A Google Timeline Extension for Location Classification

Gabriel Gladstone
dept. of Computer Engineering
Charlottesville, VA
zwy7ce@virginia.edu

***Abstract*—Researchers use location data to determine the mental health of an individual and forecast epidemics. This study presents a system to extend Google Timeline's characterization of significant locations in a participant's life. The study analyzed the online data of 1 participant over 40 days to prove the effectiveness of the methodology. The study concludes that the Timeline Extension is a cheap and autonomous location labeling mechanism and can be used by researchers as a first-pass for location analysis.**

I. Introduction

Location data is tracked for a variety of commercial purposes including navigation, advertising, and fraud prevention. However, location data has also become key source of information in forecasting epidemics [6] [3] and detecting depression.

This study presents a process to automatically identify and characterize the significant locations in a person's life building off of Google Timeline geolocation data and other metadata.

II. Methodology

A. Study Design

The study takes place from January 12, 2025 to February 21, 2025. This is a 40 day period from the start of UVA's academic semester to the time of writing this report. The time period began at the moment Google Timeline data collection was enabled. The time period was also chosen because the locations visited during the school year are limited. The singular participant in this study is me, the author of this paper.

The goal of this paper was to develop a method to accurately label Google Timeline data and characterize the significant locations in a person's life. This data could be used in future studies to better understand someone's mental health and understand who they interact with, possibly for epidemic detection. All Google Timeline location data from the specified time was pulled. Significant locations were determined from Google's Visit geolocations and weighted density based clustering algorithm. Google Details and Google Nearby Search API's were used to determine significant locations and API metadata along with the Datamuse API's [2] reverse Thesaurus search were used for labeling. The system builds upon Google Timeline to create an accurate characterization of the significant locations of a person's life

B. Data Source Exploration

There are many platforms that track geolocation. Many studies have used Twitter geo-tagged posts. There are geocoding API's to easily get the latitude and longitude of a person's location. However, this study will use Google Timeline data which is on for the entire study.

C. Google Data Extraction

Google Timeline data tracks 3 types of records.

Visit records are locations Google believes a person has visited. This record comes with a unique placeID of the location that can be used to query further information about the location. These records are the standard for significant locations in a person's life, and this paper proposes an extension method to find additional significant locations.

Activity records are records that designate and start and end geolocation with the type of activity specified such as walking.

timelinePath records are records that designate a location and duration. Google maps the timeline path per day. These geolocations may overlap with other records.

All geolocations and associated metadata from these records were extracted for initial analysis. While the data covers some places off campus, the focus of this analysis will be on campus. Initial Google Timeline can be seen in 1.

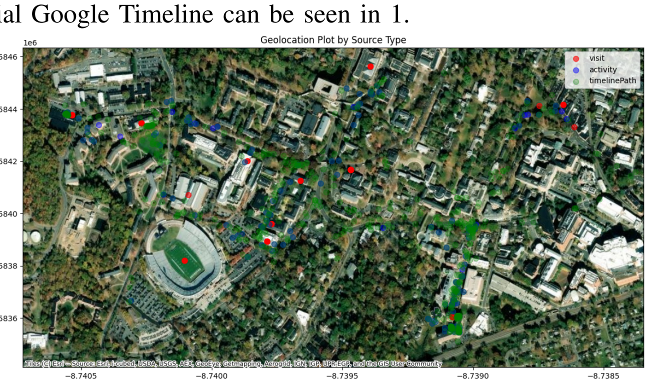

Fig. 1: On Campus Google Timeline Data

Google provides a Place Details API to extract further information about these locations [5]. However, a place_ID is required and only Visit records have a place_ID.

A query to the Details API was made for each visit record and relevant fields to best label the location were collected in a .txt file as shown in Fig. 2.

```
Name: University of Virginia
place_id: ChIJ3cRJt0iGs4kRRliLzMIIUUA
Address: Charlottesville, VA, USA
Types: university, point_of_interest, establishment
Business Status: OPERATIONAL
Open Now: N/A
Opening Hours: N/A
Editorial Summary: Founded by Thomas Jefferson in 1819, this research school features many historic buildings.
```

Fig. 2: Details API Metadata

D. *Significant Location Clustering*

As can be seen in Fig. 1, Visit records give a good representation of significant locations, but of course there are more significant locations that can be inferred from this data.

A density based clustering algorithm is employed to utilize the other records, activity and timelinePath. I thought density based clustering was the best clustering option because it is an algorithm that groups together tightly packed points and can mark outliers. However, an important factor I wanted to consider in the clustering algorithm was time spent at a location. Time spent over the course of the study varied from $\approx 18,000$ min to 0 min. I did not want these 2 extremes to be treated equally.

I used Luc Boruta's weighted extension of DBSCAN [1]. This allows me to use a custom dissimilarity matrix to find clusters. I used a method that considers the euclidean distance between 2 geolocations as well as the duration spent at the geolocations. The following factors were fine-tuned through visual observations to get the final result.

- epsilon: The radius in meters around the geolocation.
- mu: The minimum number of points or threshold for a point to form a cluster
- Time constant: The influence of duration on clustering.

The clustering results can be seen on the UVA campus overlay in Fig. 3 and in the plot of Fig. 9.

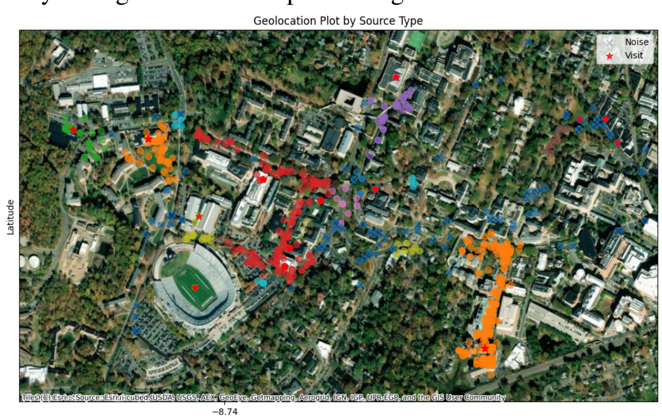

Fig. 3: Weighted DBSCAN Geolocation Clustering

As shown in Fig. 3, there is a strong overlap between a cluster and Visit record locations. However, there are some clusters that don't overlap. The overlap lends credibility to the cluster method as many of the Visit record locations are indeed the locations I most frequently visit. However, some clusters don't overlap with Visit locations so there are probably more valid significant locations that Google's Timeline didn't identify.

The algorithm also effectively eliminated non-dense location as shown in the full location data view 10.

E. *Additional Significant Location Identification*

Now that the cluster algorithm has been fine-tuned by observation and validated for accuracy. More significant locations can be found, in addition to the Google Timeline Visit Records.

I use Google Places Nearby Search [4] to find additional locations. The Nearby Search API will return relevant locations given a geolocation, radius, and other fields. My query to this API only consisted of a geolocation and radius where the nearby places returned are sorted in order of prominence, or importance in Google's index. Keyword and are good fields to get more relevant results, but for my use, this information was not available with the methods and data I collected.

To determine the Geolocation in the nearby search I found the location with the longest duration spent for each cluster.

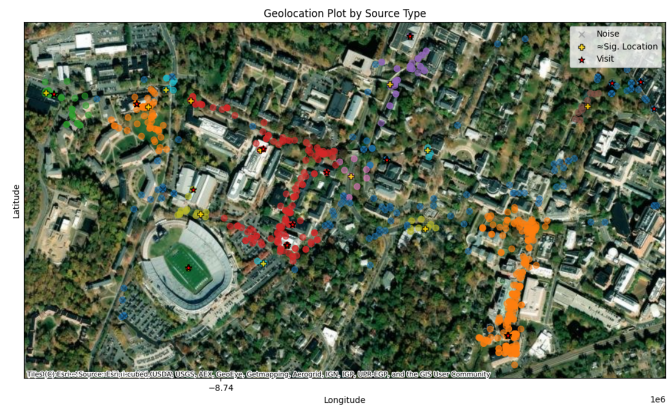

Fig. 4: Significant Location Guesses per Cluster

As can be seen in Fig. 4, 7 of the 23 guessed locations by duration overlapped with the visit search. Since each visit location already has a unique placeID that can be queried by Google's Place API, only non-overlapping geolocations were used. Thus, 16 queries were made to the Nearby Search API, to limit api costs and maximize relevancy of the query, the result radius was limited to prominent locations within 65 meters. These queries returned metadata of multiple places and important fields were collected and stored in a .txt file for later analysis (and to reduce api costs) as shown in Fig. 5. While there were many other fields provided by the API, I chose fields related to the description of the place that would best help in labeling the location. The Types field was particularly useful.

```
Name: The Brass Tap
Guessed GeoLocation: 37.590915,-77.492886
Actual GeoLocation: 37.5908117,-77.4932986
Place ID: ChIJiS53ugAVsYkR1r4kqdxEjg8
Address: 4901 Libbie Mill East Boulevard Suite 100, Richmond
Types: bar, meal_takeaway, restaurant, food, point_of_interest, establishment
Business Status: OPERATIONAL
Open Now: True
Opening Hours: N/A
Editorial Summary: N/A
```

Fig. 5: Nearby Search Metadata

The multiple locations from these guesses can be seen in Fig. 6

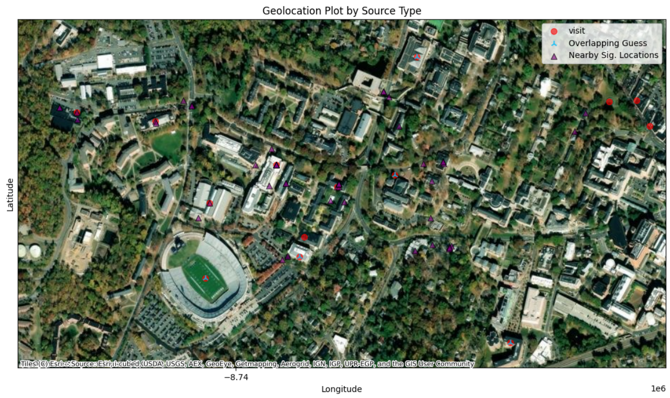

Fig. 6: Visit and Nearby Search Locations

F. Labeling

In total, 100 significant locations were identified from the Visit Places and Nearby Search queries. Now that all Google metadata was collected, the locations could be accurately labeled. I noticed that Google accurately labeled about half of the records in its Types field assigning university to University of Virginia, gym to Triangle Rock Club - Richmond, and more. However, some places were given the generic label of point of interest or establishment. For these places, another method has to be used.

For generic labeled locations, I used a reverse thesaurus from the Datamuse API [2]. A reverse thesaurus essentially lets a user input a string and get words related to that string. To create this string I concatenated the Name, Types, and Editorial Summary fields and made means-like query to get related words to the string. 100 words per query were returned and the first word was used as the label. The other words were used in a more general analysis of important locations.

III. Results

The labeled locations can be seen in Fig. 7. Only visit locations are shown for readability, but all significant locations were labeled.

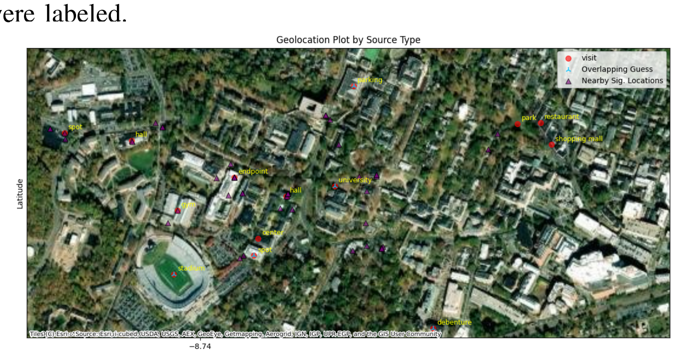

Fig. 7: Labeled On Campus Significant Locations

Many of the results are accurate, such as Scott Stadium being labeled as a Stadium or Rocky Top Climbing labeled as a gym. Some locations were given generic labels like center or locality, and some places were given incorrect labels such as debenture (because of the "Bond" in Bond House).

There were also places listed that I don't frequent. I don't eat at the Virginian on the Corner, but I do eat at Chipotle on the Corner. Error for nearby locations should be taken into consideration with use of this system.

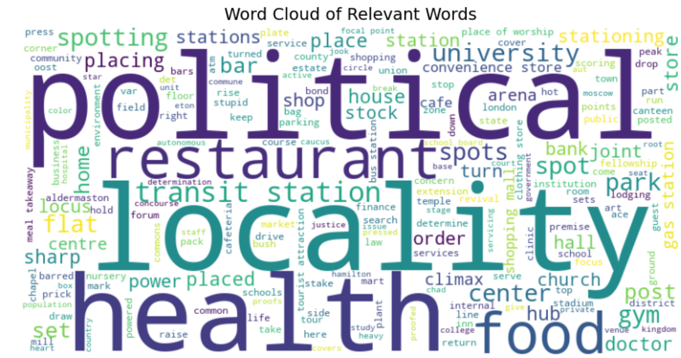

Fig. 8: Significant Locations Word Cloud

As shown in Fig. 8, from the Type field words and 1600 (16 requests x 100 per request) Datamuse API words, I created a word cloud where larger words are more significant to the individual, meaning the frequency of these words was greater throughout the location data. While there is no interpretation of the data in this study, some assumptions can be drawn about the individual to help determine there mental health. For instance, health, food, restaurant, gym, and park, can be seen in the word cloud. Maybe this means that the participant is in a good mental state because they are frequently doing something active and eating.

A. Limitations

As discussed in the results, there were many limitations to the methods of this study.

First, the density-based clustering algorithm should be fine-tuned for the studies purpose and I have not created a formal fine-tuning process.

Second, the additional labeling through datamuse did improve labeling results, but it is still not to the quality of some of the Google results.

Third, nearby locations may be misidentified due to the prominence of the location. If a user regularly frequents a small bar, the important building next to the bar may be identified which could lead to dramatically different conclusions about the mental health of the participant event if the location identified is near that actual location.

With these limitations, the Google Timeline Extension for Location Classification I described should be used as a first-pass in location and mental health analysis.

B. Future Work

I hope that researchers will use this tool in the mental health studies and improve upon it. The tool proposes an autonomous general labeling mechanism of identification, but is probably better suited for analyzing a fixed set of categories. Future studies should identify important categories to their studies such as home, school, store, etc... Specific categories will make it easier to use other online information to better label the locations. For instance, search data from a laptop will most likely not be associated with a participant shopping.

C. Conclusion

This study reviewed the Google Timeline location data of one participant. It presented a method for extending Google Visit location data and making general comments on the mental health of an individual. The study is a quick and autonomous first-pass for researchers analyzing the significant locations of individuals.

References

[1] Luc Boruta. "Indicators of Allophony and Phonemehood". PhD thesis. Université Paris-Diderot - Paris VII, Sept. 26, 2012. URL: https://theses.hal.science/tel-00746163 (visited on 02/23/2025).

[2] *Datamuse API*. URL: https://www.datamuse.com/api/ (visited on 02/23/2025).

[3] Jeremy Ginsberg et al. "Detecting influenza epidemics using search engine query data". In: *Nature* 457.7232 (Feb. 2009). Publisher: Nature Publishing Group, pp. 1012–1014. ISSN: 1476-4687. DOI: 10.1038/nature07634. URL: https://www.nature.com/articles/nature07634 (visited on 02/23/2025).

[4] *Nearby Search — Places API — Google for Developers*. URL: https://developers.google.com/maps/documentation/places/web-service/search-nearby (visited on 02/23/2025).

[5] *Place Details — Places API*. Google for Developers. URL: https://developers.google.com/maps/documentation/places/web-service/details (visited on 02/23/2025).

[6] Adam Sadilek, Henry Kautz, and Jeffrey P. Bigham. "Finding your friends and following them to where you are". In: *Proceedings of the fifth ACM international conference on Web search and data mining*. WSDM'12: Fifth ACM International Conference on Web Search and Data Mining. Seattle Washington USA: ACM, Feb. 8, 2012, pp. 723–732. ISBN: 978-1-4503-0747-5. DOI: 10.1145/2124295.2124380. URL: https://dl.acm.org/doi/10.1145/2124295.2124380 (visited on 02/23/2025).

IV. Appendix

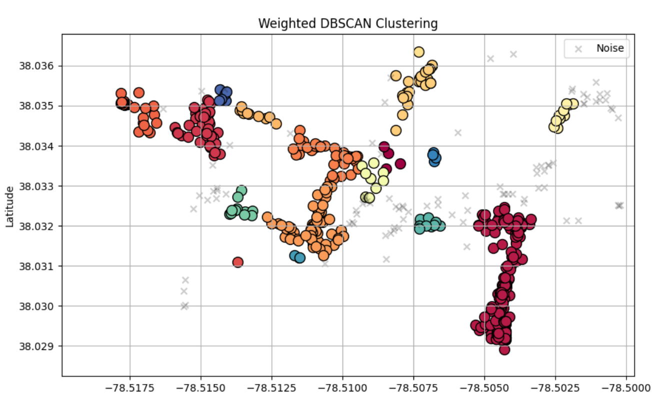

Fig. 9: Weighted DBSCAN Geolocation Clustering

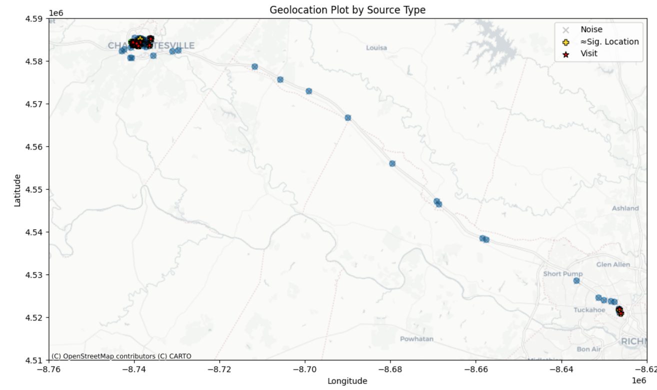

Fig. 10: Full View Geomap with Significant Location Guess per Cluster