USAGE OF FUZZY LOGIC BASED DATA MINING METHODS IN ANALYSIS OF PUBLIC TRANSPORTATION DATA

by

Ahmet Can DİKER

August, 2015

İZMİR

USAGE OF FUZZY LOGIC BASED DATA MINING METHODS IN ANALYSIS OF PUBLIC TRANSPORTATION DATA

ABSTRACT

Intelligent Transportation Systems are used to construct and manage public transportation system based on knowledge efficiently and also to increase interest of people for public transport. In this study, by examining boarding data obtained from smart cards used in public transportation system in Izmir, estimation of alighting stop for bus mode has been firstly dwelled on. Then, solution has been sought to multi-criteria fuzzy route planning problem and for the first time in the literature, fuzzy neighborhood relations between stop-stop, line-stop and line-line, and fuzzy preference degree of stop have been discussed and route planner for Izmir has been developed by using these concepts. It is thought that this study will provide contribution to researchers and recently popular issues; smart cities and sustainable mobility.

Keywords: Intelligent transportation systems, smart cart data analysis, route planning problem, fuzzy preference degree of stop

TOPLU ULAŞIM VERİLERİNİN ANALİZİNDE BULANIK MANTIĞA DAYALI VERİ MADENCİLİĞİ YÖNTEMLERİNİN KULLANIMI

ÖZ

Toplu ulaşım sistemlerinin bilgiye dayalı olarak etkin bir şekilde yapılandırılması, yönetilmesi ve ayrıca toplu ulaşıma olan ilginin artırılması için Akıllı Ulaşım Sistemleri kullanılmaktadır. Çalışmada ilk olarak İzmir toplu ulaşım sisteminde kullanılan akıllı kartlardan elde edilen biniş verileri incelenerek otobüs binişlerine ait iniş durağı tahmininin yapılması üzerinde durulmuştur. Ardından, çok-kriterli bulanık rota planlama problemine çözüm aranmış ve literatürde ilk kez durak-durak, hat-durak ve hat-hat bulanık komşuluk ilişkileri ile bulanık durak tercih derecesi ele alınmış ve bu kavramlar kullanılarak İzmir için rota planlayıcı geliştirilmiştir. Çalışmanın, araştırmacılara ve son yıllarda popüler olan akıllı şehirler ve sürdürülebilir hareketlilik konularına katkı yapacağı düşünülmektedir.

Anahtar kelimeler: Akıllı ulaşım sistemleri, akıllı kart verisi analizi, rota planlama problemi, bulanık durak tercih derecesi

1. Introduction

In large cities, motorization and urbanization increase in parallel with permanent growth of population, and as a result of this, problem of urban transportation has become one of the world's biggest problems with the impacts on economic, environmental and social issues. Intelligent Transportation Systems (ITS) play a key role to solve this problem. Advanced Public Transportation Systems (APTS) is composed of various systems and all of them use high technologies to increase efficiency, and safety of public transportation in addition; providing access to data for users. But success of APTS applications depends upon how people understand these systems, and how they use them. Advanced Traveler Information Systems (ATIS) provides information related to all traffic and weather conditions and it also helps commuters or drivers by letting them to use this information before a trip (Sussman, 2005).

In this study, entry-only smart card validation data is analyzed to estimate alighting bus stops by utilizing several methods. These methods, include three approaches; trip-chaining, random assignment, and designation to the last stop. All of them depend on assumptions, and when combined, form line-based alighting stop estimation method. To make a data mining application, this method's function is to extract alighting information as knowledge from a large dataset by using public transportation data.

Another subject in this paper is route planning problem. The problem focuses on how a journey can be planned optimally, and how a route between specified origin and destination can be designated under some criteria such as distance, time etc. The new concepts, preference degree of stops and fuzzy neighborhood relations between stop-stop, line-line and line-stop are presented. All these relations and criteria turn the problem into a multi-criteria fuzzy variation of it.

An algorithm, which is based on assumptions and explained in Section 2, is to developed for estimating alighting stop. Then, an algorithm to solve route planning problem is proposed and running mechanism of the algorithm is also explained as well, in Section 3. In Section 4, two different algorithms handled in this study are tested in terms of their applicability for multi-modal transportation system in Izmir. Finally, in Section 5, conclusion to this study and future remarks are given.

2. Smart Card Data Analysis

Reducing the number of private cars participating in urban traffic plays a key role to solve traffic problem that increases day-by-day. In this case, designing and management of sustainable public transportation system with expanded network and improved service quality are required. At this point, passenger demand or flow is one of the critical information which is valuable for assessing the cost that might affect the performance of transport planning. Assessments of the analyzing data obtained through automated system, are significant in order to determine passenger demand and hence opportunity of providing sustainable mobility.

Collecting the necessary data has great importance for the success of public transportation analysis (Ceder, 2007). In this study, smart cards are used as one of the data sources in automated systems. Automated systems are identified as follows: Automated Fare Collection Systems (AFC), location-based Automatic Vehicle Location (AVL), and Automated Passenger Counting (APC) which counts the number of passengers who board and alight by using sensors (Wilson et al., 2009).

Smart cards as an actor of AFC systems are used as data source apart from its fare collection aim. With advancing technology, data obtained through automated system have been used for planning of the public transportation since 2000 (Pelletier et al., 2011). Bagchi et al. (2003) examined nature of the smart card data for the first time by using the data taken from two bus companies in England. Morency et al. (2006) studied on boarding data of ten months period to determine the passenger behaviors in Gatineau, Quebec.

Origin Destination (OD) matrix is to determine passenger flow which is a critical information for planning of public transportation systems. Trip-chaining is well-known method in literature to estimate alighting information that it is based on examining sequential movements of passengers. Barry et al. (2002) handled the case that includes only boarding transactions on urban rail system in New York. Trépanier et al. (2007) offered a model by using a database programming approach to estimate the destination location for each individual boarding a bus in Gatineau, Quebec. Munizaga & Palma (2012) presented a methodology to estimate OD matrix from both smart card and GPS data for large transit network in Santiago, Chile.

2.1 Estimation of Alighting Stop Information via Smart Card Data

Trip-chaining approach was examined consecutive trips of passengers in different studies (Barry et al., 2002; Cui, 2006; Zhao et al., 2007). This approach is based on two main assumptions that stated as follows: the first of them and primary assumption of trip-chaining approach is that alighting point estimation of previous trip is determined as start point of the next trip. Second assumption is formed to estimate alighting place with idea that at the end of the day, riders return to place where they started the day.

Since alighting stop is estimated as next boarding stop in trip-chaining approach, the condition "The alighting is done at a stop that is on the boarding line" can't be met except the situations that the transfer is made at the same stop. Therefore line-based method may reflect the fact for public transit. Consecutive trips of passenger are considered for line-based method as trip-chaining approach. This method guarantees that the alighting stop is on the boarding line. Trépanier et al. (2007) expressed a method that the alighting stop of the passenger is the closest stop to the point of the next boarding and this point is on the boarding line. In this study, assumptions are described as follows:

Assumption 1. Passengers alight at the nearest bus stop on the boarding line to their next boarding stop.

Assumption 2. Passengers complete their trips at the stop of their first boarding at the end of the day.

In these assumptions, nearest stop means that distance between estimated alighting stop and next boarding stop is a walkable distance and this should be limited with plausible distance for walking. Distances between stops and specified points should be computed by spherical geometry because all points are defined as geographical coordinates in this study. For this purpose, Haversine formula that was proposed by Sinnott (1984) is used to compute distance more accurately between any two geographical points X and Y, as follows:

$$d(X,Y) = 2R\, sin^{-1}\left(\left[sin^2\left(\frac{\phi_1-\phi_2}{2}\right) + cos\phi_1 cos\phi_2 sin^2\left(\frac{\lambda_1-\lambda_2}{2}\right)\right]^{0.5}\right). \tag{1}$$

where R is radius of the Earth ($R = 6367450$ meters); ϕ is latitude, λ longitude; X is given $X = (\phi_1, \lambda_1)$, and Y is given $Y = (\phi_2, \lambda_2)$. Walkable distance criterion which is calculated by using Equation (1) is determined as 1,000 meters in this study. If this criterion isn't met, alighting stop can't be estimated by Assumption 1.

Let us start by giving definitions of some terms used in this study. Let s be any stop in set of the stops S ($s \in S$) and l be any line in set of the lines l ($l \in L$). Any line can be defined as an ordered set of the certain stops. Stop-Line relation is denoted by $LS(l, s)$. The sequence number of s on line l is given as $LS(l, s) \geq 1$. LS relation is given as follows:

$$LS(l,s) = \begin{cases} \geq 1, & \textit{sequence number of stop } s \text{ on line } l, \\ 0, & \textit{stop } s \textit{ is not on line } l. \end{cases} \tag{2}$$

Let $S1(l)$ denote the set of stops on the line l passing through, i.e.

$$S1(l) = \{s \in S\colon\ LS(l,s) \geq 1\}. \tag{3}$$

where the number of stops of the any line l is represented as $|S1(l)|$.

Let P be set of the passengers and $b_{pd}^{(i)} \epsilon S$ denotes boarding stop of i^{th} trip ($i = 1,2, \dots, I_{pd}$) of the day d ($d = 1, \dots, D$) for any passenger p ($p \epsilon P$), and boarding time for same transaction marks with $t_{pd}^{(i)}$. Let boarding stop of same transaction is also represented with s_l^k that denotes k^{th} stop on line l. Hence $b_{pd}^{(i)} = s_l^k$ ($k = 1,2, \dots, |s1(l)| - 1$) can be written. Let $b_{pd}^{(i+1)}$ be boarding stop of $(i+1)^{\text{th}}$ trip for the same passenger p of the day d, and alighting stop $a_{pd}^{(i)}$ is estimated to Assumption 1 as below:

$$a_{pd}^{(i)} = \operatorname{argmin}_s d\left(b_{pd}^{(i+1)}, s\right) \tag{4}$$

s.t.:

$$d\left(b_{pd}^{(i+1)}, s\right) \leq 1000, \tag{5}$$

$$LS\left(l, b_{pd}^{(i)}\right) < LS(l,s), \tag{6}$$

$$LS\left(l, b_{pd}^{(i)}\right) \geq 1, \tag{7}$$

$$s \in S. \tag{8}$$

If walking distance criterion isn't met, alighting stop can't be estimated. An example of alighting estimation method is illustrated in Figure 1.

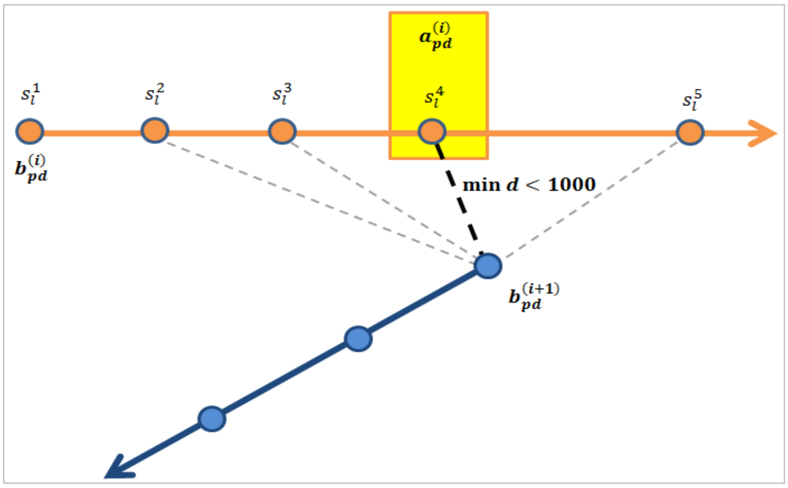

Figure 1. An example of alighting estimation method

Results of the analysis can be interpreted in detail by dividing a day into parts and thus extensive information can be used for planning of public transportation. Therefore a day is divided into q parts and parts of the day are found by any x by using following function:

$$L(x) = \begin{cases} q, & \eta_{q-1} \leq x < \eta_q \\ -1, & otherwise \end{cases}, \quad q \in \mathbb{Z}^+, \tag{9}$$

where η_{q-1} and η_q are lower and upper bounds of q^{th} part of the day respectively.

After Assumptions 1 and 2 are employed to consecutive trips of any passenger for day d, the number of passenger going from each boarding stop s_l^k to each estimated alighting stop s_l^m for all lines and their each service c are obtained and saved in $OD_{dq}\left(l, c, s_l^k, s_l^m\right)$. Alighting stop estimate can't be done by using trip-chaining approach for two cases. First one of these cases is estimation of alighting information of passengers who board daily once per day. In the literature, this case was dealt in several studies (Trépanier et al., 2007; Munizaga et al., 2014; Wang et al., 2011). Novelty of this study is that a new method called as random assignment estimation method (RAEM) is described. Second case is special that is observed in the public transport system of Izmir. This situation arises from impersonal use of cards except students, teachers, retired and disabled persons, etc. In other words, a passenger can use her/his smart card for another passenger on the same trip.

Another case except two aforementioned situations is related to alighting stop which cannot be estimated by applying Assumption 1. RAEM is also carried out for this case as it proposed above. Therefore four new assumptions related with the method are given below:

Assumption 3. Alighting stop of the passengers that board daily for once is determined by using line, boarding stop and alighting stop patterns (obtained from Assumption 1) based on RAEM.

Assumption 4. For transactions of passenger who make multiple boarding with the same card on the same service, two cases are satisfied as follows:

- First of these transactions remains in trip-chain and its alighting stop is estimated depending on trip-chaining approach (Assumption 1 and Assumption 2).
- Others are taken into account as separate boardings, and alighting stops of them are estimated by using line, boarding stop, alighting stop patterns obtained with trip-chaining approach base on RAEM.

Assumption 5. Alighting stops which can’t be estimated due to the case that distance between stops is out of limits according to walkable distance criterion mentioned above in Assumption 1 are estimated by using RAEM.

Assumption 6. Alighting stops which can’t be estimated by using RAEM are estimated as last stop on the boarding line.

The method which is formed by all assumptions that are mentioned above is called as Line-Based Estimated Method (LBEM) in the rest part of this study.

2.2 Random Assignment Estimation Method (RAEM)

Alighting stops that can't be estimated with trip-chaining approach and case handled in previous section should be estimated in attempt to equilibrate the number of both boardings and alightings in the bus. Trip chaining approach provides 3-tuples which consist of boarding stop, line and alighting stop and these may produce general information about passenger flow. RAEM is developed with the idea that passengers might be followed each other with similar flow patterns.

The total number of passenger travelling from s_l^k to s_l^m with line *l* into *q* part of the day *d* is called as movement count $MC_{dq}\left(l, s_l^k, s_l^m\right)$ and it is computed as follows:

$$MC_{dq}\left(l, s_l^k, s_l^m\right) = \sum_{c=1}^{c_d} OD_{dq}\left(l, c, s_l^k, s_l^m\right), \tag{10}$$

where c denotes any service of line l for $c = 1,2, \dots, c_d$; $s_{l_c}^k$ and $s_{l_c}^m$ are respectively k^{th} and m^{th} stops on line l of its service c for $k < m$. The average of total number of passengers going from s_l^k to s_l^m on line l into *q* part of the all day is written as below:

$$\overline{MC}_q\left(l, s_l^k, s_l^m\right) = \frac{1}{D} \sum_{d=1}^{D} MC_{dq}\left(l, s_l^k, s_l^m\right). \tag{11}$$

For any q^{th} part of the day, probability of the average value given in Equation (11) is computed as follow:

$$P_q\left(l, s_l^k, s_l^m\right) = \frac{\overline{MC}_q\left(l, s_l^k, s_l^m\right)}{\sum_{j=k+1}^{|S1(l)|} \overline{MC}_q\left(l, s_l^k, s_l^j\right)}, \tag{12}$$

The random number $u \epsilon (0,1]$ is generated to estimate alighting stops by random number generator. Alighting stop s^* that meets the condition, $s^* = s_l^{r+1}$, is estimated in following equation:

$$\sum_{m=k}^{r} P_q\left(l, s_l^k, s_l^m\right) < u \leq \sum_{m=k+1}^{r+1} P_q\left(l, s_l^k, s_l^m\right) \;\; , \;\; r = k, \dots, |S1(l)| - 1, \tag{13}$$

where $P_q\left(s_l^k, s_l^k\right) = 0$ is accepted.

The case is also considered that some alighting stop can't be still estimated despite the use of trip-chaining and RAEM. This situation occurs while there isn't any alighting stop estimate for certain boarding stop on any line by using previously obtained patterns. For this reason, Assumption 6 as it mentioned in previous section is considered for LBEM and alighting stop is estimated as the last stop of the boarding line while this stop can't be inferred by any method, and this method is also called Assignment to Last Stop Method (ALSM). All transactions about boarding data are stored in $LBEM^d$ table and LBEM algorithm is shown as follow:

LBEM Algorithm

Inital: *Estimated* and *NotEstimated* tables are created as empty.

Stage 1:

Step 1.1: $LBEM^d$ table is handled for $d = 1$.

Step 1.2: If $d \leq D$, it continues from Step 1.2; otherwise it continues from Step 1.9.

Step 1.3: For all transactions in $LBEM^d$ table of each card, ascending order numbers is given according to boarding time. It is continued with Step 1.4.

Step 1.4: Multiple boardings with the same card on the same service are determined. If they are made from bus mode, they are evaluated for each card separately as follow: first of these transactions remains at $LBEM^d$ table, others are added to *NotEstimated* table. All multiple boardings are extracted from $LBEM^d$ table and it is continued with Step 1.5.

Step 1.5: Transactions that include single boarding in a day are detected in $LBEM^d$ table. If they are made from bus mode, they are inserted in *NotEstimated* table. These transactions obtained from all modes are removed from $LBEM^d$ table and it is continued with Step 1.6.

Step 1.6: According to Assumption 2, first boarding for all cards in $LBEM^d$ table are added as last boarding to $LBEM^d$ table and it is continued with Step 1.7.

Step 1.7: All **(**boarding stop,next boarding stop) pairs are obtained from $LBEM^d$ table. Boarding places of these pairs is evaluated as follow:

- for bus stop, alighting stops that are estimated by applying Assumption 1 are added to *Estimated* table whereas alighting stops which cannot be estimated are insert in *NotEstimated* table.
- for stations or piers, transactions are extracted from analysis.

Then it is continued with Step 1.8.

Step 1.8: $d = d + 1$ is made and it is continued with Step 1.2.

Stage 2:

Step 2.1: The values are obtained from *Estimated* table by utilizing Equation (10)-(13) and RAEM is applied to *NotEstimated* table by using these values to estimate alighting stops. If there are estimated alighting stops, these data are moved to *Estimated* table and it is continued with Step 2.3; otherwise it is continued with Step 2.2.

Step 2.2: ALSM is performed in *NotEstimated* table and all alighting estimates are moved to *Estimated* table.

Step 2.3: If any data remains in *NotEstimated* table, it is continued with Step 2.2; otherwise all boardings in *Estimated* table is saved in $OD_{dq}\left(l, c, s_l^k, s_l^m\right)$ according to day and parts of the day.

2.3 Finding Fuzzy Comfort Degree of Passenger in Bus

The number of passengers in bus, also called as passenger density, is critical information for planning and operating of transportation system. It is possible to find the passenger density by using results obtained from smart card data analysis mentioned in previous section. In this study, the density is described as the total number of passengers in any bus while the bus leaves from any stop.

Let s_l^k and s_l^m imply respectively k^{th} and m^{th} stops of line l for its c^{th} service and $k < m$ is satisfied; $OD_{dq}\left(l, c, s_l^k, s_l^m\right)$ matrix which was mentioned above, is expressed as passengers' count going from s_l^k to s_l^m for q^{th} part of the day d and thus the total number of passengers who board and alight is obtained by using this matrix. For any part q of the any day d, total number of passengers who board from s_l^k on line l is marked as $B_{dq}\left(s_l^k\right)$:

$$B_{dq}\left(s_l^k\right) = \sum_{j=k+1}^{|S(l)|} OD_{dq}\left(l, c, s_l^k, s_l^j\right) \tag{14}$$

where the number of passengers who board from last stop on the line is zero i.e. $B_{dq}\left(s_l^{|S(l)|}\right) = 0$. For q^{th} part of day, the number of average daily boardings from stop s_l^k on line l $\bar{B}_q\left(s_l^k\right)$ is computed as follow:

$$\bar{B}_q\left(s_l^k\right) = \frac{1}{D} \sum_{d \epsilon D} \sum_{l \epsilon L} B_{dq}\left(s_l^k\right). \tag{15}$$

where $B_{dq}\left(s_l^k\right)$ denotes total number of passengers who board from s_l^k on line l ($l \epsilon L$) for qth part of day d ($d = 1,2,\dots,D$). The number of average daily boardings is used to calculate degree of stop activity handled in Section 3. The total number of passengers who alight at s_l^k on line l is denoted as $A_{dq}\left(s_l^k\right)$:

$$A_{dq}\left(s_l^k\right) = \sum_{j=1}^{k-1} OD_{dq}\left(l, c, s_{l,}^j, s_l^k\right) \tag{16}$$

where the number of passengers who alight at first stop on the line is zero i.e. $A_{dq}(s_l^1) = 0$. By using Equation (14) and Equation (16), while bus leaves s_l^k passenger, density in it is computed as below:

$$C_{dq}\left(s_l^k\right) = \sum_{i=1}^{k} \left(B_{dq}\left(s_l^k\right) - A_{dq}\left(s_l^k\right)\right). \tag{17}$$

Comfort degree of passenger indicates the status of crowd in the bus and it can be determined by utilizing Equation (17). In this study, it is computed depending on bus capacity and thus bus capacities should be known.

The total number of seating passenger is limited with τ_1 value. After exceeding of τ_1, it is considered that passengers are travelled comfortably up to specified upper bound of τ_2. Then, it is assumed that bus is crowded after exceeding τ_2 value. While almost no space on the bus, in other words bus is reached to its maximum capacity, τ_3 is taken as limit for this situation. As including $C_{dq}\left(s_l^k\right) = x$, fuzzy membership function depending on passenger comfort in bus for qth part of the day d is written as below:

$$\mu_{CD}^{dq}(x) = \begin{cases} 1, & x \leq \tau_1, \\ \frac{(\tau_2-\tau_1)-0.5(x-\tau_1)}{\tau_2-\tau_1}, & \tau_1 < x \leq \tau_{2,} \\ \frac{0.5(\tau_3-x)}{\tau_3-\tau_2}, & \tau_2 < x \leq \tau_3, \\ 0, & \tau_3 < x. \end{cases} \tag{18}$$

3. Route Planning Model

The route planning problem focuses on the issue that; how journey might be planned optimally for the routes between specified origin and destination under some criteria such as distance, time, traffic, etc. and various types of graphs is usually used in different kinds of systems. In most of the cases, the solution to this problem is given by shortest path algorithms on these graphs and bidirectional searches proposed by Dantzig (1963) for a road network that are generally accepted as the basis of the solution techniques (Von Ferber et al., 2005; Von Ferber et al., 2009; Derrible & Kennedy, 2011). Many of these algorithms, which are based on Dijkstra, have been modified to achieve satisfactory results when large road network is concerned because of their slow response time (Thorup, 2004; Schultes & Sanders, 2007; Geisberger et al., 2008). Bast et al. (2010) presented that although these algorithms are running successfully in the road network, they can't be applied to public transportation network directly. Solution for the route planning problem on the public transportation networks is known to be harder than road networks since transition between different modes could also be made possible and bus, metro or ferry have certain routes with timetables.

There are various criteria in real life problems where people prefer to decide upon routes according to their conditions and it is called as a multi-objective shortest path problem that usually arises in transportation problems. It has been shown that; obtaining a solution that gives the best results for all criteria may not be possible since route decision of the people depending on their current circumstances could vary according to different scenarios.

Various methods were proposed in literature that they are called speed-up techniques based on narrowing the search space although shortest path algorithms on public transportation network have already been modified and utilized on graphs (Thorup, 2004; Geisberger et al., 2008; Gutman, 2004; Goldberg et al., 2006). Since in most of these algorithms, preprocessing increases memory cost, heuristic approaches were proposed in some other studies as an alternative methodology (Liu et al., 2001; Liu, 2002; Chang et al., 2007; Bast et al., 2010). There are a few studies considering the fuzzy approach even that there are many studies on public transportation networks in the literature (Golnarkar et al., 2010; Verga et al., 2013). Golnarkar et al. (2010) and Verga et al. (2013) discussed finding the best routes between origin and destination on multi-modal transportation network by taking arcs of graph as fuzzy numbers. In those studies, time constraint has been pre-emptively discussed and representing arcs of graph with fuzzy numbers was one of the common approaches but in this study the approach consists of neighborhood relation between stops and lines that has not been addressed in the literature yet.

3.1 Fuzzy Neighborhood and Preference Degrees of Stops

Search process is usually performed between the stops that are nearest to both origin and destination points in route planning problem. This process limits the possible search domain among these stops, where algorithms may not find any solution or might be obtaining a solution with many unnecessary transfers. The main problem in this approach is that the search process is only carried out between two stops based on the idea of walking to the nearest stop. This cannot be applied in real life situations where people could easily change their mind depending on different conditions. In this paper, three decision criteria will be considered for preference of stops: walking distance to the stop, activity of the stop and count of lines passing through the stop (being a hub).

Let us start by giving definitions of some terms used in this problem. Let s be any stop in set of the stops S ($s \in S$) and l be any line in set of the lines L ($l \in L$); PL_s denotes set of lines that pass through the stop s and it is defined as follows:

$$PL_s = \{l : s\epsilon S1(l)\}, \tag{19}$$

where $S1(l)$ is given in Equation (3).

Definition 1 (a neighborhood relation between stops): Let $s_i,\ s_j \in S$ be any two stops. Then a fuzzy neighborhood relation between stops can be defined as below:

$$\mu^d(s_i, s_j) = max\left\{0,\ 1 - \frac{d(s_i, s_j)}{d_{max}}\right\} \tag{20}$$

where d_{max} is specified distance's maximum value, $d(s_i, s_j)$ is the distance between the stops s_i and s_j.

Definition 2 (a fuzzy neighbor stops set of a stop): The set of fuzzy neighbor stops of any given stop *s* is defined as

$$N_s(s') = \{(\mu^d(s,s')/s') \ |s' \in S\} . \tag{21}$$

For specified γ level, N_s^γ represents γ-level set of the fuzzy set N_s and it is described as:

$$N_s^\gamma = \{s' \in S : N_s(s') \geq \gamma\}. \tag{22}$$

N_s^γ set is called as a γ-neighbor set of the stop *s*. Degree of the fuzzy level is linked with neighborhood radius as follows: for each γ-level there exists a radius ε such that following equation is satisfied:

$$\{s' \in S : N_s(s') \geq \gamma\} = \{s' \in S : d(s,s') \leq \varepsilon\}. \tag{23}$$

Liu et al. (2001) specified that possibility of passing between lines has been determined by the presence of the exactly common stops of lines. In this study, a concept of fuzzy neighbor stops is taken into consideration to ensure transition between lines that have not common stops for these lines.

Definition 3 (a fuzzy neighbor stops set of a line): The fuzzy neighbor stops set of the line $l \in L$ is defined as follows:

$$NRS_l(s') = \max_{s \epsilon S1(l)} N_s(s'). \tag{24}$$

Definition 4 (a fuzzy neighbor lines set of a line): A fuzzy neighbor lines set of a line *l* is defined as follows:

$$N_l \ = NLS_l \circ PL_s, \tag{25}$$

in other form,

$$N_l \ (l') = \vee_{s \in S}[NLS_l(s) \wedge PL_s(l')], \forall l' \in L. \tag{26}$$

In fact, the lines which are actually not linked each other, become connected by using concept of the γ-neighborhood of a line, so a traveler can reach to destination with some transfers in related situation (Figure 2).

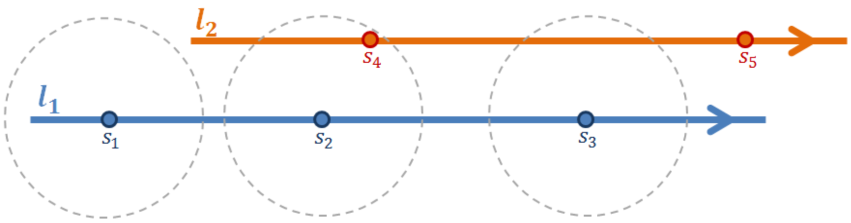

Figure 2. Example related to the fuzzy neighborhoods of line

According to the Figure 2, it is seen that $S1(l_1) = \{s_1, s_2, s_3\}$ and $S1(l_2) = \{s_4, s_5\}$. It is obvious that there is no intersection between l_1 and l_2 lines. Neighbor stops of line l_1 are s_1, s_2, s_3 and s_4 ($NLS_{l_1}^{\gamma} = \{s_1, s_2, s_3, s_4\}$). Although there is no exactly common stops between the lines l_1 and l_2, l_2 is a γ-neighbor of the line l_1 ($N_{l_1}^{\gamma} = \{l_1, l_2\}$), and it can be used as a continuation line for the line l_1.

3.2 Fuzzy Preference Degree of a Stop

Three criteria will be used to determine fuzzy preferences of stops. First of them is the walking distance which is a distance between the stop and a chosen geographical point. Suppose that the maximum feasible walking distance is denoted with the d_{max}. While the distance is considered according to any selected point X, a fuzzy membership function indicating the fuzzy neighborhood degree of any stop s, can be defined as

$$\mu_X^d(s) = max\left\{0,\ 1 - \frac{d(X,s)}{d_{max}}\right\}. \tag{27}$$

Second criterion to determine preference degree of the stop is defined as the stop activity. It is proposed to encourage travelers to use stops that are more active. $B(s)$ value representing the number of average daily boardings on stop s will be reckoned. This value is calculated by using Equation (15). The activity degree of a stop s can be described as below:

$$\mu^A(s) = \frac{B(s)}{\max_{s \epsilon S} B(s)}. \tag{28}$$

Last criterion of our interest to determine preference degree of the stop is about being hub stop. Different lines passed through a stop arise can provide to reach easier to destination. However, there is a certain degree in fuzzy logic that depends on number of lines passing through the stop if stop is represented as a hub. Suppose that PL_s is given as the set of lines passing through stop *s* and let us denote the total number of these lines as $|PL_s|$. Then a hub degree of any stop *s* can be defined as below:

$$\mu^H(s) = \frac{|PL_s|}{max_{s \in S}|PL_s|}. \tag{29}$$

Minimum Aggregation method that is one of the common methods in fuzzy logic theory can also be used in here to combine three criteria as it mentioned above.

$$\mu_X(s) = min\{\mu_X^d(s), \mu^A(s), \mu^H(s)\}. \tag{30}$$

So, $\mu_X(s)$ denotes a preference degree of the stop *s* according to the geographical point *X*. Hereby, for any geographical point *X*, if could be organized a fuzzy set of preferred stops with their preference degrees as follows:

$$S2(X) = \{(\mu_X(s)/s) |\ s \in S\} \tag{31}$$

Note that also any stop with its geographical location can be used instead of the referenced point *X*. The γ-level set $S2(X)^\gamma$ of the set $S2(X)$ at a fixed level $\gamma \in (0,1]$ can be described as

$$S2(X)^\gamma = \{s \in S :\ \mu_X(s) \geq \gamma\}. \tag{32}$$

It is obvious that the total number of stops within the $S2(X)^\gamma$ will decrease if γ-level value increases. Preference degree of the stop can present a more realistic solution than solving the problem only by choosing the nearest stop, since it does not only take into account distances, but also it includes degrees of both stop activity and being hub stop aggregated with fuzzy logic. This γ-level approach demonstrates the expectation degree of the commuters when making decision indeed and it therefore gives possibility to change their demands on the purpose of planning a route flexibly.

3.3 Multi-Criteria Fuzzy Route Planning Problem

As it is known in the fuzzy set theory, general solution of the fuzzy problem can be given as a union of the γ-level solution sets. So problem is described via its $\gamma \in (0, 1]$ fuzzy level set decomposition.

Let us first introduce connection as 3-tuples (s_1, l, s_2) that is expressed as a travel from any stop $s_1 \epsilon S$ to $s_2 \epsilon S$ traveling via a line $l \epsilon L$.

Definition 5 (a possible connection): A 3-tuple (s_1, l, s_2) is a possible connection if it satisfies following conditions:

i. $s_1 \in S1(l)$;
ii. $s_2 \in S1(l)$;
iii. $LS(l, s_1) < LS(l, s_2)$.

It is obvious that a 3-tuple (s_1, l, s_2) also is a possible connection if it satisfies the $LS(l, s_1) \geq 1$ condition.

Definition 6 (a fuzzy route): A sequence of *m* possible connections going from stop s_X to stop s_Y is a fuzzy route denoted as

$$\pi_{s_X s_Y} \triangleq \langle (s_{i_1}, l_{j_1}, s_{k_1}), (s_{i_2}, l_{j_2}, s_{k_2}), \ldots, (s_{i_m}, l_{j_m}, s_{k_m}) \rangle \tag{33}$$

if it satisfies following conditions:

i. $s_{i_1} = s_X$;
ii. $s_{k_m} = s_Y$;
iii. $(s_{i_t}, l_{j_t}, s_{k_t})$, $t = 1, \ldots, m$, are possible connections.

Moreover, the connectivity degree of the fuzzy route $\pi_{s_X s_Y}$ is

$$min_{t=2,\ldots,m}\, \mu_{s_{i_t}}(s_{k_{t-1}}) \ . \tag{34}$$

It is known that a level set decomposition approach to solve fuzzy optimization problems is a widely used approach in fuzzy sets theory. So definition of γ-connectivity for a fuzzy route is given, and then a γ-decomposition statement of the fuzzy route planning problem is formulated.

Definition 7 (a fuzzy $\boldsymbol{\gamma}$-connective route): A fuzzy route $\pi_{s_X s_Y}$ with its connectivity degree greater or equal to γ is a fuzzy γ-connective route denoted as $\pi^{\gamma}_{s_X s_Y}$.

It is obvious that a γ-connective route is a sequence of m possible connections going from stop s_X to stop s_Y which satisfies the following conditions:

i. $s_{i_1} = s_X$;
ii. $s_{k_m} = s_Y$;
iii. $\mu_{s_{i_t}}(s_{k_{t-1}}) \geq \gamma$;
iv. $(s_{i_t}, l_{j_t}, s_{k_t})$ are possible connections and $t = 2, \dots, m$.

A number of transfers on a route is another criterion handled in this study. Moreover, alternative routes that have the same number of transfers are sorted in ascending order with respect to total time spent on the route. Let $|(s_i, l, s_k)|$ denotes the total number of stops on line $l \epsilon L$ between the stops s_i and s_k.

Let travel time of possible connection from s_i to s_k with line l is denoted as $T[(s_i, l, s_k)]$ and it is computed as follow:

$$T[(s_i, l, s_k)] = \frac{D[(s_i, l, s_k)]}{\bar{v}}, \tag{35}$$

where $D[(s_i, l, s_k)]$ denotes total travel distance of possible connection and $\bar{v}$ is the average speed of vehicle. Let total walking time between all transfers for all trips in route $\pi_{s_X s_Y}$ is marked as T_W. In this case, total time spent on the route is computed as follow:

$$T(\pi_{s_X s_Y}) = \sum_{t=1}^{m} T[(s_{i_t}, l_{j_t}, s_{k_t})] + T_W , \tag{36}$$

where $s_{i_1} = s_X$ and $s_{k_m} = s_Y$ are satisfied, and m is the number of possible connections of the route.

Thereby, a fuzzy multi-criteria route planning problem in its γ-level decomposition form for any specified geographical points *X* and *Y*, can be stated as follows:

$$\max : \gamma, \tag{37}$$

$$\min \ : T(\pi_{s_X s_Y}) \tag{38}$$

s.t.

$$LS(l_{j_t}, s_{i_t}) \geq 1 \text{ , } t = 1, \ldots, m, \tag{39}$$

$$LS(l_{j_t}, s_{k_t}) > LS(l_{j_t}, s_{i_t}), \ t = 1, \ldots, m. \tag{40}$$

$$\mu_X(s_X) \geq \gamma, \tag{41}$$

$$\mu_Y(s_Y) \geq \gamma, \tag{42}$$

$$s_{i_1} = s_X \ , \tag{43}$$

$$s_{k_m} = s_Y \quad , \tag{44}$$

$$\mu_{s_{i_t}}(s_{k_{t-1}}) \geq \gamma, \ \ t = 2, \ldots, m; \tag{45}$$

$$\gamma \in (0,1], \tag{46}$$

where (37) states making choice from among possible routes as higher preference degree as possible, Equation (38) indicates selection of possible routes completed in less time. Constraints in Equations (39)-(40) indicate possibility of each connection on the route. Equations (41) and (42) are constraints to ensure that the first stop on the route should be fuzzy preferable according to the geographical point *X*, similarly last stop on the route should be preferable according to the geographical point *Y*. Constraints in Equations (43)-(45) state fuzzy γ-connectivity of a route. Solution of the problem given in Equations (37)-(46), provides choosing possible shorter time routes among preferable fuzzy routes connecting the geographical points *X* and *Y*. Moreover, the preference degree of the solution of the problem will be

$$\mu_X(s_X) \wedge \mu_Y(s_Y) \wedge \min_{t=2,\ldots,m} \mu_{s_{i_t}}(s_{k_{t-1}}). \tag{47}$$

3.4 Algorithm

In order to find γ-optimal solution, departure and arrival stops on the route must be found first, then direct connection, single transfer and double transfer must be applied respectively unless there are no results for the current case. Fuzzy preference degrees of the stops are searched around origin and destination points that are represented as O and D respectively, by satisfying the following condition: stops are incorporated if their degrees are greater than γ degree. Then, k_O count of departure stops around O and k_D count of arrival stops around D are found. The searching process is separately employed between these departure and arrival stops on each stage for k_O x k_D pairs.

Initially, direct connection is searched and thus it is examined whether an intersection exists among any line between departure and arrival stops or not. In case of a direct link absence, searching process continues with transfer case where single transfer case is handled first. It could be said that there exists a single transfer if there is a fuzzy neighborhood relation between any lines departed from one of the $s_o \in S2(O)^{\gamma}$ stops and any other line arrived to one of the $s_d \in S2(D)^{\gamma}$ stops. To satisfy this need, it is required to introduce a transfer location that has ingoing line(s) from departure stops and that has also outgoing line(s) to arrival stops and these lines must be fuzzy neighborhood, thus it connects origin to destination. There must be alighting (s_{c_1}) and boarding (s_{c_2}) stops in transfer area to ensure the possible connection exactly.

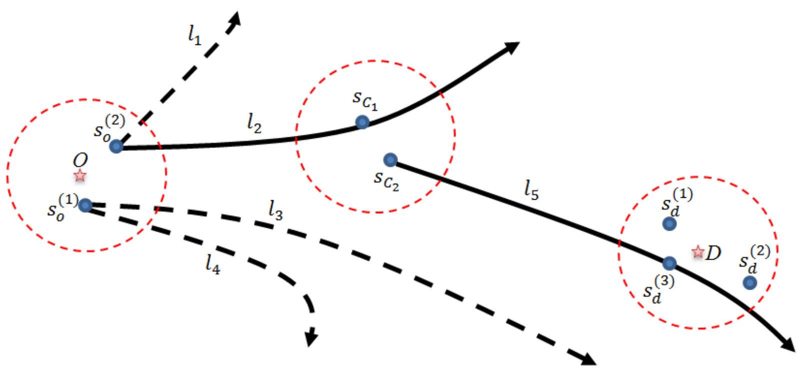

Figure 3. Example of single transfer route; $s_X^{(1)}, s_X^{(2)} \in S2(X)^{\gamma}$ and $s_Y^{(1)}, s_Y^{(2)}, s_Y^{(3)} \in S2(Y)^{\gamma}$; $s_X^{(2)}$, $s_{c_1} \in S1(l_2)$ and $LS(l_2, s_X^{(2)}) < LS(l_2, s_{c_1})$; $s_Y^{(3)}, s_{c_2} \in S1(l_5)$ and $LS(l_5, s_{c_2}) < LS(l_5, s_Y^{(3)})$; $\Pi_{XY} = \langle (s_X^{(2)}, l_2, s_{c_1}), (s_{c_2}, l_5, s_Y^{(3)}) \rangle$

It is noted that if alighting stop of previous trip is same with boarding stop of the next trip, passengers ride the line where they get off from previous line when transfer, otherwise they should walk between these stops before boarding the line. The route is determined as consecutive two possible connections which will be $\left(\mathrm{s_o}, l_1, \mathrm{s_{c_1}}\right)$ and $\left(\mathrm{s_{c_2}}, l_2, \mathrm{s_d}\right)$ respectively (Figure 3): first of them, it begins with departure stop s_o by boarding l_1 and it ends in alighting stop s_{c_1} by getting off l_1; then secondly, it continues by boarding l_2 from s_{c_2} and the route ends in $\mathrm{s_d}$ stop by getting off from l_2 line.

Double transfers should be sought if there is no single transfer. A route could be defined as a double transfer if there are three lines that have fuzzy neighborhood respectively and if they satisfy three specific conditions. These three conditions are as follows; first line of the route passes from origin to first transfer location, second line links between double transfer location directed towards a second transfer location and finally third line connects second transfer location to destination.

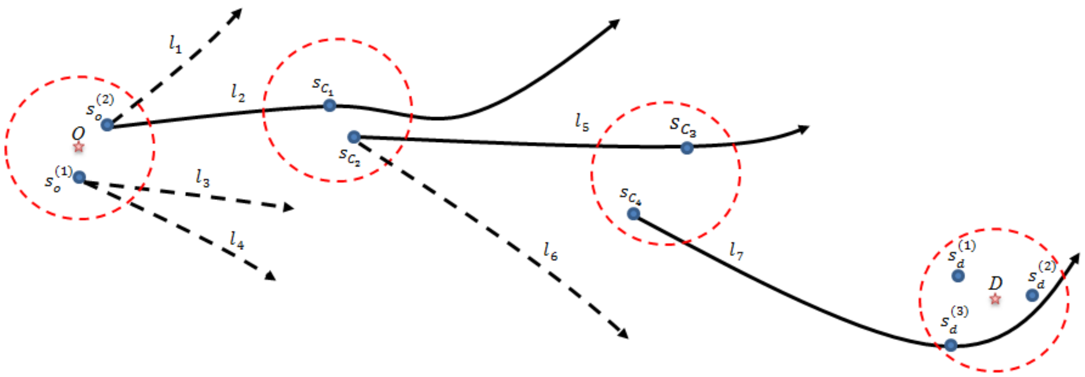

Figure 4. Example of double transfer route; $s_X^{(1)}, s_X^{(2)} \in S2(X)^\gamma$ and $s_Y^{(1)}, s_Y^{(2)}, s_Y^{(3)} \in S2(Y)^\gamma$; $s_X^{(2)}, s_{c_1} \in S1(l_2)$ and $LS\left(l_2, s_X^{(2)}\right) < LS\left(l_2, s_{c_1}\right)$; $s_{c_2}, s_{c_3} \in S1(l_5)$ and $LS\left(l_5, s_{c_2}\right) < LS\left(l_5, s_{c_3}\right)$; $s_Y^{(3)}, s_{c_4} \in S1(l_7)$ and $LS\left(l_7, s_{c_4}\right) < LS\left(l_7, s_Y^{(3)}\right)$; $\Pi_{XY} = \langle\left(s_X^{(2)}, l_2, s_{c_1}\right), \left(s_{c_2}, l_5, s_{c_3}\right), \left(s_{c_4}, l_7, s_Y^{(3)}\right)\rangle$.

Each transfer location is represented as a pair that consists of alighting and boarding stops respectively. Alighting stop $\mathrm{s_{c_1}}$ of the first transfer location links outgoing line $\mathrm{l_O}$ from $\mathrm{s_o}$ to itself that the stop will be on same line. Boarding stop $\mathrm{s_{c_2}}$ of the first transfer location is around $\mathrm{s_{c_1}}$ that it will satisfy condition (iii) in Definition 7. It connects itself to alighting stop $\mathrm{s_{c_3}}$, which is first parameter of the second transfer location, with l line going towards $\mathrm{s_{c_3}}$. Boarding stop $\mathrm{s_{c_4}}$ of the second transfer links itself to $\mathrm{s_d}$ with $\mathrm{l_D}$ that they will satisfy conditions of possible connection. Similar to the procedure described in single transfer, three possible connections will be required (s_o, l_O, s_{c_1}), (s_{c_2}, l, s_{c_3}) and (s_{c_4}, l_D, s_d) respectively and they are denoted in Figure 4.

4. Application

Izmir is the third metropolitan city of Turkey and it settles in the sea coast has co-ordinated transport which integrates bus, metro, suburban and ferry. In 1999, electronic fare collection system of Izmir was used for the first time with validating contactless smart card in Turkey. Fees are collected in this system via the contactless smart card which integrates all public transit modes is swiped to the card readers (or validators) at the buses, entrance of the stations or pier. After the card is once swiped, the system allows unlimited free boarding during 90 minutes by transfer between both same and different modes.

4.1 Smart Card Data Analysis for Izmir

Data considered in this study that it is acquired from both AFC and AVL system, covers 30 consecutive days from November 2014. Data obtained from AFC system contains about totally 44.5 million for 30 consecutive days. Therefore relational database is created by decomposing tables to get rid of redundant fields since both boarding and bus service transactions have common fields. It is necessary to identify both erroneous and missing values which are thought to affect analysis process and results before beginning data analysis. Data cleaning and editing are considerable parts of the data preparation in order to prevent data loss. The number of total cleaned data is about 0.21% of raw data for all days in this study. Thus low amounts of cleaning data cause to increase reliability of the analysis.

After data preparation, remains data includes boarding information that is totally 44,558,731 with 1,631,591 unique cards for 30 consecutive days on November 2014. Two cases, namely single boarding for a day and multiple boarding with same card on same service, are firstly determined. Data that belongs to passengers who made multiple boarding with same card on the same service generate 2.81% of the daily raw data. Transactions of passengers who made single daily boarding are 6.35% of all transactions. As it is mentioned in Section 2, it is necessary to obtain (boarding stop, next boarding stop) pairs in order to implement first phase of the LBEM which is formed with trip-chaining approach. Bus mode to all modes is 63% of whole pairs and the remaining part of these data belongs to other modes to all modes which are 37% of all pairs and these are extracted from analysis. The first stage of the LBEM algorithm is completed with operations mentioned above and alighting stop is estimated to be 85.05% of all (boarding, next boarding) pairs for bus to all modes according to trip-chaining approach. In next stage, RAEM is applied to data belongs to single boarding per day, multiple boarding with same card and not estimated by performing Assumptions 1 and 2. The data of single boarding per day is estimated to be 99.92% of whole data via RAEM and multiple boarding data is estimated to be 99.89% of whole data via RAEM. Similarly, the data which cannot be estimated by using Assumption 1 is estimated to be 99.88% of whole data via RAEM. Remaining data which cannot be estimated by trip-chaining and RAEM is estimated by using ALSM.

Results that are obtained by performing all operations in LBEM algorithm are saved in $OD_{dq}\left(l, c, s_l^k, s_l^m\right)$ matrix, where d is any day; q is any parts of the day; l is the line; c is the service; s_l^k denotes boarding stop that is k^{th} stop on line l; s_l^m denotes alighting stop that is m^{th} stop on line l. Results are stored into tables on database of Microsoft SQL Server 2012 Management System and all operations are written by SQL scripts. All operations are completed within about 28 minutes for 30 days.

4.2 Route Planner for Public Transportation in Izmir

The fuzzy neighborhood relations between stop-stop, line-stop and line-line which are very important concepts of this study distinguish our route planning algorithm from the other algorithms. Therefore stop-stop, line-stop and line-line are represented with tables for a fixed degree γ. The knowing fuzzy neighbor lines of any line decreases processing time of the route planning algorithm since all lines which can be transferred are easy to know.

Sub-procedures of route planning algorithm which are handled in Section 3 are developed with three separate queries for direct connection, single transfer and double transfer respectively. These procedures were implemented with SQL queries that are written by T-SQL programming. Before executing these, indexes were created for required fields on the tables mentioned above to speed-up the queries. The queries are executed on multimodal transportation system in Izmir consists of 6744 places (stops, stations and piers) located on 623 lines and all routes saved in database about 190 minutes.

Web platform is developed to allow to users selecting both origin and destination points and time by using ASP.NET framework and in addition, it is designed to visualize the results. The platform works as follows: user selects firstly origin and destination points on web form or map. After she/he clicks the “find routes” button, the request is sent to server-side and then function which provides connecting SQL database and executing SQL queries is called. The function returns JSON encoded string and results are visualized on Google Maps by using Google Maps API v3.0 and JQuery. Routes are reported on the web form as well and screenshot example is illustrated in Figure 5.

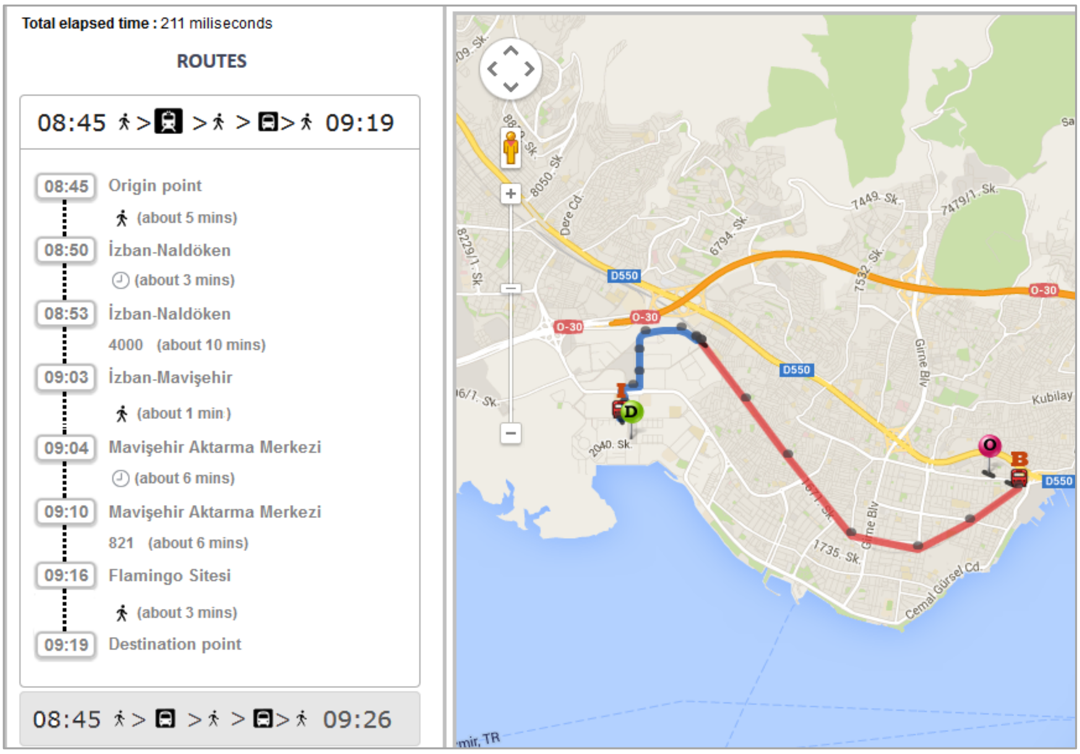

Figure 5. Example screenshot from the route planner

5. Conclusion

In this study, LBEM which is applied to smart card data used in entry-only system of public transport in order to estimate alighting stop, was examined theoretically and practically. Alighting stop information is important factor in public transportation planning since it provides benefit for reconstructing line routes, adding new stop or line to public transportation network and reorganizing timetables of lines. It is note that data cleaning and editing processes were the time-consuming part of smart card data analysis due to considerable amount of data needed for manual editing.

The concepts of the fuzzy neighborhood relations between stop-stop, line-stop, line-line are, discussed for the first time while addressing the route planning problem in Section 3. Fuzzy preference degree of the stops, is a new approach in identifying route planning problems, has been explained and it is emphasized that choosing stops with preference degree can affect while searching alternative routes. Walking distance is known to be an important criterion in trip planning but in real life situations, it is not sufficient to determine which stop is the most suitable one. In this study, in addition to this criterion, two new criteria; stop activity and being hub stop is included in preference degree. The minimum aggregation method, is one of the common methods in fuzzy logic theory and is used to compute preference degree of stop by combining three degrees mentioned above. Further researches can be done to determine this degree by using different aggregation methods.

As a consequence, studying on two important subtopics of ITS, namely APTS and ATIS, using information obtained from APTS application on ATIS implementation give this study credibility. It is intended for this study to contribute to further researches related to urban traffic problems which have been and will be trying to find solutions with advanced technologies. The proposed algorithms might also be an element of a more intelligent system which observes and makes use of human behaviors in public transportation. From this angle, our approaches might contribute to literature and systems which will be developed under the concepts of smart cities and sustainable mobility.

Acknowledgement

This study was conducted within the scope of Ph.D. Thesis in Statistics Program in Graduate School of Natural and Applied Sciences of Dokuz Eylül University. This study has also been partially supported by TUBITAK (grant No.113E535). I acknowledge Izmir Metropolitan Municipality ESHOT General Directorate, Department of Transport Planning and Statistics for their kind support.

References

Bagchi, M., Gleave, S. D., & White, P. R. (2003). Use of public transport smart card data for understanding travel behavior. *European Transport Conference*.

Barry, J., Newhouser, R., Rahbee, A., & Sayeda, S. (2002). Origin and destination estimation in New York City with automated fare system data. Transportation Research Record, 1817, 183-187.

Bast, H., Carlsson, E., Eigenwillig, A., Geisberger, R., Harrelson, C., Raychev V., & Viger, F. (2010). Fast routing in very large public transportation networks using transfer patterns, *ESA'10 Proceedings of the 18th annual European conference on Algorithms: Part I*, 290-301.

Ceder, A. (2007). *Public transit planning and operation: theory, modelling and practice* (1st ed.). Oxford: Butterworth-Heinemann.

Chang, H., Tai, Y., & Hsu J. Y. (2007). Public transportation routing using route graph. *Proceeding 12th Conference on Artificial Intelligence&Applications (TAAI 2007)*.

Dantzig, G. B. (1963). *Linear programming and extensions* (1st ed.). New Jersey: Princeton University Press.

Derrible, S., & Kennedy, C. (2011). Applications of graph theory and network science to transit network design. *Transport Reviews*, *31* (4), 495-519.

Geisberger, R., Sanders, P., Schultes, D., & Delling, D. (2008) Contraction hierarchies: Faster and simpler hierarchical routing in road networks. In C.C. McGeoch (Ed.), *Lecture Notes in Computer Science Vol.5038* (319-333), Springer Berlin Heidelberg.

Goldberg, A., Kaplan, H., & Werneck, R. (2006). Reach for A*: Efficient point-to-point shortest path algorithms. *In Workshop On Algorithm Engineering & Experiments*, 129–143.

Golnarkar, A., Alesheikh, A. A., & Malek, M. R. (2010). Solving best path problem on multimodal transportation networks with fuzzy costs. *Iran Journal of Fuzzy Systems*, *7* (3), 1-13.

Gutman, R. (2004). Reach-based routing: A new approach to shortest path algorithms optimized for road networks, *In Workshop on Algorithm Engineering and Experiments (ALENEX)*, 100–111.

Liu, C. (2002). Best-path planning for public transportation systems. *Proceeding IEEE 5th International Conference on Intelligent Transportation Systems*, 834–839.

Liu, C., Pai, T., Chang, C., & Hsieh C. (2001). Path-planning algorithms for public transportation systems. *Proceeding IEEE Intelligent Transportation Systems (ITSC 2001)*, 1061–1066.

Morency, C., Trépanier, M., & Agard, B. (2006). Analysing the variability of transit users behaviour with smart card data. *Proceedings of the IEEE ITSC 2006*, 44-49.

Munigaza, M. A., Devillaine, F., Navarrete, C., & Silva, D. (2014). Validating travel behavior estimated from smart card. *Transportation Research Part C*, *44*, 70-79.

Munizaga, M. A. & Palma, C. (2012). Estimation of a disaggregate multimodal public transport origin-destination matrix from passive smartcard data from Santiago, Chile. *Transportation Research Part C*, *24*, 9-18.

Pelletier, M. P., Trépanier, M., & Morency, C. (2011). Smart card data use in public transit: A literature review. *Transportation Research Part C*, *19*, 557–568.

Schultes, D., & Sanders, P. (2007). Dynamic Highway-Node Routing, In C. Demetrescu (Ed.). *Lecture Notes in Computer Science Vol.4525* (66-79). Springer Berlin Heidelberg.

Sinnott, R. W. (1984). Virtues of the Haversine. *Sky and Telescope*, *68* (2), 158.

Sussman J. S. (2005). *Perspectives on intelligent transportation systems* (1st ed.). New York: Springer.

Thorup, M. (2004). Compact oracles for reachability and approximate distances in planar digraphs. *Journal of the ACM, 51* (6), 993-1024.

Trépanier, M., Tranchant, N. & Chapleau, R. (2007). Individual trip destination estimation in a transit smart card automated fare collection system. *Journal of Intelligent Transportation Systems*, *11*(1), 1-14.

Verga, J., Silva, R.C., Yamakami, A., & Shirabayashi, W. V. I. (2013). Using the fuzzy sets theory in the multimodal transport network problem. *Proceeding IFSA World Congress and NAFIPS Annual Meeting (IFSA/NAFIPS)*, 801–806.

Von Ferber, C., Holovatch, T., Holovatch, Y., & Palchykov, V. (2009). Public transport networks: Empirical analysis and modeling. *The European Physical Journal B*, *68* (2), 261-275.

Von Ferber, C., Holovatch, Y., & Palchykov, V. (2005). Scaling in public transport networks. *Condensed Matter Physics*, *8* (1), 225-234.

Wang, W., Attanucci, J. P., & Wilson, N. H. M. (2011). Bus passenger origin-destination estimation and related analyses using automated data collection systems. *Journal of Public Transportation*, *14* (4), 131-150.

Wilson, N. H. M., Zhao, J., Rahbee, A. (2009). The potential impact of automated data collection systems on urban public transportation planning. In N. H. M. Wilson, A. Nuzzolo (Eds.). *Schedule-based modeling of transportation networks* (75-99). Springer US.

Zhao, J., Rahbee, A., & Wilson, N. H. M. (2007). Estimating a rail passenger trip origin-destination matrix using automatic data collection systems. *Computer-Aided Civil and Infrastructure Engineering*, *22* (5), 376-387.