Logit Regression Model to Predict Driver Left Turn Destination Lane Choice Behavior at Urban Intersections

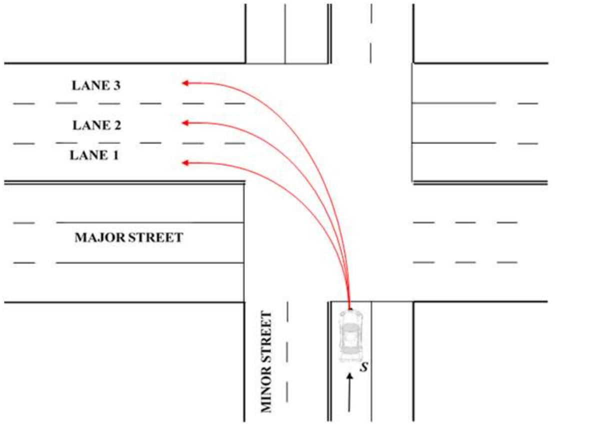
Eric Proffer[1], John Frazier[1], Matthew Vechione[2], and Okan Gurbuz[3]

[1]Undergraduate Research Assistant, The University of Texas at Tyler; [2]Assistant Professor, The University of Texas at Tyler; [3]Assistant Research Scientist, Texas A&M Transportation Institute

Objective: To improve previous research of drivers' left-turn behavior and to better predict the destination lane choice of left-turns at urban intersections

Introduction

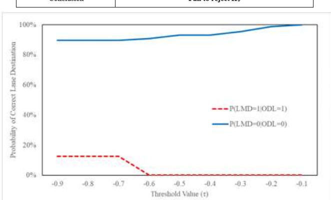
- Left turning movement is critical due to the unpredictability of driver behavior and the severity of collisions that can result due to misinterpretation of a driver's left-turn destination lane choice.
- When a driver makes a turn at an intersection, he/she has the opportunity to select a downstream destination lane. In some states in the U.S., it is required by law that drivers use a designated destination lane at intersections so as to avoid a potential collision with another concurrent turning movement.
- The destination lane choice is assumed to be chosen based on different driver behaviors, including the speed of the subject vehicle when entering the intersection, among others.
- With the increase in the number of automated vehicles that are on roads daily, an increased understanding of left-turn behaviors will allow for better prediction of a given vehicle's chosen destination lane. This increase in prediction accuracy will allow for the safer navigation of urban intersections by the automated vehicles that will continue to be introduced to the public in the near future.



Previous Work

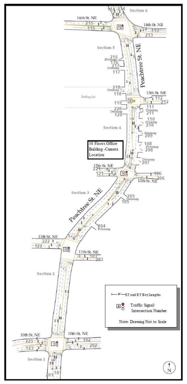
- Previous research completed by **Frazier et al. (2020)** using the same data sets found no significant difference between the afternoon and evening data sets.
- A binary logit regression model incorporating vehicle speed, vehicle type, and vehicle turn direction was performed.
- In predicting the accuracy of vehicles turning into destination lanes 2 and 3, the model performed poorly, with an **accuracy of 12.5%**.
- In predicting the accuracy of vehicles turning into destination lane 1, the model performed well, with an **accuracy of 89.7%**.

	Dataset A	Dataset B
Destination Lane		
Sample mean	0.09	0.08
Sample std. dev.	0.29	0.28
Sample size	247	228
t-value	0.052	
Conclusion	**Fail to reject H_0**	



Data Collection

- The data used for this research was from the Next Generation Simulation (NGSIM) vehicle trajectory data sets, both of which were along Peachtree St., an urban arterial, located in Atlanta, GA.
- Since previous research determined there was no significant difference between the two data sets, they were combined for the development of this model.
- All standard passenger vehicles were considered.
- Any vehicle that made a left-turn movement from a minor road to the major road (Peachtree St.) from any of the five intersections was first selected.
- For each identified subject vehicle, the **speed and acceleration of the vehicle** was recorded as the vehicle was approaching (i.e. entering) the intersection.
- At the time of the vehicle's turn, the **headway** (i.e. distance) between the subject vehicle and both the observed preceding and following vehicles was recorded, as well as the **downstream volume** of the major road section that the subject vehicle turned into.
- Subject vehicles were monitored until exiting the testing area. The vehicles' **downstream turn direction**, as well as the **downstream turn distance** was also recorded.



Cambridge Systematics, Inc. (2007)

Decision Parameter	Speed	Downstream Turn Value	Downstream Turn Distance	Acceleration	Headway to Preceeding Vehicle	Headway to Following Vehicle	Volume Downstream	Destination Lane
Unit	ft/s	1 if right, 0 otherwise	ft	ft/s^2	ft	ft	Number of Vehicles	1 for Lanes 2 or 3, 0 otherwise
Sample Size	73	73	73	73	73	73	73	73
Min	0	0	147	-10.96	34.92	0	1	0
Max	27.96	1	1913	12.27	1654.16	980.68	21	1
Mean	14.00	0.23	1523.93	-0.80	-387.05	115.84	9.05	0.29
Std. Deviation	6.95	0.43	550.26	5.66	294.80	185.05	4.86	0.46

Binary Logit Model

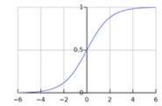
- All data was compiled into one table.
- Then, a stratified 80/20 split was made to separate the data into a training and test data set.
- A binary logit model was developed using the training data set.
- The coefficients provide insights on drivers' left turning behavior and which factors have the greatest impact on destination lane choice.

Overall Data Set	
No. of left-turning movements or subject vehicles	73
No. of subject vehicles with output of 0 (chose lane 1 as destination lane)	52
No. of subject vehicles with output of 1 (chose lane 2 or 3 as destination lane)	21
Training Data Set	
No. of left-turning movements or subject vehicles	59
No. of subject vehicles with output of 0 (chose lane 1 as destination lane)	42
No. of subject vehicles with output of 1 (chose lane 2 or 3 as destination lane)	17
Test Data Set	
No. of left-turning movements or subject vehicles	14
No. of subject vehicles with output of 0 (chose lane 1 as destination lane)	10
No. of subject vehicles with output of 1 (chose lane 2 or 3 as destination lane)	4

Decision Parameter	Coefficient	t-statistic	Std. Error	P-value
Speed	-0.066	-1.29	0.051	0.197
Downstream Turn Value	1.011	1.18	0.854	0.236
Downstream Turn Distance	0.000	0.33	0.001	0.744
Acceleration	-0.047	-0.82	0.057	0.415
Headway to Preceding Vehicle	0.000	-0.10	0.001	0.919
Headway to Following Vehicle	-0.002	-0.63	0.003	0.526
Volume Downstream	0.132	1.55	0.085	0.121
Number of Observations	59			
LR chi^2(7)	9.32			
Prob > chi^2	0.2307			
Pseudo R^2	0.1315			

Logit Model Output

- The binary logit model output shows the likelihood of choosing lanes 2 or 3. These values are transformed to fall within [0,1] using $y = \frac{1}{(1+e^{-x})}$
- These transformed values are evaluated using various thresholds, where they are assigned to be a binary value of {0,1}.

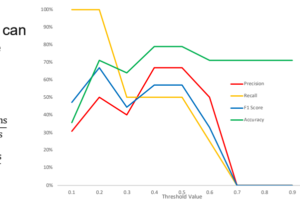


- Multiple threshold values were applied, starting with a value of 0.1 and increasing by 0.1 up to a final threshold value of 0.9. Outputs greater than or equal to a threshold become 1, while outputs less than a threshold become 0.
- The threshold value chosen must minimize the number of major errors (accidents), while also remaining accurate in predicting the turning movements of the subject vehicles (high F1 score).
- Performance at each threshold is determined using an F1 score that accounts for accuracy and severity of errors.
- The best threshold value for F1 score is **0.2**.
- As the chosen threshold increased, a decrease can be seen in precision, recall, and F1 score of the model.
- At a threshold of 0.7 or higher, the model is no longer feasible, with a precision, recall, and F1 score of 0%.

$$\text{Precision} = \frac{\text{\# of Correct Lane 2 or 3 Predictions}}{\text{\# of Total Lane 2 or 3 Predictions}}$$

$$\text{Recall} = \frac{\text{\# of Correct Lane 2 or 3 Predictions}}{\text{\# of Observed Lane 2 or 3 Totals}}$$

$$\text{F1 Score} = \frac{2 * (\text{Precision} * \text{Recall})}{\text{Precision} + \text{Recall}}$$



Threshold Value Results

Threshold = 0.2					
Observed Destination Lane		Logit Model's Decision (Model Output = 1)	Logit Model's Decision (Model Output = 0)	Total	Recall
	Lanes 2 or 3 (ODL=1)	4	0	4	100.0%
	Lane 1 (ODL=0)	4	6	10	60.0%
	Total	8	6	14	71.4%
Precision	ODL = LMD	50.0%	0.0%		Accuracy
F1 Score	66.7%				

- At each threshold, there are four possible model results:
  - Correctly output lane 1 (Observed vehicle chose lane 1)
  - Incorrectly output lane 1 (Observed vehicle chose lanes 2 and 3): Major Error: Automated vehicle may be involved in an accident
  - Correctly output lanes 2 and 3 (Observed vehicle chose lanes 2 and 3)
  - Incorrectly output lanes 2 and 3 (Observed vehicle chose lane 1): Minor Error: Automated vehicle will wait on a left-turning vehicle when it has the ability to turn right.
- A **final threshold value of 0.2** was chosen, as it provides the **best F1 score**, while also predicting **zero major errors**, meaning no accidents are caused by an incorrect model prediction.

Conclusions

- A **binary logit regression model** has been developed using new data to more accurately predict the left turn destination lane chosen based on the input decision variables. The model can **more accurately predict** the chosen left turn destination lane, with an F1 score of 67% while preventing any accidents due to an incorrect prediction.
- The **most significant decision parameters** incorporated into the binary logit regression model are **vehicle speed**, **downstream turn value**, and the **volume downstream** from a subject vehicle's turn location.
- The **safest and most accurate threshold value** for the binary logit regression model with regards to F1 score is **0.2**.