An Approach of Analyzing Movement Patterns Using Word Embeddings from Geo-tagged Tweets

Masaharu Hirota
Department of Information Science, Okayama University of Science
1-1 Ridaicho, Kita-ku, Okayama-shi, 700-0005, Japan

Tetsuya Oda
Department of Information and Computer Engineering, Okayama University of Science
1-1 Ridaicho, Kita-ku, Okayama-shi, 700-0005, Japan
E-mail: hirota@mis.ous.ac.jp, oda@ice.ous.ac.jp

Abstract

Many people share content about daily events on social media sites. Understanding people's movement patterns using the contents benefits numerous applications, such as tourism recommendation, city planning, and geo-targeting. This study proposes a new approach for clustering user trajectories to discover movement patterns. Our approach generates feature vectors for movement patterns using the word embedding model to learn movements between a pair of areas quantized by their latitude and longitude. Then, our approach uses multiple models to learn each area of different sizes and integrates the generated embedding vectors. The vectors represent the relationships between the movements from one area to the next. We demonstrated that clustering results by our proposed method.

Keywords: trajectory, location information, clustering, Twitter.

1. Introduction

Because of the wide dispersion and expeditious improvement of various devices such as smartphones and tablets, diverse and vast data are generated on the web. Social media sites have become popular because users can quickly post data and various messages. Twitter[1], a social media site that provides a micro-blogging service, is used as a real-time communication tool. Numerous tweets have been posted daily by vast numbers of users. Therefore, Twitter is used to obtain, from a large amount of information posted by many users, real-time information corresponding to the real world.

Understanding human mobility patterns is important as basic information for various applications such as urban planning and transportation network analysis. To analyze these patterns, we need information on the vast movement trajectories. However, it is difficult to collect a large amount of data on the trajectory. Social media information, such as Twitter, can be used as an alternative source of information for such analysis. For example, some tweets have geo-tag with location information based on latitude and longitude. We can regard a sequence of consecutive geo-tagged tweets posted by a user as the movement trajectory. Here, information from social media such as Twitter is not a sufficient sampling of the population about age, race, ethnicity, and other attributes of people. However, the information has been used in various studies as a valuable information to analyze people's movement patterns.

Clustering of movement trajectories[2,3] is one of the methods to analyze the people's movement patterns from a large number of movement trajectories. By applying the

© The 2022 International Conference on Artificial Life and Robotics (ICAROB2022), January 20 to 23, 2022

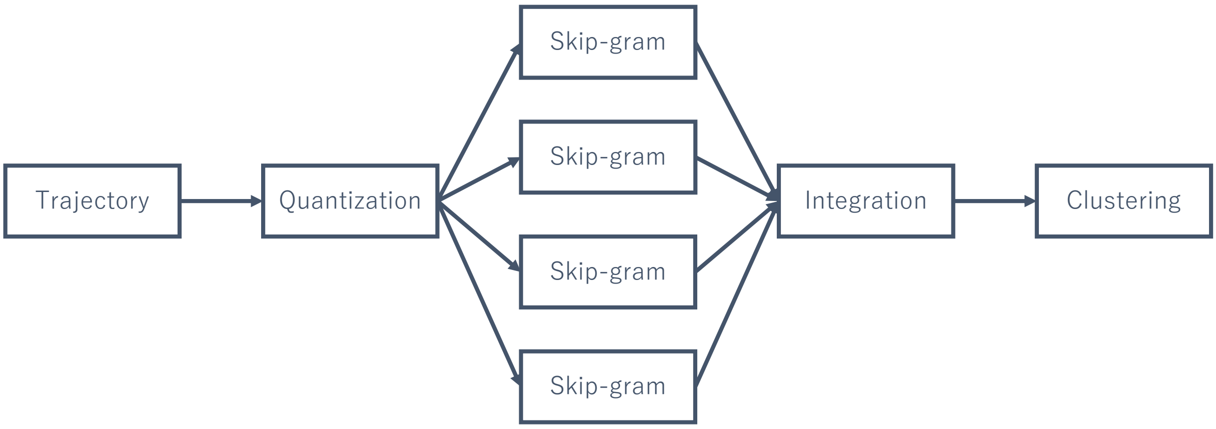

Fig. 1. Overview of our proposed method.

clustering method to the trajectories, similar trajectories are grouped called clusters. We can then regard each cluster as people's movement patterns.

This paper proposes a new approach to clustering trajectories represented by geo-tagged tweets to analyze movement patterns. Our proposed method uses a word embedding model by an improved skip-gram model[4], inspired by the skip-gram model[5], for learning user movements. The model learns movements between a pair of areas quantified by the latitude and the longitude. Then, by applying clustering to the trajectories based on the vectors generated by multiple models, we analyze the movement patterns of users between the quantized regions.

2. Literature Review

This section describes several studies that use the movement patterns obtained from social media sites such as Twitter.

Studies have been conducted to analyze behavior, destinations, and purposes of movement by analyzing the movement patterns using tweets. Jurdak et al.[6] showed that movement patterns differ according to their daily travel distance trends by analyzing tweets posted in Australia. Salas-Olmedo et al.[7] analyzed the spatial distribution of users of public spaces by analyzing the movement represented by geo-tagged tweets in a medium-sized city.

There is some study based on Twitter on the impact of the COVID-19 pandemic on the mobility of people. Huang et al.[8] proposed an index to evaluate the degree of change in people's movement and analyzed the impact of COVID-19. Jiang et al.[9] analyzed the impact of COVID-19 on changes in movement patterns for land use in New York City. These studies show that the change of movement patterns depends on the phase of the COVID-19 epidemic and the response to it by national.

3. Proposed Method

This section describes a method for clustering movement trajectories. Figure 1 shows the overview of the proposed method. The proposed method consists of the following steps.

(i). Quantize movement trajectories based on latitude and longitude.
(ii). Create multiple improved skip-gram models by learning the quantized trajectories.
(iii). Input the trajectories to each model and integrate obtained vectors.
(iv). Cluster the trajectories based on the vectors.

3.1. *Quantization of user-posted locations*

In this section, we describe how to quantize tweets. First, we sort the series of tweets of a user in post time by ascending order. Next, we regard the posting times of those tweets as different movement trajectories if the difference between the posting times of two consecutive tweets is more than 3 hours. This reason is that when a user posts tweets over a long time, we should regard them as different movements. We also delete the trajectory if the number of tweets is smaller than 5.

The next step is to convert the movement trajectories, which consist of latitude and longitude, into a continuous string. The improved skip-gram model is a method for learning continuous 1D features. Also, this process simplifies the learning while accounting for the differences due to GPS errors and the number of digits in the latitude and the longitude.

This study converts the latitude and longitude annotated to a tweet into quadkey[10]. The quadkey is a unique identifier represented by the string representation of a standard map tile on a particular zoom level. In addition, depending on the string length of the quadkey, it is possible to control the degree of accuracy loss. This study generated each quadkey with zoom levels from 14 to 17 from the tweets.

3.2. *Improved skip-gram*

Next, we describe improved skip-grams, a model for learning the movement trajectory. The conventional skip-gram learns $w(t-c)\cdots w(t-1)$ and $w(t+1)\cdots w(t+c)$ around a word $w(t)$ in a sentence. Next, the improved skip-gram learns only $w(t+1)\cdots w(t+c)$ for $w(t)$. The difference between these models is that the conventional skip-gram learns the movement to and from a region represented by w(t). In contrast, the improved skip-gram learns only the movement from one region to another represented by w(t). In this study, we create improved skip-gram models for each level of quadkey.

3.3. *Clustering trajectory*

We describe the procedure of movement trajectory clustering using the generated models. First, we convert the latitude and longitude of each trajectory into a sequence of multi-level quadkey codes. Then, each level of quadkey is input to the corresponding improved skip-gram model to generate multiple vectors. The trajectory vector is the average of the vectors obtained from the quadkeys included in the trajectory for each level of quadkey.

We apply k-means clustering, a popular clustering algorithm, to the vectors to cluster movement trajectories. In this study, we regard the result as the clustering result of the trajectory.

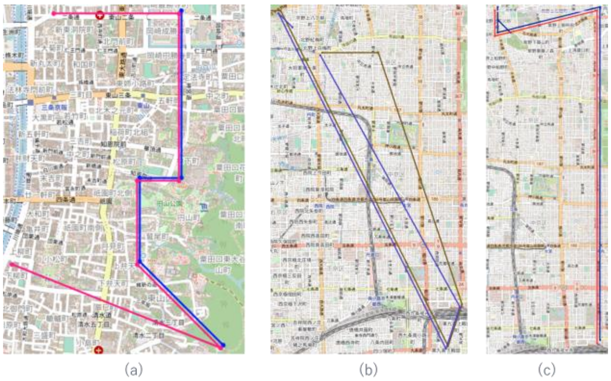

Fig. 2. Clustering results.

4. Visualization

4.1. *Dataset*

In this experiment, we used tweets, including latitude and longitude information. The tweets were posted in Kyoto, Japan, between May 21, 2015, and December 31, 2018. The number of tweets was 3,622,043, and the number of movement trajectories was 8,379.

The parameters of each improved skip-gram are that the vector dimension is 50 and that the window size is 3. We set the number of clusters k for the k-means clustering to 300.

4.2. *Result*

Figure 2(a) ~ (c) shows three clusters in the clustering result of our proposed method. Each line in those figures represents movement trajectories. A distinct color represents each trajectory. We used the OpenStreetMap[11].

In Figure 2(a), the cluster includes the trajectories for uses who mainly visited around Kiyomizu temple, Chion-in, and Okazaki Park. In Figure 2(b), the cluster includes the trajectories for uses that are mainly visited around Ritsumeikan University and Kyoto Station. In Figure2(c), the cluster includes the trajectories for uses that are mainly visited around Funaokayama Park and Kyoto Station. The trajectories included in each cluster differ mainly in the places visited. In those figures, the trajectories included in each cluster are different movement patterns. Therefore, we conclude that our proposed method has successfully clustered the trajectories.

5. Conclusion

This paper proposed a new approach for analyzing movement patterns by clustering movement trajectories

© The 2022 International Conference on Artificial Life and Robotics (ICAROB2022), January 20 to 23, 2022

using a word embedding algorithm. We presented three clusters of trajectories using our method. One of the limitations is that our method does not consider the direction of the trajectory. As a result, the proposed method regards the trajectories visited the same point as one cluster, even if they have the opposite movement direction. Therefore, future work will include the improvement of a method of word embedding.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number JP19K20418 and Grant for Promotion of OUS Research Project (OUS-RP-20-3).

References

1. "Twitter", https:/twitter.com/
2. S.J. Corrado, T.G. Puranik, O.J. Pinon, D.N. Mavris, “Trajectory Clustering within the Terminal Airspace Utilizing a Weighted Distance Function”, Proceedings, Vol. 59, No. 1, 2020.
3. H. Li, J. Liu, K. Wu, Z. Yang, R. W. Liu and N. Xiong, “Spatio-Temporal Vessel Trajectory Clustering Based on Data Mapping and Density”, Proceedings, IEEE Access, vol. 6, pp. 58939-58954, 2018.
4. M. Hirota, T. Oda, M. Endo, and H. Ishikawa, “Generating distributed representation of user movement for extracting detour spots” in Proceedings of the 11th International Conference on Management of Digital EcoSystems, pp. 250–255, 2019.
5. T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, “Distributed representations of words and phrases and their compositionality”, Advances in Neural Information Processing Systems, pp. 3111–3119, 2013.
6. R. Jurdak, K. Zhao, J. Liu, M. AbouJaoude, M. Cameron, D. Newth, “Understanding Human Mobility from Twitter”, PLOS ONE, Vol. 10, No. 7, pp. e0131469, 2015.
7. M.H. Salas-Olmedo, C. Quezada, “The use of public spaces in a medium-sized city: from Twitter data to mobility patterns”, Journal of Maps, Vol. 13, No. 1, pp. 40-45, 2017.
8. X. Huang, Z. Li, Y. Jiang, X. Li, D. Porter, “Twitter reveals human mobility dynamics during the COVID-19 pandemic”, PLOS ONE, Vol. 15, pp. 1-21, 2020.
9. Y. Jiang, X. Huang, Z. Li, “Spatiotemporal Patterns of Human Mobility and Its Association with Land Use Types during COVID-19 in New York City”, ISPRS International Journal of Geo-Information, Vo. 10, No. 5, pp. 334, 2020.
10. “Quadkey”, https://docs.microsoft.com/en-us/azure/azure-maps/zoom-levels-and-tile-grid?tabs=csharp#quadkey-indices
11. “OpenStreetMap”, https://www.openstreetmap.org

Authors Introduction

Dr. Masaharu Hirota

He received a Doctor of Informatics degree in 2014 from Shizuoka University. After working for the National Institute of Technology, Oita College, he has worked in the Faculty of Informatics, the Okayama University of Science, since April 2017. His research interests include data engineering, GIS, and social media. He is a member of ACM, DBSJ, and IPSJ.

Dr.Tetsuya Oda

He received a Ph.D. degree in 2016 from Fukuoka Institute of Technology (FIT). After working for the Faculty of Information Engineering, FIT, he has worked in the Faculty of Engineering, the Okayama University of Science, since April 2017. His research interests include wireless mesh networks, intelligent systems and robotics. He is a member of AAAI, ACM, DBSJ, IEEE and IPSJ.