Multimodal Learning Models for Traffic Datasets

Anusha Neupane
aneupane1@aggies.ncat.edu
North Carolina A&T State University
Dept. of Civil, Archi., & Envir. Eng.

Venktesh Pandey
North Carolina A&T State University
Dept. of Civil, Archi., & Envir. Eng.
Greensboro, NC, USA

Niharika Deshpande
North Carolina A&T State University
Dept. of Comput. Data Sci. & Eng.
Greensboro, NC, USA

Hyoshin Park
hpark1@ncat.edu
North Carolina A&T State University
Dept. of Comput. Data Sci. & Eng.

ABSTRACT

Predictive routing is effective in knowledge transfer. However, it ignores information gained from probability distributions with more than one peak. We introduce traffic multimodal information learning, a new class of transportation decision-making models that can learn and transfer online information from multiple simultaneous observations of a probability distribution with multiple peaks or multiple outcome variables from one time stage to the next. Multimodal learning improves the scientific and engineering value of autonomous vehicles by determining the best routes based on the intended level of exploration, risk, and limits.

CCS CONCEPTS

• **Computing methodologies → Knowledge representation and reasoning**.

KEYWORDS

Kullback-Leibler divergence, Information gain, Sequential decision

ACM Reference Format:
Anusha Neupane, Venktesh Pandey, Niharika Deshpande, and Hyoshin Park. 2022. Multimodal Learning Models for Traffic Datasets. In *KDD '22: ACM SIGKDD Conference on Knowledge Discovery and Data Mining, August 14–18, 2022, Washington DC*. ACM, New York, NY, USA, 5 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Transportation has been crucial part of human life and it gets even salient as we grow towards technology and industries. We all have witnessed the significance of transportation in many ways and have been a part of the consequences that it creates. Many navigation tools such as Google Maps are used now days to simplify the way to the destination yet with the complexity of the multi-level roads it still gets confusing and frustrating for the users. As the traffic network system gets highly dense and complex, solving all the issues related to transportation may be difficult but it will certainly make people's lives easier, and we intend to contribute to the process by using multimodal learning in transportation. The term "Multimodal" here refers to a probability distribution with multiple modes. By determining the best routes based on the intended level of exploration, risk, and limits, multimodal learning improves the scientific and engineering value of autonomous vehicles.

In metropolitan regions, traffic congestion has a negative influence on the quality of life and economic output. It raises fuel consumption, passenger and commercial transportation costs, the number of accidents, and dangerous pollutants. According to an article based on the Texas A&M Transportation Institute's assessment [5], an average American commuter wastes 54 hours per year due to traffic delays. In cities like Los Angeles the condition is worse with the average being at 119 hours per year. The consequence of traffic congestion does not end with wasted time, it also contributes to the lost productivity and wasted fuel which can be hazardous to the environment and the economy.

In highly uncertain conditions, this reduction of entropy is vital to any optimization platform employed in the robust, efficient, autonomous exploration of the search space. To overcome [12]'s limitation in multimodal learning [3], we consider both standard deviation and entropy. We target cells with the highest importance of information distinguishing two cells with identical entropy, but different values of information. For example, the expected statistics of "speed classifications = $[0, 50, 150, 200]$" show two probability distributions "$P_A = [0.5, 0, 0, 0.5]$" and "$P_B = [0, 0.5, 0.5, 0]$" with the *same* entropy, but *different* standard deviations.

Predictive routing [1, 6] is effective in knowledge transfer, however, ignore information gained from probability distributions with *more than one peak*. Consider a network with a grid laid on top in Figure 1 (left), where each cell represents a small geographical region. To find an optimal route from an origin cell to a destination, forecasting the condition of intermediate cells is critical. Routing literature did not use a location's observed data to forecast conditions at distant non-contiguous locations' unobserved data. We aggregate the data from all the grid cells and cluster cells that have similar combinations of probability distributions. When one cell of a cluster is explored, the information gained from the explored cell can partially remove uncertainty about the conditions in distant non-contiguous unexplored cells of the same cluster. With this new framework, we explore the best options to travel with partial, sequential, and mixture of information gain.

KDD '22, August 14–18, 2022, Washington DC

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

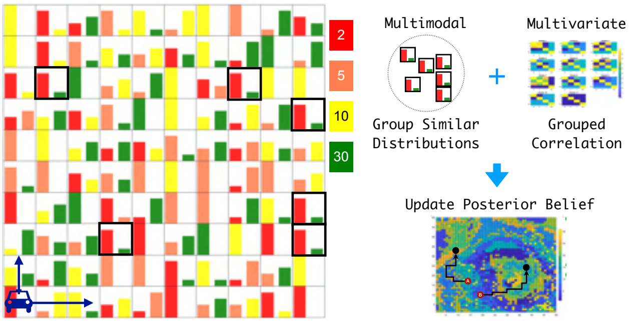

Figure 1: Each cell contains a combination of two discrete travel time distributions (i.e., 2, 5, 10, 30min) with different weights. Cells with a similar combination of distributions are clustered together based on the similarity between the combinations (e.g., 6 cells outlined in black). As users traverse the map, exploration of a cell in a cluster will remove the travel time uncertainty of other cells in the same cluster.

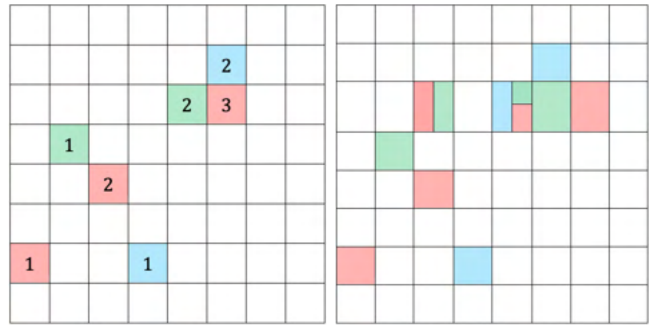

Figure 2: Various types of sequential information gain can be investigated starting from the simplest model of sequential single-type information gain in each cell (left), a sequence of multiple independent univariate observations in each cell with partial information gain for each stage of a process (right) and multi-variate probability distributions are considered in diverse environments.

Our main contribution is the development of a new family of online predictive decision making models, Temporal Multimodal Multivariate Learning (TMML), that can indirectly learn and transfer online information from multiple modes of probability distributions and multiple variables between different time stages, which can be applied to many routing problems under uncertainty such as Mars exploration [10], Hurricane sensing [6], and urban routing [11]. Preliminary remedy [10] partially filled this gap by grouping similar types of locations based on their classified output (e.g., sandy or rocky), used in optimizing vehicle routing to improve the prediction uncertainty proven to be superior to partially observable Markov decision processes.

This article is focused on a part of this research project where we focus on the process of multimodal learning and clustering for urban routing and traffic applications. In this article, we use multimodal learning to learn the spatio-temportal correlation between non-contiguous geographical locations. Once a learning framework is developed, it will then be used to design route recommendations for travelers accounting for uncertainty in travel time across a given region. The congestion on the roads of the area of research is studied using a real time data of the road in a time frame. Traffic Message Channels (TMC) can be used to determine the travel time of a vehicle and also to inform the driver about the traffic congestion on the road towards their destination.

The rest of article is organized as follows: Section 2 we will review some of the literature related to the topic of multimodal learning. Section 3 discusses the details of the multimodal learning using traffic data. Section 4 presents some analysis results and Section 5 concludes the paper.

2 LITERATURE REVIEW

Recent advances in approximating multimodal output distribution [2] have well handled prediction uncertainties rather than averaging the distribution. While those advanced multimodal learning models helped the prescriptive analytics make proactive decisions through accurate prediction of future events, sequential learning of those approximated information has depended on unimodal or bimodal probability distribution. In a sequence of information learning and transfer decisions, the traditional reinforcement learning (RL) cannot accommodate the noise in the data that could be useful for gaining information from other locations, thus cannot handle multimodal and multivariate gains in their reward transition function. Still, there is a lack of interest in learning and transferring multimodal information effectively to maximally remove the uncertainty. In this study, a new information theory overcomes the traditional entropy approach by actively sensing and learning information in a sequence. Particularly, the autonomous navigation of a team of heterogeneous unmanned ground and aerial vehicle systems in Mars and Hurricane outperforms RL through indirect learning.

We start extending standard deviation-based information theory [3] to ensure that locations with broad bimodal probability distributions are targeted over locations with narrow probability distributions. "Correlated cells" are defined as cells with a similar travel time probability distribution. The states of correlated cells are probabilistic until one of them is visited and the true state is observed. If the assumption that the cell states are correlated is true, then visiting one cell will improve the state estimate of all cells that share similar travel time probability distribution. Instead of modeling the entropy of a binary state in which cells are obstructed or clear, this new information flow framework integrates different sources of partial entropy toward spatiotemporal information gain of a partially correlated grid, nearby grid with a sight radius, and measure of variation in the correlation. Using a two-agent model, each agent performs its own exploration and develops its own route plans using separate utility functions. While one agent focuses on exploration, the other focuses on exploitation of the information gained by the exploring agent.

In the the proposed multi-agent entropy-based path planning, information is shared with other agents, influencing their route choices. The path is planned in advance and updated as information about the grid is discovered. As agents discover the state of the grid, that information is conveyed to the other agents. Each agent updates its path plan every time it moves to a new grid cell. By sharing information about the state of the grid cells, each agent helps to define the optimal parameters to be used in the other agent's utility functions. If an identical cell is visited by another agent and found to be in the same state as the original cell of that type, then all agents have confirmation that the assumption that these cells are correlated is more likely to be true.

3 MULTIMODAL TRAFFIC LEARNING

Road closures, accidents, and poor weather can all add time to our daily commute. These delays, which can surpass users' intended commute time, can result in missed meetings, canceled appointments, and more, adding up to a significant amount of money. A multi-modal urban transportation network gives travelers a variety of options for getting around. Assume we know that a highway connection A normally takes two minutes to travel without traffic, but it could take eight minutes owing to an unforeseen event (e.g., incidents). If the bimodal trip distributions for both links are similar, we can group A and A' together in the same correlated group. The literature overlooks three advantages of deploying a platoon of vehicles to A rather than B, as depicted in Figure 4 at the bottom: 1) We can update the estimated travel time on this link A so other drivers can modify either their departure time or route to utilize this 2-minute shortcut, in the case of a scenario that turned out to be 2-minutes due to the quick clearance of the event. 2) We can update travel time on other links with similar probability distributions (e.g., A'). We can send extra vehicles to this route and relieve other route congestion that turned out to be 8 minutes due to the extended clearance time of the incident if we know the overall travel time of the route is 4-minutes. 3) We update travel time on other links with the same sort of probability distributions (e.g., A'). By knowing that the total travel time of a route AA' is 16-minutes, we can notify fewer vehicles to use this route, and redistribute traffic to other routes (i.e., BC) having shorter travel times. While the existing routing literature only considers close links, no research has been done on the realization of multimodal travel time distributions based on real-world data.

Grouping similar probability distributions. Clustering identifies similar probability distributions (e.g., terrain type, speed) based on the output of the classifier Figure 2. The global correlation between non-contiguous cells of an entire map are estimated by using an expectation maximization algorithm on multimodal mixture distributions.

$$P(X|\alpha,\beta) = \prod_{i=1}^{N}\prod_{k=1}^{K}\alpha_k P(x_i|\beta_k) \tag{1}$$

where X is the data, α and β are Dirichlet parameters, $P(x_i|\beta_k)$ is the multinomial density, $i \in N$ are the observations, and $k \in K$ are clusters. Using Expectation Maximization, the optimal distribution of the data over K clusters can be determined by maximizing the lower bound of the log of the likelihood in Eq. 1. The optimal cluster number can then be determined by minimizing the Bayesian Information Criterion $BIC = Dln(N) - 2ln(\widehat{L})$, where D is the number of parameters, N is the total number of observations, and $\widehat{L}$ is the likelihood of the model.

Bayesian Posterior Update. After clustering, observations are made of the environment as the agent traverses its planned path.

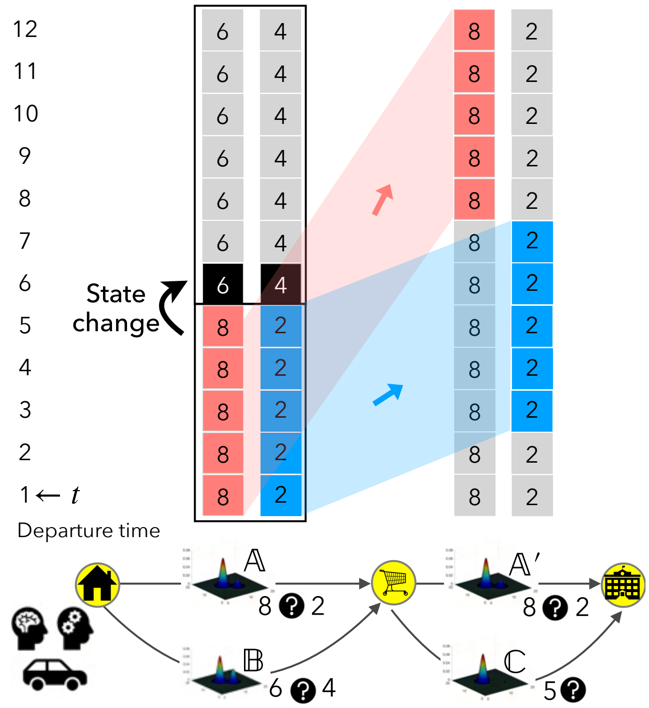

Figure 3: Temporal Multimodal Learning (TML) from correlation of time-varying bimodal distributions between links.

Variational Bayesian Inference was used to generate a posterior belief given the prior belief of terrain type distributions within each of the clusters. The posterior is approximated using the variational multinomial distribution $Q(Z)$ such as $P(Z|X) \approx Q(Z)$. over unobserved variables $z \in Z$ given data X. The dissimilarity between the approximated posterior $Q(Z)$ and the true posterior $P(Z|X)$ is then optimized by minimizing the $\mathbb{KL}$ [4] divergence.

$$D_{KL}(P||Q) = \sum_{x\in\mathcal{X}} P(x) log\left(\frac{P(x)}{Q(x)}\right) \tag{2}$$

We focus on speed and travel time information from the data collected from the Regional Integrated Transportation Information System (RITIS) [11]. RITIS combines data from a variety of agencies, systems, and even the commercial sector to enable better event reaction and planning decisions. A wide range of analytical tools and features are available in RITIS. Transportation officials, planners, researchers, and others use the RITIS as a data archiving and analytics platform. We used RITIS to collect real time data and information on Traffic Message Channel (TMC) of the area of our research.

There are 39 TMCs in the region of interest and we chose TMCs that are in the same direction. We have RITIS data for all 39 TMCs from January 1, 2021, to September 30, 2021. The data is averaged over a 10-minute interval, which divides a day's 24 hours into 10 minutes. We have built a Python program that selects speed information for a time interval of our choosing, such as 9:30 a.m. to 9:40 a.m. for a TMC of our choosing. We grouped the speed data into speed bins once the information was gathered. The speed bins range from 2 to 100 mph, with a 2mph spacing between them, bins range

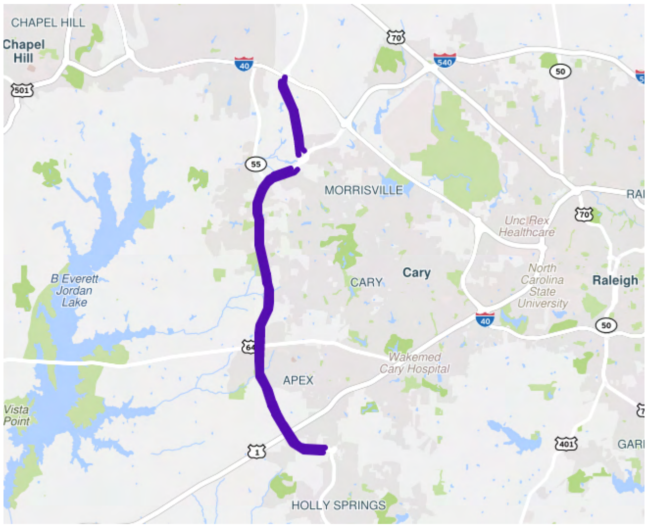

Figure 4: Data cleansing and downloading from the RITIS interface for the network of Triangle Expressway.

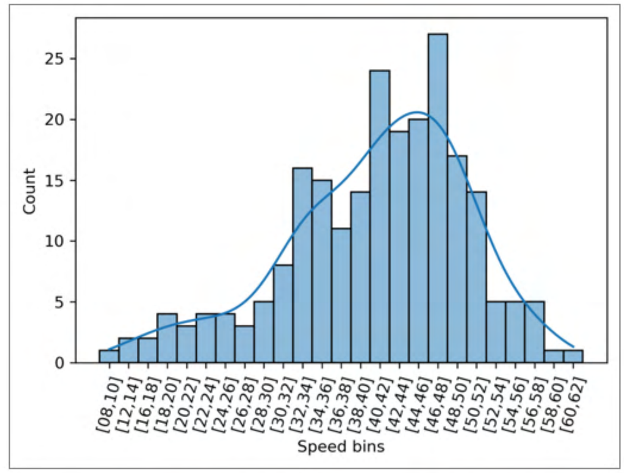

Figure 5: Histogram of the counts and the speed for the TMC within the given time interval after categorizing the data into these bins

from [2, 4], [4, 6], to [98, 100] and there are 49 speed bins in total. We plotted the histogram of the counts and got the distribution of speed for that TMC within the given time interval after categorizing the data into these bins. The goal of identifying this distribution was to create clusters of distributions that are comparable using the clustering algorithm. For a given time interval, the TMCs that belong to one cluster are spatially correlated. As a result, if we have speed information from one TMC, we can update speed information for other TMCs within that cluster, lowering uncertainty.

The distribution is then run through a MATLAB algorithm to determine whether it is multimodal or not. The term “multimodal distribution" refers to a distribution with two or more local maxima. If our distribution is multimodal, that means there is speed variance owing to traffic congestion. The following graph depicts the distribution of one TMC from 9:30 to 9:40 a.m. The following data is statistically multimodal data. However, in order to exhibit multimodality in the next image, we must manually draw the density function. In Python, we used the kernel density estimate ("kde=true") functionality to display the multimodality of distribution.

The map is segmented into regions of comparable cell types using k-means clustering. The entropy and expected value of each cell are used to build clusters. A Gap function was used to determine the best number of clusters for the data set. When entropy is given as a percentage rather than a decimal value, k-means clustering works better due to scale differences. Para

4 RESULTS

We had 39 TMC for our network and we divided them into 12 groups according to their characteristics at different times as shown in the figure 4. The correlation between the TMCs were checked or grouped using k-means clustering. K-means Clustering is a popular exploratory data analysis tool for gaining an understanding of the data’s structure. It is the task of identifying subgroups in data so that data points within the same subgroup (cluster) are extremely similar while data points within different clusters are very dissimilar. It attempts to make intra-cluster datasets as identical as possible to keep less variation and make the data points homogeneous.

Figure 6: 39 TMCs separated into 10 groups according to their characteristics. Each color represent a group.

We performed K-means and found TMCs at certain times to be correlated and assigned to the same cluster. The groups were separated to identify with different colors. We made a visual representation of the connection and similarity of the TMCs which belonged to identical groups in different time zone comparing it with TMC 12 (figure 7). The representation shows the TMCs are correlated.

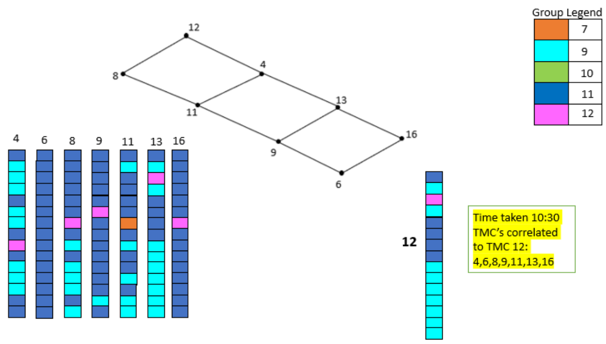

Figure 7: Connection of TMCs correlating to TMC 12

The major objective is to decrease uncertainty in speed prediction by utilizing the information gained through spatiotemporal correlation between TMCs. Figure 8 illustrates the significant percent decrease in prediction uncertainty that occurs. The reduction in uncertainty is greater when speed readings using TML are in close proximity to historical observations.

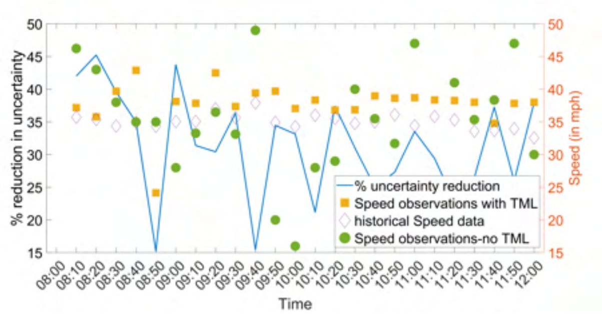

Figure 8: Spatio-temporal learning

5 CONCLUSION

Traditional RL has focused on direct learning unimodal univariate statistics. In this paper, the sequential decisions are made by combining the benefits of direct learning and the additional benefits from indirect learning through the covariance structure. The multimodal indirect information learned from one observed geographical location is transferred to other similar type locations.

The above multimodal learning example depends on the univariate travel speed outcome variable. Traditional machine learning frameworks overlook simultaneous observations of more than one outcome variable in different locations and times without lowering the prediction errors. The dynamic impact area of a prior event could predict the probability of posterior events [8, 9]. However, when frequent minor events are occurring in a sequence, due to high uncertainty, the literature could not reliably predict the dynamic spatiotemporal evolution of a mutual relationship between events [7]. Machine learning with rule extraction [10] partially alleviates Black box issues, but without an effort to reduce uncertainty by observing a ground truth, the routing solutions are still unreliable and intractable.

If there is a strong correlation structure found from simultaneous observations with more than one outcome variable, learning wind speed uncertainty can also remove the additional uncertainty of other outcome variables. The future study will address the gaps in current information provision systems by providing multimodal multivariate informatics to any user and system with an optimized policy. The assumption of exact information sharing between multiagents will be relaxed to sequential and partial information gain with multimodal distribution, then further extended to multivariate information gain. The developed algorithms are evaluated in simulated scenarios.

6 ACKNOWLEDGMENTS

Part of the research work was carried out at the NASA Jet Propulsion Laboratory (JPL), California Institute of Technology, under a contract with the National Aeronautics and Space Administration (NASA) [80NM0018D0004]. Funding for this research was provided by NSF [1910397, 2106989] and NCDOT [TCE2020-01]

REFERENCES

[1] Rutger Claes, Tom Holvoet, and Danny Weyns. 2011. A Decentralized Approach for Anticipatory Vehicle Routing Using Delegate Multiagent Systems. *IEEE Transactions on Intelligent Transportation Systems* 12, 2 (2011), 364–373.

[2] Federico Errica, Davide Bacciu, and Alessio Micheli. 2021. Graph Mixture Density Networks. In *Proceedings of the 38th International Conference on Machine Learning (ICML 2021)*. 3025–3035.

[3] Larkin Folsom, Masahiro Ono, Kyohei Otsu, and Hyoshin Park. 2021. Scalable Information-Theoretic Path Planning for a Rover-Helicopter Team in Uncertain Environments, 18(2): 1-16. *International Journal of Advanced Robotic Systems* (2021).

[4] Solomon Kullback and Richard A. Leibler. 1951. On Information and Sufficiency. *The Annals of Mathematical Statistics* 22, 1 (1951), 79–86.

[5] Jennifer Liu. 2019. Commuters in this city spend 119 hours a year stuck in traffic. https://www.cnbc.com/2019/09/04/commuters-in-this-city-spend-119-hours-a-year-stuck-in-traffic.html. (2019).

[6] Niharika Mahajan, Andreas Hegyi, Serge P. Hoogendoorn, and Bart van Arem. 2019. Design analysis of a decentralized equilibrium-routing strategy for intelligent vehicles. *Transportation Research Part C: Emerging Technologies* 103 (2019), 308–327. https://doi.org/10.1016/j.trc.2019.03.028

[7] Hyoshin Park and Ali Haghani. 2016. Real-time prediction of secondary incident occurrences using vehicle probe data. *Transportation Research Part C: Emerging Technologies* 70 (2016), 69 – 85.

[8] H. Park and A. Haghani. 2016. Stochastic Capacity Adjustment Considering Secondary Incidents. *IEEE Transactions on Intelligent Transportation Systems* 17, 10 (2016), 2843–2853.

[9] Hyoshin Park, Ali Haghani, Siby Samuel, and Michael A. Knodler. 2018. Real-time prediction and avoidance of secondary crashes under unexpected traffic congestion. *Accident Analysis & Prevention* 112 (2018), 39–49.

[10] Hyoshin Park, Ali Haghani, and Xin Zhang. 2016. Interpretation of Bayesian neural networks for predicting the duration of detected incidents. *Journal of Intelligent Transportation Systems* 20, 4 (2016), 385–400.

[11] RITIS. 2022. An Introduction to RITIS. https://ritis.org/intro. Accessed: 2022-06-15.

[12] Claude Elwood Shannon and Warren Weaver. 1949. *The Mathematical Theory of Communication*. Urbana, The University of Illinois Press. 1–117 pages.