A DISCRETE ROUTE CHOICE MODEL USING SUPPORT VECTOR MACHINE IN CONTEXT OF DHAKA CITY

*Saurav Barua[1]
[1]Department of Civil Engineering, Daffodil International University

Email: saurav.ce@diu.edu.bd

Abstract: The study delves to identify the route choice behavior between two alternative routes for a set of origin and destination point in Dhaka city. The Route 1 is Uttara to Motijheel through Rampura and the Route 2 is Uttara to Motijheel through Mohakhali. Perceived travel time, congestion level, trip purpose and familiarity of route are considered as important attributes for the route choice. A Support Vector Machine (SVM) approach is adopted to model the route choice behavior. The proposed model shows 88% accuracy in testing and precision of choosing Route 1 and Route 2 are 89% and 87% respectively. In total, 47.62% and 52.38% traffic are assigned in Route 1 and Route 2 respectively. Also found that, perceived travel time and familiarity of route are the most important attributes while making route choice decision.

Keywords: *Route choice, Support Vector Machine, Travel time, Congestion level, traffic assignment.*

1. INTRODUCTION

Route choice behavior is widely adopted for analyzing and understanding road users' behavior, develop road navigation technologies, transportation forecasting, traffic operation and design road infrastructure. It predicts road users' behavior under different scenarios and forecasts future traffic conditions on existing transportation network. Route choice is an essential part of traffic assignment, where traffic demand travelling from origin to destination are assigned to the selected routes in a road network. Transportation engineers can evaluate performance of road network and design efficient traffic control strategies by route choice modelling accurately.

Since transportation infrastructure, socio-economic aspects, perception regarding travel time and available alternatives are different, road users' behavior in developing countries are different than those of developed countries. Dhaka is a densely populated city where traffic congestion is day-to-day phenomena; understanding the road users' route choice behavior is complex. Travel cost, congestion level and availability of alternatives make the route choice decision more multifaceted along with travel time.

Various non-parametric discrete choice models have been developed to explore route choice modelling. For instance, Sander and Ahmed [1] used Neural Network (NN), Hassan and Nassir [2] adopted Fuzzy Logic and Musharraf et al. [3] utilized decision tree technique to model route choice. Those methods provide decent prediction performance. Sun and Park [4] utilized Support vector machine (SVM) to develop route choice model. Their study found that SVM provides better prediction accuracy and takes less computational time compare to Neural Network. Moreover, SVM can work with small sample size as well efficiently. In this study, stated preference questionnaire survey (QS) have been conducted and analyzed route choice behavior by support vector machine (SVM) for the selected Uttara to Motijheel route in Dhaka city. The research investigates influence of several attributes on the choice of route and its sensitivity with respect to the variability in the route attributes.

2. LITERATURE REVIEW

It is essential to determine the attributes those effects road users' route choice while modeling route choice behavior. Understanding these attributes will aid to improve transportation planning and network analysis. Route choice model identifies the type and content of traffic information that will guide road users' in their decision making. Though travel time is considered as most important attribute that affects road users' route choice—travel cost, directness, less congestion, trip purpose and urgency are also important [4]. Various researches showed that road users consider numerous criteria in choosing a route, including, experiences, habits, environment, traffic safety and cognitive limits [4, 5]. Some road users may prefer to reduce travel time, while other may feel uncomfortable with congested roads. Some road users may look for familiar landmarks, while others may wish to see scenic routes. Each criterion may relate to a different preferred route. Therefore, considering travel time as the only criterion of route choice is unrealistic and represents inaccurate traffic assignment of road network.

Studies on route choice model cover wide variety of aspects. Asakura et al. [5] considered topological aspects as key attributes of drivers' route choice behavior. Alivand et al. [6] developed path sized logit model to understand different attributes of scenic route selection. Adaptation of latest technology and ease to availability of traffic information help to change travel route choice behavior among road users. Khoo and Asitha [7] performed stated preference survey to observe influence of smart phones for traffic information and route guidance. Kim et al. [8] developed evacuation route choice model of Haeundae Beach in Korea using simulation. Some of the route choice models developed for specific vehicle classes. Zimmermann et al. [9] studied recursive logit model to identify attributes of bicycle facilities and focused on link flow and accessibility measures. Anderson et al. [10] developed route choice model for multimodal public transport network considering heterogeneity among travelers.

Support Vector Machine (SVM) is a machine learning approach introduced by Cortes and Vapnik [11]. It is used for classification, in which, a hyper plane is identified by separating data points belong to various categories with maximum margins. SVM does not have over fitting like Neural Network. It is used in several researches related to traffic route choice decision models. Sun and Park [4] evaluated performance of SVM over Neural Network route choice model and found better computational efficiency. Zeng et al. [12] adopted SVM to model route choice considering trade-off between emission reduction and travel time constraint. Said et al. [13] developed adaptive transportation decision system and predicted best transport route using Support Vector Machine (SVM).

3. METHODOLOGY

A. *Formulation of SVM*

The SVM model maps the data point into high dimensional space and determines the hyper plane which divides the data points representing different categories. The hyper plane is in following form in Equation (1).

$$(w, x) + b = 0 \quad (1)$$

w and b are the parameters, which are obtained through solving optimization problem in Equation (2).

Minimize $\frac{1}{2}(w, x)+C\sum_{i=1}^{n}\xi_i$ for $w \in H_0$, $b \in R$, $\xi \in R^n$ (2)

Subject to $y_i\,((w, \Phi(x_i)) + b) \leq 1- \xi_i$, $i = 1,\ldots n$

$\xi_i \geq 0 \quad i = 1 \ldots\ldots n$

Here, x_i and y_i are the data points in the sample. C is used to adjust the first term of the objective function and it is positive term. ξ_i is slack variable which balances the margin between the hyper plane and the nearest data points. $\Phi(x_i)$ is feature map which maps the data points into a high dimensional space H_0. H_0 is called Hilbert space.

Lagrange approach is used to solve this dual program and a Lagrange multiplier is added. The solution of the optimization problem, α_i^*, w_D^*, b_D^* can define the optimal hyper plane which is used for binary classification.

$$f(x) = (w_D^*,\Phi(x))+b_D^*=\sum_{i=1}^{n} y_i\, \alpha_i^* k\,(x, x')+b_D^* \quad (3)$$

where $k(x, x') = (\Phi(x_i), \Phi(x)) \quad x, x' \in X$

$f(x)$ is classification function and its result describes the category that predictor x belongs to $k\,(x, x')$. $k\,(x, x')$ is linear the kernel function. Proper kernel function is selected depending on the problem studied. The kernel function and two parameters are determined before training. Since the study is on two-class learning, linear kernel function is considered. The value of parameter C and kernel scale parameter are calculated through cross validation. The kernel scale parameter is not shown in the above equations. It adjusts the scale of kernel function and makes the data points stay within certain range. The dataset is randomly grouped into five divisions and each of them is considered as evaluation data once. Three random divisions are prepared and the averages of all the test performances at different values of parameters are determined. Thus the combination of C and kernel scale parameters which produces the best performance of the model is selected.

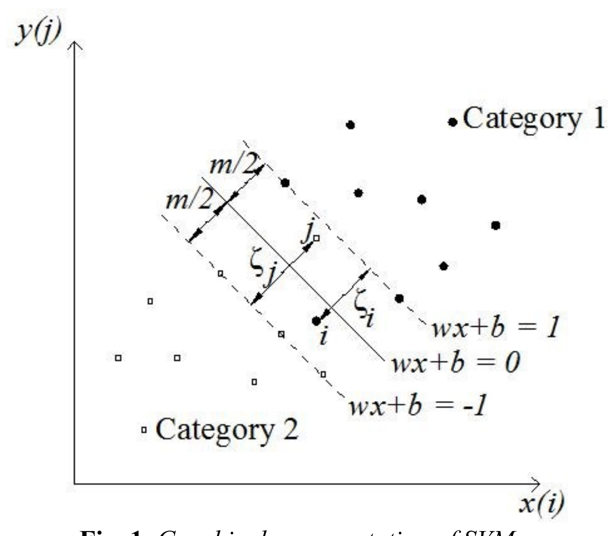

Fig. 1. *Graphical representation of SVM classification (Reference: Zhang and Xie [14])*

B. *Experiment Design*

Dhaka, the capital of Bangladesh is selected as the study area. Origin node of the studied routes is Uttara residential area and destination of the route is Motijheel business area. Two alternative routes are studied, The Route 1 is Uttara to Motijheel via Rampura and the Route 2 is Uttara to Motijheel via Mohakhali. The length of Route 1 and Route 2 are 21.6km and 20.1km respectively. The Route 1 has 8 signalized intersections and it characterized with residential area and kitchen market activities. The Route 2 has 10 signalized intersections and number of commercial and residential area situated along this route. The Database used in this research has been collected through roadside interview surveys. A predesigned questionnaire has been developed to capture the road users' preference for the study route while traveling from origin to destination. The attribute values of the route perceived by the road users are recorded. Road users are asked to rank the attributes. Four attributes are considered in this study, those are —travel time, congestion level, trip purpose and familiarity. Travel time and congestion level are categorized as very low, low, medium, high and very high; where very low is 1, low is 2, medium is 3, high is 4 and very high is 5. Trip purpose is categorized as work trip and non-work trip with numerical value 1 and 2 respectively. Familiarity of route is categorized as familiar for 1 and non-familiar for 0. Total 142 survey samples are collected among which 139 are considered for analysis, other three are dropped because of incomplete information.

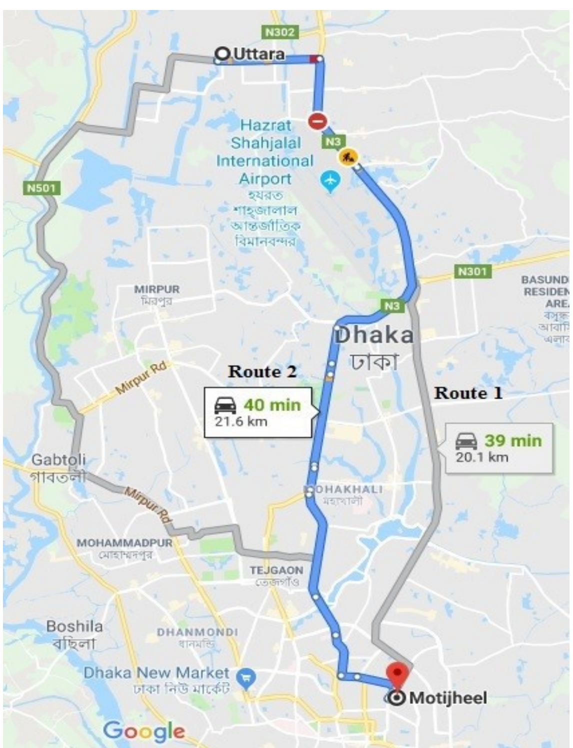

Fig. 2. *Google map Showing Studied route*

A Support Vector Machine (SVM) model is developed to map road users' route choice behavior. The input parameters are perceived travel time, perceived congestion level, trip purpose and familiarity with the route. The output of the model is choice of the route, the answer is 'yes' or 'no' type of selection for specific route. The output is numerically expressed by 1 for selection of a route and 0 for not selection of the route. The model is developed by Scikit-learn (version 0.20.3) library [15] under Python 3.0 program.

4. Results

Among total 139 data, 70% data i.e. 97 data points are used for training and rest of 30% i.e. 42 data points are used for testing. Precision of choosing Route 1 and Route 2 are 89% and 87% respectively while testing. Accuracy of estimation of the proposed SVM model in testing is the weighted average 0.88 i.e. 88%., which is shown in Figure 3. Traffic assigned in Route 1 and Route 2 are 47.62% and 52.38% respectively.

```
Confusion matrix

[[17  3]
 [ 2 20]]
              precision    recall  f1-score   support

     Route 1       0.89      0.85      0.87        20
     Route 2       0.87      0.91      0.89        22


weighted avg       0.88      0.88      0.88        42

Coefficients
[[ 0.69562186  0.36971044 -0.06499018 0.62601219]]
[-11.84743498]
```

Fig. 3. *Confusion matrix, classification report and coefficient values in Python*

The coefficient of perceived travel time, perceived congestion level, trip purpose and familiarity with route found 0.6956, 0.3697, -0.0649 and 0.626 (rounding up to three digit decimal) respectively. Among the attributes travel time is the most important and trip purpose has the least importance. Details of sensitivity analysis and rank of attributes are presented in the Table I.

Table I. RANK OF ROUTE CHOICE ATTRIBUTES AND SENSITIVITY ANALYSIS

Attributes	Rank	%Sensitivity
Perceived Travel time	1	7%
Familiarity with the route	2	6%
Perceived Congestion level	3	4%
Trip purpose	4	1%

(Table footnote: Sensitivity analysis is performed by considering 1% change in attribute dataset which will affect the route choice decision in percentage)

Road users in Dhaka make their route choice decision most depending on their perception regarding travel time. They are concerned about the travel time, however, other attributes also have important contribution. Both Route-1 and Route-2 run through commercial and shopping area. Both of the routes suffer traffic congestion during peak hour. As a result, work trip and non-work trip have less effect on the route choice model in the study. On contrary, familiarity of specific route is second most important attribute. It reveals that road users in Dhaka make significant numbers of route choice decision based on their familiarity with a specific route. Perceived congestion level is the third most important attribute for route choice decision making. Road users' perception regarding congestion varies with wide variety and depending on their cognitive level.

5. Conclusion

In this study, Support Vector Machine has been used to model the choice of the road users. Road users have a tendency of comparative attitude towards various alternative options. The model is framed to perceive road users about two routes, which are compared to make the choice. Travel time perception, congestion level perception, trip purpose and familiarity of routes are considered as decision factors for making the choice of route. These factors are converted to corresponding rating scale—which is in the form of a linguistic way and refers to road users' statements. Two alternative routes with various attributes between origin and destination have been provided for the road users in the choice set to make the choice decision. The inputs of the model are differences between the perception values of each attributes. The output is the percentage preference of one route over the other and ranking the attributes based on their influence. One outcome of this study is the percentage of road users on each route i.e. distribution of the road users between the two routes. The other outcome is observing the ranking and sensitivity of different attributes on route choice decision.

6. Acknowledgment

The authors would like to express thanks to the Committee for Advanced Studies and Research (CASR) of Bangladesh University of Engineering and Technology (BUET) for the questionnaire survey data.

References

[1] S. Van Cranenburgh and A. Alwosheel, "An artificial neural network based approach to investigate travellers' decision rules." Transportation Research Part C: Emerging Technologies, Vol. 98, pp.152-166, 2019.

[2] M.N. Hassan, T.H. Rashidi and N. Nassir, "Developing a Fuzzy Logic Based Sampling Protocol for Transit Route Choice Modeling (No. 18-06674)", Transportation Research Borad 97th Annual Meeting, 2018.

[3] M. Musharraf, J. Smith, F. Khan and B. Veitch, "Identifying route selection strategies in offshore emergency situations using decision trees." Reliability Engineering & System Safety, 2018.

[4] B. Sun and B.B. Park, "Route Choice Modeling with Support Vector Machine." Transportation research procedia, Vol. 25, pp.1806-1814, 2017.

[5] Y. Asakura, T. Yamauchi, E. Hato and M. Kashiwadani, "A route choice model considering topological aspects of a road network." Journal of the Eastern Asia Society for Transport Studies, Vol. 4(3), pp.55-67, 2001.

[6] M. Alivand, H. Hochmair and S. Srinivasan, "Analyzing how travelers choose scenic routes using route choice models." Computers, environment and urban systems, Vol. 50, pp.41-52, 2015.

[7] H.L. Khoo and K.S. Asitha, "User requirements and route choice response to smart phone traffic applications (apps)." Travel Behaviour and society, Vol. 3, pp.59-70, 2016.

[8] J. Kim, S. Lee and S. Lee, "An evacuation route choice model based on multi-agent simulation in order to prepare Tsunami disasters." Transportmetrica B: transport dynamics, 5(4), pp.385-401, 2017.

[9] M. Zimmermann, T. Mai and E. Frejinger, "Bike route choice modeling using GPS data without choice sets of paths." Transportation research part C: emerging technologies, Vol. 75, pp.183-196, 2017.

[10] M.K. Anderson, O.A. Nielsen and C.G. Prato, "Multimodal route choice models of public transport passengers in the Greater Copenhagen Area." EURO Journal on Transportation and Logistics, Vol. 6(3), pp.221-245, 2017.

[11] C. Cortes, N. Vapnik, "Support-vector networks. Machine Learning", Vol. 20 (3), pp.273–297, 1995. CiteSeerX 10.1.1.15.9362. DOI:10.1007/BF00994018.

[12] W. Zeng, T. Miwa and T. Morikawa, "Application of the support vector machine and heuristic k-shortest path algorithm to determine the most eco-friendly path with a travel time constraint." Transportation Research Part D: Transport and Environment, Vol. 57, pp.458-473, 2017.

[13] A.M. Said, E. Abd-Elrahman and H. Afifi, "A comparative study on machine learning algorithms for green context-aware intelligent transportation systems." In 2017 International Conference on Electrical and Computing Technologies and Applications (ICECTA) (pp. 1-5). IEEE, November 2017.

[14] Y. Zhang and Y. Xie, "Travel mode choice modeling with support vector machines." Transportation Research Record: Journal of the Transportation Research Board, Vol. 2076, pp.141-150, 2008.

[15] F, Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg and J. Vanderplas, "Scikit-learn: Machine learning in Python." Journal of machine learning research, 12(Oct), pp.2825-2830, 2011.

