A Smart Bus Tracking System Based on Location-Aware Services and QR Codes

Süleyman Eken  
Department of Computer Engineering  
Kocaeli University  
Kocaeli, Turkey  
suleyman.eken@kocaeli.edu.tr  

Ahmet Sayar  
Department of Computer Engineering  
Kocaeli University  
Kocaeli, Turkey  
ahmet.sayar@kocaeli.edu.tr  

Abstract—When it comes to taking the public transportation, time and patience are of essence. In other words, many people using public transport buses have experienced time loss because of waiting at the bus stops. In this paper, we proposed smart bus tracking system that any passenger with a smart phone or mobile device with the QR (Quick Response) code reader can scan QR codes placed at bus stops to view estimated bus arrival times, buses’ current locations, and bus routes on a map. Anyone can access these maps and have the option to sign up to receive free alerts about expected bus arrival times for the interested buses and related routes via SMS and e-mails. We used C4.5 (a statistical classifier) algorithm for the estimation of bus arrival times to minimize the passengers waiting time. GPS (Global Positioning System) and Google Maps are used for navigation and display services, respectively.

Keywords—QR codes; GPS; Smart bus stops; C4.5 algorithm; Smart phones, Interactive maps.

I. INTRODUCTION

Public transport has become a part of life. Most people reach from homes to workplace or school using public transportation. People can lose time in transportation because of unwanted waiting. Also, people have the right to know where the bus is now and how long time it takes bus to reach bus stop. The services provided to passengers by transport systems are very important. There are two kinds of service that all transport systems must provide: (i) route and schedule information (maps, schedules, and information on connections) (ii) basic information (fare policy, stop locations, etc.). These types of information are delivered in a variety of ways: (a) traditional delivery methods include printed maps and schedule cards, and “Rider guides.” These are often distributed physically onboard buses and at key transit locations. (b) As with other types of information, the majority of distribution has moved to the Internet. Nearly all transport systems now provide service information on their websites where users can either view it electronically or print it at home or in their office. (c) Third-party distribution systems have also become increasingly common. Most major transport systems now possess route and schedule information through Google Transit, and smaller transit systems are also moving in this direction. Many transport systems are also now making their Google Transit data publicly available for use in the development of third party smartphone applications [1]. If we look in terms of delivering service information, our study is included in the last way.

Real-time vehicle tracking and management system has been the focus of many researchers, and several studies have been done in this area. Verma and Bhatia [2] stated in their study that GPS could be used in many applications and it is possible to follow routes and locations driven a vehicle by means of GPS. They develop a web based system presenting vehicles’ locations to the user. Gong et al. [3] improved approach to predict the public bus arrival time based on historical and real-time GPS data. After analyzing the components of bus arrival time systematically, the bus arrival time and dwell time at previous stops are chosen as the main input variables of the prediction model. They concluded that their model outperforms the historical data based model in terms of prediction accuracy. Guo et al. [4] integrated the Victoria Regional Transit System with appropriate communication technologies to develop a corresponding Smartphone application. In this smart bus system, users can access real-time passenger information such as schedules, trip planners, bus capacity estimates, bike rack availability and bus stop locations, via Smartphone, on computers and at bus stops. El-Medany et al. [5] supposed cost effective real time tracking system that provides accurate localizations of the tracked vehicle by using GPS and GPRS modules. By means of GPS receiver, proposed system has ability of tracking current position of the vehicle in any specific time. They tested efficiency of the system in different areas on Kingdom of Bahrain using Google maps. There are relevant works not only highway transit systems, but also other tracking systems for ships, flights trains, and etc. Flightradar24 [6] is a flight tracker showing live traffic information from all over the world. Flightradar24 combines data from several data sources including ADS-B (Automatic dependent surveillance-broadcast), MLAT (Multilateration) and FAA (Federal Aviation Administration). The ADS-B, MLAT and FAA data is aggregated together with schedule and flight status data from airlines and airports to create a unique flight tracking experience. The National Rail Enquiries train timetable site [7] shows all trains currently on approach to a particular station. Trains and stations are shown in different colours. Trains move in approximately real time, or rather quicker if user checks the speed-up box. Marine Traffic project [8] provides free real-time information to the public, about ship movements and ports, mainly across the coast-lines of many countries around the world. The initial data collection is based on the Automatic Identification System (AIS). Similar systems have been developed for purposes of monitoring and
management of vehicles, protection against the thefts and reducing the probability of loss [9]-[11].

In this paper, we proposed a location-aware smart bus stop system that any passenger with a smart phone or mobile device can scan QR codes placed at bus stops to view bus arrival times, and buses current locations on the maps. Users can also view bus routes on the map with their geographic and non-geographic attributes. We used C4.5 algorithm for estimation of bus arrival time. GPS and Google Maps are used for displaying current locations of buses on the maps, together with the related route information. If users are registered to the system, they can be informed of routes and bus arrival times via SMS and e-mails.

The rest of the article is structured as follows. In the second section, system architecture is proposed. Third section presents user services and user interfaces in the proposed system. Conclusion and some future enhancements are given at the last section.

II. SYSTEM ARCHITECTURE

The components and architecture of the proposed system consists of four stages as shown in Fig. 1. These are (i) scanning QR codes placed at bus stops and listing buses according to passengers’ desires, (ii) searching buses and/or bus stops and showing timetables, (iii) showing bus stops and routes on the Google map, and (iv) estimating bus interval time with machine learning algorithms and finally sending this information to users via SMS and/or e-mails. Each stage is detailed in the following subsections.

A. Usage of QR Codes in the System

With the increasing usage of smart phones and wireless network infrastructures, passengers are getting acquainted with obtaining information about timetables, bus arrival time and etc. by means of mobile phones. QR code was created as an information container forming of two-dimensional by Toyota subsidiary, Denso Wave in 1994. Data is encoded in QR optically readable format using QR code generators. So, QR code can be captured and decoded by smart phones. It is capable of handling up to several hundred times more information than the traditional bar codes unlike conventional bar codes are only capable of storing twenty digits. According to different versions of QR code, distinct information storage capacity may be used (see Fig. 2). The cost of information transfer via QR code is extremely low as compared with other technologies where specific hardware is always required [12]. Consequently, QR code is the most widely used information container that can be applied to different printed materials (e.g., posters, books or magazines) and places (e.g. bus stops, store windows, etc.).

There are many different applications [13]-[17] in different areas built based on QR codes. It is initially developed as an alternative technology to UPC bar codes. However, its applications include product tracking, item identification, time tracking, document management, general marketing, and much more. Raj et al. [18] proposed a cost effective and 3D (latitude, longitude and altitude) smart phone solution which helps in indoor navigation with the help of QR codes. Accuracy of proposed system is high and compared with Bluetooth, AGPS (Assisted Global Positioning System) and RFID (Radio-frequency identification). Costa-Montenegro et al. [19] present QR-Maps, a simple and efficient tool that can be used in smartphones to obtain accurate indoor user locations. This tool employs QR-Codes containing a short text which indicates locations shown within a custom Google map.

We use QR codes for the purpose of providing locations of bus stops. QR codes contain bus stop location information such as latitude and longitude (see Fig. 3). When QR codes, which are placed at bus stops, are scanned by passengers, they can view estimated bus arrival time, and bus’s current location. Users can also view routes of selected buses on the map.
B. Google Map API and GPS

To locate any bus on a map, its location in latitude and longitude (lat, long) coordinate values have to be obtained. These values are obtained from GPS receiver in each bus. Google serves its map images via a simple REST (Representational State Transfer) API enhanced with JavaScript and AJAX technologies. The system utilizes map API in open source JAVA framework with the standard libraries. A sample http query to request a map is given below.

```
http://maps.google.com/maps/api/staticmap?
```

After the question mark, query parameters and their constraints are appended. Parameters’ numbers and their values change depending on what users need from map services. The parameter lists are separated by the ampersand (&) symbol. An example query for a specific map is given below.

```
http://maps.google.com/maps/api/staticmap?center=Kocaeli +University,+İzmit,+Turkey&zoom=14&size=512x512&map
```

C. Estimation of Bus Arrival Time

As we know the classes of training set, we can use different algorithms to estimate the classes of new instance set by discovering the way the attributes-vector of the instances behaves. One of these algorithms is Decision Trees (DT’s). A tree is either a leaf node labeled with a class linked to two or more sub-trees. If we classify some instance, firstly we have to get its attribute vector and then apply this vector to the tree. To complete the classification process, the tests are performed into these attributes until reaching one or other leaf.

J. Ross Quinlan [20] developed a decision tree algorithm known as ID3 (Iterative Dichotomiser) in early 1980s. This algorithm is expanded on earlier work on concept learning systems described by E.B. Hunt, J. Marin, and P. T. Stone [21]. Quinlan presented C4.5 algorithm (statistical classifier), is based on the ID3 algorithm, tries to find simple DT’s [22].

C4.5 algorithm has a few premises cases:

- If all samples (cases) belong to same class, it creates a leaf and the leaf is returned to choose that class.
- In case of none of the features provide any information gain; C4.5 creates a decision node higher up the tree using the expected value of the class.
- In case of instance of previously-unseen class is encountered; C4.5 creates a decision node higher up the tree using the expected value.

Pseudocode of C4.5 algorithm is as following:

1. Check for premises cases
2. For each attribute, calculate the potential information provided by a test on the attribute. Also calculate gain information that would result from a test on the attribute.
3. Depending on the current selection criterion, find the best attribute to split on.
4. Recur on the sub-lists obtained by splitting on best attribute, and add other nodes as children of best attribute node

In the proposed system, C4.5 algorithm is used for estimation of bus arrival times. In order to use this algorithm, training data set reflecting past experiences is prepared as seen in Fig. 4. Columns of the data set are as follows: data id, late time (how long time a bus is late), reason for being late (breakdown, traffic accident, and passenger density), weather condition (rainy, normal, and snowy), scheduled time table, route number and bus stop id.

```
<table>
<thead>
<tr>
<th>Data Set</th>
<th>LateTime</th>
<th>Reason</th>
<th>WeatherCon</th>
<th>BusStopTime</th>
<th>RouteNum</th>
<th>BusStopId</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6</td>
<td>Traffic</td>
<td>Rainy</td>
<td>10:00</td>
<td>23</td>
<td>21</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>Normal</td>
<td>Snowy</td>
<td>10:00</td>
<td>23</td>
<td>21</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>Passenger</td>
<td>Normal</td>
<td>10:00</td>
<td>23</td>
<td>21</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>Traffic</td>
<td>Snowy</td>
<td>10:00</td>
<td>23</td>
<td>21</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>Normal</td>
<td>Normal</td>
<td>10:00</td>
<td>23</td>
<td>21</td>
</tr>
<tr>
<td>6</td>
<td>10</td>
<td>Passenger</td>
<td>Normal</td>
<td>10:00</td>
<td>23</td>
<td>21</td>
</tr>
<tr>
<td>7</td>
<td>10</td>
<td>Normal</td>
<td>Normal</td>
<td>10:00</td>
<td>23</td>
<td>21</td>
</tr>
<tr>
<td>8</td>
<td>10</td>
<td>Passenger</td>
<td>Normal</td>
<td>10:00</td>
<td>23</td>
<td>21</td>
</tr>
<tr>
<td>9</td>
<td>10</td>
<td>Normal</td>
<td>Normal</td>
<td>10:00</td>
<td>23</td>
<td>21</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>Accident</td>
<td>Rainy</td>
<td>10:00</td>
<td>23</td>
<td>21</td>
</tr>
</tbody>
</table>
```

Fig. 4. Data set example used for estimation of bus arrival time

To estimate expected arrival times at any specific bus stop, it is required to make some arrangements on data set. Depending on the schedule and physical conditions for a route, late times might change in a very large range. Let’s assume it has a range from 1 to 30 minutes, and chunk this range into 6 sub-ranges, such as 1-5 min for class 1, 6-10 min for class 2, etc., Remainder of this section explains the application of C4.5 algorithm on a given dataset (see Fig. 4) to estimate actual arrival time.

If S is any set of samples, let freq (Ci, S) stand for the number of samples in S that belong to class C_i (out of k possible classes), and |S| denotes the number of samples in the set S. Let’s explain C4.5 on our simple data set. According to data given Fig. 4, k equals to 6 and |S| equals to 10. Then the entropy of the set S (Info(S)):

```
Info(S)= - \sum_{i=1}^{k} \left( \frac{freq(C_i, S)}{|S|} \right) \cdot \log_2 \left( \frac{freq(C_i, S)}{|S|} \right)
```

```
Info(S)=-(1/10)*log_2(1/10)-(4/10)*log_2(4/10)-(3/10)*log_2(3/10)-(1/10)*log_2(1/10)-(1/10)*log_2(1/10)=
```

```
2.046
```

After set T has been partitioned in accordance with n outcomes of one attribute test X:

```
Info_X(T) = - \sum_{i=1}^{k} \left( \frac{|T_i|}{|T|} \right) \cdot Info(T_i)
```

Gain(X) = Info(T) - Info_X(T). It represents information gain.

Attribute selection is done with the highest gain value.

```
Gain(reason) = 2.046-0.875=1.171bits
```
Similarly, Gain(weatherCon) is calculated as 1.010 bits.

Attribute “reason” gives the highest gain of 1.171 bits in our example, and therefore this attribute will be selected for the first splitting. After splitting on best attribute, we add other nodes as children of best attribute node and we do same operations recursively. To make a decision tree model more readable, a path to each leaf can be transformed into an IF-THEN production rule.

III. USER SERVICES AND INTERFACES

General flow of the proposed application is as follows:

- Passengers with smart phones or mobile devices with the QR code readers scan QR codes placed at bus stops. After scanning QR codes, they are decoded via ZXing (Zebra Crossing) library [23].
- Route numbers of buses are listed for corresponding bus stops (see Fig. 5). Passengers choose one or more of them to view interested and estimated bus arrival times and current location of the selected bus (see Fig. 7).
- Passengers can view routes of buses on maps. As seen in Fig. 6, the green markers represent buses, the red markers represent bus stops and the blue line points the route of the bus. Users can utilize Google map’s mapping tools such as zooming in/out, panning, dragging and dropping to get better view for the selected route and buses.
- If users are registered to the system, they can be informed of the routes and the busses they are interested in, through e-mails and SMSs.

In this paper, we have presented a smart bus tracking system. It is based on GPS, GSM, QR coding and Google’s map technologies. The proposed system, basically tracks the busses, estimates their arrival times at specific bus stops and informs the users through e-mails and SMSs. The system prevents passengers unnecessarily to wait at bus stops and enables them to use their time more efficiently.

In the future, we plan to enhance the system with some other estimation tools and statistical analysis. This might be used not only by public users but also by decision makers in the local municipalities. Moreover, since the system is developed with open standards and open sources, it is easily extended with future technologies according to users’ needs.

REFERENCES


