intervals (e.g., every second, every 10 meters). On the other hand, social media posts are asynchronous (the user decides when to post) and usually sparse, but they have associated textual contents, that may serve as annotations to help understand movements.

The problem considered in this paper is the fusion of (portions of) trajectories with (portions of) trails, based on spatiotemporal proximity, to produce trajectories annotated with trails contents. Its inputs are a set of raw trajectories and a set of social media trails. Its outputs are pairs of the form $\langle traj, trail \rangle$, where $traj$ is a moving object's trajectory or a continuous subsequence of its points, and $trail$ is a sequence of social media user's footprints (posts). Each returned pair $\langle traj, trail \rangle$ must have trajectory points and trail footprints that are close in space and time, as illustrated in the upper portion of Figure 7.

4. Proposed Method for Fusing Trajectories with Trails

The proposed method efficiently determines the best matching pairs $\langle traj, trail \rangle$ from a large dataset of trajectories and trails, by using a spatiotemporal distance function that allows ranking the matchings, among those that satisfy at least the minimum matching criteria. Figure 6 presents an overview of the proposed method. It is a process with three phases: trajectories preprocessing, trajectories compression, and fusion of trajectories with trails. The preprocessing phase can clean and structure raw trajectories in a sequence of episodes, for example. The compression stage compresses the structured trajectories in a representation that can be analyzed more efficiently than the mere aggregation of the trajectory points in episodes. The fusing phase calculates the global matching coefficient for pairs $\langle traj, trail \rangle$ that may be related and selects the best ones. This method is flexible in the sense that it allows different algorithms for performing specific tasks in each phase, according to the dataset and application peculiarities. The main focus of this paper is the fusion phase, that is divided in four steps, described in the following subsections.

4.1. Select Candidate Pairs

Comparing every pair $\langle traj, trail \rangle$ is not viable for large datasets, due to the amount of time needed for doing so. Thus, we consider a temporal window $[t_i, t_f]$ around each trajectory and trails, where t_i is the initial instant of the trajectory trail less a threshold, and t_f is the final instant of the trajectory or trail plus the same threshold. Using these temporal windows it is possible to efficiently select only the pairs $\langle traj, trail \rangle$ that temporally overlap. It is also possible to use spatio-temporal windows and joins based on their intersections processed with efficient spatiotemporal data access methods to determine the candidate matching pairs. The matching based on enclosing windows is expected to generate a relatively small set of pairs $\langle traj, trail \rangle$, compared to the Cartesian product of the trajectories and trails datasets. These pairs are then evaluated in mored detail to calculate

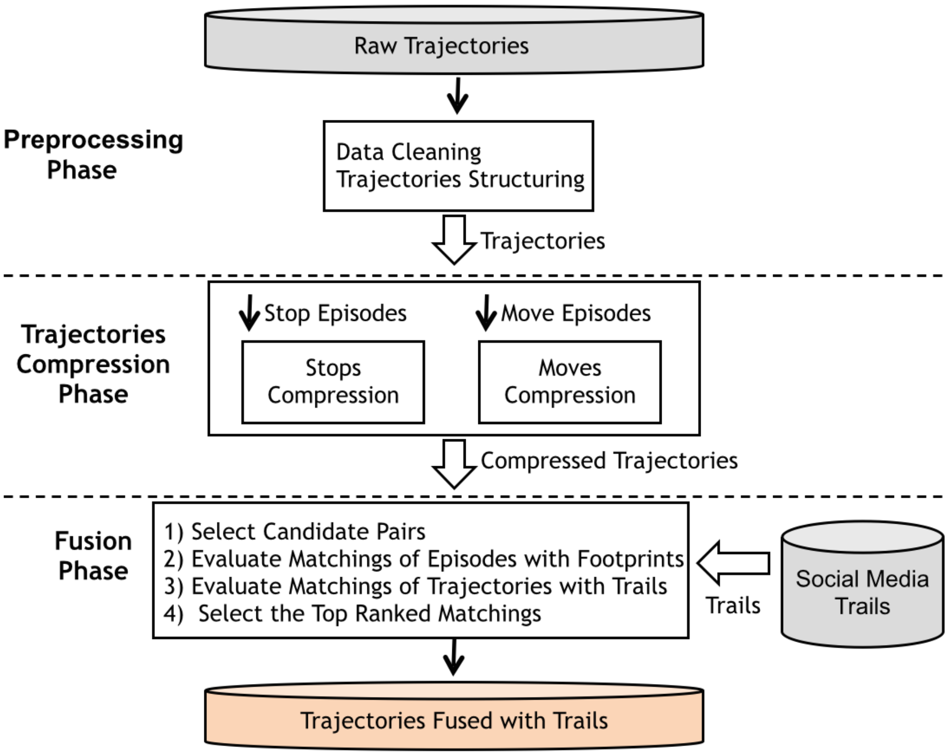

Figure 6. General process of the proposed method

first the local matching coefficients between trajectory episodes and trail footprints, and then the global matching coefficients of the respective $\langle traj, trail \rangle$ pairs.

4.2. Evaluate Matchings of Episodes with Footprints

Consider a candidate pair $\langle traj, trail \rangle$ such that $traj$ is a time ordered sequence of episodes $e_1, \ldots, e_n$ of a moving object's structured trajectory, and $trail$ is a time ordered sequence of footprints (posts) $f_1, \ldots, f_m$ of a social media user's trail ($m, n \geq 1$). The **Temporal Matching Coefficient (TMC)**, given by Equation 1, measures the temporal compatibility between episode e_i and footprint f_j ($1 \leq i \leq n$, $1 \leq j \leq m$).

$$TMC(e_i, f_j) = \begin{cases} 0 & \text{if } (f_j.t \in e_i.t) \\ |\Delta t(e_i.t, f_j.t)| & \text{if } (|\Delta t(e_i.t, f_j.t)|) \leq \tau_t \\ \infty & \text{otherwise.} \end{cases} \quad (1)$$

TMC is 0 if the time stamp $f_j.t$ of the footprint f_j is inside the time span of episode e_i. Otherwise, TMC is the time difference between the trajectory and the trail, if this difference is less or equal an predetermined time threshold τ_t. If this difference is greater than τ_t then TMC is set to infinity. This coefficient guaranties that the footprint is temporally close to the trajectory episode that it may be associated to. It is crucial in the proposed method, because its goal is to associate episodes with social media posts that occur around the same time, and the time stamps are usually reliable in both data sources.

The **Spatial Matching Coefficient (SMC)** of Equation 2 employs a distance metric of the Minkowsky family (L_p) to measure the spatial compatibility between a trajectory episode e_i and a trail footprint f_j.

$$SMC_k(e_i, f_j) = \sqrt[k]{|(e_i.x - f_j.x)|^k + |(e_i.y - f_j.y)|^k} \quad (2)$$

Variations of Equations 1 and 2 could use, for example, the log of δ_t and Lp, respectively, instead of the bare distances. It can help to adjust the measurement scales in a more suitable manner to capture the temporal and/or spatial compatibilities for certain datasets and application domains.

Finally, the **Local Matching Coefficient (LMC)** between trajectory episode e_i and trail footprint f_j is calculated as stated by Equation 3, which simply sums the values $TMC(e_i, f_j)$ with $SMC(e_i, f_j)$, calculated by using Equations 1 and 2, respectively.

$$LMC(e_i, f_j) = TMC(e_i, f_j) + SMC(e_i, f_j) \quad (3)$$

4.3. Evaluate Matchings of Trajectories with Trails

The Global Matching Coefficient (GMC) of a pair $\langle traj, trail \rangle$ is calculated by using the LMC between each spatiotemporally close episode and footprint of $traj$ and $trail$, respectively. In this work, we use Dynamic Time Warping (DTW) [Rakthanmanon et al. 2013] for doing this task. DTW is an efficient algorithm to calculate the proximity between two temporal sequences, by computing optimal matchings between their component points. The sequences are warped non-linearly in the time dimension to determine a measure of their similarity, independent of certain non-linear variations in the time dimension. Although DTW measures a distance-like quantity between two given sequences, it does not guarantee the triangular inequality property.

In this work $GMC(traj, trail)$ is the DTW proximity between $traj$ and $trail$, i.e., the optimal sum of $\sum_{e_i \in traj} LMC(e_i, f_j)$ between episodes of a trajectory and trail footprints. DTW has two binding possibilities to an episode and a footprint, regarding annotation purposes. These cases are: (i) bind an episode with 1 or more footprints of a trail (B1+); or (ii) bind more than one episode of a trajectory with the same footprint (B+1). In case B1+ we annotate the episode with all footprints binded to it, and in case B+1 we annotate the episode with the closest footprint. These binding cases are illustrated in the figure 7.

4.4. Select the Top Ranked Matchings

After computing the GMC between $\langle traj, trail \rangle$ pairs, these pairs are regarded as edges of a bipartite graph $BG(V, E)$ where $V = TrailsSet \cup TrajectoriesSet$, and the weight of each edge $\langle traj, trail \rangle$ is the value $GMC(traj, trail)$. Then, a greedy algorithm that orders the edges in descending order of their weights, and takes the edge (pair $\langle traj, trail \rangle$) with the lowest weight (value $GMC(traj, trail)$) to annotate episodes of the trajectory $traj$ with the textual information associated to the footprints of $trail$.

5. Experiments

We have implemented the method proposed in this work as a prototype. The implementation of this prototype was done in Java version 1.7.0. The database management system

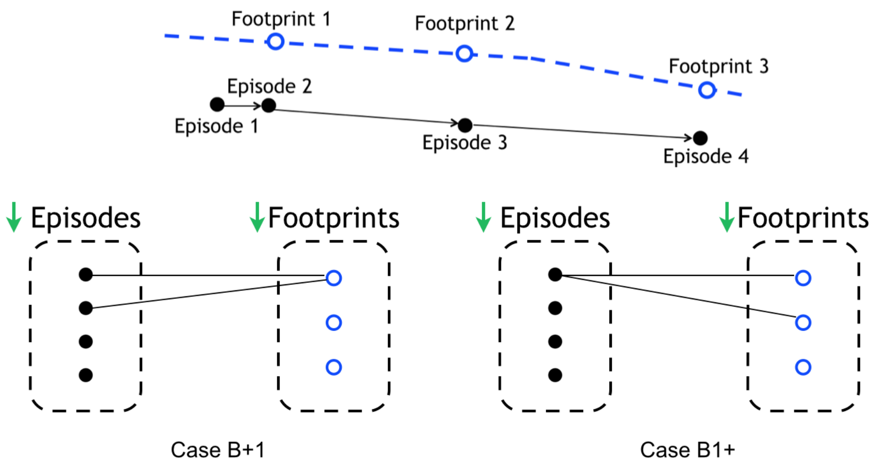

Figure 7. Illustration of the binding cases for a pair $\langle traj, trail \rangle$**.**

Postgres[1] version 9.1.3 and its spatial extension PostGIS[2] version 2.1 were used to hold trajectory and social media data. They provide support to efficiently access and process this data, with operators such as `CONTAINS`, `OVERLAPS`, and proximity joins that use these operators on geometric representations of spatiotemporal data indexed with GIST[3].

The proposed method needs two datasets as its inputs: a trajectories dataset and a trails dataset, that must be spatially and temporally compatible (i.e., contain data of the same geographic area, preferably collected during the same time period). However, finding compatible datasets has not been a trivial task. Currently, we only have access to trajectory databases that were collected a few years ago, while social media APIs only allow collecting data of recent posts, or even those occurring at real time. Thus, we are doing efforts in parallel to collect trajectories and posts that are compatible in space and time. In fact, we are pursuing work with volunteers to collect some subsets of trajectories connected to specific sequences of social media users' posts, to serve as ground truth to evaluate the quality of our method. Meanwhile, we have done preliminary experiments to evaluate our method with partially compatible datasets, such as collections of trajectories and social media data of the same geographic area, but not the same year. Then, in the experiments done so far, we consider datasets with temporal compatibilities such as the same days of the year, months or seasons, but in different years.

The Dataset of Raw Trajectories (DRT) was collected by using GPS on taxis moving in the metropolitan region of Fortaleza, during the period between July 3 2012 and October 20 2012. For this experiment we selected 10 taxis drivers, and segmented the data in such a way that each trajectory corresponds to a taxi ride for a passenger, generating a total of 8,253 trajectories. The Dataset of Social Media Trails (DSMT) contains the posts of 11,974 Twitter users, who sent tweets from the metropolitan region of Fortaleza between July 3 2014 and October 15 2014, making a total of 339,713 footprints.

[1] http://www.postgresql.org
[2] http://postgis.net
[3] http://gist.cs.berkeley.edu

5.1. Data Processing

In the Pre-processing Phase we deleted from the DRT all the trajectories that had less than 10 points, and assured that they were segmented by the taxis ride. These procedures reduced the number of trajectories of the DRT to 6,429. Then, we applied the algorithm CB-SMoT [Bogorny et al. 2011] to generate the Database of Structured Trajectories (DST), containing 13,167 stops in 4,962 trajectories. All the stops generated have their duration equal or bigger than 15 seconds, average movement speed of 0.5 km/h, and maximum instantaneous speed between two points of 1 km/h. The other 1,467 trajectories did not meet these requisites to generate valid stops.

In the Trajectory Compression Phase, we computed the centroids of the stops produced in the previous phase for each structured trajectory in DST. In the experiments done so far we did not use the moves of these trajectories.

Finally, in the Fusion Phase we set the temporal threshold for the temporal window to 5 minutes. It is important to denote that our trajectory dataset do not match the trail dataset exactly in time, we disregard the the year of the trajectories data. In addition, as these datasets do not have a matching segmentations, we allow matching to be done between subsequences of trails and subsequences of trajectories. Therefore, the pairs $\langle traj, trail \rangle$ do not have to match as a whole.

The average number of candidate trails to match each trajectory in the DST were 2 trails, after applying the temporal window. Thus, trails that are out of the temporal window of each trajectory are not taken into account in the computation of the GMC. Consequently, the fusing algorithm can run more efficiently by only considering time compatible trails to match with each trajectory.

We used the distance metrics L1 and L2 to calculate the spatial compatibility coefficient (SMC_k), i.e, we made experiments with $k = 1$ and $k = 2$ in Equation 2. We built as an output of the Global Matching Coefficient (GMC) computation an Associative Database (ADB) that is composed by a structured trajectory from DST, a trail id from DSMT, a GMC_1 value considering SMC_1, and a GMC_2 value considering SMC_2.

5.2. Results

We verified that 87% of the greedy algorithm choices for binding pairs $\langle traj, trail \rangle$ were the same using either SMC_1 or SMC_2. The mean execution time for the method was 8 minutes, obtained in set of 20 executions with the same conditions. We computed SMC_1 and SMC_2 in each one of these executions.

The proposed method was able to annotate around 32% of the structured trajectories that had at least one stop, using either GMC_1 or GMC_2. The execution of the method using GMC_1 and GMC_2 generated 1,621 annotated trajectories with at least one episode annotated with the textual contents of a social media post.

6. Conclusion and Future Works

This paper proposes a method to fuse moving objects' trajectories with social media users' trails (sequences of posts). The proposed process adds the textual information of posts contents to structured trajectories to produce annotated trajectories. This fusion is performed in three phases: preprocessing, trajectory structuring, and data fusion. The last phase relies on the spatiotemporal proximity of individual posts of a trail with trajectory episodes. The main contributions of this proposal are: (i) for the best of our knowledge, it the first one to annotate trajectories via fusion with social media posts; (ii) the proposed method is totally automatic; and (iii) it is convenient and efficient enough to be used with vast amounts of trajectories and social media data.

Future works include: (i) conduct further information fusing experiments with other trajectories and social media data collections; (ii) extend the current version of the method to also annotate moves; (iii) optimize the proposed method to run faster; (iv) evaluate the quality of the annotated trajectories generated by different versions of the proposed method, and schemes for tunning its parameters; and (v) employ the resulting textually annotated trajectories to feed a variety of semantic enrichment, analysis, and mining methods.

Acknowledgments

This work was supported by the European Union IRSES-SEEK (grant 295179) and FP7-DATASIM (grant 270833) projects, CNPq (grant 478634/2011-0), CAPES, and FEESC.

References

Alvares, L. O., Bogorny, V., Kuijpers, B., de Macedo, J. A. F., Moelans, B., and Vaisman, A. (2007). A model for enriching trajectories with semantic geographical information. In *the 15th Intl. Symp. on Advances in Geographic Information Systems*, pages 22:1–22:8, New York, NY, USA. ACM.

Azmandian, M., Singh, K., Gelsey, B., Chang, Y.-H., and Maheswaran, R. T. (2012). Following human mobility using tweets. In Cao, L., Zeng, Y., Symeonidis, A. L., Gorodetsky, V., Yu, P. S., and Singh, M. P., editors, *ADMI*, volume 7607 of *LNCS*, pages 139–149. Springer.

Bogorny, V., Avancini, H., de Paula, B. C., Kuplich, C. R., and Alvares, L. O. (2011). Weka-STPM: a software architecture and prototype for semantic trajectory data mining and visualization. *T. GIS*, 15(2).

Broll, G., Cao, H., Ebben, P., Holleis, P., Jacobs, K., Koolwaaij, J., Luther, M., and Souville, B. (2012). Tripzoom: An app to improve your mobility behavior. In *11th Intl. Conf. on Mobile and Ubiquitous Multimedia*, MUM '12, pages 57:1–57:4, New York, NY, USA. ACM.

Castanedo, F., García, J., Patricio, M. A., and Molina, J. M. (2010). Data fusion to improve trajectory tracking in a cooperative surveillance multi-agent architecture. *Information Fusion*, 11(3):243 – 255. Agent-Based Information Fusion.

Cheng, Z., Caverlee, J., Lee, K., and Sui, D. Z. (2011). Exploring millions of footprints in location sharing services. In Adamic, L. A., Baeza-Yates, R. A., and Counts, S., editors, *ICWSM*. The AAAI Press.

Doulamis, A., Pelekis, N., and Theodoridis, Y. (2012). Easytracker: An android application for capturing mobility behavior. *2012 16th Panhellenic Conference on Informatics*, 0:357–362.

Fileto, R., Krüger, M., Pelekis, N., Theodoridis, Y., and Renso, C. (2013). Baquara: A holistic ontological framework for movement analysis using linked data. In *Conceptual Modeling*, volume 8217 of *LNCS*, pages 342–355. Springer Berlin Heidelberg.

Gabrielli, L., Rinzivillo, S., Ronzano, F., and Villatoro, D. (2013). From tweets to semantic trajectories: Mining anomalous urban mobility patterns. In Nin, J. and Villatoro, D., editors, *CitiSens*, volume 8313 of *LNCS*, pages 26–35. Springer.

Kisilevich, S., Krstajic, M., Keim, D., Andrienko, N., and Andrienko, G. (2010). Event-based analysis of people's activities and behavior using flickr and panoramio geotagged photo collections. In *14th Intl. Conf. on Information Visualisation*, pages 289–296, Washington, DC, USA. IEEE Computer Society.

May, C. and Fileto, R. (2014). Connecting textually annotated movement data with linked data. In *IX Regional School on Databases*, ERBD, São Francisco do Sul, SC, Brazil (in Portuguese). SBC.

Parent, C., Spaccapietra, S., Renso, C., Andrienko, G. L., Andrienko, N. V., Bogorny, V., Damiani, M. L., Gkoulalas-Divanis, A., de Macêdo, J. A. F., Pelekis, N., Theodoridis, Y., and Yan, Z. (2013). Semantic trajectories modeling and analysis. *ACM Computing Surveys*, 45(4):42.

Pelekis, N. and Theodoridis, Y. (2014). *Mobility Data Management and Exploration*. Springer.

Rakthanmanon, T., Campana, B., Mueen, A., Batista, G., Westover, B., Zhu, Q., Zakaria, J., and Keogh, E. (2013). Addressing big data time series: Mining trillions of time series subsequences under dynamic time warping. *ACM Trans. Knowl. Discov. Data*, 7(3):10:1–10:31.

Rinzivillo, S., de Lucca Siqueira, F., Gabrielli, L., Renso, C., and Bogorny, V. (2013). Where have you been today? annotating trajectories with daytag. In *SSTD*, pages 467–471.

Spaccapietra, S., Parent, C., Damiani, M. L., de Macedo, J. A., Porto, F., and Vangenot, C. (2008). A conceptual view on trajectories. volume 65, pages 126–146, Amsterdam, The Netherlands, The Netherlands. Elsevier Science Publishers B. V.

Wakamiya, S., Lee, R., and Sumiya, K. (2012). Crowd-sourced urban life monitoring: urban area characterization based crowd behavioral patterns from twitter. In *ICUIMC*, page 26.

Xiu-li, Z. and Wei-xiang, X. (2009). A clustering-based approach for discovering interesting places in a single trajectory. In *Intelligent Computation Technology and Automation, 2009. ICICTA '09. 2^{nd} Int. Conf. on*, volume 3, pages 429–432.

Yan, Z., Chakraborty, D., Parent, C., Spaccapietra, S., and Aberer, K. (2013). Semantic trajectories: Mobility data computation and annotation. *ACM TIST*, 4(3).

Yin, Z., Cao, L., Han, J., Luo, J., and Huang, T. S. (2011). Diversified trajectory pattern ranking in geo-tagged social media. In *SDM*, pages 980–991. SIAM / Omnipress.

Zigkolis, C., Papadopoulos, S., Kompatsiaris, Y., and Vakali, A. (2011). Detecting the long-tail of points of interest in tagged photo collections. In Martinez, J. M., editor, *CBMI*, pages 235–240. IEEE.