uberNow Documentation

Release 1.1.0

Anirban Roy Das

Apr 10, 2017

Contents

It is an app to notify users via email as to when to book an uber as per given time to reach a specified destination from a specified source.

Home Page : https://www.github.com/anirbanroydas/uberNow

CHAPTER 1

Details

Author Anirban Roy Das

Email anirban.nick@gmail.com

CHAPTER 2

Documentation:

Overview

uberNow is a small web app which takes four inputs from the user and notifies the user via email as to when to book an Uber.

All you(the user) have to do is give the 4 required inputs and submit the request. The app will take care of the rest.

The inputs are:

- 1. source
- 2. destination
- 3. email address
- 4. time to reach destination

It uses Tornado as the application server. The web app is created using the sockjs protocol. **SockJS** is implemented in many languages, primarily in Javascript to talk to the servers in real time, which tries to create a duplex bi-directional connection between the **Client(browser)** and the **Server**. Ther server should also implement the **sockjs** protocol. Thus using the sockjs-tornado library which exposes the **sockjs** protocol in Tornado server.

It first tries to create a Websocket connection, and if it fails then it fallbacks to other transport mechanisms, such as **Ajax**, **long polling**, etc. After the connection is established, the tornado server **(sockjs-tornado)** calls Uber Apis and Google Maps Apis to process the requests.

Problem Statement

Short Version

Given four inputs, (a). **Source** (b). **Destination** (c). **Email Address** (d). **Time of the day to reach the destination**, send an email to the given email address about booking its uber at that very moment so that you have ample time to reach your destination which takes into consideration the following conditions:

- Time the uber driver will take to reach your source.
- Time it will take to reach your destination in the current live traffic conditions.
- Whether you will be actually be able to book an uber at that very moment or not or will it take some time before you can actually book a cab.
- The actual time when you reach your destination should be as accurate or close to the pre defined time as possible.

Long Version with example

Given four inputs, (a). **Source** (b). **Destination** (c). **Email Address** (d). **Time of the day to reach the destination**. Assume that source and destination are provided as latitude/longitude and not as addresses(to keep things simple). The app needs to find the exact time a user needs to book an uber to be at that destination at that time. An email needs to be sent to the above email ID at this time saying "Time to book an uber!"

Example: I am in Koramangala(12.927880, 77.627600) and need to be in Hebbal(13.035542, 77.597100) at 8:00 PM for a meeting. The app will have to email me at 6:43 PM because the uberGO will take 9 mins to reach me at 6:43 PM(as per uber) and it takes 68 mins to drive from Koramangala to Hebbal(as per google maps)

Assumptions:

- To keep things simple, whatever the uber api and google maps api tell is true.
- Time to arrive at the destination is within today. The time entered is in IST.
- Not considering cases where it's already too late to book the uber. Assuming that there is still some time left.
- The maximum deviation of driving times at any time of the day is 60 mins. That means, if it takes 40 mins to drive from Koramangala to Hebbal now, assume that it's not more than 100 mins at any time of the day.
- The cab is UberGO.
- Accuracy is the most important factor.
- Second important factor is optimization of requests to the APIs. Pinging the APIs every minute to see if you should leave now is not the best solution.

Clarifications:

1. Lets say you want to reach destination B at 8:00 pm starting from source A at 7:30 pm, the travel time is 25 minutes but the next time you pool google api, it says now the travel time is 40 mins. In such case its already to late to book an uber.
2. Lets say, first time you pool google api and it says that it will take 1 hour from source A to destination B. But next time you check the google api, it says 1:45 mins (Huge Traffic Jam) and then next time it say 2:15 mins (It started raining :( ). You can assume that the maximum deviation is 1 hour at any time of the day i.e max it will take 2 hours to reach destination.

Algorithm & Logic

This project does not have a definitive solution. Its an open ended problem. So the attempt to solve the problem is to find a good solution that solves the problem with as much less complexity and as much better scalability as possible. Since this is an open ended problem, the solutions and algorithms put forward here may not be convincing to you but this is the one that is being used to solve this project presently. You can fork the project and apply your own logic to it. Send a PR and we will see if it improves upon the present logic.

Before moving forward lets discuss the problem first and why a straight definitive solution is not possible here.

Discussion

The Uber Api only allows to check for drivers and book can instantly. Now whenever you try to book anduber its not always certain that the cab will be present in your area and thus you have to keep checking once in a while to book a cab if no cabs present at the time you are trying to book. There is not provision to book an uber for a furture time. Its always on the spot instant booking.

This means being certain about uber cab booking is difficult and you have to have hitory of booking data to be certain that at a particular time and place what is the probability that you will get a successful booking. Also you sometimes have to book the cab in advance even if taht means reaching destination before but its always better thatn reaching destination after the specified time. That means piniding uber api in intervals is a safe (not the best) way to start.

Method 1 - Optimized Prediction (by using Google Maps optimized traffic model for future traffic predictions)

In this method, since google gives you few traffic models via which you can predict future traffic duration to reach from point A to point B, we can pre decide the time.

Example : Request by User 1 -

- Source - Bellandur
- Destination - Koramangala
- Time of giving request - 2pm
- Time to reach destination - 8pm.

Lets say, calls to **uber apis** as **UP**, and calls to **google maps apis** as **GP**, where **UP1**, **UP2**, **UP3...UPi** means the number of calls to Uber Apis, similarly **GP1**, **GP2**, **GP3**. All these numbering are per request basis. I mean suppose **user 1** sent the above request, then **UP1, UP2, UP3...GP1, GP2, GP3...etc** will be for **user 1**. For another user, these UPi's and GPi's set will be different though their numbering for each user will be same.

Step 1

- UP1 at 2pm => result - 7 mins.
- GP1 at 2pm =>

Now GP1 will do 2 different types of requests, i.e based on 2 different traffic models of google Maps.

1. **Traffic Model = Best_Guess:** Takes departure time as input and sends the duration to reach as a best guess which considers past history at that time and current traffic conditions.
2. **Traffic Model = Pessimistic:** Takes departure time as input and sends the duration to reach as a pessimistic time which considers the past history and gives a time which is higher than most days.

NOTE : We can use both models to build the predictions simultaneously and use one of them for few days and the other for some days and see which one gives us better results and then with continuous use, we will be able to choose one over the other eventually.

- **GP1.1** => Best Guess with dep. time = now(2pm) => result = 57mins
- **GP1.2** => Pessimistic with dep. time = now(2pm) => result = 70 mins

These our base rough estimations so that we can start our predictions from some base values.

- **best_guess_suitable_starting_times** = BST = { }
- **pessimistic_suitable_starting_times** = PST = { }

Step 2 (our real predictions start from this step)

New dep. time = 8pm - [57mins (best_guess) or 70mins (pessimistic)] lets say, dep_t_b_g = 7.03pm and dep_t_p = 6.50pm

- **GP2.1** => Best Guess with dep. time as 7.03pm => result = 59mins

Difference to reach before 8pm = 57-59 = -2 (negative means its going to cross 8pm by 2 mins, so not a suitable starting time)

- **GP2.2** = Pessimistic with dep. time as 6.50pm => result = 72mins

Difference to reach before 8pm = 70-72 = -2 (negative means its going to cross 8pm by 2 mins, so not a suitable starting time)

Step 3

New dep. time = 8pm - [59mins (best_guess) or 72mins (pessimistic)] lets say, dep_t_b_g = 7.01pm and dep_t_p = 6.48pm

- **GP3.1** => Best Guess with dep. time as 7.01pm => result = 60mins

Difference to reach before 8pm = 59-60 = -1 (negative means its going to cross 8pm by 1 mins, so not a suitable starting time)

- **GP3.2** => pessimistic with dep. time as 6.48pm => result = 70 mins

Difference to reach before 8pm = 72-70 = 2 (positive means its going to reach 8pm earlier by 2 mins, so its a suitable starting time. Thus at this we will mark MIN diff for now as = 2 min)

- **PST** = { 6.48 : 2, }

Step 4

New dep. time = 8pm - [60mins (Best Guess) or 70mins (Pessimistic)] dep_t_b_g = 7pm and dep_t_p = 6.50 pm

- **GP4.1** => best_guess with dep. time as 7pm => result = 58mins

Difference to reach before 8pm = 60-58 =2 (positive means its going to reach 8pm earlier by 2 mins, so its a suitable starting time. Thus at this we will mark MIN diff for now as = 2 min)

- **BST** = { 7.00 : 2, }
- **GP4.2** => pessimistic with dep. time as 6.50pm => result = 69mins

Difference to reach before 8pm = 70-69 =1 (positive means its going to reach 8pm earlier by 1 min, so its a suitable starting time. Thus at this we will mark MIN diff for now as = 1 min, since its less than the current MIN of 2 mins)

- **PST** = { 6.48 : 2, 6.50 : 1, }

Step 5

New dep. time = 8pm - [58mins (best_guess) or 69mins (pessimistic)] dep_t_b_g = 7.02pm and dep_t_p = 6.51pm

- **GP5.1** => best_guess with dep. time as 7.02min => result = 58mins

Difference to reach before 8pm = 58-58 =0 (positive means its going to reach 8pm earlier by 0 mins, so its a suitable starting time. Thus at this we will mark MIN diff for now as = 0 min, since its minimum than than the last minimum of 2 mins)

- **BST** = { 7.00 : 2, 7.02 : 0, }

Here, in case of best_guess we found that if we leave at 7.02 pm we will reach at 8pm exact, i.e. the difference to reach before 8pm is minimal here, in this special case its 0. So we will stop our best case.

- **GP5.2** => pessimistic with dep. time as 6.51pm => result = 66mins

Difference to reach before 8pm = 69-66 =3 (positive means its going to reach 8pm earlier by 0 mins, so its a suitable starting time. Thus we will not change the minimum since its higher than the current MIN of 1min)

- **PST** = { 6.48 : 2, 6.50 : 1, 6.51 : 3, }

Hence we will continue the pessimistic until we get few possible suitable starting times as k-minimum of the list, where k can be 3 to 5 or more.

While in the process, anytime its possible that we get a negative time difference continuously for more than 3-5 times, then we will start going back in time by 5 mins and start the cases again.

NOTE : I tested the above method with different source and destination pairs and different reaching time per pair. So I tested for several combinations, and in all the cases I found this method to converge and give a result having minimum time difference as 0 at some point in time in more than majority of the cases.

You can say this method to be some kind of **heuristic** approach which has shown positive results although I tested it with only 30 combinations with consistent results.

Now what we are left with are **BST** and **PST**, now we have to somehow consider the **Uber APIs** and consider the **uber ETAs**.

NOTE : Uber APIs are real time, we cannot pool uber apis for future time like we did for google apis.

For this we can start pooling Uber APIs with our base uber eta case, which was 7 mins.

So for **BST = { 7.00 : 2, 7.02 : 0, }** we will start pooling uber at times equal to:

- **UP2.1** = 7.02pm - 7mins = 6.55pm

Now if eta is 2 mins, then we find that, we will have to start at 6.57. but we have our data as to start at 7.02 pm. So 6.55 pm is not that suitable.

We need a time, so that the eta for uber should reach as close to 7.02pm, i.e lets say if eta is 5mins at 6.55pm, then we can start our trip at 7.00pm which is closer to 7.02 pm than the earlier 6.57pm.

So 6.55 pm in this case when eta is 5 mins is better suitable time than the earlier one when eta was 2 mins.

So how to figure this out?

Heuristic :

Suppose initially we find **ETA** to be 2 mins. And our base case eta was 7mins.

- pessimistic eta = 7mins
- best_guess eta = avg of previous etas = (7-2)/2 = 4.5 mins

So using **best_guess** we can again pool uber api for eta at (7.02pm - 4.5mins)= 06:57:30pm.

We need to consider if the time to pool uber api has already gone or not, if yes, then we simple pool uber api at current time, else we call at that future time, in this case at 06.57.30pm.

If eta at 6:57:30pm is 2 mins again, then we can start our trip at 6.59:30 which is close to 7.02pm positively by 2.5 mins.

Now in all this, we have to simultaneously consider the other options in the BST, for example in this case 7:00pm.

Now we see that if we start at 6.57:30 pm then eta is 2 mins and uber will reach at 6.59:30pm which is only 30 seconds earlier to the 7:00pm option, in which we will reach our final destination earlier by 2 mins at 7.58pm.

So we can either choose this option or pool the uber apis again for a new eta.

Lets say we already pooled at 6.57:30pm with eta 2.

- pessimistic eta is still 7min
- best_guess eta = avg of previous = (4.5-2)/2 = 3.25 mins

So a next suitable uber api pooling time would be 7.02pm - 3.25mins = 6.58:45 pm

Now if this time has not been gone than we can pool at this time and get new data or we can just choose the other above option.

NOTE : At anytime, when we see that our etas are becoming longer, we can just stop processing further and send that current time as the suitable time so as to not affect our actual destination reaching time by very big margins.

NOTE : This whole method was optimized prediction using google's traffic models.

There is another basic predictions which is not going to use google's future time traffic predictions and pool google apis always in real time. This basic prediction will be more difficult and will consider the assumption of deviation to be at most by 1 hour.

Method 2 - Machine Learning with predictive model

This is still under consideration but yet to be implemented. Here we try to learn from the uber api calls and train on the dataset of the avaialble cabs at any particular time at any particular place and increase the accuracy of makeing uber api calls and this weiill allow us to not make uber api calls in intervals and be more precise about when to make the uber api calls.

This method need access to uber api calls data and time to lear and train on the data and develop a better predictive model using it. Its not implemented yet. Happy to chat and discuss over this. I will work on this when I get time. This method needs more time and dedication. The first method is easy and thus I did it first and haven't got time around to implement the second method so far.

Features

Technical Specs

sockjs-client (optional) Advanced Websocket Javascript Client used in **webapp example**

Tornado Async Python Web Library + Web Server

sockjs-tornado SockJS websocket server implementation for Tornado

Uber Time-Estimation Apis HTTP Rest APIs to estimate time required to book an uber at given time

Google Maps Distance-Matrix Apis HTTP Rest Apis to calculate distance and duration to reach from source to destination

pytest Python testing library and test runner with awesome test discobery

pytest-flask Pytest plugin for flask apps, to test fask apps using pytest library.

Uber's Test-Double Test Double library for python, a good alternative to the mock library

Jenkins (Optional) A Self-hosted CI server

Travis-CI (Optional) A hosted CI server free for open-source projecs

Docker A containerization tool for better devops

Features Specs

- Web App
- Email Notification
- Uber Booking Reminder
- Microservice
- Testing using Docker and Docker Compose
- CI servers like Jenkins, Travis-CI

Installation

Prerequisite (Optional)

To safegurad secret and confidential data leakage via your git commits to public github repo, check `git-secrets`.

This git secrets project helps in preventing secrete leakage by mistake.

Dependencies

1. Docker
2. Make (Makefile)

See, there are so many technologies used mentioned in the tech specs and yet the dependencies are just two. This is the power of Docker.

Install

- **Step 1 - Install Docker**

  Follow my another github project, where everything related to DevOps and scripts are mentioned along with setting up a development environemt to use Docker is mentioned.

  - Project: https://github.com/anirbanroydas/DevOps
  - Go to setup directory and follow the setup instructions for your own platform, linux/macos

- **Step 2 - Install Make**

```
# (Mac Os)
$ brew install automake

# (Ubuntu)
$ sudo apt-get update
$ sudo apt-get install make
```

- **Step 3 - Install Dependencies**

  Install the following dependencies on your local development machine which will be used in various scripts.

1. openssl
2. ssh-keygen
3. openssh

CI Setup

If you are using the project in a CI setup (like travis, jenkins), then, on every push to github, you can set up your travis build or jenkins pipeline. Travis will use the `.travis.yml` file and Jenknis will use the `Jenkinsfile` to do their jobs. Now, in case you are using Travis, then run the Travis specific setup commands and for Jenkins run the Jenkins specific setup commands first. You can also use both to compare between there performance.

The setup keys read the values from a `.env` file which has all the environment variables exported. But you will notice an example `env` file and not a `.env` file. Make sure to copy the `env` file to `.env` and **change/modify** the actual variables with your real values.

The `.env` files are not commited to git since they are mentioned in the `.gitignore` file to prevent any leakage of confidential data.

After you run the setup commands, you will be presented with a number of secure keys. Copy those to your config files before proceeding.

NOTE: This is a one time setup. **NOTE:** Check the setup scripts inside the `scripts/` directory to understand what are the environment variables whose encrypted keys are provided. **NOTE:** Don't forget to **Copy** the secure keys to your `.travis.yml` or `Jenkinsfile`

NOTE: If you don't want to do the copy of `env` to `.env` file and change the variable values in `.env` with your real values then you can just edit the `travis-setup.sh` or `jenknis-setup.sh` script and update the values their directly. The scripts are in the `scripts/` project level directory.

IMPORTANT: You have to run the `travis-setup.sh` script or the `jenkins-setup.sh` script in your local machine before deploying to remote server.

Travis Setup

These steps will encrypt your environment variables to secure your confidential data like api keys, docker based keys, deploy specific keys.

```
$ make travis-setup
```

Jenkins Setup

These steps will encrypt your environment variables to secure your confidential data like api keys, docker based keys, deploy specific keys.

```
$ make jenkins-setup
```

Usage

After having installed the above dependencies, and ran the **Optional** (If not using any CI Server) or **Required** (If using any CI Server) **CI Setup** Step, then just run the following commands to use it:

You can run and test the app in your local development machine or you can run and test directly in a remote machine. You can also run and test in a production environment.

Run

The below commands will start everythin in development environment. To start in a production environment, suffix `-prod` to every **make** command.

For example, if the normal command is `make start`, then for production environment, use `make start-prod`. Do this modification to each command you want to run in production environment.

Exceptions: You cannot use the above method for test commands, test commands are same for every environment. Also the `make system-prune` command is standalone with no production specific variation (Remains same in all environments).

- **Start Applcation**

```
$ make clean
$ make build
$ make start

# OR

$ docker-compose up -d
```

- **Stop Application**

```
$ make stop

# OR

$ docker-compose stop
```

- **Remove and Clean Application**

```
$ make clean

# OR

$ docker-compose rm --force -v
$ echo "y" | docker system prune
```

- **Clean System**

```
$ make system-prune

# OR

$ echo "y" | docker system prune
```

Logging

- To check the whole application Logs

```
$ make check-logs

# OR

$ docker-compose logs --follow --tail=10
```

- To check just the python app's logs

```
$ make check-logs-app

# OR

$ docker-compose logs --follow --tail=10 identidock
```

API

This contains all the modules and classes used to make the app. You can go through each of them for better understanding of the project.

Main View

IndexHandler

BookingHandler

Api Calls Module

Api

Testing

Now, testing is the main deal of the project. You can test in many ways, namely, using `make` commands as mentioned in the below commands, which automates everything and you don't have to know anything else, like what test library or framework is being used, how the tests are happening, either directly or via `docker` containers, or may be different virtual environments using `tox`. Nothing is required to be known.

On the other hand if you want fine control over the tests, then you can run them directly, either by using `pytest` commands, or via `tox` commands to run them in different python environments or by using `docker-compose` commands to run differetn tests.

But running the make commands is lawasy the go to strategy and reccomended approach for this project.

NOTE: Tox can be used directly, where `docker` containers will not be used. Although we can try to run `tox` inside our test contianers that we are using for running the tests using the `make` commands, but then we would have to change the `Dockerfile` and install all the `python` dependencies like `python2.7`, `python3.x` and then run `tox` commands from inside the `docker` containers which then run the `pytest` commands which we run now to perform our tests inside the current test containers.

CAVEAT: The only caveat of using the make commands directly and not using `tox` is we are only testing the project in a single `python` environment, nameley `python 3.6`.

- To Test everything

```
$ make test
```

Any Other method without using make will involve writing a lot of commands. So use the make command preferrably

- To perform Unit Tests

```
$ make test-unit
```

- To perform Component Tests

```
$ make test-component
```

- To perform Contract Tests

```
$ make test-contract
```

- To perform Integration Tests

```
$ make test-integration
```

- To perform End To End (e2e) or System or UI Acceptance or Functional Tests

```
$ make test-e2e

# OR

$ make test-system

# OR

$ make test-ui-acceptance

# OR

$ make test-functional
```

CHAPTER 3

Indices and tables

- genindex
- modindex
- search