A Framework for Mining Semantic-Level Tourist Movement Behaviours from Geo-tagged Photos

Guochen Cai, Kyungmi Lee, and Ickjai Lee[✉]

College of Business, Law and Governance, Information Technology Academy,
James Cook University, Cairns, QLD 4870, Australia
`ickjai.lee@jcu.edu.au`

Abstract. This study investigates tourist movement patterns on the type of place semantic-level. We extract the semantic common movement patterns that a group of tourists have similar movement trajectories on the semantic level, and find out semantic trajectory patterns which are sequences of the type of place objects with transit time. Using real geo-tagged photos, we find out interesting common movement patterns and trajectory patterns. These results provide richer information and understanding of tourist movement behaviour on the type of place semantic-level.

Keywords: Semantics · Trajectory data mining · Movement behaviors · Geo-tagged photos

1 Introduction

Advances in Web 2.0/3.0 technologies promote a rapid increase of user-generated contents. These contents reflect their actual lives, ideas and activities. User-generated contents can be associated with location and time information through geo-tagging services the websites provide. These geo-referenced contents indicate tourists' movement patterns that help understand their behaviours.

This study aims to provide a framework to extract semantic-level tourist movement behaviours from geo-tagged contents. Previous studies mainly focus on spatial level movement behaviours by using only geometric features of trajectory data. These spatial level behaviours fail to reflect contextual semantic-level patterns. A semantic-level movement behaviour is a trajectory behaviour whose predicate bears on some contextual data [8] such as the type of place semantic information. This kind of behaviour provides detailed and fine-tuned information in the contextual semantic-level, and can help specific applications provide targeted services to tourists based on the semantic-level behaviours, like the type of place context semantics.

2 Literature Review

On-line user-generated contents are being constantly generated providing a data-rich environment for various tourist behaviour mining activities [5]. Especially,

the social media data associated with geographic location and time information, that are annotated by geo-tagging services, are useful sources for spatio-temporal patterns of tourist movements.

One recent research is to find tourist spatial travel behaviours of location preferences. [7] discovers POI that a high density of tourists visits, and further finds out their visitation association rules. Another recent novel research is to learn tourist spatio-temporal dynamics from geo-tagged social media data. The geo-tagged data form tourist trajectories, spatio-temporal movements, by chronologically connecting all geo-tagged data. These trajectory data are then used to analyse tourist movement behaviours. [5] uncovers dynamic visitor traffic flow, inbound and outbound trajectories between cities. Another popular travel behaviour is the spatial movement pattern that a group of tourists share. [2] extracts tourist spatio-temporal sequential patterns whilst [9] mines tourist popular travel routes. Each route is a sequence of spatial region of areas. A popular travel route shows the common path that a group of tourists share.

A common drawback of these previous studies is that they only focus on spatial level travel behaviours that are based on geometric feature of spatial information. This spatial geometric feature only is not sufficient for some specific applications [8]. For instance, some applications require more meaningful background semantic information, like movement behaviours between different types of place, that the geometric feature merely cannot provide. The semantic-level behaviours provide more detailed and fine-tuned information about tourist movements on semantic-level that are valuable to domain experts like tourism industry and urban management.
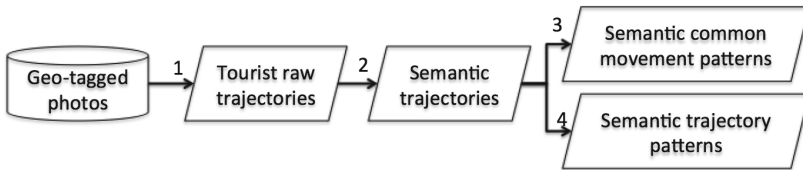

Fig. 1. Conceptual framework.

3   Framework and Approaches

Figure 1 shows our proposed framework. From geo-tagged photos, collected from a photo-sharing platform Flickr (https://www.flickr.com/) in this study, the framework firstly creates raw trajectories and then classifies them into tourist and non-tourist trajectories in Step 1 based on the time span of trajectory. The time span of a trajectory is calculated by using the time gap between the last photo and the first photo. We consider a photo-taker as a tourist if the time span of trajectory is less than 31 days. We use these tourist trajectories in this study. In Step 2, we enrich trajectories with semantics and transform trajectories into semantic trajectories. A semantic trajectory is a sequence of semantic

annotations. Using the semantic Region-of-Interest (RoI) mining method [3], we detect semantic RoIs from trajectories. Each semantic RoI is annotated with a type of place semantics. We use this type of place as the basic semantics of travel to learn tourist movement behaviours on the type of place semantic-level. As a result, a basic semantic trajectory is a sequence of type of place annotations. In this study, we enrich basic semantic RoIs with more city, temporal and weather condition annotations by using geographic information database, visit time of RoI and weather observation database respectively, and RoIs become multidimensional. In particular, temporal semantic contains two features: date type and day time. Date type is the day of week, weekday and weekend, and day time is the time period of the day. Step 3 requires a clustering approach to group similar trajectories into the same common movement pattern. This framework adopts the EXTRACTDBCAN-Clustering method [1] to generate clusters from the ordering results and uses the SemT-OPTICS algorithm [3] to deal with enriched semantic trajectories. Since Steps 1–3 in our framework are based on [1,3], we recommend readers to refer to them for more details.

A semantic trajectory pattern is a sequence of visited objects with the transit time between two neighboring objects. Step 4 is to find these semantic trajectory patterns. The proposed method adopts $\mathcal{TAS}$ algorithm [4] which is a projection-based method built on the PrefixSpan method [6] designed for sequential patterns. $\mathcal{TAS}$ algorithm uses T-sequence data type instead of normal sequence used in projections. A T-sequence is a projected sequence enriched with an annotation sequence where the annotation sequence includes records of occurrences of the prefix in the original sequence. Our method adopts the T-sequence data type, but we use a progressive increase approach to calculate frequent interval time and semantic trajectory patterns. In addition, the proposed framework utilities arbitrary combination of dimensions when it generates trajectory patterns. Semantic trajectories are associated with four additional semantic dimensions. We find not only the trajectory patterns associated with a set of all four dimensions, but also patterns with subsets of four dimensions.

4   Experimental Results

4.1   Dataset

In this study, we use real geo-tagged photos collected from Flickr for Queensland area in Australia for a period between April 2014 and March 2015. We collected 64,733 photo data, and generate 1404 valid raw trajectories that the length is greater than 1. Then we select the trajectories with time span less than 31 days as tourist trajectories. Finally, we obtain 770 tourist trajectories. To enrich additional semantic annotations to trajectories, We use the Australia gazetteer data and cities 1000 dataset from GeoNames (http://www.geonames.org/) as our geographic information database. For the weather information database, we use the the observation stations database and daily weather observation database from Bureau of Meteorology Australia (http://www.bom.gov.au/climate).

For parameters, the methods we used to extract tourist travel behaviour of movement patterns require several parameters, respectively. It is a non-trivial problem to choose best values of parameters that produce meaningful and insightful RoIs and patterns. The following default values were chosen for this particular experiment for experimental purposes. For semantic trajectory generation, the semantic RoI mining method relies on the minimum support ($MinSup$) value for a cell to become a RoI and also on the size of cell ($CellSize$) that is used to partition the study region. We choose a value of 0.004 which means 0.4 Km for $cellSize$ and a value of 0.007 which means 0.7% for $MinSup$. For semantic trajectory patterns, the semantic trajectory pattern mining algorithm requires parameters $MinSup$ and time tolerance (tau) which is the acceptable range for a time interval. We use a value 2 days for tau. For SemT-OPTICS algorithm in semantic common movement pattern mining, we use all basic and additional semantics in this study. We select the basic semantics PLACE_TYPE as default compulsory dimension, and other four semantics, CITY, DAY_TYPE, DAY_TIME and WEATHER, as optional dimensions. A set of compulsory dimensions plays an important or crucial role in patterns. For simplicity, we set each optional dimension with an average weight value, which is used to compute the similarity between trajectories, a value of 0.25. We set values for ohter parameters element matching score threshold ($elematThreshold = 0.3$) and ratio threshold ($rThreshold = 0.3$) for our similarity method, basic parameters $minPts = 4$ and $epsilon = 1$ that OPTICS algorithm requires, and set a value for parameter $epsilon = 0.5$ for extractDBSCAN-clustering method.

We find 72 semantic RoIs. A semantic RoI is a spatial region with a type annotation indicating the type of the place in the region. Each type of place annotation is represented as a feature code defined in Geonames geographic database. We obtain 399 semantic trajectories and 204 valid semantic trajectories (length > 1). A sample of final semantic trajectory is presented as below.

$$< (HTL_{[GoldCoast][weekday][Clear][evening]})(BCH_{[GoldCoast][weekday][Clear][dawn]}) >$$

4.2   Semantic Common Movement Patterns

In this section, we present some tourist semantic common movement patterns. Applying the SemT-OPTICS algorithm on the semantic trajectory dataset and extractDBSCAN method to ordering list, we found 5 clusters. We choose the first semantic trajectory cluster to represent the common movement pattern of that cluster. Table 1 shows two common movement patterns. The general form of pattern is a semantic trajectory. The first common movement pattern is a trajectory of starting at a hotel in Gold Coast and then going to a populated place and a park in Brisbane and visiting a hotel and a beach in Gold Coast. All of these visits are in clear weekday, but are at different day time that first four visits are at dawn in the day whilst the last visit is in the afternoon. This pattern shows a tourist common movement between Gold Coast and Brisbane which are nearby cities. Another common movement pattern is in Cairns that tourists start from a hotel in the morning and then go to the pier (where fleet

Table 1. Semantic common movement patterns (HTL: Hotel, RSTN: Railroad station, PPLX: Section of populated place, BCH: Beach, PRK: Park, PIER: Pier).

Cluster	Common movement pattern
cluster 1	$[(\text{HTL}_{[GoldCoast][weekday][Clear][dawn]}) \rightarrow$ $(\text{PPLX}_{[Brisbane][weekday][Clear][dawn]}) \rightarrow$ $(\text{PRK}_{[Brisbane][weekday][Clear][dawn]}) \rightarrow$ $(\text{HTL}_{[GoldCoast][weekday][Clear][dawn]}) \rightarrow$ $(\text{BCH}_{[GoldCoast][weekday][Clear][afternoon]})]$
cluster 2	$[(\text{HTL}_{[Cairns][weekday][Lightrain][morning]}) \rightarrow$ $(\text{PIER}_{[Cairns][weekday][Lightrain][morning]}) \rightarrow$ $(\text{HTL}_{[Cairns][weekday][Lightrain][evening]})]$

stations are) and back to hotel in the evening at last. This pattern supports a popular full-day tour to Great Barrier Reef (departing from the fleet station in the early morning and returning late in the afternoon).

4.3   Semantic Trajectory Patterns

We present several typical tourist travel behaviours of semantic trajectory patterns in this section including basic semantic trajectory patterns and multidimensional semantic trajectory patterns. A semantic trajectory pattern shows a sequence of visited types of place with transit time that occurs in a density of tourist trajectories. Multidimensional semantic trajectory patterns contain more semantics and provide much richer information about tourists frequent trajectory patterns. Due to the limitation of space, we list some multidimensional semantic trajectory patterns for the basic pattern PRK $\xrightarrow{[0,8]}$ HTL associated with various combinations of additional dimensions including dimension WEATHER and combinations of dimensions DAY_TYPE, CITY and WEATHER. We observe that additional semantics provide more information to basic patterns. For a combination of CITY dimension, we know this pattern is in Gympie, a regional town in the Wide Bay-Burnett region of Queensland. For combinations of dimensions DAY_TYPE, CITY and WEATHER together, we obtain more day type and weather information about the semantic trajectory pattern (Table 2).

Table 2. Multidimensional semantic trajectory patterns for basic pattern: PRK $\xrightarrow{[0,3]}$ HTL.

Combination of dimensions	Semantic trajectory patterns
WEATHER	$\text{PRK}_{[clear]} \xrightarrow{[0,8]} \text{HTL}_{[clear]}$
DAY_TYPE, CITY, WEATHER	$\text{PRK}_{[weekday][Gympie][clear]} \xrightarrow{[0,8]}$ $\text{HTL}_{[weekday][Gympie][clear]}$

5   Conclusions

We presented a study of converting geo-tagged photos into trajectories and further into semantics enriched trajectories to extract semantic-level tourist movement patterns. The proposed framework enabled us to find out tourist common movement patterns on the type of place semantic-level with additional city, temporal and weather condition semantic information. In addition, we discover tourist semantic trajectory patterns, frequent visitation sequences of the type of place with transit time information. The trajectory patterns are also associated with various combinations of additional semantics that supply further insights into movements in different semantics contexts. Overall, these detected patterns are indicative of semantic-level tourist movement behaviours, and provide richer knowledge about travel movements on semantic-level.

References

1. Ankerst, M., Breunig, M.M., Kriegel, H.P., Sander, J.: Optics: ordering points to identify the clustering structure. ACM SIGMOD Rec. **28**(2), 49–60 (1999)
2. Bermingham, L., Lee, I.: Spatio-temporal sequential pattern mining for tourism sciences. Procedia Comput. Sci. **29**, 379–389 (2014)
3. Cai, G., Lee, K., Lee, I.: Discovering common semantic trajectories from geo-tagged social media. In: Fujita, H., Ali, M., Selamat, A., Sasaki, J., Kurematsu, M. (eds.) IEA/AIE 2016. LNCS (LNAI), vol. 9799, pp. 320–332. Springer, Heidelberg (2016). doi:10.1007/978-3-319-42007-3_27
4. Giannotti, F., Nanni, M., Pedreschi, D.: Efficient mining of temporally annotated sequences. In: SDM, pp. 348–359. SIAM (2006)
5. Girardin, F., Dal Fiore, F., Blat, J., Ratti, C.: Understanding of tourist dynamics from explicitly disclosed location information. In: Symposium on LBS and Telecartography, vol. 58. Citeseer (2007)
6. Han, J., Pei, J., Mortazavi-Asl, B., Pinto, H., Chen, Q., Dayal, U., Hsu, M.: Prefixspan: mining sequential patterns efficiently by prefix-projected pattern growth. In: proceedings of the 17th International Conference on Data Engineering, pp. 215–224 (2001)
7. Lee, I., Cai, G., Lee, K.: Mining points-of-interest association rules from geo-tagged photos. In: 2013 46th Hawaii International Conference on System Sciences (HICSS), pp. 1580–1588. IEEE (2013)
8. Parent, C., Spaccapietra, S., Renso, C., Andrienko, G., Andrienko, N., Bogorny, V., Damiani, M.L., Gkoulalas-Divanis, A., Macedo, J., Pelekis, N., et al.: Semantic trajectories modeling and analysis. ACM Comput. Surv. (CSUR) **45**(4), 42 (2013)
9. Zheng, Y.T., Zha, Z.J., Chua, T.S.: Mining travel patterns from geotagged photos. ACM Trans. Intell. Syst. Technol. (TIST) **3**(3), 56 (2012)