Trajectory-based Visual Analytics for Anomalous Human Movement Analysis using Social Media

Conference Paper · May 2015

DOI: 10.2312/eurova.20151102

Junghoon Chae
Oak Ridge National Laboratory

Cui Yuchen
Beijing University of Civil Engineering and Architecture

Abish Malik
Purdue University

David S. Ebert
Purdue University

EuroVis Workshop on Visual Analytics (2015)
E. Bertini and J. C. Roberts (Editors)

Trajectory-based Visual Analytics for Anomalous Human Movement Analysis using Social Media

Junghoon Chae[1] , Yuchen Cui[1] , Yun Jang[2] , Guizhen Wang[1] , Abish Malik[1] and David S. Ebert[1]
[1]Purdue University, USA, [2]Sejong University, South Korea,

Abstract

The rapid development and increasing availability of mobile communication and location acquisition technologies allow people to add location data to existing social networks so that people share location-embedded information. For human movement analysis, such location-based social networks have been gaining attention as promising data sources. Researchers have mainly focused on finding daily activity patterns and detecting outliers. However, during crisis events, since the movement patterns are irregular, a new approach is required to analyze the movements. To address these challenges, we propose a trajectory-based visual analytics system for analyzing anomalous human movements during disasters using social media. We extract trajectories from location-based social networks and cluster the trajectories into sets of similar sub-trajectories in order to discover common human movement patterns. We also propose a classification model based on historical data for detecting abnormal movements using human expert interaction.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Visual Analytics, Social Media Analysis, Trajectory Analysis, Data Clustering, Anomaly Analysis—

1 Introduction

Analysis of human mobility patterns is important for urban planning [ZLYX11], traffic management [WZP12], and understanding the pandemic spread of diseases [EGK*04]. For crisis and disaster events, movement analysis, such as where people move to/from and how they respond to disasters, is critical for evacuation management. Unfortunately, finding meaningful data is challenging and collecting relevant data can be costly. However, the rapid development and increasing availability of mobile communication and location acquisition devices allow people to share location data using location-based social networks (LBSNs). Particularly, trajectories—sequences of georeferenced data nodes of each user—extracted from such LBSNs provide opportunities and solutions to challenges in human movement analysis [AAB*09, FAAJ13, GRRV14].

Previous studies have mainly focused on finding regular movement patterns using spatial data. They have demonstrated that human movements are normally influenced by geographic constraints, life patterns, and spatial and temporal events, such as local festivals and holiday seasons [AAH*11, FLS10]. However, during disaster events, since human movement patterns (e.g., volume and direction of movements) are unusual compared to normal situations, a new approach is required to analyze these movements. Therefore, to address these challenges, we propose a trajectory-based visual analytics system for anomalous human movement analysis during crises using LBSNs. Our system generates trajectories using geo-location information of chronologically ordered Twitter messages (tweets) for each person. The extracted raw trajectories, however, do not have enough location points. So, we supplement this sparse data in the trajectories using route information between each position sample. We group the individual trajectories into classes of similar sub-trajectories using a trajectory clustering model based on the partition-and-group framework [LHW07], enabling users to discover common sub-patterns, rather than just seeing common holistic patterns. We also incorporate a novel classification model based on historical data for abnormal movement detection using human expert interaction. We allow users to utilize their domain knowledge of their regions of interest to interactively identify and compare abnormal trajectory patterns with normal trajectory patterns.

2 Related Work

As many social networks move towards LBSNs, researchers have proposed various approaches to analyze spatiotemporal social media data. Adrienko et al. [AAB*13b] describe a visual analysis approach for exploring tweet text and spatiotemporal patterns. Chae et al. [CTB*12, CTJ*14]

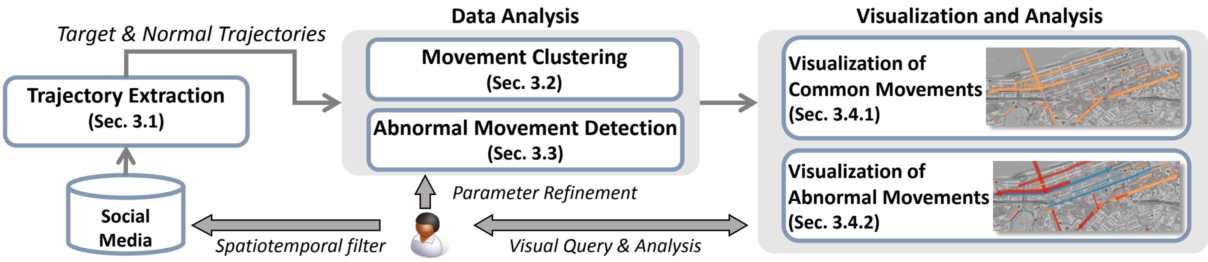

Figure 1: *Overview of our iterative analysis scheme for common human movement discovery and anomaly analysis.*

propose multiple visualizations of spatiotemporal social media analysis for disaster management. They focus on detecting abnormal events and finding crowded places using social media, but have limitations in clustering trajectories and anomalous movement detection. Approaches putting less focus on visualizations and more on data mining mechanisms have been proposed by some studies [WZP12, BMdCM11, CML11] to discover human movement patterns based on LBSNs. For the research on collective movement, clustering is a popular approach in looking for common patterns. Andrienko et al. [AAB*13a] propose a wide range of clustering-based analytics models and combine those with visualization techniques. Their clustering models group similar trajectories as a whole and extract common trips. In this work, we focus on finding common sub-trajectories. Our clustering of sub-trajectories (as opposed to whole trajectories) enables the extraction of similar portions of trajectories, even when no overall clusters may exist.

Existing anomaly detection models [LYC10, KNT00, BKNS00] for trajectory data have mainly focused on identifying outliers from a target dataset. The models look for major patterns and determine whether each trajectory belongs to the majority according to specific criteria. However, during abnormal situations, even the major behaviors can be unusual compared to normal situations. To address this challenge, our work focuses on the anomalous human behavior analysis through the combination of user expert knowledge and automatic anomaly detection models. The research [AAW07, AA11, AAB*13c] dealing with GPS data for collective movement analysis takes advantage in high spatial density compared to density of LBSNs. However, it is difficult to collect data for areas of interest and the data usually has no other context. In order to resolve these issues, we utilize additional context (i.e., tweet text) from LBSNs and visually incorporate the information to enhance the human movement analysis by improving situational awareness.

3 System Overview

Our system is designed for exploring and discovering common and anomalous movement patterns using LBSNs data. The system consists of three major components: a trajectory extraction module, a data analysis module, and a visualization module as illustrated in Figure 1. Our system allows users to select an initial spatiotemporal filer. The users then select a normal time frame. Here, we expect the users to have knowledge about the geographical and temporal characteristics of their locations of interest. The *trajectory extraction module* (Section 3.1) generates two different sets of trajectories: target and normal trajectories. The *data analysis module* consists of two components: movement clustering (Section 3.2) and abnormal movement detection (Section 3.3). The *visualization module* (Section 3.4) allows the users to explore the trajectories, common routes, and abnormal movements. Users can iteratively make visual queries and refine the parameters used for clustering and anomaly detection to optimize the results.

3.1 Trajectory Extraction

Our system extracts trajectories from location-based social media data. Users first select a geographical boundary and a target time window of interest. The users also select one or more past time windows representing normal situations to compare against the target time frame. The trajectory extraction module then requests two sets of tweets from the database for the two selected time windows. The module generates two sets of trajectories: target and normal trajectories using location information of the chronologically ordered tweets for each person. The generated trajectories, however, are usually sparse and incomplete. In order to reduce the spatial sparseness of the raw trajectories, we fill the trajectories with supplementary points between each point pair of the trajectories. These points are calculated by shortest-path-based route directions.

3.2 Movement Clustering

Discovering common movements is a critical process for exploration and analysis of a large volume of trajectory data. Clustering is a popular approach in looking for common patterns in the trajectory data. In this work, we utilize a modified partition-based clustering model [LHW07], in order to find common sub-trajectories. For each given trajectory, this model first partitions a trajectory into a set of line segments, and then groups the line segments into clusters of similar line segments. Clustering the line segments (as opposed to whole trajectories) enables the extraction of similar portions of trajectories. For example, the three trajectories (green, black, red) in Figure 3 have different origins and destinations, but there is a common path in all three trajectories (blue).

Clustering the line segments requires a distance function

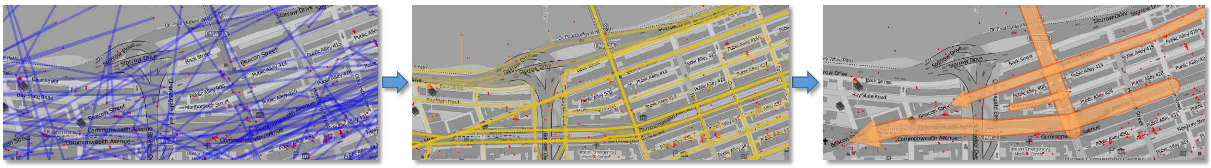

Figure 2: *The process of discovering common human movement patterns using location-based social networks data. Visualization of raw trajectories (left), supplemented trajectories (center), and sub-trajectory clusters (right).*

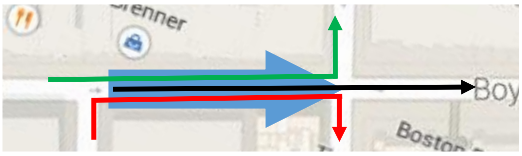

Figure 3: *Discovering a common sub-trajectory from three different trajectories.*

measuring the distance between line segments. We use the distance function based on a line segment Hausdorff distance [CLG03]. We modify the distance function to consider the directions of the two line segments, as the existing model does not consider the directions.

3.3 Abnormal Movement Detection

In this work, we propose a classification model based on historical data for abnormal movement detection using human expert interaction. We allow the users to utilize their domain knowledge of the geographical and temporal characteristics of the location where an abnormal event of interest occurs. The users select a target time window for the abnormal situation and also choose another time window representing a normal/regular situation for the location. We extract two sets of trajectories for the two different time frames and cluster each set of trajectories into two sets of trajectory clusters: target T and comparable C clusters.

Next, we calculate target and comparable representative trajectories from T and C. A target representative trajectory $RT_{ti} \in T$ is classified as an outlier if there is no close comparable representative trajectory $RT_{ci} \in C$. More specifically, we identify outlying line segments $L_{ti} \in RT_{ti}$, which are determined by the distance from neighboring RT_{ci}. We set the default distance threshold D to 200 meters based on our experimental results. Then, intuitively, a target representative trajectory RT_{ti} is outlying if the percentage of the total length of outlying line segments is higher than P. The default value of P is set to 30% based on our experimental results. Our system allows the users to interactively adjust the two parameter values, D and P in order to refine their results.

3.4 Visualization and Analysis

Our design goal is to show trajectories extracted from geo-tagged social media of each person. However, when the number of trajectories shown over the map increases, visual clutter issues arise that hinder the discovery of flow patterns. To reduce clutter, we use a modified partition-based trajectory clustering model for discovering common sub-trajectory patterns. The discovered common movements have multiple attributes to be analyzed, such as cluster size, direction, and length. The users need to not only identify abnormal movement patterns, but also understand how abnormal and normal movement patterns differ.

3.4.1 Visualization of Common Movements

Figure 2 shows the process of discovering common movement patterns. If we display the raw trajectories, it is difficult to understand the realistic movement patterns because of the high degree of sparseness of the trajectories as shown in Figure 2 (left). Therefore, we reduce the sparseness of the raw trajectories using the method described in Section 3.1 and display the supplemented trajectory on the map in Figure 2 (center). Next, we cluster the trajectories into sets of similar sub-trajectories and generate representative trajectories for each cluster as described in Section 3.2. The representative trajectories represent common movement behaviors in Figure 2 (right).

We provide visual cues to show multiple attributes for a representative trajectory. We use a polyline with an arrow head to display the length and the direction of the representative trajectory. The thickness of the line represents the size (i.e., the number of sub-trajectories belonging to a cluster) of the corresponding cluster. Figure 2 (right) shows the representative trajectories within the region. Placement order of the trajectories depends on the length; the longest trajectory is placed at the bottom and the shortest one at the top, to avoid obscuring the shorter trajectories.

3.4.2 Visualization of Abnormal Movements

Our analytics model identifies abnormal representative trajectories from target trajectories by comparing with normal ones. We define target outliers as the abnormal representative trajectories and target normal trajectories as the rest of the target representative trajectories. The target normal trajectories are close to the normal representative trajectories. We visualize the three different types of representative trajectories: target outlier, target normal, and normal trajectories using a similar visual encoding scheme described in Section 3.4.1. We use different colors to distinguish between the different types: target outlier with red, target normal with orange, and normal with blue in Figure 4. We can see that the trajectories (1), (2), and (3) are close to the normal trajectory (4), but they head toward the opposite/perpendicular direction. These are, therefore, classified as outliers. The tra-

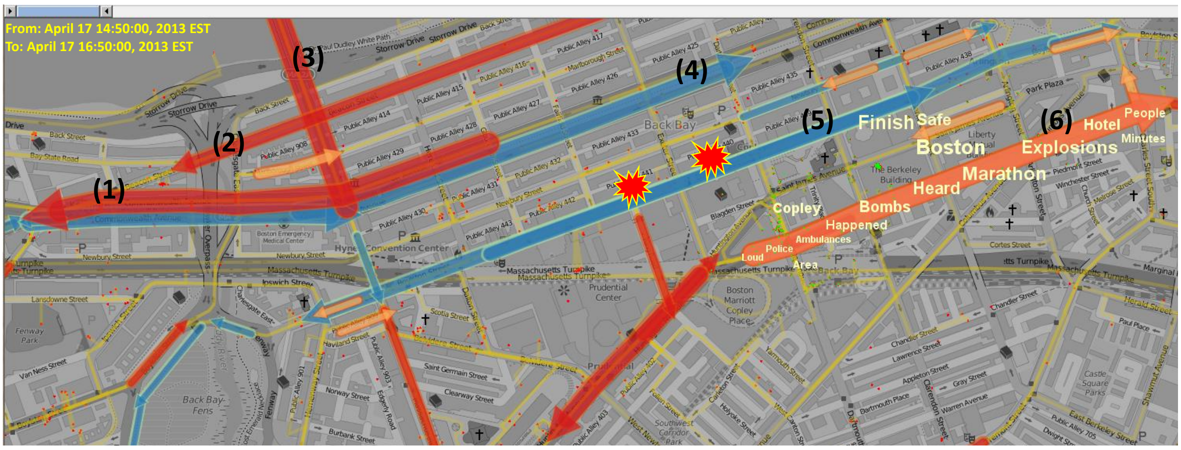

Figure 4: *The trajectories (red and orange) shows the human movement patterns around the finish line at the Boston Marathon 2013 during 2 hours after the explosions. The date/time selection slider can seen at the top and the selected time frame is annotated at upper-left corner. The trajectories (blue) represent the movements for the normal situation (the same time period of the same event in 2014). The two markers indicate the locations of the two explosions.*

jectory (6) is not classified as an outlier because it is close to the normal trajectory (5) and also has the same direction.

Analyzing the spatial behaviors alone is limited in achieving situational awareness of local events—they cannot answer why people are moving and what situations is occurring. To address this challenge, we extract most frequent keywords from the tweets used to generate a trajectory cluster and in the surrounding regions. To display the extracted keywords, we utilize tag cloud visualization. Tag clouds have been used to represent the most frequent words in order to summarize text collections [VW08]. Also, tag clouds can be exploited for analytics tasks, such as topic-based document navigation and labeling geographical points of interest [TBK*12,LSC08]. In this work, we display the keywords along their corresponding representative trajectory. The font size of each keyword encodes the frequency to show the popularity of the keyword. An example of showing extracted keywords display along the trajectory (6) is shown in Figure 4.

4 Case Study: Boston Marathon Explosion

On April 15, 2013, two bombs exploded near the finish line during the Boston Marathon at 2:49 pm EDT. Figure 4 shows two markers that indicate the locations of explosions. The target trajectories (shown in red and orange) in Figure 4 show the movements during the Boston Marathon bombing using Twitter data for the 2 hours after the explosions. Emergency managers who have knowledge about geographical and temporal characteristics of the location around the explosions can select a normal time frame. The normal trajectories (shown in blue) represent the normal movements using the data from next years' Boston Marathon event (we use next year's data for illustrative purposes instead of the previous year's data due to the unavailability of data for the previous year in our database). The system utilizes the normal trajectories in order to determine that each of the target trajectories is abnormal (shown in red) or normal (shown in orange). The target trajectories show that people were dispersed from the locations of the explosions and did not use the roads where the explosions occurred. Also, the abnormal trajectories (1), (2), and (3) in Figure 4 show that participants and spectators moved in the opposite direction of the finish line or crossed the bridge in order to get away from the location of impact. Furthermore, we see the keywords extracted along the trajectory (6) are shown as a tag cloud. Since the trajectory is close to the explosion locations, the extracted keywords along the trajectory show a strong relationship to the accident.

5 Conclusions and Future Work

We presented a trajectory-based visual analytics system, making it possible for emergency managers to: 1) generate trajectories using geo-tagged tweets, 2) discover human common movement patterns, and 3) detect fine-grained abnormal movement patterns. We also developed a classification model using human expert interaction to identify abnormal movements. We have limitations in reducing the visual clutter of trajectories and adding annotations on the map. In future work, we will improve the classification model to automatically identify the abnormal movements.

Acknowledgement

This material is based upon work supported by the U.S. Department of Homeland Security under Award Number 2009-ST-061-CCI001-06. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security. Jang's work was supported in part by Basic Science Research Program through the National Research Foundation of Korea (NRF-2013R1A1A1011170).

References

[AA11] ADRIENKO N., ADRIENKO G.: Spatial generalization and aggregation of massive movement data. *IEEE Trans. Visualization and Computer Graphics 17*, 2 (Feb 2011), 205–219. 2

[AAB*09] ANDRIENKO G., ANDRIENKO N., BAK P., KISILEVICH S., KEIM D.: Analysis of community-contributed space-and time-referenced data (example of panoramio photos). In *Proceedings of the 17th ACM SIGSPATIAL international conference on advances in geographic information systems* (2009), ACM, pp. 540–541. 1

[AAB*13a] ANDRIENKO G., ANDRIENKO N., BAK P., KEIM D., WROBEL S.: *Visual analytics of movement*. Springer Publishing Company, Incorporated, 2013. 2

[AAB*13b] ANDRIENKO G., ANDRIENKO N., BOSCH H., ERTL T., FUCHS G., JANKOWSKI P., THOM D.: Thematic patterns in georeferenced tweets through space-time visual analytics. *Computing in Science and Engg. 15*, 3 (May 2013), 72–82. 1

[AAB*13c] ANDRIENKO N., ANDRIENKO G., BARRETT L., DOSTIE M., HENZI P.: Space transformation for understanding group movement. *IEEE Trans. Visualization and Computer Graphics 19*, 12 (Dec 2013), 2169–2178. 2

[AAH*11] ANDRIENKO G., ANDRIENKO N., HURTER C., RINZIVILLO S., WROBEL S.: From movement tracks through events to places: Extracting and characterizing significant places from mobility data. In *IEEE Symposium on Visual Analytics Science and Technology* (2011), IEEE, pp. 161–170. 1

[AAW07] ANDRIENKO G., ANDRIENKO N., WROBEL S.: Visual analytics tools for analysis of movement data. *ACM SIGKDD Explorations Newsletter 9*, 2 (2007), 38–46. 2

[BKNS00] BREUNIG M. M., KRIEGEL H.-P., NG R. T., SANDER J.: Lof: Identifying density-based local outliers. *SIGMOD Rec. 29*, 2 (2000), 93–104. 2

[BMdCM11] BRAGA R. B., MEDEIROS DA COSTA S. D. M., MARTIN H.: A trajectory correlation algorithm based on users' daily routines. In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (2011), ACM, pp. 501–504. 2

[CLG03] CHEN J., LEUNG M. K., GAO Y.: Noisy logo recognition using line segment hausdorff distance. *Pattern recognition 36*, 4 (2003), 943–955. 3

[CML11] CHO E., MYERS S. A., LESKOVEC J.: Friendship and mobility: user movement in location-based social networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining* (2011), ACM, pp. 1082–1090. 2

[CTB*12] CHAE J., THOM D., BOSCH H., JANG Y., MACIEJEWSKI R., EBERT D., ERTL T.: Spatiotemporal social media analytics for abnormal event detection and examination using seasonal-trend decomposition. In *IEEE Symposium on Visual Analytics Science and Technology* (2012), pp. 143–152. 1

[CTJ*14] CHAE J., THOM D., JANG Y., KIM S., ERTL T., EBERT D. S.: Public behavior response analysis in disaster events utilizing visual analytics of microblog data. *Computers & Graphics 38* (2014), 51–60. 1

[EGK*04] EUBANK S., GUCLU H., KUMAR V. A., MARATHE M. V., SRINIVASAN A., TOROCZKAI Z., WANG N.: Modelling disease outbreaks in realistic urban social networks. *Nature 429*, 6988 (2004), 180–184. 1

[FAAJ13] FUCHS G., ANDRIENKO G., ANDRIENKO N., JANKOWSKI P.: Extracting personal behavioral patterns from geo-referenced tweets. In *AGILE* (2013). 1

[FLS10] FUJISAKA T., LEE R., SUMIYA K.: Discovery of user behavior patterns from geo-tagged micro-blogs. In *Proceedings of the 4th International Conference on Uniquitous Information Management and Communication* (2010), ACM, p. 36. 1

[GRRV14] GABRIELLI L., RINZIVILLO S., RONZANO F., VILLATORO D.: From tweets to semantic trajectories: mining anomalous urban mobility patterns. In *Citizen in Sensor Networks*. Springer, 2014, pp. 26–35. 1

[KNT00] KNORR E. M., NG R. T., TUCAKOV V.: Distance-based outliers: algorithms and applications. *The International Journal on Very Large Data Bases 8*, 3-4 (2000), 237–253. 2

[LHW07] LEE J.-G., HAN J., WHANG K.-Y.: Trajectory clustering: a partition-and-group framework. In *ACM SIGMOD international conference on Management of data* (2007), ACM, pp. 593–604. 1, 2

[LSC08] LUBOSCHIK M., SCHUMANN H., CORDS H.: Particle-based labeling: Fast point-feature labeling without obscuring other visual features. *IEEE Trans. Visualization and Computer Graphics 14*, 6 (2008), 1237–1244. 4

[LYC10] LIAO Z., YU Y., CHEN B.: Anomaly detection in gps data based on visual analytics. In *IEEE Symposium on Visual Analytics Science and Technology* (2010), IEEE, pp. 51–58. 2

[TBK*12] THOM D., BOSCH H., KOCH S., WOERNER M., ERTL T.: Spatiotemporal anomaly detection through visual analysis of geolocated twitter messages. In *IEEE Pacific Visualization Symposium (PacificVis)* (2012), pp. 41 – 48. 4

[VW08] VIÉGAS F. B., WATTENBERG M.: Timelines tag clouds and the case for vernacular visualization. *interactions 15*, 4 (2008), 49–52. 4

[WZP12] WEI L.-Y., ZHENG Y., PENG W.-C.: Constructing popular routes from uncertain trajectories. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining* (2012), ACM, pp. 195–203. 1, 2

[ZLYX11] ZHENG Y., LIU Y., YUAN J., XIE X.: Urban computing with taxicabs. In *Proceedings of the 13th international conference on Ubiquitous computing* (2011), ACM, pp. 89–98. 1