

Enhancing Kepler.gl for processing Google Maps Timeline data

Tianqi Xia^{ab}, Hiroshi Kanasugi^{ab*}, Zipei Fan^b, Ruochen Si^b, Chuyao Feng^b, Hiroshi Takayasu^{bc}, Ryosuke Shibasaki^b

^a LocationMind Inc. Tianqi Xia, xia@locationmind.com, Hiroshi Kanasugi, knsq@locationmind.com

^b The University of Tokyo, Zipei Fan, fan@csis.u-tokyo.ac.jp, Ruochen Si, si@csis.u-tokyo.ac.jp, Chuyao Feng, fengcy2020@csis.u-tokyo.ac.jp, Ryosuke Shibasaki, shiba@csis.u-tokyo.ac.jp

^c East Japan Railway Company, Hiroshi Takayasu, h-takayasu@jreast.co.jp

* Corresponding author

Keywords: Human mobility visualization, Kepler.gl, Google Maps Timeline

Abstract:

With the development of location-based service (LBS), mobility data has been applied in a lot of studies on different themes. In the previous studies, mobility analysis mainly relies on the data collected by LBS apps or the mobile network operator. Though the data is quite huge for analyzing large-scale mobility trends, this data is not suitable for research focused on small-scale and rich-semantic mobility information since most of the data providers cannot share personal information to the researchers and the data itself are too big to be processed. On the other hand, though small-scale analysis can be conducted by asking the subjects to collect location by themselves with some tracking apps, keeping tracking with these apps is sometimes annoying that the user might forget to open the app. In addition, it is difficult for them to process location data to remove private information.

The usage of Google Maps Timeline (GMT) makes it easier for such small-scaled analysis. GMT will automatically track volunteers' location information after they grant GMT service and the location tracking function of any other google apps. The quality and quantity are impressive if a user using Google map frequently or allow Google apps to collect location data in background. GMT mainly include raw location history and the semantic location history computed from the raw one, which is regarded as very sensitive according to the survey conducted by Ministry of Internal Affairs and Communications¹². What's worse, location history with rich semantic information can be used to mine the detailed personal information such as home and work address (Curran, 2018), which is even more sensitive from the same survey. As a result, the data requires desensitization and filtering before uploading from the volunteer's client (Table 1).

	Sensitive items	Desensitization method
Raw Location history	Location coordinates	Convert to mesh grid/region
	Time	Convert to time index
Semantic location history	Origin/destination location address	Generalize address
	Origin/destination location coordinates	Convert to mesh grid/region
	Departure/arrive time	Convert to time index
	Transportation mode	Target filtering
	Facility name	Convert to facility type

Table 1. Sensitive data in GMT and their desensitization

Though GMT could be visualized or processed by existing official or third-party systems, these apps have several drawbacks such as data analysis and visualization with other data source is impossible and batch data filtering is not available through the interface. On the other hand, Kepler.gl (Kepler) is an open-source location data visualization tool developed by Uber(He, 2018). With Kepler volunteers can process trajectory data, plot trajectories via animation and export them with several filters. In addition, instead of providing the service as a software, Kepler is an app on web

¹ <https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h26/html/nc133210.html>

² Considering the different privacy preference and the status of using GMT, we also prepared GMT related a questionnaire: <https://forms.gle/FSFrN2tZ8S4bJGU9A>

browser with most of the data processing happening in the front end, which means that the private information will not be transferred to the server.

However, it is difficult to directly implement Kepler in GMT with the following problems: first, in the current stage Kepler cannot directly handle the Google Maps Timeline data as it is not a standard data format; second, the filtering function in kepler.gl is too limited to remove the private information; finally, data editing and converting function in Kepler is very limited that to process the sensitive data. Fortunately, Kepler is an open-source application with MIT license, which make it possible for researchers to adapt it for their studies. Thus, this study proposes a system for analyzing Google takeout data using Kepler. The current issues and the functions to be enhanced is listed in Table.2 and are introduced in the remaining of the paper. Notice that address analysis and geocoding still requires server request.

(1) File reading and visualization enhancement: since Kepler can only visualize spatial data with several layer types, there are two strategies for kepler.gl to visualize GMT data - convert GMT to standard layers or create new layers which are compatible with GMT data. Considering the complexity of create new layers, in this framework Kepler is adapted to have the ability of converting one GMT file into several target layers for visualization. In detail, GPS location will be converted to point layer and trip layer while semantic location will be converted to arc layer or line layer.

(2) Data filtering enhancement: in this system, filtering function will be enhanced for the following attributes. For timestamp data, a filter from calendar is implemented to select the date range and a time range selection in implemented to select the timespan in each day. In addition, for localization in Japan, we also plan to implement a filtering to exclude weekends and holidays. For spatial data, a filter that can use the selected polygon data in other dataset is implemented to implement spatial filtering with complicated polygons. An address-based filtering will also be implemented for extracting the locations with specific address. We will also consider a filter based on Japanese prefecture for localization in Japan.

(3) Data processing enhancement: In this system, besides the file reading and layer interaction (join) processing parts mentioned above, data processing enhancement also includes desensitization and specification part. In this system, desensitization includes mainly the two parts: For GPS location, the desensitization mainly includes the aggregation of GPS points to spatial and temporal units. For semantic location, the desensitization mainly includes the generalization of the origin and destination and the generalization of address. In addition, the system should also implement the function to convert the facility names to the facility types. On the contrary, specification function aims to convert abstract location into more detailed spatial objects. This includes the enhancing geocoding for address tables and in the future, we will also consider implementing map matching for generating more detailed road network-based mobility.

Functions	Current features in kepler	Enhanced system	Server request
File reading	Cannot support GMT data	Can read GMT data as tables	No
Data visualization	Cannot support GMT data	Visualize GMT data as multiple layers	No
Data filtering	Spatial filtering	Only support basic spatial filtering by edited spatial region	No
	Temporal filtering	Only support a range	No
	Address filtering	Not available	Maybe
Data processing	Geocoding	Cannot save as a new column of the data	Yes
	Table join	Unavailable	No
	Address simplification	Unavailable	Maybe
	Data aggregation	Unavailable	No
	POI converting	Unavailable	Convert detailed POI names to POI types

Table 2. Functions to be enhanced in Kepler

This abstract introduces an enhanced application based on Kepler to process GMT data. In the current stage this application is still on developing and the full application is hopefully to be demonstrated during ICC conference.

References

- Curran, D., 2018. Are you ready? Here is all the data Facebook and Google have on you. *The Guardian*, 30(03), p.2018.
- He, S., 2018. From Beautiful Maps to Actionable Insights: Introducing Kepler. gl, Uber's Open Source Geospatial Toolbox. *Uber Engineering*, 29.