Explain urban regional function

Submission

No Institute Given

Abstract. How to identify different functional regions and give them reliable explanations are the main objectives of this work.
(15–250 words.)

Keywords: Explanation · Urban computing · Online content · Sentiment analysis · Latent Factor · Matrix Factorization.

1 Introduction

Motivation. Urban computing, defined in [1] as a process of acquisition, integration and analysis of big and heterogeneous data generated by diverse sources in urban spaces.One of the most important data source is human mobility data, e.g. pick-ups and drop-offs of taxicabs. Such mobility data can contribute to improve living quality of residents, e.g. optimizing urban planning [2], easing traffic congestion [3], decreasing energy consumption [18], reducing air pollution [19] and so on. Therefore, mining and modeling mobility data has attracted attentions from both academia and industry.

Identifying functional regions has been a critical step towards efficient government administration and policy making for decades. In early work, a functional region was defined as a geographical region where the majority of local population recruit and are employed within the region [14]. Due to the availability of large-scale mobility data, an increasing amount of data-driven approaches [9] have been proposed. Most of them use clustering methods on commuting data, such as origin-destination pairs of labour market data [12] while others use remote-sensor image data [4]. Later work focuses on city-level functional region identification, where a functional region is a region in a city which supports different needs of peoples urban lives [15, 16]. To obtain accurate identification, recent research tends to use complex models for various form of data, e.g. latent factor models based on Dirichlet Multinomial Regression are applied on mobility data of taxicabs and points of interests (POIs) [15, 16], clustering algorithm based on the 'modularity function' is applied on telecommunication [13, 6],unsupervised semantic labeling framework based on the Latent Dirichlet Allocation is applied on remote-sensor data [5].

A severe drawback of existing research is the lack of explanation for the identifications of functional regions. Though existing applications [16] have achieved high identification accuracy,the complex nature of these models weakens the interpretability to end users and system designers.There is an emerging trend in studying explainable AI [34]. In geographical systems, explainable system is still in its initial stage [35, 36]. To the best of our knowledge, none of the previous urban functional identification provide explanation for the results of region segmentation and functional labels.

In this paper,to make the results more convincing, we build a system that provides explanations for our functional region clustering results. We build our system upon two data sets. One is mobility data 3 million including pick-up and drop-off coordinates and time,which provide detailed moving pattern for the dynamic urban system to discovering regions with different functions. And the other is comment data with shop address that reveal sentiment tendencies of the shop.We could exact labels as the form of 'character-opinion-sentiment' from and pick labels with strongest sentiment as our explanation. We combined the urban computing using mobility data with sentiment analysis using geo-content. As shown in Fig.1, our system not only segments regions of a city based on the functionals, but also delivers explanations of the functionals. To be specific, we first associate each functional with an urban activity, next we extract the most representative urban feature for each activity, finally we learn/highlight/assign the sentiment towards typical urban features for each functional region. For example,Fig.1 given a map of city A, reg region I corresponds to residential region while blue region II corresponds to commercial region. We found that people are most interested on traffic for a commercial region. And a residential region usually has sentiment positive on vanishing noise.

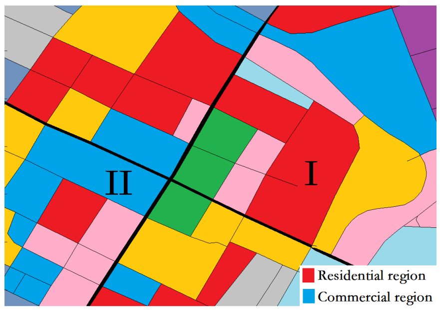
Challenge. In order to solve the problem,we face two challenges.

The first one of it is to extract the urban feature from comment data. Urban feature is a set of characters with special geographical attributes, e.g., traffic, noise, neighbors and so on. Part of our content data comes from social media without detailed coordinates.They are full of sentiment tendencies but have difficulty to induced to a particular region. The rest are comments of shops with the shop address.The feature extracted from it is more related to shop feature than urban feature. Due to the scenario, existing sentiment extraction model [37] can not be used directly in our models.In order to get the location information from these content data, features with typical urban scene are selected from all features extracted by a toolkit called Sentires [32]. These selected urban feature

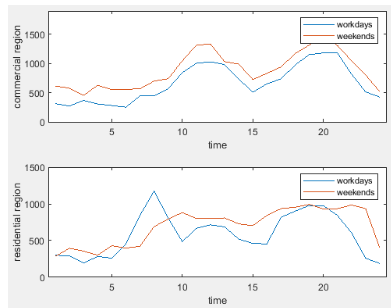
And the second challenge is how to explain the features. If the extracted urban features used for explanations can not correspond to the labeled activity used to identify functional regions, the explanations will not persuasively explain the reasons for the identification, or even will not be generated.

Contribution. Our paper has an outstanding result in following several aspect:

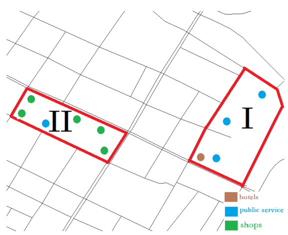
- To the best of our knowledge,we are the first to give explanation to urban functional region identification , which could improve the trustiness and satisfaction of user.
- We combined the sentiment and human mobility as explanation,find the urban feature within a region.



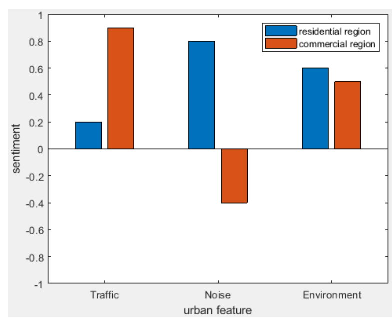
(a) An example of our identification result



(b) moving patterns



(c) Sentiment labels with their probability



(d) POIs in different functional region

Fig. 1. The explanation provided for functional region identification. Region I in Fig.1(a) is labeled as residential region while II is commercial region. Fig.1(b) revealed differences of the mobility patterns between residential region and commercial region,workdays and weekends. Fig.1(c) shows the distribution of POIs within the two regions,e.g. there is more shops in commercial region. And from fig.1(d) we find that people have a more interest in a positive sentiment in traffic for commercial and pay more attention to noise for residential region.

- The explanation is given by latent factor model. We use matrix factorization to decrease the sparseness of matrix.

Structure. The remainder of the paper is organized as follows. We give an overview for related work in Section 2. In Section 3, we introduce our novel framework as well as proving their feasibility. Section 4 presents the superior result in experiments of our models. Finally, we made a conclusion and looked forward to our future work in Section 5.

2 Related Work

Two lines of work are related to this paper: urban computing and online sentiment analysis .

2.1 Urban Computing

Urban computing [1] tackles the major issues that cities face by analyzing human mobility collected from different sensors. Major sources of human mobility data are check-ins in POI [26], pick-up and drop-off behavior of taxicabs [8, 3] in different locations and trajectories.

The simplest form of mobility data is **check-ins data**, which are collected from locating sharing services. Check-ins data usually includes a set of point revealing users' current location. Statistical association analysis is commonly conducted on check-ins. For example, radius of gyration is measured in [24], which is extended by combining with lexicon into demographics [26]. A few recent work adopts model based approaches, e.g. context-aware tensor factorization that take account of contextual factors that influence consumers refueling decision [18].

The second form of mobility data is **origin-destination pairs**, i.e. a pick-up point and a drop-off point of a taxi trajectory. Clustering methods including Newman modularity cluster algorithm [13] are applied in functional region identification [6]. Recently, latent factor models are proposed to treat a region as document and infer functional-specific [15].Latent activity is imported in topic model in [16] to define the specific functions of different regions.

Alternatives for origin-destination pairs are **trajectory data**. Applications on trajectory data include recommendation, e.g. to recommend more suitable place to drivers [8, 10] and traffic planning, e.g. to find reachable region within a given temporal period [20], or to predict travel time [21].

However, existing urban computing systems extensively rely on complex machine learning algorithms hence they act as blank-boxes for end users. The lack of explanation weakens the persuasiveness and trustworthiness of the system for users. Our work is to make up for this drawback by providing intuitive explanations of the results for users or system designers

2.2 Geographical Analysis of Online Sentiment

Recently, an emerging research interest is witnessed in exploring the geographical factors that affect online sentiment. Empirical studies have been conducted on large-scale human mobility data, such as check-in [26]and trajectory [25],to find the geographical content analysis with sentiment. Associations are found between online sentiments and geographical factors, e.g happy regions are more likely to connect with each other [23],a high check-in density region usually presents a more positive mode [26], the whole process and development of a organized movement could be tracked on the social media [22] and so on.

However, most existing work of this line employ simple statistical analysis to uncover the associations. Such a coarse-grained analysis is distorted by latent variables, such as activity of the region. Our work is the first to incorporate activity to obtain a fine-grained analysis.

2.3 Explanation for Complex Models

3 System Overview

We proposed a novel framework to make an explanation for the functional region identification.

4 Urban Feature Extraction

5 Explanation Model

$$R = b_u + U^T V + \beta()$$

5.1 Functional Region Identification

5.2 Preliminary: Latent Factor Model and Matrix factorization

To identify different functional region,we need to train a model

kernel function is a method that map low-dimension data into high-dimension data.It

5.3 Preliminary: Kernel Function

6 Experiments

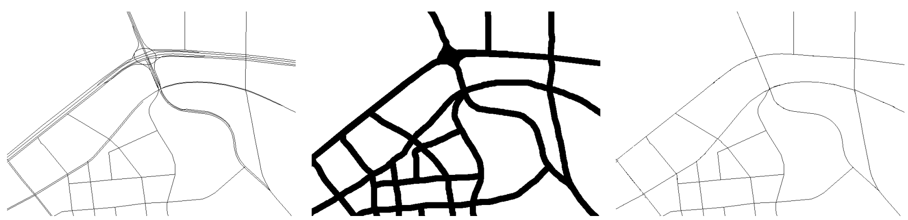
6.1 Data Set

The data set for our experiments including both mobility data and content data. The mobility data including the pick-up and drop-off time and coordinates in November of the year 2016 provided by DiDi, the biggest taxi platform in China.It contributed to the movement pattern of human mobility. And the online content is crawled from a website with many comments similar to Yelp called DazhongDianping and a social media named Weibo , which helps sentiment analysis of the regions.

6.2 Preparation

To form the basic region of city,we segment the urban area of city into region units by the major road network and make a map simplification. The longitude of map range is [103.93,104.21] and latitude range is [30.56,30.79], which covers the main area of a city. Raster-based model is more computationally efficient and succinct for territorial analysis, which is suitable for our map scenario. We downloaded the major road network of this region[1] and rasterized the area into a 2000×2400 grid. In the grid, the road network is converted to a binary image, as 1 stands for the road while 0 stands for the blank areas.

The main road data is present in Fig.2(a), including motorway, trunk, primary, secondary, tertiary and their links. But the Fig.2(a) is full of some unnecessary details, such as the lanes of a road and the overpasses, which disturb the distribution of regions. As explained in [16], the dilation and thinning process are operated on the original road data to remove these details and simplify the map.



(a) Original segment of city (b) After dilation operator (c) After thinning operator

Fig. 2. The preparation process of road network

6.3 Baseline and other comparison

6.4 Evaluation Metrics

We evaluate the functional region through two approaches. One is the percentage of regions that could find explanation label by our framework; while the other is the quality and accuracy of the explanation given by our framework.

For the first approach, we define the percentage of explainable region as follow:

$$explanation\ percentage = \frac{|regions\ with\ explanation\ label|}{|all\ regions|}$$

For the second approach,

[1] http://www.bigemap.com/

6.5 result

7 Conclusion

Function regions identification is an important part of urban computing. But the relative research lack of persuasive explanation. In this paper, we proposed a novel framework to give persuasive explanation for functional regions identification.

References

1. Zheng Y, Capra L, Wolfson O, et al. Urban Computing:Concepts, Methodologies, and Applications[J]. Acm Transactions on Intelligent Systems & Technology, 2014, 5(3):1-55.
2. Zheng Y, Liu Y, Yuan J, et al. Urban computing with taxicabs[C]// International Conference on Ubiquitous Computing. ACM, 2011:89-98.
3. Meng, C.; Yi, X.; Su, L.; Gao, J.; Zheng, Y. City-wide Traffic Volume Inference with Loop Detector Data and Taxi Trajectories. Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems,ACM,2017, 1:1-1:10
4. Vatsavai R R, Bright E, Varun C, et al. Machine learning approaches for high-resolution urban land cover classification:a comparative study[C]// International Conference and Exhibition on Computing for Geospatial Research & Application, Com.geo 2011, Washington, Dc, Usa, May. DBLP, 2011:1-10.
5. Vatsavai R R, Cheriyadat A, Gleason S. Unsupervised Semantic Labeling Framework for Identification of Complex Facilities in High-Resolution Remote Sensing Images[C]// IEEE International Conference on Data Mining Workshops. IEEE Computer Society, 2010:273-280.
6. Ratti C, Sobolevsky S, Calabrese F, et al. Redrawing the Map of Great Britain from a Network of Human Interactions[J]. Plos One, 2010, 5(12):e14248.
7. Yuan J, Zheng Y, Zhang L, et al. Where to find my next passenger[C]// Proceedings of the 13th international conference on Ubiquitous computing. ACM, 2011:109-118.
8. Ge Y, Liu C, Xiong H, et al. A taxi business intelligence system[C]// ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, Ca, Usa, August. DBLP, 2011:735-738.
9. Farmer, C. J. (2009). Data driven functional regions. In Proceedings of 10th International Conference on GeoComp
10. Nandani Garg and Sayan Ranu. Route Recommendations for Idle Taxi Drivers: Find Me the Shortest Route to a Customer!. In KDD 18: The 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, August 19C23, 2018, London, United Kingdom. 2018.
11. Goddard J B. Functional Regions within the City Centre: A Study by Factor Analysis of Taxi Flows in Central London[J]. Transactions of the Institute of British Geographers, 1970, 49(49):161-182.
12. Karlsson C, Olsson M. The identification of functional regions: theory, methods, and applications[J]. Annals of Regional Science, 2006, 40(1):1-18.
13. Newman M E J. Detecting community structure in networks[J]. European Physical Journal B, 2004, 38(2):321-330.

14. R.M. Ball. The use and definition of Travel-to-Work Areas in Great Britain: Some problems[J]. Regional Studies, 1980, 14(2):125-139.
15. Yuan J, Zheng Y, Xie X. Discovering regions of different functions in a city using human mobility and POIs[C]// ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2012:186-194.
16. Yuan N J, Zheng Y, Xie X, et al. Discovering Urban Functional Zones Using Latent Activity Trajectories[J]. Knowledge & Data Engineering IEEE Transactions on, 2015, 27(3):712-725.
17. Yuan N J, Zheng Y, Xie X. Discovering Functional Zones in a City Using Human Movements and Points of Interest[J]. 2018.
18. Zhang F, Yuan N J, Wilkie D, et al. Sensing the Pulse of Urban Refueling Behavior:A Perspective from Taxi Mobility[J]. Acm Transactions on Intelligent Systems & Technology, 2015, 6(3):1-23.
19. Shang J, Zheng Y, Tong W, et al. Inferring gas consumption and pollution emission of vehicles throughout a city[J]. 2014:1027-1036.
20. Wu G, Ding Y, Li Y, et al. Mining Spatio-Temporal Reachable Regions over Massive Trajectory Data[C]// IEEE, International Conference on Data Engineering. IEEE, 2017:1283-1294.
21. D Wang,J Zhang,W Cao,J Li,Y Zheng.When Will You Arrive? Estimating Travel Time Based on Deep Neural Networks.In International AAAI Conference on Web and Social Media, 2011.
22. R. Alvarez, D. Garcia, Y. Moreno, and F. Schweitzer. Sentiment cascades in the 15m movement. EPJ Data Science, 4(1):1C13, 2015
23. Alshamsi A, Awad E, Almehrezi M, et al. Misery loves company: happiness and communication in the city[J]. Epj Data Science, 2015, 4(1):7.
24. Z. Cheng, J. Caverlee, K. Lee, and D. Sui. Exploring millions of footprints in location sharing services. In International AAAI Conference on Web and Social Media, 2011.
25. Gonzlez M C, Hidalgo C A. Understanding individual human mobility patterns[J]. Nature, 2008, 453(7196):779-782.
26. Gallegos L, Huang A, Huang A, et al. Geography of Emotion: Where in a City are People Happier?[C]// International Conference Companion on World Wide Web. International World Wide Web Conferences Steering Committee, 2016:569-574.
27. Zhao S, King I, Lyu M R. A Survey of Point-of-interest Recommendation in Location-based Social Networks[J]. 2016.
28. Yongfeng Zhang, Yi Zhang, and Min Zhang. SIGIR 2018 Workshop on Explain-Able Recommendation and Search (EARS 2018). In Proceedings of The 41st International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 18). 2018.
29. Zhang Y, Chen X. Explainable Recommendation: A Survey and New Perspectives[J]. arXiv preprint arXiv:1804.11192. 2018.
30. Herlocker J L, Konstan J A, Riedl J. Explaining collaborative filtering recommendations[C]// Acm Conference on Computer Supported Cooperative Work. 2000:241-250.
31. Ferwerda B, Swelsen K, Yang E. Explaining Content-Based Recommendations[J]. Bruceferwerda Com.
32. Zhang Y, Zhang M, Liu Y, et al. Boost Phrase-level Polarity Labelling with Review-level Sentiment Classification[J]. Computer Science, 2015. Zhang Y, Lai G, Zhang M, et al. Explicit factor models for explainable recommendation based on phrase-level sentiment analysis[C]// International Acm Sigir Conference on Research & Development in Information Retrieval. ACM, 2014:83-92.

33. Costa F, Ouyang S, Dolog P, et al. Automatic Generation of Natural Language Explanations[J]. 2017.
34. Preece A. Asking 'Why' inAI: Explainability of intelligent systems - perspectivesand challenges. Intell Sys Acc Fin Mgmt. 2018;25:63C72.
35. Korpan, R., Epstein, S.L., Aroor, A., Dekel, G.: WHY: Natural explanations from a robot navigator. In: AAAI 2017 Fall Symposium on Natural Communication for Human-Robot Collaboration.
36. Jose M. Alonso, Corrado Mencar. Building Cognitive Cities with Explainable Artificial Intelligent Systems. In Proceedings of the First International Workshop on Comprehensibility and Explanation in AI & ML, 2017.
37. Lu Y, Castellanos M, Dayal U, et al. Automatic construction of a context-aware sentiment lexicon:an optimization approach[C]// International Conference on World Wide Web, WWW 2011, Hyderabad, India, March 28 - April. DBLP, 2011:347-356.