Proceedings of the 11th International Conference on GeoComputation

University College London, UK
20th – 22nd July 2011

GeoComputation 2011

The art and science of solving complex spatial problems with computers

STANDARD

Spatio-Temporal Analysis of Network Data and Route Dynamics

http://standard.cege.ucl.ac.uk

Proceedings of the 11th International Conference of Geocomputation 2011

University College London
20th – 22nd July 2011

Editors: Tao Cheng
Paul Longley
Claire Ellul
Andy Chow

GeoComputation 2011
The art and science of solving complex spatial problems with computers

UCL

Trajectory Data Mining: Classification and Spatio-Temporal Visualization of Mobile Objects

A. Nara & Paul, M. Torrens

Arizona State University, 975 S. Myrtle Ave., Coor Hall, Tempe, AZ, 85287
Telephone: +1-480-965-7533
Email: {anara; paul.torrens}@asu.edu

1. Introduction

Trajectory-based data mining is a very active research topic in the field of Knowledge Discovery in Databases (KDD) in response to the influx of mobile object data. Using a set of spatio-temporal sequences of mobile object data collected from various types of Location Aware Technologies (LATs) or generated by simulation models, trajectory data mining discovers spatio-temporal knowledge through exercises including pattern detection, clustering, classification, generalization, outlier detection, and visualization. Potential applications across various fields include, for example, vehicle and pedestrian traffic control (e.g., transportation management and facilities design); Location-Based Services (LBS) (e.g., navigation assistance and mobile advertising); weather forecasting (e.g., hurricane trajectory prediction and risk analysis); law enforcement (e.g., video surveillance for criminal activities); animal conservation (e.g., tracking at-risk animal populations); and logistics for goods and human.

In recent years, many approaches have been proposed and applied to various fields to investigate patterns and trends from massive datasets of mobile objects (e.g., Gaffney et al. 2007; Lee, et al. 2007; Andrienko et al. 2009; Guo et al. 2010). Research challenges identified in previous works include characterization, generalization, and visualisation of massive and complex trajectories to discover interesting patterns, trends, and useful knowledge across scales.

In this paper, we propose a trajectory data mining framework that employs trajectory partitioning and clustering algorithms to extract behavioural patterns of mobile objects, as well as visual analysis to display extracted patterns and trends in space and time. As a case study, we developed an Agent-Based Model of pedestrian evacuation based on the social force model and generated crowd evacuation dynamics on a street corridor. The proposed framework successfully differentiated and visualized spatio-temporal clusters of local movement behaviours including smooth evacuation and bottleneck.

2. Methodology

To investigate movement behaviours in trajectory datasets, our proposed trajectory data mining framework includes three methodological steps, trajectory partitioning, trajectory clustering, and spatio-temporal visualization of trajectory clusters.

- Step1: Trajectory partitioning
  - Distance-Threshold approach

- Step 2: Trajectory clustering
  - Quantification of sub-trajectory
  - Principal Component Analysis (PCA)
  - K-means cluster analysis
- Step3: 3D visualization of trajectory clusters
  - Spatio-Temporal Kernel Density Estimate (STKDE) and volume rendering technique

A set of trajectory dataset is described as {Trajectory Set: $TR_{set} = TR_1, TR_2, TR_3, \ldots, TR_i$, where i denotes the number of mobile objects}. Each trajectory is composed of a sequence of three-dimensional points $\{\{TR_i = p_1, p_2, p_3, \ldots, p_j$, where j denotes the number of points in the trajectory i }, $\{p_j = x, y, t\}\}$. The trajectory partitioning process partitioned an entire trajectory into trajectory partitions (sub-trajectories), the process of which is a key to extract local movement behaviours. In this study, a Distance-Threshold approach was employed. It uses a distance threshold value to partition a trajectory into sub-trajectories. This is based on the assumption that in many situations human movements involve stopping/staying when a person changes its behaviour. Such behaviours can be seen at multiple scales; for example, when a pedestrian decelerates and ultimately stops to make a sharp turn or to avoid collisions with other pedestrians; a commuter stays at home, walks to a bus stop, waits for a bus, and stays at its office to work; and a person may relocate and find a new home to stay with its life events. Methodologically, partitioning a trajectory based on staying behaviour can be simply achieved by introducing a Distance-Threshold (Th_d). If a distance of each segment in a trajectory is less than Th_d, then the segment is assigned as *STAY* and a trajectory is partitioned by the segment. If consecutive segments are assigned to *STAY*, those segments are considered as one sub-trajectory in order to differentiate short and long staying behaviours.

For each trajectory partition ($TR_{par(i)}$), multi-dimensional vectors are calculated to characterize the sub-trajectory. The vector values include total duration (d_t), total horizontal distance (d_x), total vertical distance (d_y), total two-dimensional distance (d_{2D}), velocity vector on x-axis (v_x), velocity vector on y-axis (v_y), and velocity (v), horizontal beeline distance (d_{sx}), vertical beeline distance (d_{sy}), two-dimensional beeline distance (d_{s2D}), area of minimum bounding box (mbb), and sum of cosine of turning angle between two consecutive segments (sct). All of these vector values are then normalized with mean equals to 0 and variance equals to 1.

To reduce the dimensionality of multiple vectors of sub-trajectories, PCA is employed. PCA is a multivariate statistical technique to the dimensionality of a dataset consisting of interrelated variables by finding a new set of variables, i.e., Principal Components (PCs), which is smaller than the original set of variables but still containing most of the information in the original dataset. Eigenvalues of PCs measure the amount of variation, and this study uses PCs if their eigenvalues are greater than 1. PC scores of each sub-trajectory for each PC (Eigenvalue $\geq$ 1) are computed, and then they are used as a new input dataset for sub-trajectory clustering.

To classify sub-trajectories for extracting local movement behaviours, the K-means clustering algorithm develped by Hartigan & Wong (1979) is applied. To estimate the optimal value of k in K-means clustering, clustering algorithms are run with different values of k (min:2, max:20), and the optimal value of k is selected by the Gap Statistic (Tibshirani, Walther, & Hastie, 2001).

The Space-Time Kernel Density Estimattion (STKDE) (Brunsdon et al. 2007) and volume rendering technique (Levoy 1988; Nakaya & Yano 2010) are used for visualising cluster density distribution in space and time. The interactive approach of volume rendering is achieved using an open source visualization software, ParaView (Henderson 2007).

3. Results

As a case study to examine the proposed trajectory data mining framework, data regarding pedestrian evacuation dynamics was analyzed. The trajectory data was generated by an ABM based on the social force model (Helbing and Molnár, 1995). In its simplest form, there are three forces formulated as follows.

$$m_i \frac{dv_i}{dt} = m_i \frac{v_i^0(t)e_i^0(t) - v_i(t)}{\tau_i} + \sum_{j(\neq i)} f_{ij} + \sum_{w} f_{iw}$$

The first force is a driving force toward a desired destination described by a pedestrian i of mass m_i, of desired velocity v_i^0, of desired direction e_i^0, and of actual velocity v_i with a certain characteristic time τ_i. The second force is a repulsive force, $\sum_{j(\neq i)} f_{ij}$, describing the interaction effects with other agents j ($j \neq i$), and the third force is a repulsive force, $\sum_{w} f_{iw}$, to avoid walls and obstacles. Pedestrians in this basic form of the social force model walk unidirectionally, i.e., each pedestrian travels between an origin and a destination. This is too simplistic, so to overcome the deficiency, the idea of multiple waypoints is implemented. In the algorithm, each pedestrian i owns a sequenced list of waypoints and walks toward the first waypoint in the list. When it reaches at the waypoint within a certain buffer zone described by a two-dimensional vector $bZ(bx, by)$, the waypoint is removed from the list and the pedestrian walks toward the first waypoint in the new list until reaching the final destination.

In this study, pedestrian evacuation dynamics on a diagonal corridor was simulated. In the simulation, pedestrians evacuate from North, West, and South corridors to an East exit. Table 1 represents initial settings for model environment and parameters used for the social force model. To analyze trajectory data of simulated pedestrian evacuation dynamics, locations (x,y) of pedestrians and corresponding time stamps were output at every one second (=30 frames). As a result of the Gap Statistic, we obtained five sub-trajectory clusters as the optimal k value.

Figure 1 illustrates the clustering result of sub-trajectories using the Distance-Threshold partitioning approach (k=5). Figure 2 presents the culster profiles describing movement characteristics within clusters. The vertical axis represents independent variables for corresponding cluster IDs (k=5) and the horizontal axis shows the average of normalized value of independent variables within a cluster. Figure 3 visualises sub-trajectory cluster density distributions in space and time estimated by STKDE. This explains when and where a particular pattern of movement behaviour occurred.

These results showed that sub-trajectories of Cluster 1 and 2 are identified as smooth evacuation behaviours because both have higher average velocity values and continuous trajectories without staying or stopping. In addition, these clusters are found beneath Cluster 4 near the intersection area and on the East corridor in the STKDE map indicating that pedestrians who reached at the corner of the intersection earlier have successful evacuation. On the other hand, Cluster 4 and 5 are partitioned by Cluster 3 that represents staying or stopping behaviours near the corners of the intersection. This explains the evacuation bottleneck due to the overcrowding.

Model environment	Number of pedestrians	120
	Area width	800
	Area height	700
	Simulation Tick	1 frame
Parameters for social force model	Pedestrian's mass m_i	1
	Pedestrian's desired velocity v_i^0	1.3
	Characteristic time τ_i	2

Table 1. Settings of pedestrian evacuation model.

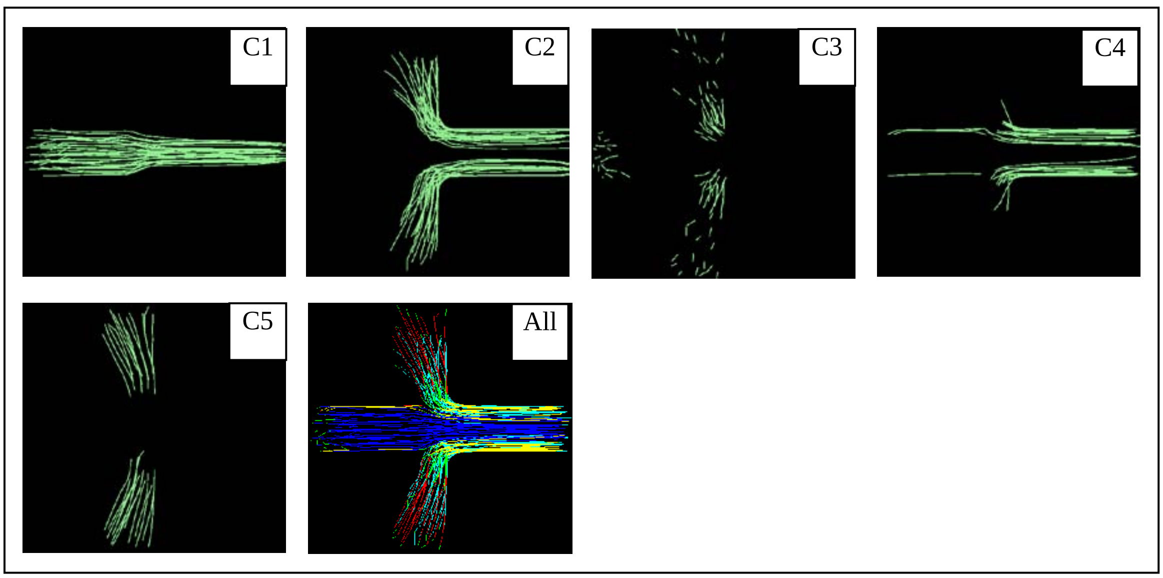

Figure 1. 2D images of sub-trajectories by each cluster (k=5)

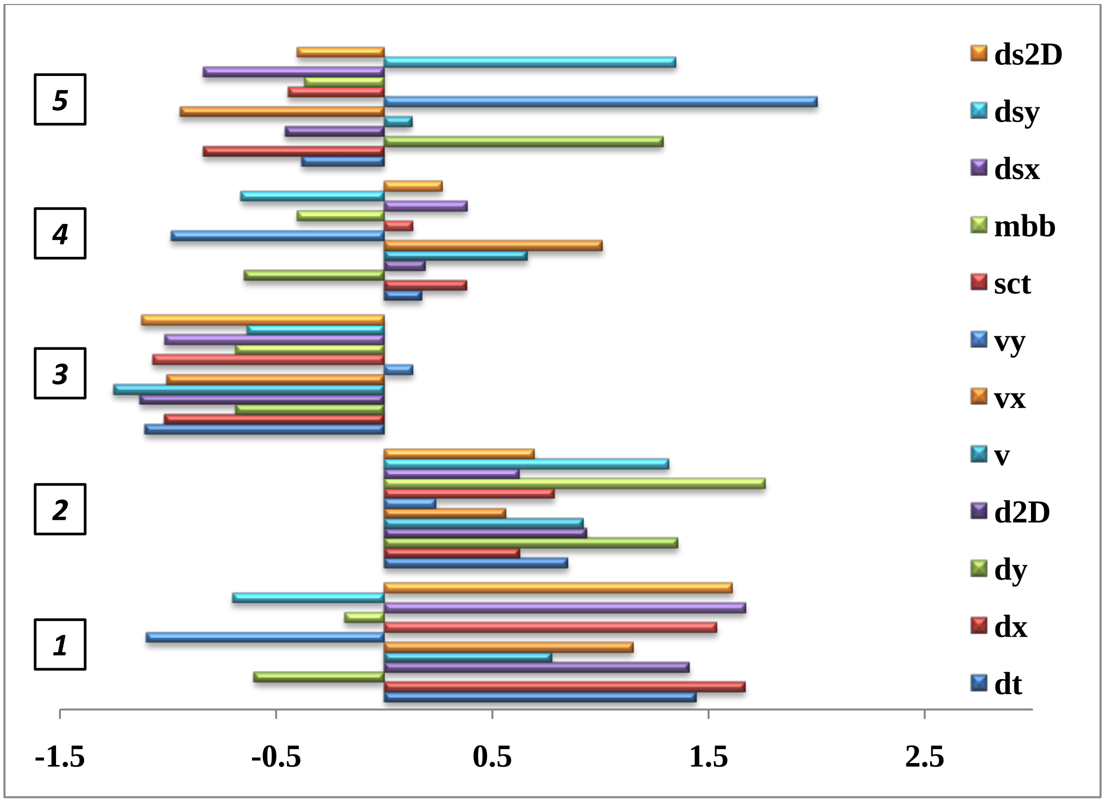

Figure 2. Sub-trajectory cluster profiles (*k*=5)

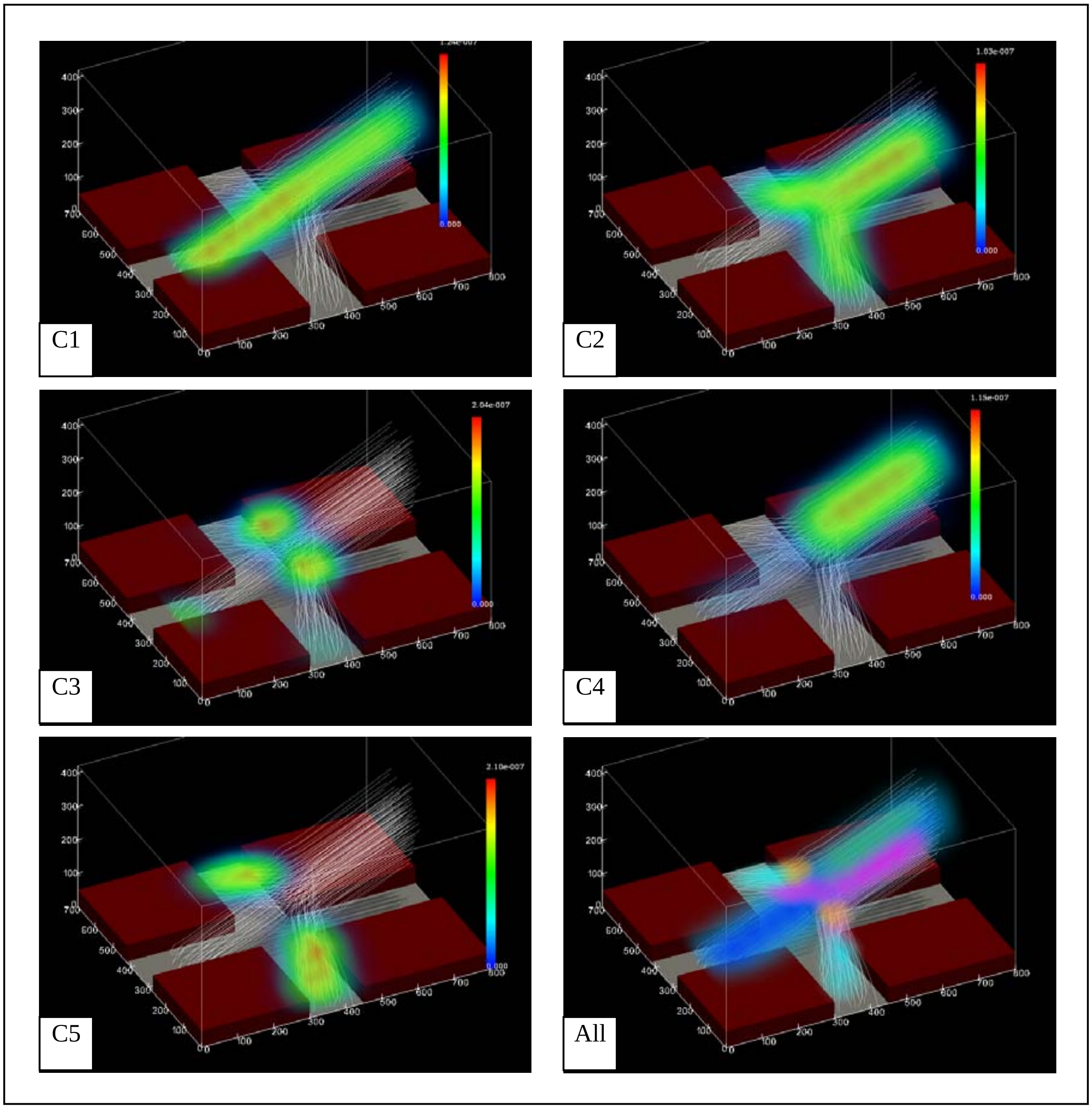

Figure 3. Sub-trajectory cluster distributions in space and time (*k*=5)

4. References

Andrienko, N., & Andrienko, G. (2011). Spatial Generalization and Aggregation of Massive Movement Data. *IEEE Transactions on Visualization and Computer Graphics*, *17*(2), 205-219.

Brunsdon, C., Corcoran, J., & Higgs, G. (2007). Visualising space and time in crime patterns: A comparison of methods. *Computers, Environment and Urban Systems, 31*, 52-75.

Gaffney, S., Robertson, A., Smyth, P., Camargo, S., & Ghil, M. (2006). *Probabilistic Clustering of Extratropical Cyclones Using Regression Mixture Models.* Technical Report, UCI-ICS 06-02, University of California, Irvine.

Guo, D., Liu, S., & Jin, H. (2010). A Graph-based Approach to Vehicle Trajectory Analysis. *Journal of Location Based Service, 4*(3), 183-199.
Helbing, D., & Molnár, P. (1995). Social force model for pedestrian dynamics. *Physical Review E, 51*, 4282-4286.
Henderson, A. (2007). *ParaView Guide, A Parallel Visualization Application.* Kitware Inc.
Lee, J. G., Han, J., & Whang, K. Y. (2007). Trajectory clustering: A partition-and-group framework. *Proceedings of the ACM SIGMOD International Conference on Management of Data*, (pp. 593-604). Beijing, China.
Levoy, M. (1988). Volume rendering: Display of surfaces from volume data. *IEEE Computer Graphics & Applications, 8*(3), 29-37.
Nakaya, T., & Yano, K. (2010). Visualising crime clusters in a space-time cube: an exploratory data-analysis approach using space-time kernel density estimation and scan statistics. *Transactions in GIS*, *14*(3), 223-239.
Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a dataset via the Gap Statistics. *Journal of the Royal Statistic Society: B, 2*, 411-423.