Mining Resident Activity Pattern Based on Spatio-temporal Trajectory Data

Jian Sun
Essence Information Technology Co.,Ltd.
Hangzhou, China
whoxjn6301@126.com

Zhentao Zhang
Essence Information Technology Co.,Ltd.
Hangzhou, China
zhentaozhang@126.com

Haibo Chen*
School of Science Hangzhou
Zhejiang Sci-Tech University
Hangzhou, China
* Corresponding author: chenhaibo@zstu.edu.cn

Daolei Liang
School of Science Hangzhou
Zhejiang Sci-Tech University
Hangzhou, China
liangdaolei@zstu.edu.cn

***Abstract*—The research on residents' activity pattern based on spatio-temporal trajectory data is helpful to the optimization of urban operation. At present, one of the difficulties in the research is how to determine the regularity of residents' activities when the labeled samples are sparse. We propose an improved periodic decision algorithm of sliding window, combined with feedforward neural network to search outlier activity patterns. Experimental results show that our method can effectively classify the overall travel features and quantify individual activity abnormalities.**

Keywords-residents activity pattern; periodic algorithm; trajectory; outlier patterns

I. INTRODUCTION

The travel features analysis based on the activity trajectory of urban residents is responsible for extracting the individual behavior from the time-space series of personnel, and for measuring periodicity, outlier, clustering, prediction, etc., so as to provide personalized services to residents or improve urban operation efficiency[1]. Supervised learning method needs to reasonably classify residents' activity patterns and label them on the existing data[2]. However, due to the diversity of urban residents' activities and the differences between cities, the effectiveness and generalization of learning are difficult to be achieved. We propose a new method based on improved SMCA, combined with feedforward neural network to solve the above shortcomings. The innovations of our method are as follows:

1) Combined with vector map, we carry out feature engineering on the raw trajectories, and vectorize the position set of dynamic length.

2) For each dimension of the activity vector, we use the improved SCMA algorithm based on sliding window to calculate its periodicity. The moving period features of all residents are clustered to obtain the feature class. In this way, the label of residents' movement feature is obtained, and this labeled class varies with different urban trajectory data which is adaptive.

3) We use three-layer feedforward network to train the labeled data, and get the supervised learning model. We use clustering model for outlier detection and feedforward neural network for behavior feature classification. Experiments show that the model can effectively detect the period of residents' behavior and find the outlier behavior.

The paper is organized as follows: section 2 summarizes the current research methods, our method is put forward in section 3, section 4 is the experiment and result analysis, and section 5 is the summary.

II. RELATED WORK

Spatio-temporal data has been ubiquitous in urban information applications. There are many literatures on pattern mining from the perspective of model, algorithm and application. Almost all studies can be attributed to the similarity (or difference) analysis of spatio-temporal characteristics.This similarity analysis can be a direct space-time distance analysis, such as dynamic time warning (DTW)[3-4] to calculate the geometric similarity under time constraint, can be semantic similarity, such as unsupervised clustering based on behavior and acceleration[5] , or be position based similarity, in which neighborhood of every point is considered[[6].

Clustering, learning and prediction based on the above similarity measures are the research focus of spatio-temporal mining. Spatio-temporal pattern clustering[7-11] can be divided into event based , geo-reference based, moving object based, trajectory based, semantic based. In semantic-based mining, domain-specific data is considered in algorithms. of spatio-temporal clustering analysis.

At present, most of the semantic information is gesture information[12], such as the direction and speed obtained by the trajectory, and there is also a small part of event information, such as stop point, point of interest[12], etc. However, very few domain semantics are used. Domain semantics refers to information related to activity goals, such as goods carried by trucks, ship operation types, pedestrian travel purposes, etc. These information lacks labels and cannot be applied in learning, which is also one of the focus problems.

Spatio-temporal pattern mining[13] is a natural result and important application after similarity detection and cluster analysis. This includes the periodicity, outlier and synergy of

2473-3547/21/$31.00 ©2021 IEEE
DOI 10.1109/ISCID52796.2021.00075

activity features. The current methods mainly focus on geometric feature patterns, lack of enough abstraction, and the high-level activity feature pattern mining methods still need to be improved.

III. SPATIO-TEMPORAL TRAJECTORY PATTERN MINING

A. *Semantic feature transformation*

According to the activity mode of residents' mobile location information mining, the location information needs to be transformed into semantic information. For urban residents, semantic information is their activity. We first give the definitions required.

Definition 1 spatio-temporal series: given a discrete-time set t, for $t \in T$, there is a two-dimensional random vector P_t, then all of the random vectors are called spatio-temporal random series, referred to as spatio-temporal series for short: $\{P_1,P_2,...,P_T\}$.

Definition 2 semantic series: given an event set M, (A, P) is an semantic event, where A is an event type, $A \in M$, $P = \{P_1, P_2,..., P_T\}$ is a spatio-temporal series. Given a sufficiently long spatio-temporal series $\{P_1, P_2,..., P_{T1}, P_{T1+1},..., P_{T2}, P_{T2+1},..., P_{TN}\}$, if a decomposition on it corresponds to an event, all decompositions form a semantic event series$\{(A_1,P_{T1}),(A_2,P_{T2}),...,(A_N,P_{TN})\}$.

Define 3 semantic time cycle and cycle patterns: For any semantic series, the given length is n, if $p,o \in Z^+$, $0 < p < n$, $0 < o < p$. C = (p, o, V) represents a similar cycle: (p * s + O), p is the cycle of event V, and o is the starting point of the cycle. For example, one-dimensional sequence t = 132113412341, C=(4,1,3).

For a random given semantic series of size n, if there are m C_I cycles with period P, $0 \le i \le m$. it is defined as a periodic pattern PC = (P, m,C).

The activities mode of urban residents are not standardized, so it is impossible to obtain the functional relationship between samples and labels through supervised learning. We first convert the original spatio-temporal trajectory into spatio-temporal semantic series, then classify these semantic series through spatio-temporal clustering, and the classification results are used as labels of spatio-temporal trajectory. Finally, the samples are trained by feedforward neural network to obtain the activity characteristic model of urban residents. The model can be used to classify trajectories and detect outliers of trajectories. The whole process is shown in Fig 1.

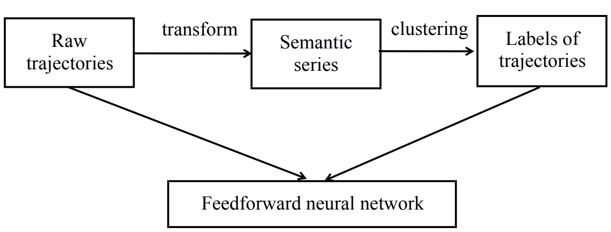

Figure 1 The process of trajectory pattern mining

By matching the trajectory coordinates with the map, the original trajectory sequence can be transformed into semantic sequence. For urban residents, semantic events include transportation, work, shopping, education, home, sports and entertainment.

B. *Periodic algorithm*

Firstly, the periodic pattern algorithm scans the semantic sequence obtained in the previous section, generates a location list, determines the time point according to the event, and establishes the sequence location shown in the following table 1.

TABLE I. SAMPLES OF EVENT SERIES

Event	*Time Series*
road	8,17,26,41,49,57,59,66,74,81,89,97,105,112
work	9,29,54,78,107,130,154,176,201,223,248
shopping	155,319,407,549,601,780,1027,1206,1334
education	211,370,484,610,695,819,964,1040,1191,1255
home	1,19,28,39,52,60,70,83,91,104,126,135,145,150

Each row in Table 1 is an event and its time series, in which each time refers to the number of hours from 0:00 on the start date.

SMCA algorithm[14] can find all possible periods of each time series, and selects the appropriate threshold to determine the period. However, due to the noise of real data, the periodicity of event patterns is often uncertain, which leads to the uncertainty of SMCA algorithm. This paper presents an improved periodic search algorithm of semantic time series based on sliding window.

```
Input:  semantic event series ES={t1,t2,...,tL}
Output: Periodic mean and variance
Procedure of periodic search:
    Initialize hash array count
    For i = L:2
       P := ti - ti-1
       count[P] += 1
    End For
    For P in count
        If count[P] >= min then
          For j=1:L
             pos = j%P
             If abs(j - result[pos].last - P) < th then
                 result[pos].S.push(j - result[pos].last - P))
                result[pos].last = j
             endif
          EndFor
        EndIf
    EndFor
    return argmin(result)
```

Figure2 Periodic search algorithm of semantic time series

The algorithm is shown in Figure 2. Periodic search algorithm of semantic time series scans the location list of the event for all events, and establish a sliding window with the maximum window size from the last time point. A new data structure is established to store the latest position and repetition times of the time point of the modified time series under cycle C. Where Q_n is the current time point of the time series and the last time point under the potential cycle of C. when the values of Q_n remainder J and remainder J are equal, define Q_n synchronization. If the interval between two synchronizations is j, then define that the two synchronizations are continuous. Divide the interval of continuous synchronization into effective segments, and use a new field to record the number of effective segments. Scan the time points of the time series in turn, and record the corresponding rep times under each possible remainder under cycle C. Repeat the above steps for the potential cycle.

C. Feedforward neural network

For each spatio-temporal trajectory data, we can get the periodic value of each events, and all the periodic characteristics of behavior form a multi-dimensional vector.

We use DBSCAN[15] algorithm to cluster the periodic vector of samples, and the classification obtained is used as the sample label. The assumption here is that there must be a certain periodic law in the activities of urban residents. This method effectively solves the type label problem of resident trajectory. Because the label is obtained by periodic algorithm, it will adapt to data changes. Finally, a supervised learning model can be obtained by training a three-layer feedforward neural network.

IV. EXPERIMENT

A. Experimental design

We selected the 300 day mobile location records of 5000 mobile phone users in Hangzhou, China as the original data.

We use the method in Section to convert spatio-temporal location data into semantic series, which includes the following discrete information:

Name: character id.

Time: the time is divided into several sections, including early morning (5-6:30), morning (6:30-8:30), morning (8:30-10:30), near noon (10:30-11:30), noon (11:30-13:00), afternoon (13-15:30), near evening (15:30-17:30), evening (17:30-19:30), night (19:30-22:00), late night (22-02), dawn (2-5).

Event: event type, including resident (the mobile phone is active but the location remains unchanged), silent (the mobile phone is inactive and the location remains unchanged), driving, bicycle and walking.

Category: category of activity places, including roads, work, shopping, education and training, home life, physical exercise, leisure places and entertainment places

Place name: site name. The specific name corresponding to each site type.

Further, according to the periodic search algorithm, the cycle and regularity of each recorded Road, work, shopping, education and training, home life, physical exercise, leisure and entertainment are obtained. The regularity is distributed in (0,1). The raw result is shown in Table 2.

TABLE II. RESULTS OF PERIODIC SEARCH ALGORITHM

Trajectory ID	***Road***	***Work***	***Shop***	***Edu***	***home***	***Sports***
1	6	20	142	0	9	12
2	6	20	130	0	10	0
3	9	18	62	159	10	277
4	14	26	301	79	7	0
5	8	24	167	115	5	314
6	8	20	160	0	12	0

We selected 4000 as the training set and the rest as the sample test set. Put in the neural network algorithm to find the weight to judge the regular activities, use the training set to generate the network model through the neural network algorithm, substitute the test set into the generated neural network operation to obtain the sample regularity results, and compare and calculate the network accuracy in the remaining samples, so as to obtain the effectiveness of the algorithm.

B. Experimental analysis

In order to verify the feasibility and effectiveness of the algorithm and accelerate the learning process, we choose the simplest single-layer neural network for experiments.

Firstly, we use the normalization of all samples introduced in section 3 to improve the operation accuracy. That is, in order to accelerate the learning process, when entering the transport layer, the feature vector is generally standardized first. The method of this paper is to normalize them between 0 and 1 by using the weighted sum of the inputs of sigmoid function.

In this experiment, the lower limit of regularity is selected as 0, because it is random simulation data, the possibility of regularity 0 exists, and it should be judged in real life. By continuously adjusting the parameters of the algorithm, it is found that when the periodicity is outside, the corresponding regularity will be greatly reduced. Therefore, after judging the periodic mode in section 3, it can be limited to the range of [5,22], which can improve the operation efficiency of the algorithm.

Input the training set samples, and use the neural network learning algorithm to generate the weight corresponding to each type of activity. The weight is updated once every iteration. The termination condition is to reach a preset number of cycles, or the predicted error rate is lower than a certain threshold, or the update of weight is lower than a certain threshold.

When all samples are input, the neural network learning process ends. The results are shown in table 3.

TABLE III. WEIGHTS OF EVENT

Event	*Weight*
road	-1.318
work	-28.389
shopping	-8.547
education	-7.336
home	-4.87
sports	-8.51
Entertainment	-2.43

In table 3, we can intuitively observe that among the weights for judging a person's regularity, the weight of work is very large, which can almost directly determine a person's overall regularity. The weight of the road is the lowest, which is also more consistent with the actual life. The road ratio of each person can not coincide periodically. There is a certain section that will be repeated in unit time. However, due to the participation of any other road, the regularity of the road will become very unstable, and the road will not be used to judge a person's regularity in place judgment, Instead, the regularity is judged according to the location type of the destination, so it is not very desirable to judge the regularity by road.

V. CONCLUSIONS

We implements a method of urban residents' activity mining based on spatio-temporal trajectory data. Aiming at the problem of lack of labels in residents' activity samples, the method generates sample labels through three steps: trajectory semantic transformation, cycle discovery algorithm of semantic events and unsupervised clustering of semantic event time series, which are used as the input of feedforward neural network, so as to realize a spatio-temporal mining algorithm of hybrid model.

This method only considers the periodicity of activity events and does not consider the sequence relationship of activities, so the classification label can not reflect all the characteristics of residents' activities, which is also the work to be completed in the next work.

REFERENCES

[1] *Wang S , Cao J , Yu P . Deep Learning for Spatio-Temporal Data Mining: A Survey[J]. IEEE Transactions on Knowledge and Data Engineering, 2020, PP(99):1-1.*

[2] *Atluri G , Karpatne A , Kumar V . Spatio-Temporal Data Mining: A Survey of Problems and Methods[J]. ACM Computing Surveys, 2017.*

[3] *Keogh et al. 2005. Exact indexing of dynamic time warping. Knowledge and information systems 7, 3 (2005), 358–386.*

[4] *[1] Li Z , Zhang F , Nie F , et al. Speed up dynamic time warpingof multivariate time series[J]. Journal of Intelligent and Fuzzy Systems, 2019, 36(3):2593-2603.*

[5] *Ansari M Y , Ahmad A , Khan S S , et al. Spatiotemporal clustering: a review[J]. Artificial Intelligence Review, 2019(2).*

[6] *W. Feng et al. 2015. STREAMCUBE: hierarchical spatio-temporal hashtag clustering for event exploration over the twitter stream. In Data Engineering (ICDE), 2015 IEEE 31st International Conference on. IEEE, 1561–1572.*

[7] *Deng M , Liu Q L , Wang J Q , et al. A general method of spatio-temporal clustering analysis[J]. Science China Information Sciences, 2013, 56(10):1-14.*

[8] *Z. Li. 2017. Semantic Understanding of Spatial Trajectories. In International Symposium on Spatial and Temporal Databases.Springer, Cham, 398–401.*

[9] *Zhang D , Lee K , Lee I . Periodic Pattern Mining for Spatio-Temporal Trajectories: A Survey[C]// International Conference on Intelligent Systems & Knowledge Engineering. IEEE, 2016.*

[10] E. Eftelioglu et al. 2016. Ring-Shaped Hotspot Detection. IEEE Transactions on Knowledge and Data Engineering 28, 12 (2016),3367–3381.

[11] M. L. Damiani. 2016. Spatial trajectories segmentation: trends and challenges. In Proceedings of the 5th ACM SIGSPATIAL International Workshop on Mobile Geographic Information Systems. ACM, 1–1.

[12] T. Cheng et al. 2014. Spatiotemporal data mining. In Handbook of Regional Science. 1173–1193.

[13] [1] Hui Y , Parthasarathy S . Mining Spatial and Spatio-Temporal Patterns in Scientific Data[C]// International Conference on Data Engineering Workshops. IEEE, 2008.

[14] Kuo-Yu Huang and Chia—Hui Chang, SMCA: A general model for mining asynchronous periodic patterns in temporal databases, IEEE Transactions 0n Knowledge and Data Engineering, June, 2005, 17(6)：774～784.

[15] Borah B , Bhattacharyya D K . An improved sampling-based DBSCAN for large spatial databases[C]// International Conference on Intelligent Sensing & Information Processing. IEEE, 2004.