

Group Behaviour Analysis of London Foot Patrol Police

Jianan Shen, Tao Cheng

SpaceTimeLab for Big Data Analytics, Department of Civil, Environmental and Geomatic Engineering, University College London,
Gower Street, London WC1E 6 BT, UK
Nov 03, 2014

Summary

The main objective of this research is to propose a method for group movement pattern generalisation and classification. To this end, DBSCAN is used on stay points for POI identification. Then, movement features are extracted and selected for the behaviour classification. A kernel-based Support Vector Machine method is developed to infer the working types of the officers based on the selected features depicting their movement histories. By analysing the geo-tagged police data, we demonstrate how this method can be used to reveal user information, especially interest information based on their POIs and spatial-temporal movement patterns.

KEYWORDS: Policing, Travel Pattern, Machine Learning, SVM Classification, POI

1. Introduction

The more and more ubiquitously used GPS-integrated devices have generated increasingly large movement history data than ever before. Such movement history data have enabled researchers to analyse and visualise movement trajectories, home range and POI (Points of Interests) distribution, and mobility patterns. Police foot patrol activity, as a special kind of human movement phenomenon, can also be analysed in similar methods with certain modifications.

In this research, we intend to present a framework of patrol pattern analysis and classification of police working type, which is capable of 1) extracting police POIs and traveling sequences from GPS trajectory data; 2) selecting and summarising individual and group movement features for description and comparison of officers undertaking different works; 3) SVM classification of different types of officers based on the selected moving history features.

2. Case Study and Data

The case study takes place in the Camden Borough of London(Figure 1). Five major police stations are located (Places No. 1, 7, 8, 9 and 11 in Figure 1) in this region, namely, West Hampstead, Hampstead, Kentish Town, Albany Street and Holborn.

Automatic Personnel Location Systems (APLS) data set provided by the Metropolitan Police recorded officers' location stamps collected by GPS-integrated radio sets portable on every officer in patrol. The length of the data is 84 days. The 241525 records in the dataset recorded information such as call signs, device IDs, as well as the locations and times of all 745 officers active in that period. Usually, the sampling rate of the system is one update every 10 minutes.

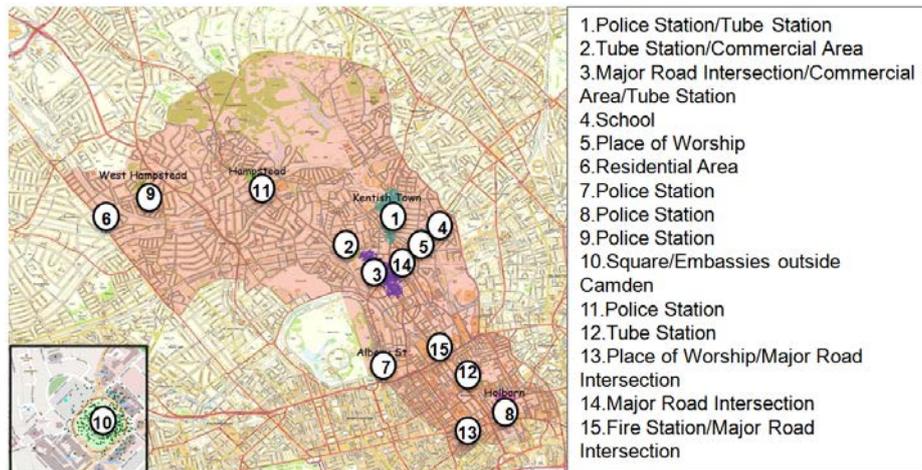


Figure 1 The POIs identified in Camden, including 10 non-police-station POIs (1 of them locates outside Camden) and 5 police stations.

3. Methodology and Results

3.1. POIs and Traveling Sequences Extraction

In discovering POIs, the location and time of officers' stopping behaviour is of more interests than the moving factors (Palma, 2008). To this end, DBSCAN is used for the clustering of stay points where officers stop or move slowly for more than 20 minutes. The massive activities and stops observed nearby police stations are common reflection of police daily routine and office works, and hence disturbed the searching for the POIs during the patrol activities. Therefore, outliers and records generated within 400m radius from the police stations are removed before the density based clustering. Only trajectory moving out of the radius will be considered as a start of a journey and when the time interval between two records exceed 2 hours, the series will be separated as two trips. By adjusting the parameters according to the quality of the data set and the method suggested by Zhou (2012), 10 non-police-station POIs are discovered. For instance, POI No. 2 represents a tube station, POI No.3 is considered to be a major road intersection.

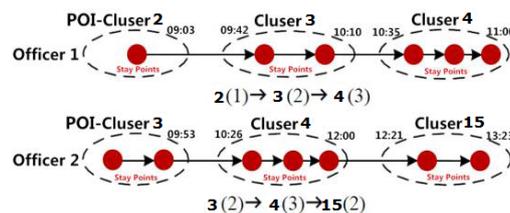


Figure 2 Individual movement history expressed by sequences of visited POIs

By identifying building information on Ordnance Survey Maps and communicating with the police. The main topics of these POIs are identified as schools, places of worships, tube stations and major road intersections. The movement histories of the officers can then be simplified and expressed by a sequence of common POIs the officers visited and the time they arrive and leave each POI attached in the sequence (Figure 2). Similarly, it can also be expressed as a series of topics an officer went through during his/her patrol journey. For example, the journey of Officer 2 in Figure 2 can be summarised as "Commercial Area > School > Fire Station" and the attached time information about how long s/he stayed in each POI. The information of time series, POI topics and trajectories will be further processed to support the classification works.

3.2 Individual and Group Movement Features

After acquiring the semantic meaning of discovered POIs in previous step, individual and group movement can be depicted and summarised with 14 features that are selected or extracted from the raw GPS log as well as the summarised POI visiting history. The features include: mean and standard deviation of speed, mean and standard deviation of patrol covering area, proportion of stopping time, proportion of time spent in police stations and POIs of different topic types, proportion of time spent in weekends and nights and the activeness expressed by the number of records generated by the officer in unit time and so on.

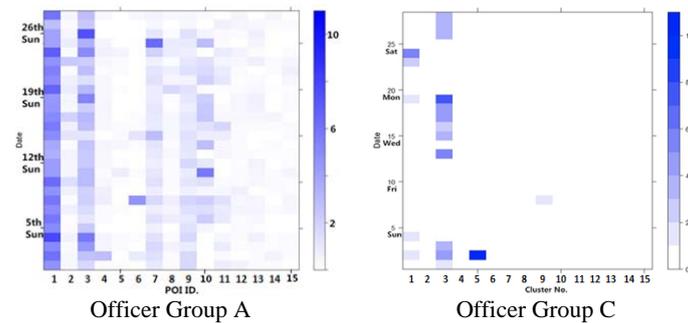


Figure 3 The comparison of POI visiting frequency between two groups consisting of two different types of officers in one month

Figure 4 shows 10 example features of the 3 groups. The values of features are z-score normalised for further classification works. It can be seen in the chart that the selected features of the 3 groups vary greatly. For example Group B has strong daytime activeness proportion and seldom work at night while the standard deviation of Group A is much larger than other groups.

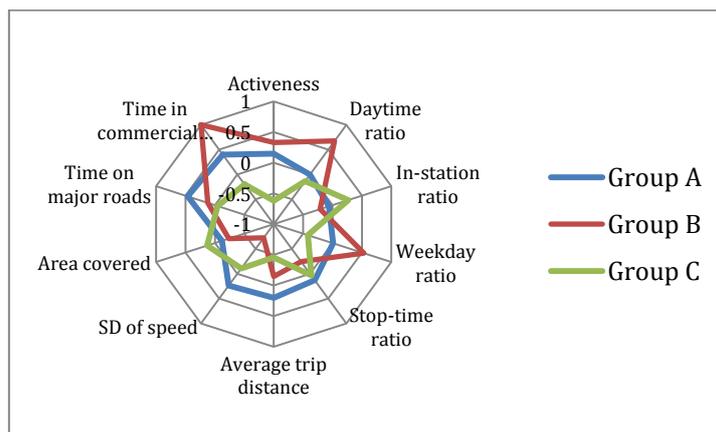


Figure 4 Radar chart comparison of 10 example features depicting the patrol behaviour of different groups

3.3. SVM Classification of Officer works using RBF Kernel

Many researches proposed SVM and other machine learning methods analysing online user behaviors. Some researchers developed similar methods on the analysis of travelling behaviour in real life, such as Baraglia (2013).

By using the parameter selection method developed by Schölkopf (2002), RBF kernel is used in the SVM classification of officer working types based on the individual movement features extracted in previous section. The model is trained with officer movement data of which the working types have already been labeled and is used to classify the data generated in 10 days of the next month. The accuracies of this attempt are logged in Table 1.

Table 1 Classification accuracy by training sets of different sizes

Size of training sets	Training Error	Accuracy of classification
10 days	13.52%	68.89%
20 days	15.00%	72.34%
30 days	14.83%	73.42%

This classification is based on manually-selected and unoptimised feature sets. This method will be further improved with optimum feature selection and using more advanced kernels.

4. Summary and future work

In this research, a method for identifying user (officer) type based on travel (patrol) history is proposed. The framework include density-based POI clustering, feature selection and validation, as well as machine learning classification. The Camden APLS data enabled the study that others cannot proceed with due to the lack of modern GPS-enabled policing equipments. The method revealed the movement features of different officers in space and in time. It can also be used in other time series geo-tagged data for automatic movement pattern generalisation, traveler interest and routine mining, similarity analysis and so on.

Further works may include improving the performance of SVM in unbalanced data sets (the officer numbers in different types vary greatly) and comparing SVM with other methods such as KNN and Artificial Neural Networks.

5. Acknowledgements

This work is part of the project - Crime, Policing and Citizenship (CPC): Space-Time Interactions of Dynamic Networks (www.ucl.ac.uk/cpc), supported by the UK Engineering and Physical Sciences Research Council (EP/J004197/1). The data provided by Metropolitan Police Service (London) is highly appreciated.

6. Biography

Jianan Shen is a PhD student in UCL SpaceTimeLab. He studied Information Engineering in NUDT in China (BEng, 2013). He is now working on the Crime, Policing and Citizenship Project sponsored by the EPSRC, UK. His research interests include movement pattern and trajectory analysis with Machine Learning approaches.

Tao Cheng is a Professor in GeoInformatics, and Director of SpceTimeLab for Big Data Analytics (<http://www.ucl.ac.uk/spacetimelab>), at University College London. Her research interests span network complexity, Geocomputation, integrated spatio-temporal analytics and big data mining (modelling, prediction, clustering, visualisation and simulation), with applications in transport, crime, health, social media, and environmental monitoring.

References

Baraglia R, Muntean C I, Nardini F M and Silvestri F (2013). LearNext: learning to predict tourists movements. *In Proceedings of CIKM*, 751-756.

Palma A (2008). A Clustering based Approach for Discovering Interesting Places in Trajectories. *Symposium on Applied Computing*, March 16-20, Fortaleza, Brazil. 863-868.

Schölkopf B and Smola A J (2002). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press.

Zhou C (2004). Discovering Personal Gazetteers: An Interactive Clustering Approach. *Geographic Information Science*, 266-273.