TransBigData: A Python package for transportation spatio-temporal big data processing, analysis and visualization

Qing Yu[1] **and Jian Yuan**[*1]

1 Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, 4800 Cao'an Road, Shanghai 201804, People's Republic of China

Summary

In recent years, data generated in transportation field has begun to explode. Individual continuous tracking data, such as mobile phone data, IC smart card data, taxi GPS data, bus GPS data and bicycle sharing order data, also known as "spatio-temporal big data" or "Track &Trace data"(Harrison et al., 2020), has great potential for applications in data-driven transportation research. These spatio-temporal big data typically require three aspects of information(Zhang et al., 2021): Who? When? Where? They are characterized by high data quality, large collection scope, and fine-grained spatio-temporal information, which can fully capture the daily activities of individuals and their travel behavior in the city in both time and space dimensions. The emergence of these data provides new ways and opportunities for potential transportation demand analysis and travel mechanism understanding in supporting urban transportation planning and management(Chen et al., 2021; Zhang et al., 2020). However, processing with these multi-source spatial temporal big data usually requires a series of similar processing procedure (e.g., data quality accessment, data proprocessing, data cleaning, data gridding, data aggregation, data visualization). There is an urgent need for a one-fit-all tool that can adapt to the different processing demand of different transportation data in this field.

Statement of need

`TransBigData` is a Python package developed for transportation spatio-temporal big data processing, analysis and visualization. `TransBigData` provides fast and concise methods for processing common transportation spatio-temporal big data such as Taxi GPS data, bicycle sharing data and bus GPS data. `TransBigData` provides a variety of processing methods for each stage of transportation spatio-temporal big data analysis. The code with `TransBigData` is clean, efficient, flexible, and easy to use, allowing complex data tasks to be achieved with concise code. It has already been used in a number of scientific publications(Li et al., 2021; Yu, Zhang, Li, Sui, et al., 2020; Yu, Zhang, Li, Song, et al., 2020; Yu et al., 2021).

For some specific types of data, `TransBigData` also provides targeted tools for specific needs, such as extraction of Origin and Destination(OD) of taxi trips from taxi GPS data and identification of arrival and departure information from bus GPS data.

Currently, `TransBigData` mainly provides the following methods:
- *Data Quality*: Provides methods to quickly obtain the general information of the dataset, including the data amount the time period and the sampling interval.

---

*corresponding author

38 - *Data Preprocess*: Provides methods to fix multiple types of data error.

39 - *Data Gridding*: Provides methods to generate multiple types of geographic grids (Rectan-
40 gular grids, Hexagonal grids) in the research area. Provides fast algorithms to map GPS data
41 to the generated grids (Figure 1).

42 - *Data Aggregating*: Provides methods to aggregate GPS data and OD data into geographic
43 polygon.

44 - *Trajectory Processing*: Provides quick methods to re-organize the data structure and imple-
45 ment data augmentation from various data formats, including generating trajectory linestring
46 from GPS points, and trajectory densification, etc.

47 - *Data Visualization*: Built-in visualization capabilities leverage the visualization package ke-
48 plergl to interactively visualize data on Jupyter notebook with simple code.

49 - *Basemap Loading*: Provides methods to display Mapbox basemap on matplotlib figures
50 (Figure 2)

51 The latest stable release of the software can be installed via pip and full documentation can
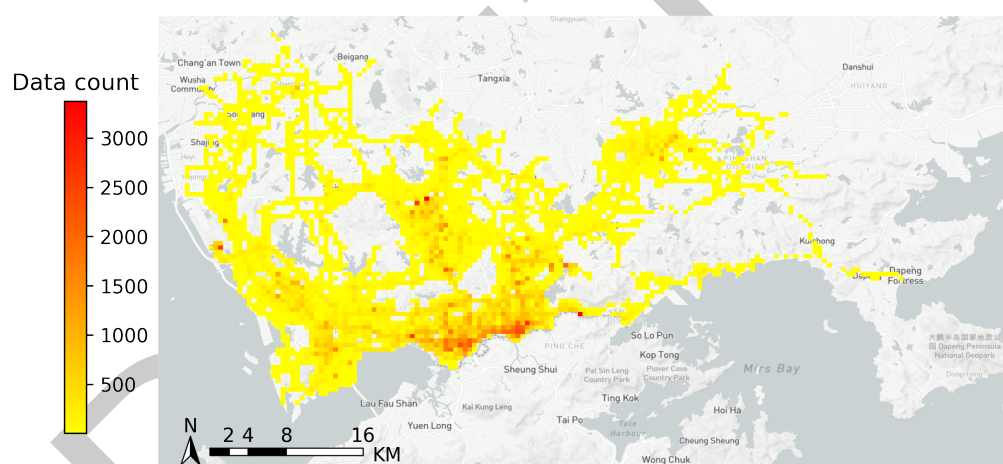52 be found at https://transbigdata.readthedocs.io/en/latest/.

Figure 1: `TransBigData` generates rectangular grids and aggregate GPS data to the grids.
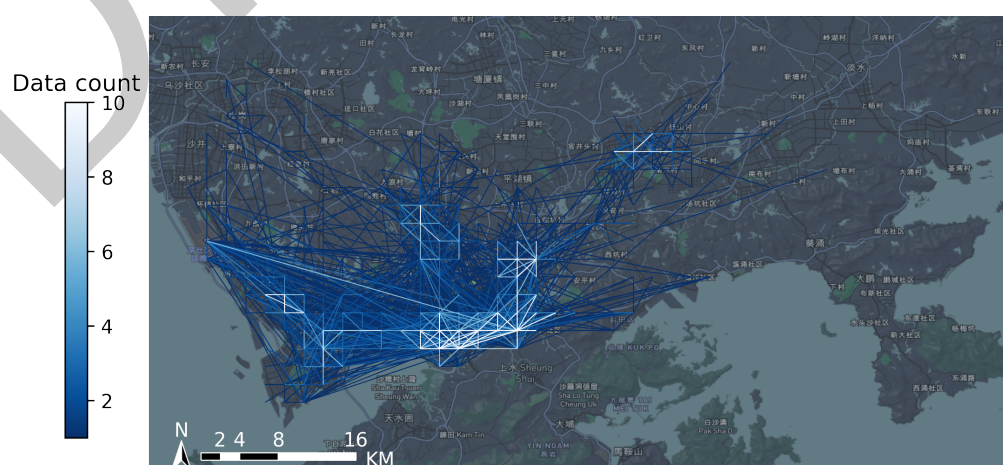

Figure 2: `TransBigData` visulizes taxi trips OD and display basemap on matplotlib figures.

References

Chen, J., Li, W., Zhang, H., Cai, Z., Sui, Y., Long, Y., Song, X., & Shibasaki, R. (2021). GPS data in urban online ride-hailing: A simulation method to evaluate impact of user scale on emission performance of system. *Journal of Cleaner Production*, *287*, 125567. https://doi.org/10.1016/j.jclepro.2020.125567

Harrison, G., Grant-Muller, S. M., & Hodgson, F. C. (2020). New and emerging data forms in transportation planning and policy: Opportunities and challenges for "Track and Trace" data. *Transportation Research Part C: Emerging Technologies*, *117*, 102672. https://doi.org/10.1016/j.trc.2020.102672

Li, Y., Li, W., Yu, Q., & Yang, H. (2021). Taxi global positioning system data in urban road network: A methodology to identify key road clusters based on travel speed–traffic volume correlation. *Transportation Research Record*, 03611981211036684. https://doi.org/10.1177/03611981211036684

Yu, Q., Li, W., Yang, D., & Zhang, H. (2021). Partitioning urban road network based on travel speed correlation. *International Journal of Transportation Science and Technology*, *10*(2), 97–109. https://doi.org/10.1016/j.ijtst.2021.01.002

Yu, Q., Zhang, H., Li, W., Song, X., Yang, D., & Shibasaki, R. (2020). Mobile phone GPS data in urban customized bus: Dynamic line design and emission reduction potentials analysis. *Journal of Cleaner Production*, *272*, 122471. https://doi.org/10.1016/j.jclepro.2020.122471

Yu, Q., Zhang, H., Li, W., Sui, Y., Song, X., Yang, D., Shibasaki, R., & Jiang, W. (2020). Mobile phone data in urban bicycle-sharing: Market-oriented sub-area division and spatial analysis on emission reduction potentials. *Journal of Cleaner Production*, *254*, 119974. https://doi.org/10.1016/j.jclepro.2020.119974

Zhang, H., Chen, J., Li, W., Song, X., & Shibasaki, R. (2020). Mobile phone GPS data in urban ride-sharing: An assessment method for emission reduction potential. *Applied Energy*, *269*, 115038. https://doi.org/10.1016/j.apenergy.2020.115038

Zhang, H., Song, X., & Shibasaki, R. (2021). *Big Data and Mobility as a Service*. Elsevier. ISBN: 978-0-323-90170-3