Proceedings of The National Conference
On Undergraduate Research (NCUR) 2019
Kennesaw State University
Kennesaw, Georgia
April 11-13, 2019

Smart Transportation: Bridging the Gap in Traffic Information Systems

Philip Kenny, Daniel Henriques, Max Acosta
School of Computer Science
Kean University
Union, NJ 07083 USA

Faculty Advisor: Dr. Daehan Kwak

Abstract

With the ever-proliferation of vehicles, traffic congestion is a ubiquitous problem which degrades the social, environmental, and economic life for smart cities. Smart cities initiatives require intelligent transportation systems (ITS) to provide reliable and accurate traffic information to improve the traffic flow. During recent years, most users have transitioned from traditional navigation devices (e.g. TomTom) to online (or mobile) maps, such as GoogleMaps. Online maps have become the de facto standard for navigation and it is essential for route recommendations and travel time estimations (ETA) to be reliable as it is a major factor in route choice. However, differences in the recommend route among map providers and variable travel times (differences in ETA among maps) can introduce uncertainty in route choice making it challenging to determine the true best route. In this study, a web mining system is developed in Python running on the Amazon EC2 cloud to crawl and scrape traffic data to conduct a comparative analysis for five popular online maps (GoogleMaps, HERE, MapQuest, BingMaps, Waze). Origin/destination coordinates are selected from one city with one of the worst traffic hotspots in the U.S. provided by INRIX. Through descriptive analysis, it is identifiable that online maps potentially introduce uncertainty in route-choice decisions. To reduce the users' uncertainty, a user-friendly dashboard with data visualization capabilities for individual users and researchers is also developed to query and collect traffic data for routes of interest via our web mining system. The disciplinary significance of this work, on a macro-level, by continually collecting data, a base is built for future development that will benefit interdisciplinary research in the domains of computer science, data science, and transportation. On a micro-level, this study provides additional solutions to reduce the users' uncertainty in route choice.

Keywords: intelligent transportation systems (ITS), traffic information systems, route choice

1. Introduction

Within the transition to online map applications, such as Google Maps, HERE We Go, MapQuest, BingMaps, and Waze, it is essential for the user to be given the most accurate results as possible to reduce uncertainty [1]. Accurate results are also needed, as variable travel times between online maps can furtherly make users more uncertain of their route taken [2, 3]. Despite efforts to assist the route decision-making, users tend to ignore the route recommendations these online maps offer. In fact, Zhu [4] shows that only 13% of commuting trips coincide with the fastest route, which consequently complicates the goal of sustainable travel.

Albeit the helpfulness with the utilization of mobile maps for their users, there is no sufficient evidence displaying the comparison of results given by each. The comparison of these mobile maps provide information that is vital to the success of sustainable traveling, by displaying the most efficient route and ETA for users. The differentiation in route recommendations between mobile maps vary depending on multiple factors: routing algorithms used, time queried, current traffic patterns, and current available routes. Examples of identical route recommendations and different route recommendations queried at the same time for different mobile maps are shown in Figure 1. Comparing these differences and representing them is integral to assisting an online map user, by displaying comparative results between each online map which will reduce uncertainty in the route given.

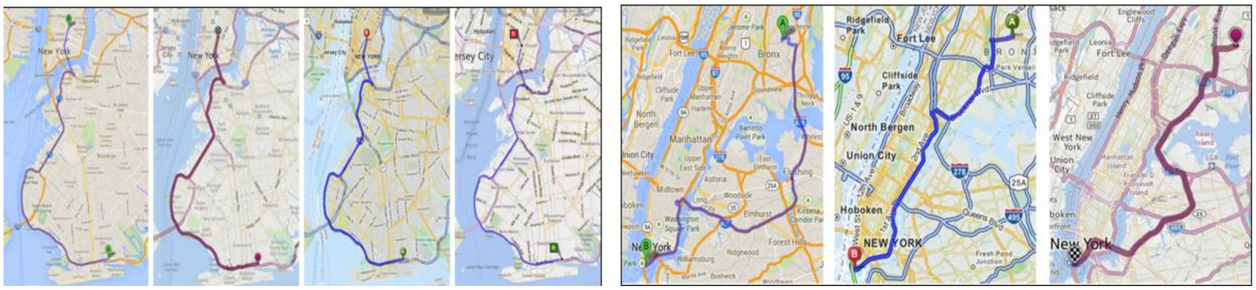

Figure 1. Example of identical route recommendations at the same time (left) and the example of different route recommendations at the same time (right)

1.1 Rationale

In order to achieve the goal of a comparative study of popular mobile maps, there are multiple factors to be considered. To differentiate query results from these mobile maps accurately, a static set of source and destination coordinates are required. There must be a way to query mobile map information both simultaneously and repeatedly throughout the day to gain sufficient evidence for a comparison. After this, necessary information must be scraped from these sources to retrieve important information to be compared, such as recommended route number, estimated time of arrival, distance from origin to destination, and a Timestamp of when the data was queried. This step is repeated depending on how many route recommendations there are from each mobile map. Then information received will additionally be stored on a database that will be used for final comparison of these mobile mapping systems.

The purpose of this system is to discover the variance in data queried from each mobile mapping system over an extended period of time. ETA is the most important factor in the data returned from each system, as it will determine the most efficient path from origin to destination. Comparing this information, the results will determine which mapping service leads to the quickest form of travel. These results are integral to the future of smart cities, which require the means of the fastest travel. The results are additionally important for sustainability due to the number of commutes taken by people that do not utilize the fastest route. Using results given by our research, drivers can improve their commute with the fastest route, and simultaneously achieve the goal of more sustainable transportation.

2. Materials and Methods

This research achieves the goal throughout using current technologies and automation. The materials that are used are objective for each part of the rationale. First, a foundation for our system is implemented throughout using an Amazon EC2 Cloud instance. This instance is essential to the system as it is responsible for running the web scrapers that will crawl data from mobile maps, exporting this travel information to a MySQL database, and then iterating this cycle on an interval to repeat the previous steps as shown is Figure 2.

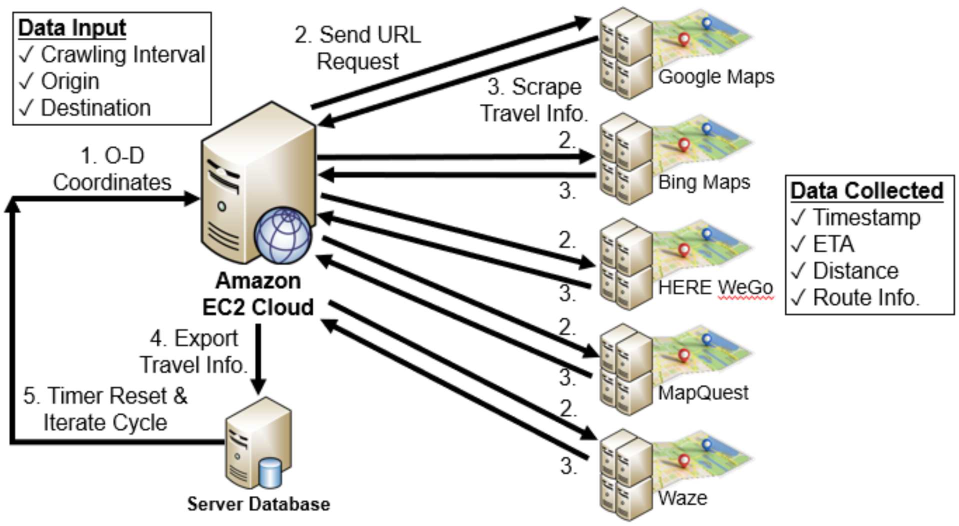

Figure 2. System architecture

The server database shown in Figure 2 is able to run multiple Python scripts at once by utilizing a terminal multiplexer. Terminal multiplexers are able to emulate terminals within a server, that can be attached or detached from their sessions. Using this, we create a terminal instance is created for each Python script that is used. This is an efficient solution for viewing the logged data that is scraped from the mobile map's web-app, and for monitoring the status of the scripts that are being run. Therefore, these instances provide a useful monitoring tool that ensures that there is an equal amount of data is being uploaded to each table. This helps leads to accurate results within the final analysis of the crawled data.

The server is the baseline for the system that runs our code developed in Python. We developed a lightweight script using Python that had libraries essential to web scraping purposes, such as Selenium. Selenium is an API that allows the automation of web browsers and allows developers to program interactions with a browser as a real user would. Using Selenium, an algorithm is developed that would load and open an online map's web-application with origin and destination coordinates. Then the program will crawl route information, consisting of the recommended route number, estimated time of arrival, distance from source to destination, and a Timestamp of when the data was queried. This is performed by crawling the xpath location of travel data on the map's web-application, which will always differ depending on online map used (Figure 3), and then putting this data into our variables. Therefore, there will be five Python scripts necessary for each static route used, because of the differentiation in the mobile map's xpaths, and the different database tables to be uploaded to. Consequently, this extracted data will be formatted to meet the requirements of the server's database, and then uploaded to populate our database using the PyMySQL API. This function will be repeated based off how many routes are currently available per map service, and then a timer will be set for the next iteration of web scraping.

```python
def route1(timeOfRetrieval):
    # Selects route 1 and retrieves route info
    route = browser.find_element_by_xpath("//*[@id='directionsPanelRoot']/div[2]/ul/li[1]/a/table/t
    route = (route.text).splitlines()

    # Retrieves time data from Route 1
    eta = browser.find_element_by_xpath("//*[@id='directionsPanelRoot']/div[2]/ul/li[1]/a/table/tr/
    eta = (eta.text).splitlines()

    # Retrieves distance data from Route 1
    distance = browser.find_element_by_xpath("//*[@id='directionsPanelRoot']/div[2]/ul/li[1]/a/tabl
    distance = (distance.text).splitlines()

    route = route[0]
    eta = eta[0]
    distance = distance[0]

    # Prints out Route 1 results
    print("Route 1: " + route + " | Time: " + eta + " | Distance: " + distance)

    upload(1, eta, distance, route, timeOfRetrieval)
```

Figure 3. Python code implementing travel information crawler

The last section of materials used is a MySQL relational database that is hosted on an EC2 instance. This is used to store all of the data that is extracted from each mobile map. There is one database for each set of origins and destinations, and there is a total of five tables within that database consisting of each online map's results. After the databases were populated, this data is exported to csv files as shown in Figure 4 and then compared for final analysis.

routenum	eta	distance	route	time
1	48 min	34.1 miles	I-95 N; Jeffer	Mon Mar 4 13:45:00 2019
2	51 min	33.3 miles	I-95 N	Mon Mar 4 13:45:00 2019
3	56 min	33.6 miles	I-95 N	Mon Mar 4 13:45:00 2019
1	39 min	33.3 miles	I-95 N	Mon Mar 4 14:00:00 2019
2	44 min	33.6 miles	I-95 N; Jeffer	Mon Mar 4 14:00:00 2019
1	34 min	33.3 miles	I-95 N	Mon Mar 4 14:15:00 2019
2	37 min	33.6 miles	I-95 N; Jeffer	Mon Mar 4 14:15:00 2019

Figure 4. Sample dataset from Waze table

2.1 Experiment Setup

The experiment set up using the implemented system consisted of the following:

- 30 days for getting results, 15-minute intervals between running all scrapers simultaneously.
- Selected one static origin and destination pair based on the INRIX® US Traffic Hotspots [5].
- Location that was chosen for this comparison was Washington D.C.
- Developed five scrapers with Python code used to crawl data from their corresponding online map.
- Scraped Data from the web applications of Google Maps, Bing Maps, HERE WeGo, MapQuest, Waze.
- Used one database that contains five database tables for storing data that was crawled from the scrapers.

This experiment was developed under the purpose of comparing each map recommended routes with their data given in one of the worst cities for traffic, Washington DC. Making an analysis on the data from this experiment is integral to finding the most efficient routes and online map service used to improve social, environmental, and economic life for Washington DC, and other smart cities.

2.2 Procedure

The system is developed in a manner of having placeholder variables for any experiment conducted with the goal of comparing online map data. The procedure used for our experiment consists of translating descriptive features from our experiment, into variables to replace the placeholders within our system. These placeholder variables consist of: cycle interval, origin and destination coordinates, and database used for the system. After replacing these variables, the procedure for this experiment will consist of the following:

1. Running Python code to request Washington D.C. route, and scraping travel data from the queried response.
2. Making travel data of a route into variables, and then uploading them into the Washington D.C. database.
3. Repeat previous step if multiple route recommendations exist.
4. Reset the 15-minute interval, and then iterate the cycle.

This procedure is executed for the 30 days of the experiment to populate our database with enough data to analyze.

3. Results

The results compiled were from exporting database tables into csv files, and then modifying them with Microsoft Excel functions. This was performed by gathering the ETA from the best route in each query, creating new rows of the data's mean value per each interval, and then extracting these rows with the best ETA into another file. Therefore, the file will contain the mean of each interval from 00:00 – 23:45, and then the file is used to create a visualization for a final descriptive analysis. This is done for each exported database table, and then graphs were created for comparing the best ETA for weekdays as shown in Figure 5 and comparing the best ETA for weekends as shown in Figure 6 in Washington D.C.

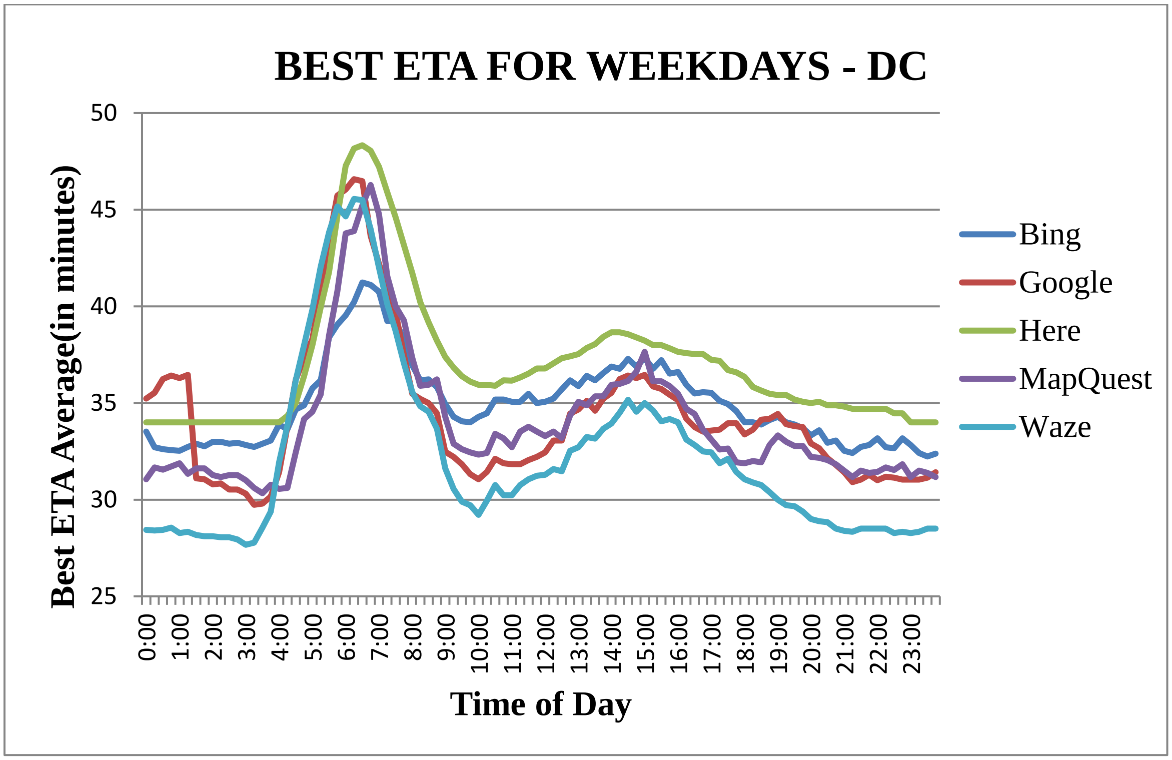

Figure 5. Visual representation of data of the best ETA for weekdays in Washington D.C.

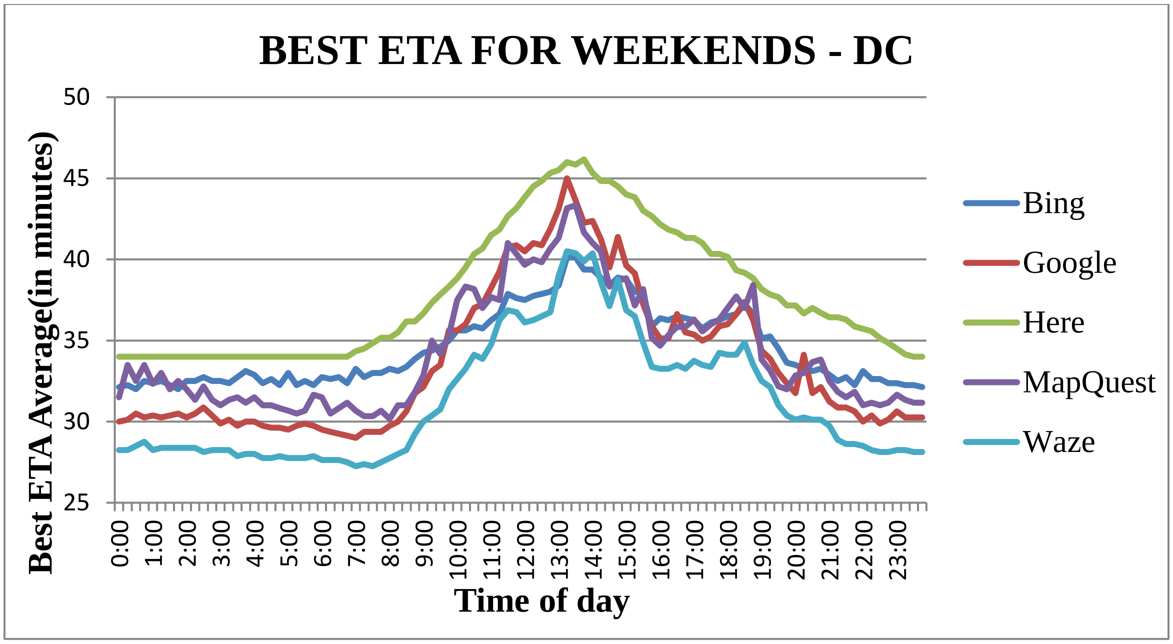

Figure 6. Visual representation of data of the best ETA for weekends in Washington D.C.

Comparing the results from the graph, it is visible that Waze had the overall shortest ETA, and HERE We Go consistently had the longest ETA. Google Maps and MapQuest had very similar results, and would often cross during rush-hour traffic of the day. The variability in ETAs are outcomes of multiple factors, speculating to be the routing algorithms used by map service, current traffic patterns used, time queried, and paths discovered by the map service. In regards with Waze, the speculated reason why it had the best ETA, stems from live updates of traffic patterns from local users of Waze. Waze is the largest social-based traffic and navigation application, that allows traffic and hazards to be reported by its users. This results in Waze calculating other faster routes, which leads to an overall shorter ETA. Therefore, online map applications which receive live updates, can reroute drivers and get them to their destination sooner. This is an important concept that should be used to further the goal of sustainable and efficient traveling.

4. Conclusion

In conclusion, online maps have become the *de facto* standard for navigation apps, and their utilization has an impact to address the key challenges into reducing traffic navigation. Our system developed was created for the purpose of analyzing and comparing these online maps. We analyze data queried from five popular online maps to check for discrepancies between the map providers and establish whether uncertainty is introduced in route decision-making. By shedding light on the characteristics of ETA variability, and differences in optimal routes offered against maps, for future work we plan on impacting individuals by reducing the users' uncertainty in route choice. On the broader impact we plan on continually collecting data to build a base for interdisciplinary research in Computer Science, Data Science, and Transportation.

5. Acknowledgments

This work is supported by the National Science Foundation under Grant No. 1551221, Computing Alliance of Hispanic-Serving Institutions (CAHSI), and Kean Students Partnering with Faculty (SpF) Summer Research Program.

6. References

1. Kwak, Daehan, Daeyoung Kim, Ruilin Liu, Badri Nath, and Liviu Iftode. "DoppelDriver: Counterfactual actual travel times for alternative routes." In 2015 IEEE International Conference on Pervasive Computing and Communications (PerCom), pp. 178-185. IEEE, 2015.
2. Wang, Lin, and Ju, Da Young. "Concurrent use of an in-vehicle navigation system and a smartphone navigation application." Social Behavior and Personality: an international journal 43, no. 10 (2015): 1629-1640.
3. Sikka, Nikhil. "Understanding travelers' route choice behavior under uncertainty." PhD thesis, University of Iowa, 2012.
4. Zhu, Shanjiang, and Levinson, David. "Do people use the shortest path? An empirical test of Wardrop’s first principle." PloS one 10, no. 8 (2015): e0134322.
5. Bob Pishue, "US Traffic Hotspots: Measuring the impact of congestion in the United States," INRIX Research, September 2017. http://www2.inrix.com/us-traffic-hotspot-study-2017.