FUZZY-LOGIC APPROACH TO ESTIMATE THE PASSENGERS’ PREFERENCE WHEN CHOOSING A BUS LINE WITHIN THE PUBLIC TRANSPORT SYSTEM

Vitalii Naumov¹,*, Baurzhan Zhumanbayev², Dinara Agabekova³, Zhumazhan Zhanbirov⁴, Igor Taran⁵

¹Faculty of Civil Engineering, Cracow University of Technology, Cracow, Poland
²Kazakh Academy of Transport and Communications, Almaty, Kazakhstan
³Eurasian Technological University, Almaty, Kazakhstan
⁴Central-Asian University, Almaty, Kazakhstan
⁵Department of Transport Management, Dnipro University of Technology, Dnipro, Ukraine

*E-mail of corresponding author: vnaumov@pk.edu.pl

Resume
For the developed system of public transport, the passengers, as the customers, have a variety of alternatives when choosing the transport mode or even the route for the given mode of public transport. The estimation of the passengers' preference is the key task for transportation planners for solving the wide range of optimization problems in the field of public transport. A methodology for estimation of the passengers' preference when choosing the bus line within a public transport system is developed in this paper. The proposed approach is based on the fuzzy-logic mathematical apparatus and uses the surveys' data to calculate the membership functions defining the passengers' preference. The case study of the passengers' survey, held in Talas (Kazakhstan), is used to illustrate the developed methodology.

1 Introduction

The passengers’ preference is the fundamental feature that conditions the choice of possible alternatives by travelers within an existing transport system [1]. These alternatives to be chosen by the transport system users, usually refer to the modes of transport or the travel routes (transport lines in the case of a public transport system).

The traditional approach to model the travelers’ preferences when selecting the mean of transport and trip path is based on the utility theory and, in practice, supposes the estimation of the utility function on the grounds of some empirical information.

The utility theory application in practical transportation studies may be found in recent publications [2-9]. The paper [2] compares different model specifications of travel time reliability in public transport route choices, the considered models are estimated based on empirical observations and used in order to evaluate the utility function. The authors of the paper [3] propose the mode choice model for public transport that integrates structural equation and discrete choice models with categorized latent variables; the objective of the presented study was to develop an improved disaggregate model that better ex-plains travel behavior within the public transport system. The study [4] uses empirical data of university students’ transport choices in the Bilbao area, the authors illustrate how the utility that individuals get from the mode of transport can be modeled and estimated based on this information. The model described in publication [5] allows bicycle-sharing system operators to plan services more effectively by examining the impact of travel distance, land use, built environment and access to public transportation infrastructure on users’ destination preferences: the authors of the paper propose to generate utility profiles as a function of distance and other attributes. Authors of [6] propose the approach to evaluate the passengers’ comfort based on the preference survey where crowding levels are presented as illustrations; the survey data are used by the authors to estimate discrete choice models and obtain a subjective evaluation of passenger density through the parameters of the utility function. The study [8] depicts the study case of using the EVA mode choice model in the city of Ljubljana, Slovenia; the authors have designed the stated preference survey and estimated different types of utility functions; the utility functions obtained based on the survey results are used further as initial data for the PTV Visum software to simulate the operation of the city transport system.
The results of the assessment of the travelers’ preference are used for modelling the demand for transport services [10-11], but also for optimization of transport processes [12-13], as well as for designing and simulation of the transport system [14-16].

As far as the traveler preference cannot be determined unequivocally (as it depends on the big number of non-deterministic parameters with the values assessed subjectively by different persons), the fuzzy logic mathematical apparatus is widely used to define the preferences of passengers and their satisfaction with the provided transport services [17-20]. The authors of the paper [17] have developed an approach combining characteristics of the analytic hierarchy process, entropy weight method and the fuzzy comprehensive evaluation method to improve the accuracy of passenger satisfaction evaluation for public transport. A similar methodology is described in [18]: the authors propose a decision support model for measuring the public transport level by using a combination of the fuzzy logic and the analytic hierarchy process. A method, used to evaluate the passenger satisfaction with the public transportation system that is based on the Pythagorean fuzzy sets and multi-objective optimization, is described in [19].

A method to describe the travel comfort characteristics with synthetic indices, based on the individual comfort indices of travel components, is proposed in [20]; the authors use a fuzzy approach to evaluate the conditions of travelers’ comfort.

This paper contributes to direction of using the fuzzy logic for assessment of the travelers’ preferences. The aim was to develop a simple but reliable method that uses the travelers’ survey data to calculate the membership functions describing the basic preferences of the passengers of a public transport system: pricing, comfortability and travel speed.

2 Proposed method of the passenger preferences assessment

The carrier’s goal is the most complete coverage of the existing and potential market, for which it is necessary to attract passengers moving from and to the points covered by the public transport route. If the fulfillment of the passenger’s need for movement is possible in the only way - by using one accessible bus route, then the carrier gets the maximum possible share of this market sector. However, if the need for travel can be realized in more than one way (the trip can be implemented by more than one route in the public transport system), the passenger has a choice. In this case, there will be a conflict situation between carriers who can potentially serve the same trip. Depending on the strategy chosen by the carrier, the passenger gives preference to the corresponding route.

The assignment of a trip to public transport routes depends on preferences of passengers, but also on importance of this trip for the carrier (meaning that the carrier determines the strategy of behavior depending on the attractiveness of a particular trip).

Passenger preferences when choosing a route depend on three main indicators: delivery speed (travel time), comfort and the price of services. All of these parameters are determined to some extent by the bus model. The delivery speed depends on the design features (however, the traffic speed is also affected by the congestion of the road network and the methods of organizing traffic). Comfortability directly depends on the passenger capacity of the bus, the design features of the cabin and vehicles’ suspension. The price of the service is determined based on the cost of the provided transport services, which depends on the performance characteristics of a particular bus model. Thus, we can claim that the preferences of passengers when choosing a particular route depend on the bus models used by carriers.

It is convenient to describe the degree of preference by passengers of a particular bus model by means of a fuzzy subset that characterizes the belonging of a given bus model to the set of optimal models. In this case, the degree of preference for different origin-destination (O-D) pairs, or for the same O-D pair at different hours of the day (day of the week), can differ significantly. For example, for O-D pairs associated with the sleeping areas of a city, the significance of the travel speed is different depending on the time of day. In general, each O-D pair can be associated with different subgroups of consumers of transport services, based on the purpose of the trip, while the composition of the O-D pairs will determine the type of a membership function.

If the preference by the criterion of the travel speed is described with the membership function \( \mu_v \), by criterion of the travel comfort - with function \( \mu_k \) and by criterion of the travel price - with function \( \mu_T \), where \( \mu_v \in [0;1], \mu_k \in [0;1], \mu_T \in [0;1] \), then the general preference of a passenger can be expressed through a fuzzy subset \( \mu \), which is a combination of subsets \( \mu_v, \mu_k \) and \( \mu_T \):

\[
\mu = w_v \cdot \mu_v + w_k \cdot \mu_k + w_T \cdot \mu_T,
\]

where: \( w_v, w_k \) and \( w_T \) are weight coefficients for the membership functions of the travel speed, comfortability and the tariff, respectively.

To assess the preferences of passengers, it is necessary to determine the type of membership functions of a fuzzy subset of optimal bus models and standardize them for the main groups of passengers. In order to determine the membership functions, use of methodology for analyzing the results of an expert survey is proposed. For this, the following stages should be performed by a researcher:

1. Collecting the data and calculating significance of features for the respondents divided into main social groups of travelers (e.g., students, retirees, etc.).
2. Estimating the sufficiency of the collected data.
3. Assessing the consistency of the respondents’ opinions.
4. Evaluating the empirical values of the membership functions for categories of the features’ significance per each social group represented in the respondent’s sample.
5. Evaluating the functional dependencies that approximate the membership functions (at this stage, the quality of the obtained mathematical models should be checked and the rescaling of the dependencies can be performed).

At the first stage, respondents (public transport customers) are invited to assess the significance of the level of tariffs, comfort and travel speed based on a 10-point scale. When processing the questionnaires for each of the characteristics, the features’ significance \( z_i \) for the respondents is determined in unit fractions as follows:

\[
z_i = \frac{B_i}{B_V + B_K + B_T},
\]

where: \( B_i \) is the number of points given by a respondent to the \( i \)-th feature; \( B_V; B_K; B_T \) are the grades given by the passengers to travel speed, comfortability and the tariff, respectively.

At the second stage, after a preliminary assessment of the respondents’ opinions, the sufficiency of the number of interviewed respondents should be assessed. For this, the average sampling error \( \Delta_i \) must be estimated according to the significance of the \( i \)-th characteristics to passengers participated in the survey [21]:

\[
\Delta_i = \sqrt{\frac{1}{N (N - 1)} \sum_{j=1}^{N} (z_{ij} - \bar{z}_i)^2},
\]

where: \( \bar{z}_i \) is the average value of the \( i \)-th feature’s significance for a passenger; \( N \) is the number of respondents (passengers interviewed).

The required (sufficiently big) size \( N_i \) of the sample for the \( i \)-th feature is estimated in the following way [21]:

\[
N_i = \frac{N}{d_i^2 + 1},
\]

where: \( d_i \) is the ratio of the sampling error margin to the standard deviation of the studied values (the features’ significance for the passengers):

\[
d_i = \frac{\Delta_i}{\sigma_i},
\]

where: \( \sigma_i \) is the standard deviation for the significance of the \( i \)-th feature.

At the third stage, it is proposed to assess the consistency of the respondents’ opinions using the Kendall’s coefficient of concordance \( W \) [21]:

\[
W = \frac{12 \cdot S}{N^8 (M^8 - M)},
\]

where: \( S \) is sum of squares of deviations of all the rank estimates of each object of examination (the preference feature) from the corresponding mean value; \( M \) is the number of the objects of expertise (the preference features being assessed by the respondents).

The closer the concordance coefficient value to 1, the better is the consistency of the respondents’ opinions. If the \( W \) value is less than 0.5, the number of the respondents participated in the survey should be increased.

At the fourth stage, based on the survey’s data, it is proposed to define the passenger’s preference as the membership function \( \mu_{ik} \) for the \( i \)-th social group by the \( j \)-th feature in the \( k \)-th category in the following way:

\[
\mu_{ik} = \frac{N_{ijk}}{\max N_{ijk}},
\]

where: \( N_{ijk} \) is the number of respondents of the \( i \)-th social group who rated the \( j \)-th feature as a value in the \( k \)-th range defining the corresponding category.

Finally, at the fifth stage, it is proposed to estimate the analytical form of the membership functions based on the pairs of empirical values and the corresponding categories \( \langle \mu_{ik}, k \rangle \). The functional dependences may be assessed for each of the features (tariff, comfort and travel speed) by using the least squares method for the polynomial model reflecting the shape of the dependence:

\[
\mu_i(k) = \sum_{k=0}^{P} a_{ik} \cdot k^p,
\]

where: \( k \) is the ordinal number of the category for the \( i \)-th feature described by the membership function \( \mu_i \); \( a_{ik} \) are the coefficients of the polynomial model defining the functional dependence; \( P \) is the complexity of the polynomial model.

It should be noted that the polynomial models fitted to the empirical data must be complex enough to provide the desired quality of estimations, e.g. the highest degree of the polynomial must be big enough to guarantee the value of the determination coefficient at least at the level of 95% (in practice, the higher is the polynomial model’s complexity, the better is the model’s fitness to the empirical data).

The coefficients \( a_{ik} \) of the polynomial models for determining the dependences of the membership functions on the natural values of the selected features (tariff, comfort level and travel speed) can be determined from the ratio:

\[
a_{ik} = a_{ik} \left( \frac{K}{\max f_i} \right)^p,
\]

where: \( \max f_i \) is the maximum value of the \( i \)-th feature; \( K \) is the number of the defined categories of the
Children were not considered as the respondents in the conducted survey, as they usually do not travel alone and, as a consequence, do not make the trip choice decisions by themselves.

The survey was conducted at the bus stops of the Talas public transport system. The respondents answered the group of questions related to their social status (age, source of income, the average income, etc.). In addition, the respondents were asked to assess their feature’s significance for the traveler (the bigger is the number of categories, the more precise estimation of the membership functions’ dependencies will be obtained).

### 3 Case study: estimation of passengers’ preference in Talas, Kazakhstan

The passenger preference survey was conducted in October 2019 in the city of Talas (Taraz), Kazakhstan. Within the conducted survey 390 inhabitants were interviewed and the following social groups were identified: adults of working age, students and pensioners (retirees). The division of the respondents on the subsets representing the mentioned social groups was conditioned by the existing pricing policy: the ticket price for students is partly refunded, adults pay the full price and retirees can use the public transport system under conditions of a full refund of the travel costs. Furthermore, the representatives of the listed social groups have different average incomes; this factor conditions the trip preferences of the passengers including the route and transport choice preferences.

#### Table 1 Results of the conducted survey

<table>
<thead>
<tr>
<th>range of the indicator's significance</th>
<th>tariff level</th>
<th>comfort level</th>
<th>speed level</th>
</tr>
</thead>
<tbody>
<tr>
<td>adults of working age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0…0.1</td>
<td>11</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>0.1…0.2</td>
<td>26</td>
<td>22</td>
<td>17</td>
</tr>
<tr>
<td>0.2…0.3</td>
<td>27</td>
<td>28</td>
<td>22</td>
</tr>
<tr>
<td>0.3…0.4</td>
<td>28</td>
<td>30</td>
<td>26</td>
</tr>
<tr>
<td>0.4…0.5</td>
<td>38</td>
<td>48</td>
<td>61</td>
</tr>
<tr>
<td>total in group</td>
<td>130</td>
<td>130</td>
<td>130</td>
</tr>
<tr>
<td>students</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0…0.1</td>
<td>1</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>0.1…0.2</td>
<td>12</td>
<td>52</td>
<td>13</td>
</tr>
<tr>
<td>0.2…0.3</td>
<td>48</td>
<td>70</td>
<td>15</td>
</tr>
<tr>
<td>0.3…0.4</td>
<td>30</td>
<td>17</td>
<td>52</td>
</tr>
<tr>
<td>0.4…0.5</td>
<td>91</td>
<td>39</td>
<td>100</td>
</tr>
<tr>
<td>total in group</td>
<td>182</td>
<td>182</td>
<td>182</td>
</tr>
<tr>
<td>retirees</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0…0.1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>0.1…0.2</td>
<td>2</td>
<td>2</td>
<td>11</td>
</tr>
<tr>
<td>0.2…0.3</td>
<td>6</td>
<td>21</td>
<td>26</td>
</tr>
<tr>
<td>0.3…0.4</td>
<td>21</td>
<td>26</td>
<td>28</td>
</tr>
<tr>
<td>0.4…0.5</td>
<td>48</td>
<td>28</td>
<td>11</td>
</tr>
<tr>
<td>total in group</td>
<td>78</td>
<td>78</td>
<td>78</td>
</tr>
<tr>
<td>total in survey</td>
<td>390</td>
<td>390</td>
<td>390</td>
</tr>
</tbody>
</table>

#### Table 2 Calculation results for the sufficient number of respondents

<table>
<thead>
<tr>
<th>parameter</th>
<th>feature</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>tariff level</td>
</tr>
<tr>
<td>standard deviation</td>
<td>0.0107</td>
</tr>
<tr>
<td>average error of the sample</td>
<td>0.0005</td>
</tr>
<tr>
<td>margin of the sample’s error</td>
<td>0.0016</td>
</tr>
<tr>
<td>the ratio of the sample’s error margin to the standard deviation</td>
<td>0.1527</td>
</tr>
<tr>
<td>statistically sufficient number of respondents</td>
<td>371</td>
</tr>
</tbody>
</table>
preference related to the transport line choice for three basic features: the level of tariff, comfortability and travel speed. For each feature, the grade was assigned according to a 10-points scale (the more preferable is the transport line's feature, the higher is the grade). Based on grades given by each respondent, the features' significance was calculated by using Equation (2).

After processing the survey results, it was stated that the value $z_i$ for $i$-th feature is not greater than 0.5.

To study the preferences of passengers, 5 ranges were defined in the range of possible values of the significance with a step of 0.1:

- 1 category: the features' significance $z_i$ in the range $[0; 0.1]$;
- 2 category: the features' significance $z_i$ in the range $(0.1; 0.2]$;
- 3 category: the features' significance $z_i$ in the range $(0.2; 0.3]$;
- 4 category: the features' significance $z_i$ in the range $(0.3; 0.4]$;
- 5 category: the features' significance $z_i$ in the range $(0.4; 0.5]$.  

All respondents were divided into three groups according to social status, after which the preferences were investigated for each attribute in each of the groups. Results of the survey are presented in Table 1.

Results of calculation of a sufficient number of respondents for each feature of the passengers' preference are shown in Table 2.

As can be seen from Table 2, a sufficient number of respondents is smaller than the number of respondents who took part in the survey. It means that the analyzed
the results of a questionnaire, according to Equation (6), is equal to 0.827, which indicates a high agreement of opinions of respondents who participated in the survey. The value of the coefficient of concordance, based on the sample is big enough to ensure the statistical significance of the membership functions being estimated. The value of the coefficient of concordance, based on the results of a questionnaire, according to Equation (6), is equal to 0.827, which indicates a high agreement of opinions of respondents who participated in the survey.
4 Results and discussion

Using the standard MS Excel functions, the trend models for the membership functions as the third-degree polynomial models were defined. The third power of the polynomial models was accepted as the satisfactory precision, as far as determination coefficient for the obtained models was not lower than 0.90. A better approximation could be performed, if the higher power of polynomial models were applied; however, that may lead to overestimation of the functional dependencies representing the membership functions.

The polynomial models of the membership functions representing features of the passengers’ preference depending on the categories of values are shown in Figures 1-3.

Results of calculations for the polynomial models coefficients, to determine the dependence of the membership function on the value of the corresponding feature, are shown in Table 3.

When calculating the coefficients \( a_0 \), values of natural indicators presented in Table 4 were considered.

The polynomial models in their recalibrated form may be used for assessment of the passengers’ preferences based on the features’ values presented in the corresponding units (not on their significance for the public transport users). However, the features’ values should be contained in the range of possible values shown in Table 4.

To evaluate the membership function \( \mu \), depicting the final preference of passengers, the weight coefficients for each of the selected features must be determined. Values of the weight coefficients, as the arithmetic mean values of \( z \), for all respondents, were defined. In accordance with the survey data, the following values were obtained \( w_0 = 0.321 \), \( w_r = 0.322 \) and \( w_r = 0.357 \). As it can be noted, the selected preference features were evaluated by the respondents almost equally, although the tariff was assessed as a bit more significant feature.

It should be underlined that the functional dependencies described in this paper represent preferences of the public transport system customers in the Talas city. There is no evidence that these models may be used for assessing the preferences of the public transport users in other regions. However, the presented methodology is transferable and can be implemented for the survey-based studies of transport preferences in any system of public transport.

5 Conclusions

The proposed approach makes possible determining the shape of membership functions of a fuzzy subset of optimal bus models for the main categories of passengers. The obtained functional dependencies describing the membership functions allow researchers to estimate the preferences of passengers for specific origin-destination pairs. Results of evaluating the passenger preferences are the initial data for solving a wide range of problems in the field of transport planning, such as defining utility functions for the route-choice and mode-choice tasks in the advanced simulation models of a city transport system, choosing strategies of transport companies servicing a public transport system, assessing the quality of services provided within a city public transport system, etc.

The considered example of assessing passenger preferences, based on the results of a survey in the Talas city, suggests that the developed approach is a convenient tool using which a statistically significant assessment of passenger preferences can be obtained with minimal effort. Special attention, however, should be paid to the design of a survey used as initial data in the proposed method: the number of respondents representing considered social groups should correspond to the existing proportion of these groups in the city population. Furthermore, the number of respondents should be big enough to ensure the statistical significance of the models for assessing the passengers’ preferences.

As directions for the further research, one should mention checking the conformity of the obtained polynomial models based on survey data in other cities, as well as using a bigger number of categories when calculating empirical values of membership functions. Another issue to be solved in the future concerns the complexity of the polynomial models approximating the membership functions, although in the case-study described in this paper the third-degree polynomial models have guaranteed the satisfactory precision.

References

FUZZY-LOGIC APPROACH TO ESTIMATE THE PASSENGERS' PREFERENCE WHEN CHOOSING A BUS...