Traffic simulation and safety assessment requirements for enhancing road safety prediction tools

Maria G. Oikonomou[1[0000-0001-6837-0188]], Apostolos Ziakopoulos[1[0000-0001-6252-6743]], Shanna Lucchesi[2,3[0000-0001-7823-7241]], Monica Olyslagers[3] and George Yannis[1[0000-0002-2196-2335]]

[1] National Technical University of Athens, Department of Transportation Planning and Engineering, 5 Heroon Polytechniou str., Athens, Greece, GR-15773,
[2] Transport System Laboratory, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil
[3] International Road Assessment Programme, The Ring, Bracknell, United Kingdom, RG12 1AX.
moikonomou@mail.ntua.gr

Abstract. Improving road safety prediction tools requires assessing established traffic simulation tools and safety assessment methods. Enhancing these tools with innovative data sources and methods can significantly reduce urban crashes and their impact. To achieve this, it is imperative to identify the requirements and gaps of relevant stakeholders in terms of professional road safety analysis tools. The present study aims to utilize association rule mining to determine underlying profiles of local stakeholders who are identified as hands-on practitioners. To accomplish this objective, a dedicated survey was conducted, and the data were analyzed to discover meaningful links among stakeholder characteristics through the characteristics mined using the Apriori algorithm. The results provide a quantification of the frequency and relationships between stakeholder responses, indicating connections between education levels, work regions, experience levels, and stakeholder needs related to road safety prediction tools. The study insights offer a quantitative perspective on the interconnections and dependencies among different stakeholder attributes, shedding light on potential patterns and preferences that can guide decision-making in the context of road safety improvements.

Keywords: Traffic simulation, Road safety assessment, Proactive Road safety tools, Apriori algorithm, Association rule mining.

1 Introduction

There is an essential need for the improvement of road safety prediction tools through the assessment of established and proven traffic simulation tools and road safety assessment methods, with the aim of subsequently enhancing these tools using innovative data sources and methods.

Incorporating behavior into road safety assessment poses a significant data challenge. A survey of National Road Administrations revealed that 60% reported limited data availability to assess factors related to user behavior [1]. Advanced crash count prediction models have been developed that distinguish between engineering-related and behavioral-related crashes [2, 3]. On the topic of road design, data scarcity often hampers the reliable implementation of road geometric designs, leading to limited safety performance. Traffic microsimulation proves ideal for such investigations, being widely utilized and enabling the testing of various configurations in safe environments [4, 5]. On the other hand, it is argued [6] that there are still no suitable methodologies and reliable surrogate indicators for simulation-based safety studies due to the absence of complete models for simulating potential crashes. To date, few studies have proposed alternative methodologies [7].

It is imperative to enhance these tools with innovative data sources and novel methods to maximize road safety gains. Infrastructure improvements should carefully account for induced demand, which arises from the intricate interconnectivity of transportation systems [8]. While Machine Learning (ML) methods are gaining prominence in travel behavior research, it is essential to foster greater cohesion in parallel research methodologies [9]. Additionally, alternative approaches, such as the Star Rating protocol developed by the International Road Assessment Programme, offer a proactive means to assess road safety even in the absence of crash data [10].

This research aims to support the improvement of road safety prediction tools that will result in valuable outcomes for stakeholders, leading to a real and observable impact in reducing crashes and mitigating their consequences. Specifically, the objectives are to utilize association rule mining to determine underlying profiles of local stakeholders that are identified as hands-on practitioners.

2 Methodology

2.1 Stakeholder Questionnaire

In order to capture the current practices, needs and gaps from the perspective of transport managers and municipalities, an online stakeholder survey was designed. The survey included questions related to several new metrics, models (e.g., behavioral) and techniques integration (e.g., road safety assessment, traffic microsimulation) considering factors such as human behavior, modal shift, and improved data exploitation through machine learning methodologies.

The total number of complete stakeholder responses received was 50, while the distribution of respondents covered 36 different cities. Specifically, the majority of respondents (41 in total) work in 29 European cities. In addition, most respondents (22%) are private sector employees involved in road safety and 17% of respondents use road safety assessment methodologies in their daily activities, while 10% use macroscopic traffic simulation. It is important to note that expert respondent samples are considerably more difficult to gather compared to layman samples, and as such the present sample size is considered decent.

The answers analyzed within this study, which are described in Table 1, concerned the area where stakeholders work, their education level, their experience on their role, their needs related to traffic microsimulation as well as their needs on road safety assessment methods, their expected frequency of using an integrated analytic tool and their evaluation of how much an integrated analytic tool could impact real-crash numbers. In Table 1, the variables analyzed and their short description are included. In Appendix, descriptive statistics are included.

Table 1. List of examined variables.

Variable	Description	Values
AREA	Work area	1 (Central or Eastern Europe), 2 (Southern Europe), 3 (Western Europe), 4 (Northern Europe), 5 (Rest of the world)
EDUCATION	Education level	1 (Secondary education diploma), 2 (University/College-level degree/diploma), 3 (Master's degree), 4 (Doctoral degree/Postdoctoral studies)
EXPERIENCE	Work experience	1 (<5 years), 2 (5-10 years), 3 (10-15 years), 4 (15-20 years), 5 (>20 years)
MICROSIM_ALL_NEEDS	Importance of incorporating in traffic simulation: (i) infrastructure safety information; (ii) modal shift information; (iii) induced demand models; (iv) human behaviour models and (v) accuracy improvement	1 (Not important) - 10 (Very important)
ROAD_AS_ALL_NEEDS	Importance of incorporating in road safety assessment methods: (i) traffic microsimulation information; (ii) modal shift information; (iii) induced demand models and (iv) AI/ML models	1 (Not important) - 10 (Very important)
INTEG_TOOL_USE	Expected usage of an integrated analytic tool	1 (Never) - 10 (Very Frequently)
INTEG_TOOL_IMPACT	Expected impact of an integrated analytic tool on real crash numbers	1 (Not at all) - 10 (Extremely)

2.2 Apriori Algorithm

In order to analyze the survey data and discover meaningful links of stakeholder characteristics, association rule mining was implemented with the Apriori algorithm. This algorithm, a seminal contribution to data mining, was introduced by Agrawal & Srikant [11]. Its primary purpose is to unearth frequent item-sets within transactional datasets, paving the way for the discovery of association rules. Association rules, characterized by an antecedent part (e.g., "if X occurs") and a consequent part (e.g., "...then Y occurs as well"). The rules generated by the Apriori algorithm are assessed based on three key parameters:

- Support: This parameter measures the probability that both X and Y occur together.
- Confidence: Confidence expresses the conditional probability that Y occurs given that X has occurred. It quantifies how often Y follows X.
- Lift: Lift measures the ratio between support and confidence. A lift value of 2, for example, indicates that the likelihood of encountering X and Y together is twice the likelihood of encountering Y alone.

3 Results and Discussion

3.1 Traffic simulation user profiles

In order to discover meaningful associations of stakeholder characteristics related to their opinions on the requirements of traffic simulation, at least a minimum support and confidence value was required, to filter out rules that either concern less frequent characteristics or are less informative. After several trials during the modelling process, the selected values were 0.05 and 0.40 for support and confidence respectively. The resulting association rules with the top 5 scoring lift values are presented in Table 2. Lift values indicate the strength of association between the antecedent and consequent of each rule, and higher lift values suggest stronger associations.

Table 2. Top 5 association rules for stakeholder needs related to traffic simulation.

Rules	Antecedent		Consequent	Sup-port	Confi-dence	Cov-erage	Lift	Count
[1]	{MICROSIM_ALL_NEEDS=5}	=>	{EDUCATION=3}	0.06	0.750	0.08	1.293	3
[2]	{MICROSIM_ALL_NEEDS=8}	=>	{EDUCATION=3}	0.10	0.833	0.12	1.437	5
[3]	{AREA=3}	=>	{MICROSIM_ALL_NEEDS=10}	0.08	0.667	0.12	1.961	4
[4]	{EXPERIENCE=1}	=>	{MICROSIM_ALL_NEEDS=10}	0.08	0.500	0.16	1.471	4
[5]	{MICROSIM_ALL_NEEDS=7}	=>	{EDUCATION=2}	0.08	0.500	0.16	1.471	4

Based on the provided results, the first rule reveals that, stakeholders who state that there is a moderate need (5 out of 10) of traffic simulation tool enhancement hold a master degree title with a 75% confidence. Similarly, those who state that there is an adequate need (8 out of 10), also hold a master degree title with an 83% confidence, based on the second rule. The third rule indicates that, there is a 67% confidence that a stakeholder who works in Western Europe believes that there is a highly need of simulation tool enhancement. Based on the fourth rule, for stakeholders who have experience under 5 years in their role, there is a 50% confidence that highly need a traffic simulation enhancement. Finally, it seems that stakeholders stated that there is a 7 out of 10 need of the enhancement hold a University or College-level degree or diploma title with a 50% confidence.

3.2 Safety assessment user profiles

Similarly, to uncover significant associations among stakeholder characteristics regarding safety impact assessment methodologies, the chosen parameter values were set at 0.05 and 0.40 for support and confidence, respectively. The resulting association rules, featuring the top 5 lift values, are detailed in Table 3.

Table 3. Top 5 association rules for stakeholder needs related to road safety assessment.

Rules	Antecedent		Consequent	Sup-port	Confi-dence	Cov-erage	Lift	Count
[1]	{ROAD_AS_ALL_NEEDS=7}	=>	{AREA=2}	0.06	0.500	0.12	1.471	3
[2]	{ROAD_AS_ALL_NEEDS=7}	=>	{EDUCATION=3}	0.08	0.667	0.12	1.149	4
[3]	{AREA=3}	=>	{ROAD_AS_ALL_NEEDS=10}	0.10	0.833	0.12	2.604	5
[4]	{ROAD_AS_ALL_NEEDS=5}	=>	{EXPERIENCE=2}	0.06	0.500	0.12	2.778	3
[5]	{ROAD_AS_ALL_NEEDS=5}	=>	{EDUCATION=3}	0.08	0.667	0.12	1.149	4

The first rule reveals that, for stakeholders who state that there is a 7 out of 10 need of road safety assessment methodologies enhancement there is a 50% confidence that they work in Southern Europe. Similarly, those who identify the same need, hold a master degree title at 67% confidence level, based on the second rule. The third rule indicates that, there is an 83% confidence that a stakeholder who works in Western Europe believes that there is a highly need of road safety assessment enhancement. For stakeholders who believe that is a moderate need of this enhancement, have an experience between 5 to 10 years in their role with a 50% confidence level as well as hold a master degree title at 67% confidence level, based on the fourth and fifth rule respectively.

3.3 Integrated Analytic Tool Expected Usage

In a similar vein as the previous analyses, in order to reveal noteworthy links among stakeholder attributes concerning the expected usage of an integrated analytic tool, support and confidence parameters were established at values of 0.05 and 0.30, respectively. The ensuing association rules, which encompass the top 5 lift values, can be found in Table 4.

Table 4. Top 5 association rules for stakeholder expectation to use an integrated analytic tool.

Rules	Antecedent		Consequent	Support	Confidence	Coverage	Lift	Count
[1]	{INTEG_TOOL_USE=6}	=>	{EDUCATION=3}	0.06	0.750	0.08	1.293	3
[2]	{INTEG_TOOL_USE=9}	=>	{EXPERIENCE=4}	0.06	0.100	0.20	2.308	3
[3]	{INTEG_TOOL_USE=9}	=>	{EDUCATION=3}	0.06	0.600	0.10	1.034	3
[4]	{INTEG_TOOL_USE=10}	=>	{AREA=3}	0.06	0.500	0.12	4.167	3
[5]	{AREA=3}	=>	{INTEG_TOOL_USE=10}	0.06	0.500	0.12	4.167	3

It seems that stakeholders who claim there is a moderate possibility (6 out of 10) of using an integrated analytic tool hold a master's degree title with a 75% confidence level. On the other hand, those who state that it is very likely to use this kind of tool (9 out of 10) have between 15 to 20 years of experience in their roles and also hold a master's degree title with 10% and 60% confidence, respectively. Finally, there is a 50% confidence that a stakeholder who is certain they will use a tool like this works in Western Europe, and conversely, a stakeholder working in Western Europe will use the tool.

3.4 Integrated Analytic Tool Expected Impact on Safety

In the pursuit of uncovering significant associations among stakeholder characteristics related to the expected impact of an integrated analytic tool on safety, parameter values of 0.05 and 0.40 were opted for support and confidence, respectively. The resulting association rules, highlighting the top 5 lift values, have been elaborated upon in Table 5.

Table 5. Top 5 association rules for stakeholder expectation of an integrated tool impact.

Rules	Antecedent		Consequent	Sup-port	Confi-dence	Cov-erage	Lift	Count
[1]	{INTEG_TOOL_IMPACT=8}	=>	{EDUCATION=3}	0.06	0.500	0.12	0.862	3
[2]	{AREA=3}	=>	{INTEG_TOOL_IMPACT=7}	0.06	0.500	0.12	3.125	3
[3]	{INTEG_TOOL_IMPACT=9}	=>	{EXPERIENCE=3}	0.08	0.667	0.12	3.333	4
[4]	{EXPERIENCE=3}	=>	{INTEG_TOOL_IMPACT=9}	0.08	0.400	0.20	3.333	4
[5]	{INTEG_TOOL_IMPACT=9}	=>	{AREA=1}	0.06	0.500	0.12	1.563	3

Based on the first rule, stakeholders who believe that an integrated analytic tool will possibly impact real-crash numbers, they hold a master degree title with a 50% confidence. The second rule indicates that, there is a 50% confidence that a stakeholder who works in Western Europe believes that there is a moderate possibility of this kind of tool to impact crashes. In addition, stakeholders who state that is very likely for an integrated analytic tool to impact crash numbers, they have experience between 10 to 15 years at a 67% confidence level and they work in Central or Eastern Europe at a 50% level, based on the third and fifth rule, respectively. Based on the fourth rule, for stakeholders who have experience between 10 to 15 years in their role, that an integrated tool is very likely to affect road safety with a 40% confidence.

4 Conclusions

Overall, the association rules indicate a relationship between education levels, work regions, experience levels, and stakeholder needs related to road safety prediction tools. These findings transcend mere data correlations; they offer a profound understanding of the interplay among these diverse stakeholder attributes. Consequently, these findings are able to offer a quantitative perspective on the interconnections and dependencies among different stakeholder attributes, shedding light on potential patterns and preferences that can guide decision-making in the context of road safety improvement. In essence, this study not only provides valuable insights into the associations between stakeholder attributes but also serves as a crucial resource for crafting more effective, data-driven road safety strategies that can ultimately improve safety and make roadways safer for all.

It should be highlighted that this profiling is not a mandate to exclude stakeholders with different characteristics from using analytic tools. Rather, it is a way to: (i) gauge the most likely stakeholder audiences, (ii) tailor the developed tools to their needs but also their capacities and (iii) develop additional venues, resources and tools to reach different stakeholder profiles.

Despite the valuable insights gained from the analysis, several limitations must be acknowledged. Firstly, this study is confined to the scope of available data, potentially overlooking other influential factors in road safety. Additionally, the integrated tools will be developed within PHOEBE in the future, and as such their characteristics are not readily available for the stakeholders to interact with. While association rules unveil correlations, they do not establish causality, prompting the need for further research into causal relationships among the identified associations. To address these limitations and contribute to the advancement of road safety research, future studies can explore causal modeling techniques, longitudinal analyses, cross-cultural comparisons, and the integration of real-time data through advanced analytics. By embracing these research directions, we can enhance our understanding of road safety dynamics and pave the way for more effective, data-driven interventions, ultimately fostering safer roadways for all.

Acknowledgement

The present research was carried out within the research project “PHOEBE - Predictive Approaches for Safer Urban Environment”, which has received funding from the European Union’s Horizon Europe research and innovation programme under grant agreement No 101076963.

References

1. Yannis, G., Dragomanovits, A., Laiou, A., Richter, T., Ruhl, S., la Torre, F., Domenichini, L., Graham, D., Karathodorou, N., & Li, H.: Use of Accident Prediction Models in Road Safety Management - An International Inquiry. Transportation Research Procedia, 14, 4257–4266 (2016).
2. Afghari, A. P., Washington, S., Haque, M. M., & Li, Z.: A comprehensive joint econometric model of motor vehicle crashes arising from multiple sources of risk. Analytic methods in accident research, 18, 1-14. (2018).
3. Shaon, M. R. R., Qin, X., Afghari, A. P., Washington, S., & Haque, M. M.: Incorporating behavioural variables into crash count prediction by severity: A multivariate multiple risk source approach. Accident Analysis & Prevention, 129, 277-288. (2019).
4. Ghanim, M., Kharbeche, M., Hannun, J., Hannun, J., & Shamiyeh, K.: Safety and operational performance of signalized roundabouts: a case study in Doha. Procedia Computer Science, 170, 427-433 (2010).
5. Bahmankhah, B., Macedo, E., Fernandes, P., & Coelho, M. C.: Micro driving behaviour in different roundabout layouts: pollutant emissions, vehicular jerk, and traffic conflicts analysis. Transportation research procedia, 62, 501-508 (2022).
6. Astarita, V., Festa, D. C., Giofrè, V. P., & Guido, G.: Surrogate safety measures from traffic simulation models a comparison of different models for intersection safety evaluation. Transportation research procedia, 37, 219-226 (2019).
7. Oikonomou, M. G., Ziakopoulos, A., Chaudhry, A., Thomas, P., & Yannis, G.: From conflicts to crashes: Simulating macroscopic connected and automated driving vehicle safety. Accid. Anal. Prev. 187, 107087 (2023).
8. Cascetta, E.: Transportation Systems Analysis (Vol. 29). Springer US (2009).
9. Hillel, T., Bierlaire, M., Elshafie, M. Z. E. B., & Jin, Y.: A systematic review of machine learning classification methodologies for modelling passenger mode choice. Journal of Choice Modelling, 38, 100221 (2021).
10. Lawson, S.: Crash rate–Star Rating comparisons. Review of available evidence. Working Paper 504.2, EuroRAP, Brussels. (2011).
11. Agrawal, R., & Srikant, R.: Fast algorithms for mining association rules. In Proc. 20th int. conf. very large data bases, Vol. 1215, pp. 487-499 (1994).

Appendix

In the following Table A1, descriptive statistics of each variable used in the analysis i.e., sample size (N) and percentage (%) are included.

Table A1. Descriptive statistics.

Variable	Levels	Code	N	Percentage
AREA	Central and Eastern Europe	1	16	32%
	Southern Europe	2	17	34%
	Western Europe	3	6	12%
	Northern Europe	4	3	6%
	Rest of the world	5	8	16%
	Total	-	50	(100.0%)
EDUCATION	Secondary education diploma	1	1	2%
	University/College-level degree/diploma	2	10	20%
	Master's degree	3	29	58%
	Doctoral degree/Postdoctoral studies	4	10	20%
	Total	-	50	(100.0%)
EXPERIENCE	<5 years	1	8	16%
	5-10 years	2	9	18%
	10-15 years	3	10	20%
	15-20 years	4	8	16%
	>20 years	5	15	30%
	Total	-	50	(100.0%)
MICROSIM_ALL_NEEDS	1 ("Not important")	1	1	2%
	2	2	0	0%
	3	3	0	0%
	4	4	0	0%
	5	5	4	8%
	6	6	3	6%
	7	7	8	16%
	8	8	6	12%
	9	9	11	22%
	10 ("Very important")	10	17	34%
	Total	-	50	(100.0%)
ROAD_AS_ALL_NEEDS	1 ("Not important")	1	1	2%
	2	2	0	0%
	3	3	0	0%
	4	4	1	2%
	5	5	6	12%
	6	6	1	2%
	7	7	6	12%
	8	8	8	16%
	9	9	11	22%
	10 ("Very important")	10	16	32%
	Total	-	50	(100.0%)
INTEG_TOOL_USE	1 ("Never")	1	3	6%
	2	2	1	2%
	3	3	1	2%
	4	4	3	6%
	5	5	3	6%
	6	6	4	8%
	7	7	12	24%
	8	8	12	24%
	9	9	5	10%
	10 ("Very Frequently")	10	6	12%
	Total	-	50	(100.0%)
INTEG_TOOL_IMPACT	1 ("Not at all")	1	3	6%
	2	2	0	0%
	3	3	1	2%

4	4	1	2%
5	5	9	18%
6	6	7	14%
7	7	8	16%
8	8	6	12%
9	9	6	12%
10 ("Extremely")	10	9	18%
Total	-	50	(100.0%)