**Travel Time through ETC**

A

1. Introduction

Recently travel time prediction has been an important issue in intelligent transportation system, especially in handling congestion problems. According to 2015 urban mobility scorecard [1], in 2014 the congestions in America had caused $160 billion congestion cost and wasted Americans 6.9 billion hours since of the congestions. To reduce the congestion cost and the waste of time, travel time prediction is needed. With that, travelers can make smart decisions about when to travel and on what routes to travel if the prediction is accurate.

In the past, it is hard to make travel time prediction since the traffic data is hard to collect for prediction. Fortunately, the advance of Internet of Things (IoT) makes it easier to collect traffic data in recent years. For example, the electronic toll collection (ETC) system is set up commonly in highway traffic system in many countries, and its popularity is still growing. ETC can collect numerous traffic data for the tolls and traffic control. From 2010 to 2015 in US, the electronic transponders on America’s roads increased 19.3M (million), while the ETC accounts increased from 19.9M to 32.7M [2]. Figure 1 shows the increase of toll accounts and transponders in United States, compared to the statistical records in the two years in 2010 and 2015.

|  |
| --- |
|  |

Figure 1. ETC accounts by state in U.S between 2010 and 2015

Since the traffic data is collected much easier than ever, many studies had addressed travel time prediction in the last decade. Travel time prediction aims at measuring future travel time for the same trip. For example, suppose node A is as a starting point and node B is as a destination. In general, people consider historical data in a period (e.g. a month away from the prediction time) of traffic data of those vehicles who took a trip from nodes A to B. Figure 2 shows the data collection for the vehicles in a specific route from day 1 to day (*m* + *k*). Assume we need to predict the traffic time for the route in time slot *t* as well as in latter time slots in day *m*. According to rule of thumb, the travel time prediction for a route in time slot *m* only related to its prior travel time in the same route; the posterior traffic condition, like congestion, and accidents, is unrelated to the prediction. All the literature for the travel time prediction adopts the assumption to train their prediction model. In detail, the traffic date from day 1 to day m are the historical data, and in each of those days the time slots from (*t*-*n*) to *t* are the data used to train the model.

|  |
| --- |
|  |

Figure 2. The traffic data for the traffic time prediction

2. Material

2.1. Collected traffic data set and its preprocessing

Since of the advance of internet of things, the data collection is easier than ever. In most countries, they incorporate ETC to reduce the workload of tollgates. The incorporation of ETC can also provide the information about the route of a vehicle along with its entry to the highway, the intersection and its exit from the highway. Figure 3 shows the road network topology in a target area. Figure 3(a) is a bird-eyes view for an area, where we can see the in-flow and out-flow of the vehicles from their entries from intersections, their travelling routes across tollgates, and their exit from intersections. In fact, the ETC is tracing the vehicle in links (or segments), each of which is around 100 meters. Figure 3(b) is the route of a vehicle, which is composed by several links along the route from intersection A to tollgate 1.

|  |  |
| --- | --- |
| https://img.alicdn.com/tps/TB1cEAlPFXXXXaBXFXXXXXXXXXX-865-454.png  (a) | (b) |

Figure 3. The road network topology in an area. The bird-eyes view for the area is shown in Figure 3(a); the links of a route traced by ETC is shown in Figure 3(b).

Since ETC sensors are installed in a fixed location along the high way, the collected data is from the dataset of the sensors without route consideration. The route connection will not show is the collection file from ETC. The paper used KDDCup 2017’s traffic-flow prediction dataset (<https://tianchi.aliyun.com/competition/information.htm?raceId=231597>) [12] in an area in China, containing the information of the route, passing time, and weather along the highway. The dataset are real data, consisted of several files, only four of which are useful for travel time prediction. The other files are used for traffic volume prediction, which is beyond the scope of this paper. Figure 4 shows the description of these four files; in detail, the first file, links.csv, describes the profile of each link in the target area, including its ID, link’s width, length, and lanes, and its in-coming links and out-going links. Note that a link is one-directional; there may be more than one links (i.e., in-coming links) flow into that link, and there may be more than one links (i.e., out-going links) flow away that link. The second file, routes.csv, describes a route from an intersection to a tollgate, and all the links along the route. Since a link is around 100 meters away from each other, a route is normally consisted of several links, and their order is arranged according to their connected order in the route. The third file, weather.csv, contains a travel time for some specific vehicle in a route from an intersection to a tollgate. Each record of the file includes the start and end points of a route, the links (or segments) of a rout, the vehicle ID traveling the route, the starting time for the travelling and total time spent on the route for the vehicle. Note that the vehicle ID is masked but represented as another symbols for privacy consideration.

|  |
| --- |
|  |

Figure 4. The data for the travel time prediction from KDDCup 2017.

In general, the congestion occurs in the same time slot of each day. For example, the congestion often appears in the rush hours, AM 0700-0900 and PM0500-0700, every work day if no accident occurs. Without loss of generality, the time slot is set as one hours in the paper. Before training the model for prediction, we transform the raw data in the above four files into the needed format. The needed data, called travel time features, includes the average travel time, traffic volume, day of week, rainfall, and so on, which are listed in Table 1.

Table 1. The travel time features extracted from the four files from KDDCup 2017.

|  |  |  |
| --- | --- | --- |
| Feature | Value | Description |
| Average travel time | Float | Average travel time, measured in seconds |
| Traffic volume | Integer | Traffic volume |
| Day of week | 0~6 | 0: Monday, 1: Tuesday, …, 6: Sunday |
| Workday | 0,1 | 0: not workday, 1: workday |
| Public holiday | 0,1 | 0: not public holiday, 1: public holiday |
| Rainfall | Float | Amount of rainfall, measured in millimeter |







1. Experiments DATA

The data comes from KDDCup 2017’s traffic-flow prediction dataset, which contains six routes, namely, route A~2, A~3, B~1, B~3, C~1, and C~3, where the naming rule for a route is the concatenation of intersection ID, “~”, and then tollgate ID. For example, for two routes’ notation, A~2 and A~3, they mean that the two routes have the same starting point, i.e., intersection A, but have different destination points, i.e., tollgates 2 and 3, respectively. In addition, the dataset also includes the time records of thousands of vehicles travelling on the six routes as well as the weather information in the area from the date 2016-07-19 to the date 2016-10-24.