User Manual of the Taxi Query Generator

Version 0.1

June 20, 2013

This document is the user manual for a program called *Taxi Query Generator (TQG)* which outputs taxi queries (i.e. requests for taxis from passengers) conforming to the real query distribution over space and time in Beijing. The software leverages T-drive trajectory dataset. In the rest of this document, we explain the rationale behind the software in Sec. 1. We describe the format of the output files in Sec. 2 and detail the usage of the software in Sec. 3.

1. Rationale behind the TQG

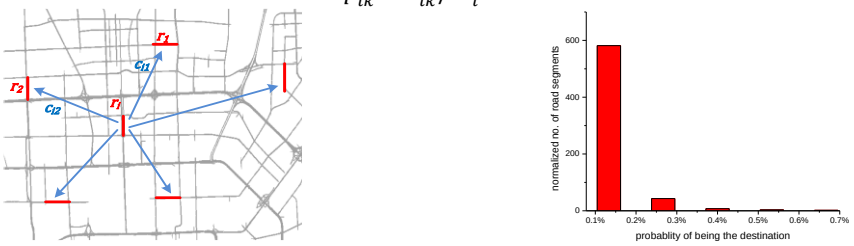
The purpose of this software is to generate taxi ride requests that are as realistic as possible. For this purpose, we first discretise one day into small time frames, denoted by f_j's. Denote all road segments by r_i's. We assign all historical ride requests into time frames.

Assume that the arrivals of ride requests on each road segment approximately follow a Poisson distribution during each time frame f_j. Thus, we can learn λ_i^j, i.e. the parameter of the Poisson distribution for road segment r_i during time frame f_j. Specifically, for each road segment r_i, we count the number of ride requests that originate from r_i within time frame f_j, denoted by c_i^j. Then we calculate λ_i^j based on c_i^j using Eq. (1) (where $len(f)$ is the length of a frame in minutes) and generate a ride request stream that follows a Poisson process with parameter λ_i^j.

In order to generate destination of requests truthfully, for each c_i^j, we decompose it into an array of numbers $\{c_{i1}^j, c_{i2}^j, \ldots c_{im}^j\}$, where $c_{ik}^j, k = 1,2,..m$ represents the number of requests which origins at road segment r_i and destines towards road segment r_k during time frame f_j, as illustrated by Fig. 1 (a). Therefore, the transition probability from r_i to r_k during time frame f_j, denoted by p_{ik}^j, can be estimated using Eq. (2). Fig. 1 (b) shows the distribution of destination road segments for requests that originate from road segment r_i. For each ride request Q generated in frame f_j with the origin road segment being r_i, the destination road segment is generated according to the transition distribution p_{ik}^j.

$$\lambda_i^j = c_i^j / len(f) \quad (1)$$

$$p_{ik}^j = c_{ik}^j / \ c_i^j \quad (2)$$



(a) Different destination road segments for requests originated from r_i

(b)Distribution of destination road segments for requests originiated from a specific road segment

Fig. 1 Transition probability of road segment r_i during a time frame f_j

Note that the taxi GPS trajectory dataset only reveals the number of ride requests that got served. In reality there are also many ride requests unsatisfied and disappeared due to the shortage of taxis. To take such ride requests into consideration, we introduce a system parameter Δ, supposing the number of real ride requests is Δ times the number of request extracted from the trajectory dataset. Fig. 2 (a) and (b) show the supposed number and the extracted number of ride requests that originate from all road segments and a specific road segment, respectively, over time of a day, where the time frame is 1 hour and Δ=2.

(a) from all road segments

(b) from a particular road segment

Fig. 2. Inflated and extracted number of ride requests during a day.

2. Output File Format

The output files have the following fields:
BIRTH_TIME, ORIGIN_LATITUDE, ORIGIN_LONGITUDE, DESTINATION_LATITUDE, DESTINATION_LONGITUDE

3. Usage

The config file contains all the input parameters that are read by the program. Please change them according to your needs.

Make sure that the "config" file and the "i" folder is in the same directory with the TaxiQueryGenerator.jar file.

Run the jar file by typing "java -jar TaxiQueryGenerator.jar".

4. Paper Citation

Please cite the following papers if you use the software:

Shuo Ma, Yu Zheng, Ouri Wolfson. T-Share: A Large-Scale Dynamic Taxi Ridesharing Service. In *Proceedings of 29th IEEE International Conference on Data Engineering, ICDE 2013*.

Shuo Ma, Yu Zheng, Ouri Wolfson. Real-time City-Scale Taxi Ridesharing. IEEE *Transactions on Knowledge and Data Engineering*, 2014