AGILE Pre-conference workshop
“Bridging space, time, and semantics in GIScience”
Wageningen, The Netherlands, 9 May 2017

Discovering Human Travel Behaviour in Urban Region

Meihan Jin
Naval Academy Research Institute in France
BCRM de Brest, Ecole navale
CC 600 - Lanveoc
F-29240
Brest, France
jmeihan@hotamail.com

Christophe Claramunt
Naval Academy Research Institute in France
BCRM de Brest, Ecole navale
CC 600 - Lanveoc
F-29240
Brest, France
christophe.claramunt@gmail.com

Abstract

As the booming of enormous moving data volume, it becomes an emergency to dig out valuable knowledge from the big data. This research proposes a travel complete behaviour framework approach in order to unveiling the people’s travel patterns behind those large data. To achieve that goal, a travel behaviour model is built as the basic knowledge acquirement principle and a series of well categorized analysing functions is designed on top of the model. Our approach contains data mining methods combining spatial and temporal analysis as well as semantic data discovery. Our work is testified on database system using real urban GPS data. The result of testing on the real data reveals certain travel patterns of citizens. The approach is proved to be applicable and inspired for multiple study realms..
Keywords: Spatio-temporal Database, Trajectory, Data Analysis

1 Introduction

With the broad application of location-based technology in worldwide, such as GPS (Global Positioning System), WiFi (Wireless Fidelity) and cellular network, an enormous booming of location based data occurs nowadays. GPS in the cars, cellular data exchange in smartphones and achievable WiFi in public regions, these continuously achieve the location data of their users.In fact, the large amount of data sources available increase the range of opportunities for geographical data analysis and mining, and this for many application domains. However, there is still a need to design and develop appropriate spatial data models as well as manipulation capabilities to take a full advantage of these new emerging big geographical data.

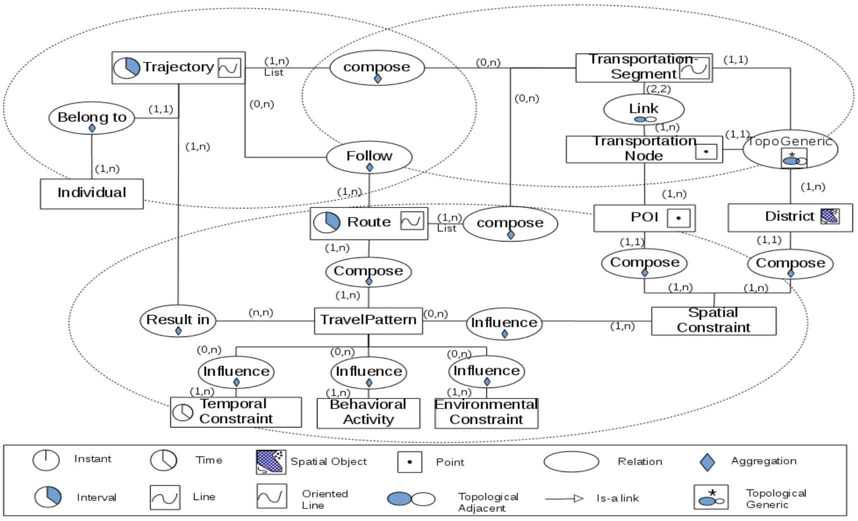
In this paper, we are aiming at discovering of human travel behaviour in the city realm by analysis on trajectory data using GIS methodology. Gaining the knowledge of that facilitates decision making in transportation, city planning, and traffic flow prediction. In order to achieve people's travel behaviour, we defined a system of analysis interfaces based on the human behaviour model we designed.

2 State of Art

According to the previous research, human's travel behaviour is worth of exploration and analysis, and existence and the existence of certain travel pattern has been proved (Hanson & Huff,1986; Jiang et al, 2012; Pentland & Liu, 1999). It is believed that analysis and prediction of individual's travel behaviour is a significant approach for urban transportation planning and policy analysis (Kitamura, 1988). For understanding human travel patterns and behaviours, various attempts have been made and mostly aggregation method is usually used to find the similarities in human behaviours. Modelling such notion of similarity is commonly based on a semantic distance measure (Parent et al., 2013). For example, Laube introduced several algorithms to detect movement patterns of dynamic objects, based on a matrix of motion attributes (Laube, 2005). The maximal length of the cumulated distances to the canter is the primary constraint used. Another clustering algorithm oriented to the representation of human trajectories in an urban network has been proposed by Buchin et al. (2011). This approach is based on the Fréchet distance derived from between two sub-trajectories computed under a constant time so as to discover a common movement pattern of a group of entities. In the work developed by Hung and Peng (2009), trajectories are modelled as Probabilistic Suffix Trees (PST), and a clustering algorithm is based on a distance function applied to the PST. The K-th distance derived from the K-nearest neighbour (KNN) (Cover & Hart, 1967) theory can also act as a foundation of a clustering algorithm (Ong et al., 2010).

Although previous work has made efforts to behaviour excavated, there is still necessity to develop a general model and system of data analysis of travel behaviour in order to derive a better understanding at the local and aggregated levels of the movement distributions in space and time.

Figure1 General Schema of Travel Behaviour Model



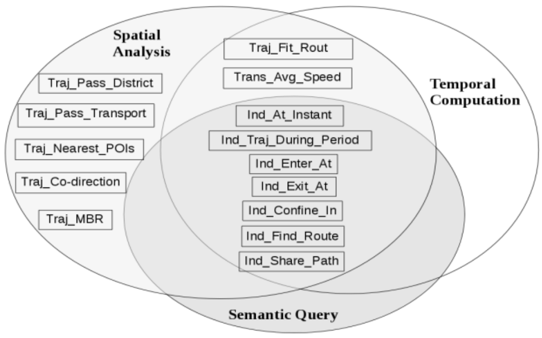
3 Human Travel Behaviour Modelling

Given the uniqueness and complexity of spatial and temporal data, the our modelling approach is based on the principles of the MADS(for Modelling of Application Data with Spatio-temporal features) model developed by Spaccapietra and Parent (2006). A concept model of travel behaviour is established on foundation of trajectory data, which is a type of spatio-temporal data produced from moving objects. In general, the travel behaviour model includes transportation network module, trajectory module and travel pattern module which contains. In the model, we consider the Individual who performs the traveling as the owner of the behaviour and the trajectory that the behaviours are extracted from. Despite of modelling trajectory along with the transportation network in the city, in order to extract travel patterns, we also modelled the object Route defined as sequence of TransportationSegments with temporal duration which the Individual appeared to have repeated more than one time. We believe that the Route is worthy of study and is the key to find out the travel habits of an Individual. For the travel pattern model, the potential factors that affect the travel decisions including spatial and temporal constraints are defined.

4 Function and Query Design

Database is one of the principal methods to preserve and manage spatio-temporal data, as a result that most of database systems nowadays can well support spatial data type and spatial data management, such as Oracle Spatial 11g, PostgreSQL/PostGIS, ArcGIS Geodatabase and more recently, MySQL. Therefore, for executing travel behaviour analysis, an object-relational travel behaviour database is designed based on the human travel behaviour model.

Figure2 Categories of Functions



While the queries for discovering knowledge of moving patterns are usually spatio-temporal related, which are relatively complicated comparing to non-spatial or non-temporal data queries, customized functions is necessary to make queries applicable and more effective. Thus, we have designed functions to solve spatio-temporal problems in analysis for travel behaviour. To begin with, categories of functions and queries should be assigned on top of well understand of travel behaviour model and definition. According to Dodge et al.(2008), people’s travel patterns can be classified into three categories: spatial patterns, temporal patterns and spatio-temporal patterns. The design of the functions and queries are based on these three categories. The definitions of the functions are partly enlighten by the work of S Sideridis et al.(2016) and the spatio-temporal computation and topological operators by Wu et al.(2014).

<u>Function 1 Ind_At_Instant</u>

The function Ind_At_Instant is a spatio-temporal function aiming at finding out where is the *Individual* at a given time.

Signature: *Ind_At_Instant: Individual x Timestamp-> Geometry(POINT)*

Function 2 Ind_Traj_During_Period

The function Ind_Traj_During_Period searches for the part of the *Trajectory* performed by a given *Individual* within a given time interval.

Signature:*Ind_Traj_During_Period: Individual x Time_Interval-> geometry(LINESTRING)*

Function 3 Ind_Enter_At

Traj_Enter_At is a spatio-temporal function that searches for the time instant when possibly a given *Individual* enters a spatial entity denoted by a geometry.

Signature:*Traj_Enter_At: Individual x Geometry-> Timestamp*

Function 4 Ind_Exit_At

The function Ind_Exit_At is relative to the function Ind_Enter_At. The objective is to search for the time instant when the given *Individual* possibly leaves a given geometry.

Signature:*Traj_Exit_At: Individual x Geometry-> Timestamp*

Function 5 Ind_Confine_In

The function Ind_Confine_In searches for the individual *Trajectories* located in a spatio-temporal Minimum Bounding Box (MBB) for a given time interval.

Signature:*Ind_Confine_In: Geometry(Polygon) x Time_Interval -> Set(Trajectory,Individual)*

Function 6 Traj_MBR

The function Traj_MBR is a spatial function that returns the minimum bounding rectangular(MBR) that the *Individual(s)* passed through while performing one or more trajectories.

Signature:*Traj_MBR: set(Trajectory) -> Region*

Function 7 Traj_Pass_District

The function Traj_Pass_District find out if a given *Trajectory* has passed by a given *District*. sSignature:*Traj_Pass_District: Trajectory x Polygon -> Boolean*

Example Query 8:

"Find the trajectories that cross the District Chaoyang identified with an Id 335"

Select trajectory.id, starttime, endtime from TrajectoryTable traj, District district where district.id = 335 and Traj_Pass_District(traj.id, district.geom);

Function 8 Traj_Pass_Transport

The function Traj_Pass_Transport determines if a given *Trajectory* has passed through a *TransportationSegment* with a given name.

Signature:*Pass_Transport: Trajectory x (LineString) -> Boolean*

Function 9 Traj_Nearest_POIs

The Traj_Nearest_POIs is a function that searches for the points of interest close to a given point.

Signature: *Traj_Nearest_POIs: Point x Distance -> set(POI)*

StartTraj: Trajectory -> Point

EndTraj: Trajectory -> Point

Traj_Nearest_POIs: StartTraj□ EndTraj(Trajectory) x Distance -> set(POI)

The function result gives a set of *POIs*.

Function 10 Traj_Direction

The function Traj_Direction is used to calculate the direction of a *Trajectory*.

Signature: *Traj_Direction: Trajectory -> Bearing*

Function 11 Traj_Co-direction

The Traj_Co-direction function evaluates if two given *Trajectories* are moving along a similar direction.

Signature: *Traj_Co-Direction: Trajectory x Trajectory -> Boolean*

Function 12 Ind_Find_Route

The Ind_Find_Route is a function whose objective is to extract the *Route(s)* performed by one-to-many *Individuals*.

Signature: *Ind_Find_Route: Set(Individual) -> Set(Route)*

Function 13 Traj_Fit_Route

The function Traj_Fit_Route determines whether a given *Trajectory* pass through a given *Route*.

Signature: *Traj_Fit_Route: Trajectory x Route -> Boolean*

Function 14 Ind_Share_Path

The function Ind_Share_Path is a function that searches for the *Path(s)* that the *Individual* has already repeated more than one time before, in which A *Path* is defined as a list of *TransportationSegments* that the *Individual* has repeated.

Signature: *Ind_Share_Path: set(individual) -> set(path)*

Function 15 Trans_Avg_Speed

The function Trans_Avg_Speed is to derive the average speed of a TransportationSegment.

Signature: *Trans_Avg_Speed: set(TransportationSegments) x Timestamp -> Speed*

5 Query Performance and Discussions

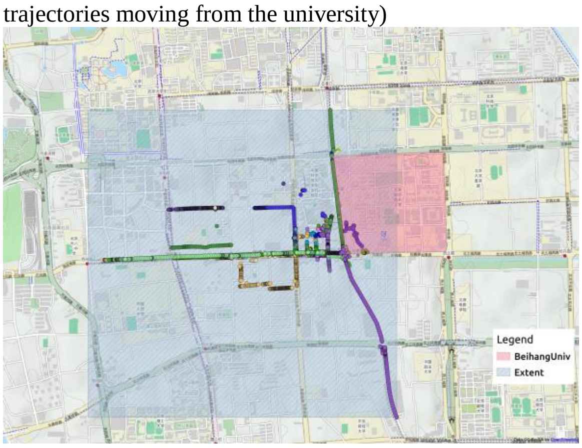
In this section, we introduce the sample queries depend on the functions discussed before. The implementation of the travel behaviour database and the functions is performed in PostgreSQL/PostGIS which is a popular open source database system with full package of basic spatial management interface. The dataset we use to test the queries and functions is real citizens' GPS data of Beijing City(Yu et,al. 2008, 2009, 2010) . The representative queries are demonstrated as follows:

Query 1 :

“Find the range of the trajectories derived from the Beihang University in the afternoon.”

select Traj_MBR(select array(select traj.id from TrajectoryTable traj, LandUse land where land.name = 'Beihang University' and st_within(startpt, land.geom) and

Figure 3: The Travel Range Derived from Beihang University (Region marked as light blue, the points are the trajectories moving from the university)



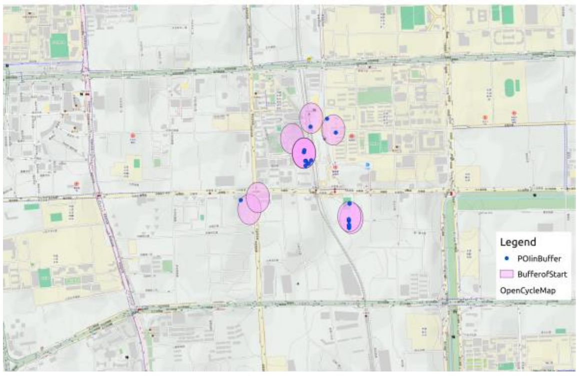
extract(hour from starttime) between 13 and 18;

This query is aiming at searching for the range of trajectories derived from a certain POI(Point of Interest) or a certain area, in this case, is a university. By finding out the range of the travel oriented from a given place, it is easy to know the moving trend of people who leave the place. In our case, by finding out the range of trajectories that start from the Beihang University in the afternoon, we can tell that people leave the university usually tend to move in south-west direction. Considering the POI object is a university, the afternoon might be the time for people finish the class and go outside the campus. Thus, it is assumed that people who work or study in the university might have bigger chance to live in the south-west from the university.

Query 2:

“Find the POIs within a distance of 50 meters destinations of the individual named 'Li Ming' traveled in the moring from 7:00 to 9:00”

Figure 4 POIs near Start Point



select * from Traj_Nearest_POIs ((select StartPoint from TrajectoryTable traj, IndividualTable indiv where indiv.name = ‘Li Ming’ and traj.individualid = indiv.id and extract (hour from timestamp starttime) between 7 and 9), 50);

As shown in Figure4, with the query result, it is obvious that the morning destination is usually confined in a small region. Given that morning travel behaviour is with high probability of going to work activity, we can estimate that this individual’s working place is in this area. Thus, the POIs in the region near destination are potential working spot. After classified these POIs, the result shows that the most possibility of this individual’s working place is a company because that high proportion of those POIs are of the type of company.

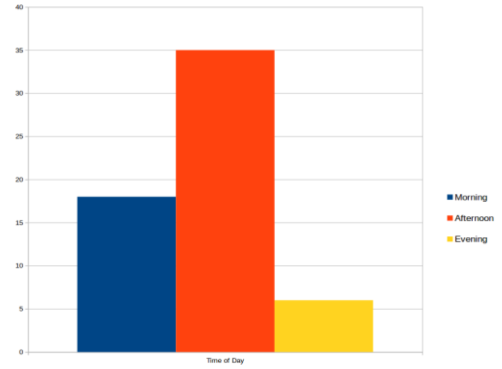
Query 3:

“How many times did 'Li Ming' passed the road “三环” in the morning, in the afternoon and in the evening individually?”

Select trajectory.id, starttime, endtime from TrajectoryTable traj, TransportationSegmentTable trans where trans.name ='三环' and tran.type = 'road' and Traj_pass_transport(traj.id, trans.geom) and extract(hour from starttime) between 7 to 12;

This query intends to dig the travel habits of an individual. By finding out the frequent path the individual takes different hours during one day may help to better understanding the travel behaviours and the traffic conditions. In Figure 4, it is obvious that the individual passed through the road called ‘三环’ most frequently in the afternoon and the least frequently in the evening. This information indicates that individual might have regular routine that passed by the road in the afternoon.

Figure 5 Chart of Numbers of Li Ming's Trajectories passed by Road '三环'



Query 4:

"Find the paths that Li Ming and Wang Lei has shared."

Select Ind_Share_Path(array(a.id,b.id)) from IndividualTable a, IndividualTable b where a.name = 'Li Lei' and b.name = 'Wang Lei'

Sources for the tables have to appear underneath them, justified and preceded by "Source:".

This query is aiming at finding the similarities of travel routines between individuals. In this query, the Path is an

Figure6 Shared Path of Li Ming and Wang Lei

object type in the travel behaviour model, which refers to the sequences of transportation segments that the individual usually take. Thus, the query indeed helps to figure out the overlapped travel routines between the two persons. Furthermore, this will make great help to find out the potential relationships between different citizens.

6 Conclusion

Human travel behaviour is a significant issue that can aid multiple realm of research. In our approach, we made an effort to discover human travel behaviour on foundation of semantic modelling. In order to pursue our goal of travel behaviour understanding, a system of travel behaviour analysis functions are designed and established A complete travel behaviour database is built in database system based on the travel behaviour model we designed. And the functions are performed on the travel behaviour database with real trajectory data.

As a result of applying the approach with real data, the individual travel routine is successfully discovered and aggregated in space and time. It proves that not only our framework can help revealing the travel behaviour, but also contain the ability to discover relationship between humans and foresee the potential traffic problem. With continuing developing our work in the future, prediction of human travel and transportation condition is possible.

References

Buchin, K., Buchin, M., Gudmundsson, J., Löffler, M., & Luo, J. (2011). Detecting commuting patterns by clustering subtrajectories. International Journal of Computational G

Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. IEEE transactions on information theory, 13(1), 21-27.

Erwig, M., Gu, R. H., Schneider, M., & Vazirgiannis, M. (1999). Spatio-temporal data types: An approach to modeling and querying moving objects in databases. GeoInformatica, 3(3), 269-296.

Güting, R. H., Böhlen, M. H., Erwig, M., Jensen, C. S., Lorentzos, N. A., Schneider, M., & Vazirgiannis, M. (2000). A foundation for representing and querying moving objects. ACM Transactions on Database Systems (TODS), 25(1), 1-42.

Hanson Susan, Huff James. (1986). Classification issues in the analysis of complex travel behavior. Transportation, 13(3), 271-293.

Hung, C. C., & Peng, W. C. (2009, November). Clustering object moving patterns for prediction-based object tracking sensor networks. In Proceedings of the 18th ACM conference on Information and knowledge management (pp. 1633-1636). ACM.

Jiang, S., Ferreira Jr, J., & Gonzalez, M. C. (2012, August). Discovering urban spatial-temporal structure from human activity patterns. In Proceedings of the ACM SIGKDD international workshop on urban computing (pp. 95-102). ACM.

Kitamura, R. (1988). An evaluation of activity-based travel analysis. Transportation, 15(1-2), 9-34.

Laube, P., van Kreveld, M., & Imfeld, S. (2005). Finding REMO—detecting relative motion patterns in geospatial lifelines. In Developments in spatial data handling (pp. 201-215). Springer Berlin Heidelberg.

Ong, R., Wachowicz, M., Nanni, M., & Renso, C. (2010, December). From pattern discovery to pattern interpretation in movement data. In 2010 IEEE International Conference on Data Mining Workshops (pp. 527-534). IEEE.

O'Sullivan, D. (2005). Geographical information science: time changes everything. Progress in Human Geography, 29(6), 749.

Parent, C., Spaccapietra, S., & Zimányi, E. (2006). Conceptual modeling for traditional and spatio-temporal applications: the MADS approach. Springer Science & Business Media.

Parent, C., Spaccapietra, S., Renso, C., Andrienko, G., Andrienko, N., Bogorny, V., ... & Theodoridis, Y. (2013). Semantic trajectories modeling and analysis. ACM Computing Surveys (CSUR), 45(4), 42.

Pentland, Alex, & Liu, Andrew (1999). Modeling and prediction of human behavior. Neural computation, 11(1), 229-242. international workshop on urban computing. ACM, 2012. p. 95-102.

Spaccapietra, S., Parent, C., Damiani, M. L., de Macedo, J. A., Porto, F., & Vangenot, C. (2008). A conceptual view on trajectories. Data & knowledge engineering, 65(1), 126-146.

Sideridis, S., Pelekis, N., & Theodoridis, Y. (2016). On querying and mining semantic-aware mobility timelines. International Journal of Data Science and Analytics, 2(1-2), 29-44.

Wu, J., Claramunt, C., & Deng, M. (2014, November). Modelling movement patterns using topological relations between a directed line and a region. In Proceedings of the 5th ACM SIGSPATIAL International Workshop on GeoStreaming (pp. 43-52). ACM.

Yu Zheng, Lizhu Zhang, Xing Xie, Wei-Ying Ma. Mining interesting locations and travel sequences from GPS trajectories. In Proceedings of International conference on World Wild Web (WWW 2009), Madrid Spain. ACM Press: 791-800.

Yu Zheng, Quannan Li, Yukun Chen, Xing Xie, Wei-Ying Ma. Understanding Mobility Based on GPS Data.