Vision Paper: Causal Inference for Interpretable and Robust Machine Learning in Mobility Analysis

Conference Paper

Author(s):
Xin, Yanan (iD); Tagasovska, Natasa; Perez-Cruz, Fernando; Raubal, Martin (iD)

Vision Paper: Causal Inference for Interpretable and Robust Machine Learning in Mobility Analysis

Yanan Xin
ETH Zürich, Institute of Cartography and Geoinformation
Zürich, Switzerland
yanxin@ethz.ch

Natasa Tagasovska
Prescient Design, Genentech | Roche
Basel, Switzerland
natasa.tagasovska@roche.com

Fernando Perez-Cruz
Swiss Data Science Center
Zürich, Switzerland
fernando.perezcruz@sdsc.ethz.ch

Martin Raubal
ETH Zürich, Institute of Cartography and Geoinformation
Zürich, Switzerland
mraubal@ethz.ch

ABSTRACT

Artificial intelligence (AI) is revolutionizing many areas of our lives, leading a new era of technological advancement. Particularly, the transportation sector would benefit from the progress in AI and advance the development of intelligent transportation systems. Building intelligent transportation systems requires an intricate combination of artificial intelligence and mobility analysis. The past few years have seen rapid development in transportation applications using advanced deep neural networks. However, such deep neural networks are difficult to interpret and lack robustness, which slows the deployment of these AI-powered algorithms in practice. To improve their usability, increasing research efforts have been devoted to developing interpretable and robust machine learning methods, among which the causal inference approach recently gained traction as it provides interpretable and actionable information. Moreover, most of these methods are developed for image or sequential data which do not satisfy specific requirements of mobility data analysis. This vision paper emphasizes research challenges in deep learning-based mobility analysis that require interpretability and robustness, summarizes recent developments in using causal inference for improving the interpretability and robustness of machine learning methods, and highlights opportunities in developing causally-enabled machine learning models tailored for mobility analysis. This research direction will make AI in the transportation sector more interpretable and reliable, thus contributing to safer, more efficient, and more sustainable future transportation systems.

CCS CONCEPTS

• **Computing methodologies → Neural networks**; **Causal reasoning and diagnostics**; • **Information systems → Spatial-temporal systems**.

KEYWORDS

causal inference, mobility analysis, interpretability, robustness, machine learning

1 ADVANCEMENTS AND CHALLENGES IN DEEP LEARNING-BASED MOBILITY ANALYSIS

Urban mobility has been rapidly increasing in volume as the progress of urbanization accelerates. Despite the advantages and opportunities mobility has brought to our society, there are also severe drawbacks such as the transport sector's role as one of the main contributors to greenhouse-gas emissions. These drawbacks have posed great challenges to achieving several of the Sustainable Development Goals as formulated by the United Nations Development Programme (UNDP, 2015), including good health and well-being, sustainable cities and communities, and climate action [31].

Mobility data analysis plays an integral part in addressing these issues, supported by the research on developing innovative spatial and computational methods [31]. Recent research on computational methods for mobility analysis has given increasing attention to black-box deep learning methods, because of their superior predictive power compared to traditional methods in many mobility-related applications. Some of these applications include but are not limited to traffic forecasting, travel demand prediction, traffic signal control, traffic incident inference, transport mode detection, and traffic congestion detection. Detailed discussions of these applications can be found in multiple survey papers [15, 19, 37].

Admittedly, achieving accurate predictions using these deep learning methods is sufficient in certain use cases. However, a lack of interpretability and robustness can substantially slow the deployment of these AI-powered algorithms in practice [17]. Beyond predicting mobility patterns, it is more useful to interpret these predictions so that we can understand what factors influence them and how we can apply policy interventions to achieve more desirable

outcomes, such as avoiding a traffic jam. Besides, models need to be robust against dynamic changes in the deployment environment, for example, a change in mobility patterns over time, and be able to detect such changes for model updates.

The lack of robustness undermines the reliability of a model and makes the system vulnerable to unforeseen input or adversarial attacks. Furthermore, the decision-making process of many deep learning models is unclear to both model developers and end users, making it difficult to derive actionable information that could be used for policymaking and to keep the public informed of the rationale behind the decisions. The need to improve the interpretability and robustness of machine learning methods is also manifested by the growing number of governmental regulations that demand an explanation for machine-derived decisions [7].

Research on interpretable and robust machine learning has gained increasing attention in the last few years, proliferating multiple methodologies [7, 22]. Among these methods, the intersection of causal inference and machine learning has recently gained traction and advanced significantly. According to Judea Pearl's three-level causal hierarchy, i.e., association, intervention, and counterfactual, machine learning methods are limited to the association level of the causal hierarchy [27], since they rely purely on associative relationships extracted from observational data. To address the issues of model explainability and robustness, higher levels of causal reasoning (i.e., intervention and counterfactual) are indispensable. Causal relationships are by definition interpretable and invariant. Causal links offer explanations of the model and/or its predictions; and underlying causal structure holds true across different problem setups and environments, thus can be leveraged to improve the transferability and robustness of machine learning models.

Although many interpretable and robust machine learning methods have been developed, most of these are tailored for image data and sequential data [7, 34], which cannot address certain challenges of mobility data analysis. The main challenge of applying known causal modeling approaches to mobility data or in general spatial data is that the spatiotemporal dependency often violates the independent observations assumption used by many causal discovery and inference approaches, such as Rubin's causal model, fast causal inference algorithm, invariant causal prediction, etc. [2, 21]. Besides, the large sample size, dimensionality (space, time, attributes), and dynamic properties of mobility data also pose great challenges in inferring causal effects. Some efforts have been devoted to developing spatially-explicit causal inference approaches, such as the spatial difference-in-differences approach [10] and the current fast causal inference method [21]. Although those causal methods are not designed for machine learning models, they provide good starting points in developing causally-enabled machine learning methods for spatiotemporal data. In addition, a few recent studies used causality-inspired machine learning methods for agent-based motion forecasting, and showed that utilizing causality holds great promises in improving generalization and interpretability [18, 20]. Overall, leveraging and adapting causal models for mobility data analysis, especially in the combination of complex machine learning models, is still in its very infancy.

2 CAUSAL INFERENCE FOR INTERPRETABLE AND ROBUST MACHINE LEARNING

2.1 Causality for Interpretability

To further improve the applicability of data-driven models across industries and domain sciences, the machine learning community shows an increasing effort in improving model robustness, reliability, and interpretability. This tendency triggered a line of research that discovered pitfalls in terms of inability to recognise an out-of-distribution data example due to the lack of uncertainty quantification [16, 36], as well as inability to handle and generalize to distributional shifts [12]. Hence, these motivated a significant interest in how humans and machines perceive/understand the world and how methods derive their predictions, namely, aiming for causal reasoning in machine learning. Incorporating causal knowledge into data-driven approaches has been of interest since the 1980s, when Judea Pearl [25] pioneered the fusion of graphical models and structural equations, providing for the first time a mathematical framework for causal discovery and inference [26]. This introduced a new subfield in machine learning with a focus on learning graphical models from observational and interventional data [29, 35]. However, such methods have recently hit a barrier due to overly intense computational operations (exponentially increasing subsets of variables to be evaluated as potential causal models), as well as a lack of ground truth reference data in real-life scenarios. This follows from the fact that a trustworthy causal model should be validated by one, if not multiple domain experts [23]. Nevertheless, the causal framework has continued its quest of improving machine learning models, taking a different turn, by leveraging different interventional distributions and using the concept of counterfactuals for interpretability. However, since causal inference is very ambitious and controversial (due to subjective views), these techniques are thought as "geared towards causality" but not necessarily able to infer the underlying true causal effects [6]. Still, they do something more intelligent aiming towards causality than an analysis based on a standard statistical model and often have more valuable interpretation than standard machine learning methods. As such they have an important role in the crucial quest for interpretable machine learning.

2.2 Causality for Robustness

A widely shared view of robustness in modelling is perceived through invariance or stability [6, 28]. Causal relationships are by definition invariant, meaning they hold true across different circumstances and domains. This is a desirable property for machine learning systems, where we often predict on data that we have not seen in training; we need these systems to be adaptable and robust. Peters et al. [28] have proposed to exploit the invariance of causal models for machine learning inference. The key underlying idea is to gather all sub-models that are statistically invariant across environments (e.g., different geographical locations or time periods, etc.) in a suitable sense. The causal sub-model, with a direct causal effect on the target variable, will be one of these invariant sub-models. As such, Peters et al. [28] introduced a new approach to make use of causal inference in methods that exploit the invariance principle to propose: an optimization framework that under mild assumptions

guarantees generalization [4], improved reinforcement learning frameworks [1], as well as leveraging it for domain adaptation and transfer learning [33].

3 OPPORTUNITIES FOR CAUSALLY-ENABLED MACHINE LEARNING METHODS IN MOBILITY ANALYSIS

In this section, we emphasize key challenges of deep learning-based mobility analysis that require interpretability and robustness, and highlight opportunities for causally-enabled machine learning approaches to address those challenges.

Opportunity 1: Identify influences of mobility data representation on predictions.

A critical challenge of deep learning-based mobility analysis has been the lack of understanding of how data representation influences the predictions. Mobility data can be represented in different formats, spatial, and temporal scales due to the many data collection methods and various needs of transportation applications. The representation of mobility data has been shown to have a key impact on the analytical results, such as the predictability of individual mobility [30], characterization of aggregate mobility behavior [24], transport mode detection [5], and traffic forecasting [11].

The impact of such data representation on model predictions has been examined for mobility analysis using more traditional machine learning models [5], but little is known about mobility analysis using deep learning models. Causal intervention can offer insights into the sensitivity of deep learning models towards different data representations through carefully-designed interventions. Based on mobility rules, we can design causal graphical models and simulate distributional changes per single or multiple variables [28]. These changes encode the representational change and allow us to systematically evaluate the influence of these changes on the predictions of deep neural networks. These understandings are critical in developing benchmark mobility data sets and evaluation metrics to compare different deep learning methods used for mobility analysis. Moreover, the understanding of data representation also provides crucial data specifications that improve the robustness and interoperability of a model.

Opportunity 2: Evaluate causal effects of input features and derive actionable insights.

The second critical challenge is the lack of methods for identifying the causal effects of input features used in deep learning models. Since mobility data are context-dependent, contextual features have shown to play a key role in mobility applications [3]. However, feature selection can be time-consuming and sometimes requires domain knowledge, and thus more contextual variables than needed are often used as input to machine learning models for mobility modelling. Deep learning models excel in capturing statistical correlations and complex relationships among variables, thus often producing state-of-the-art predictive performance. Yet, such prediction is still based on statistical correlations, which alone cannot identify causes of an event. This offers little actionable information for intervention to avoid an undesirable outcome, such as a traffic incident or traffic congestion [17]. Besides, due to the complex structure of deep learning models, it is difficult to evaluate

the importance of a feature, which makes it challenging to conduct feature selection.

As a promising solution, causality can be used to guide the selection of explainable and robust features [39]. Besides, counterfactual reasoning can provide explanations for black-box models retrospectively [38] and offer suggestions to actions that lead to desirable alternative outcomes. For a prediction for a certain input, we can generate multiple minimum perturbed samples of that input that cause the prediction value to change. This will give us an insight, for example in the application of traffic forecasting, into how we could have changed the on-ramp flow (e.g., adding ramp metering) such that the volume of traffic at that time point would have been reduced.

Opportunity 3: Increase robustness of deep neural networks against spatio-temporal domain shifts.

Mobility data constantly change over space and time, thus data used during test time might come from a different underlying distribution than the training data, resulting in a discrepancy between the training and testing data distributions, called a domain shift. When domain shifts happen, the model will likely fail in maintaining its performance, which calls for further updates of the model. It is critical to identify domain shifts in time to minimize the potential loss or damage caused by the poor performance of the model, and to improve the reliability of the model when transferring it across different geographic regions or time periods.

Leveraging the invariance nature of causal features, we can identify domain-invariant features to improve the robustness to spatio-temporal domain shifts. For example, Invariant Causal Prediction methods can be used to distinguish between so-called core/causal features and style/spurious ones [13, 28]. In the case of traffic forecasting, one could expect road density to be a core feature while the density of Points of Interest only projects spurious dependence on traffic volume due to its strong correlation with road density. Having sufficient and diverse data will allow us to distinguish the causal features by appearing as significant predictors across multiple environments when predicting the same target.

Opportunity 4: Increase robustness of deep neural networks against geometric change.

Raw mobility data are sequences of time-stamped geolocations depicting the trajectories of humans. These trajectories can be represented as individual linear features or road network graphs. To efficiently store, analyse, and visualize mobility data, trajectory simplification is often applied to reduce redundant information. This is particularly useful for routing and navigation applications in mobile devices that have limited storage and battery capacity [14].

Most trajectory simplification algorithms remove points or segments of trajectories based on either the importance of geometrical or semantic information [9, 32]. When trajectory data are aggregated into road network graphs, map compression can be further used to simplify the road network graph by deleting certain nodes and edges [14]. During these trajectory simplification or map compression processes, the geometry of trajectory data is inevitably distorted. These trajectory simplification methods do not consider the influence of the geometric change on the predictive accuracy of machine learning models. To ensure that machine learning models keep a desired level of predictive accuracy when different trajectory

simplification algorithms are used, domain-invariant/causal representations can be used to guide the simplification process such that the information loss of causal features is minimized. By extracting features that are invariant across different levels of simplifications, we can improve the robustness of neural networks against the geometric change of trajectory data in general, thus making neural networks robust to adversarial attacks.

4 CONCLUSION

In this paper, we envision opportunities of utilizing causal inference to enhance the interpretability and robustness of deep learning methods and address specific challenges in mobility analysis. This research direction will enhance our understanding of factors that influence the performance of machine learning models and help design benchmark data sets and more suitable criteria for evaluating deep learning models used for mobility analysis. It also holds great promises in guiding the development of learning algorithms that are robust to domain shifts. The outcomes can assist transportation planners to evaluate the risk and maturity of deploying a machine learning model in practice. In addition, the research will enrich our understanding of the type of explanations that are technically feasible and help policymakers refine AI regulations (e.g., Ethics Guidelines for Trustworthy Artificial Intelligence [8]) related to mobility applications.

ACKNOWLEDGMENTS

REFERENCES

[1] Kartik Ahuja, Karthikeyan Shanmugam, Kush Varshney, and Amit Dhurandhar. 2020. Invariant risk minimization games. In *International Conference on Machine Learning*. PMLR, 145–155.
[2] Kamal Akbari, Stephan Winter, and Martin Tomko. 2021. Spatial Causality: A Systematic Review on Spatial Causal Inference. *Geographical Analysis* (2021).
[3] Gennady Andrienko, Natalia Andrienko, and Marco Heurich. 2011. An event-based conceptual model for context-aware movement analysis. *International Journal of Geographical Information Science* 25, 9 (2011), 1347–1370.
[4] Martin Arjovsky, Léon Bottou, Ishaan Gulrajani, and David Lopez-Paz. 2019. Invariant risk minimization. *arXiv preprint arXiv:1907.02893* (2019).
[5] Oliver Burkhard, Henrik Becker, Robert Weibel, and Kay W Axhausen. 2020. On the requirements on spatial accuracy and sampling rate for transport mode detection in view of a shift to passive signalling data. *Transportation Research Part C: Emerging Technologies* 114 (2020), 99–117.
[6] Peter Bühlmann and others. 2020. Invariance, causality and robustness. *Statist. Sci.* 35, 3 (2020), 404–426. Publisher: Institute of Mathematical Statistics.
[7] Diogo V Carvalho, Eduardo M Pereira, and Jaime S Cardoso. 2019. Machine learning interpretability: A survey on methods and metrics. *Electronics* 8, 8 (2019), 832.
[8] European Commission, Content Directorate-General for Communications Networks, and Technology. 2019. *Ethics guidelines for trustworthy AI*. Publications Office. https://doi.org/doi/10.2759/346720
[9] David H Douglas and Thomas K Peucker. 1973. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica: the international journal for geographic information and geovisualization* 10, 2 (1973), 112–122.
[10] Jean Dubé, Diègo Legros, Marius Thériault, and François Des Rosiers. 2014. A spatial difference-in-differences estimator to evaluate the effect of change in public mass transit systems on house prices. *Transportation Research Part B: Methodological* 64 (2014), 24–40.
[11] Alireza Ermagun and David Levinson. 2018. Spatiotemporal traffic forecasting: review and proposed directions. *Transport Reviews* 38, 6 (2018), 786–814.
[12] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. 2016. Domain-adversarial training of neural networks. *The journal of machine learning research* 17, 1 (2016), 2096–2030. Publisher: JMLR. org.

[13] Christina Heinze-Deml, Jonas Peters, and Nicolai Meinshausen. 2018. Invariant causal prediction for nonlinear models. *Journal of Causal Inference* 6, 2 (2018). Publisher: De Gruyter.
[14] Amruta Khot, Abdeltawab Hendawi, Anderson Nascimento, Raj Katti, Ankur Teredesai, and Mohamed Ali. 2014. Road network compression techniques in spatiotemporal embedded systems: A survey. In *Proceedings of the 5th ACM SIGSPATIAL International Workshop on GeoStreaming*. 33–36.
[15] Nishant Kumar and Martin Raubal. 2021. Applications of deep learning in congestion detection, prediction and alleviation: A survey. *Transportation Research Part C: Emerging Technologies* 133 (2021), 103432.
[16] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. 2017. Simple and scalable predictive uncertainty estimation using deep ensembles. In *NeurIPS*.
[17] Ibai Laña, Javier J Sanchez-Medina, Eleni I Vlahogianni, and Javier Del Ser. 2021. From data to actions in intelligent transportation systems: A prescription of functional requirements for model actionability. *Sensors* 21, 4 (2021), 1121.
[18] Yuejiang Liu, Riccardo Cadei, Jonas Schweizer, Sherwin Bahmani, and Alexandre Alahi. 2022. Towards Robust and Adaptive Motion Forecasting: A Causal Representation Perspective. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 17081–17092.
[19] Massimiliano Luca, Gianni Barlacchi, Bruno Lepri, and Luca Pappalardo. 2021. A survey on deep learning for human mobility. *ACM Computing Surveys (CSUR)* 55, 1 (2021), 1–44.
[20] Daniel McDuff, Yale Song, Jiyoung Lee, Vibhav Vineet, Sai Vemprala, Nicholas Alexander Gyde, Hadi Salman, Shuang Ma, Kwanghoon Sohn, and Ashish Kapoor. 2022. Causalcity: Complex simulations with agency for causal discovery and reasoning. In *Conference on Causal Learning and Reasoning*. PMLR, 559–575.
[21] Konrad P Mielke, Tom Claassen, J Huijbregts, Aafke M Schipper, and Tom M Heskes. 2020. Discovering cause-effect relationships in spatial systems with a known direction based on observational data. In *International Conference on Probabilistic Graphical Models*. PMLR, 305–316.
[22] Christoph Molnar. 2020. *Interpretable machine learning*. Lulu. com.
[23] Joris M Mooij, Jonas Peters, Dominik Janzing, Jakob Zscheischler, and Bernhard Schölkopf. 2016. Distinguishing cause from effect using observational data: methods and benchmarks. *Journal of Machine Learning Research* 17 (2016), 1–102.
[24] Tuhin Paul, Kevin Stanley, Nathaniel Osgood, Scott Bell, and Nazeem Muhajarine. 2016. Scaling behavior of human mobility distributions. In *The Annual International Conference on Geographic Information Science*. Springer, 145–159.
[25] Judea Pearl. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Series in Representation and Reasoning.
[26] Judea. Pearl. 2009. *Causality*. Cambridge University Press.
[27] Judea Pearl. 2019. The seven tools of causal inference, with reflections on machine learning. *Commun. ACM* 62, 3 (2019), 54–60.
[28] Jonas Peters, Peter Bühlmann, and Nicolai Meinshausen. 2016. Causal inference by using invariant prediction: identification and confidence intervals. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* 78, 5 (2016), 947–1012.
[29] Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. 2017. *Elements of Causal Inference: Foundations and Learning Algorithms*. MIT Press (available on-line).
[30] Weicheng Qian, Kevin G Stanley, and Nathaniel D Osgood. 2013. The impact of spatial resolution and representation on human mobility predictability. In *International Symposium on Web and Wireless Geographical Information Systems*. Springer, 25–40.
[31] Martin Raubal. 2020. Spatial data science for sustainable mobility. *Journal of Spatial Information Science* 20 (2020), 109–114.
[32] Kai-Florian Richter, Falko Schmid, and Patrick Laube. 2012. Semantic trajectory compression: Representing urban movement in a nutshell. *Journal of Spatial Information Science* 4 (2012), 3–30.
[33] Mateo Rojas-Carulla, Bernhard Schölkopf, Richard Turner, and Jonas Peters. 2018. Invariant models for causal transfer learning. *The Journal of Machine Learning Research* 19, 1 (2018), 1309–1342. Publisher: JMLR. org.
[34] Thomas Rojat, Raphaël Puget, David Filliat, Javier Del Ser, Rodolphe Gelin, and Natalia Díaz-Rodríguez. 2021. Explainable artificial intelligence (xai) on timeseries data: A survey. *arXiv preprint arXiv:2104.00950* (2021).
[35] Bernhard Schölkopf. 2019. Causality for Machine Learning. *arXiv preprint arXiv:1911.10500* (2019).
[36] Natasa Tagasovska and David Lopez-Paz. 2019. Single-Model Uncertainties for Deep Learning. In *NeurIPS*.
[37] Matthew Veres and Medhat Moussa. 2019. Deep learning for intelligent transportation systems: A survey of emerging trends. *IEEE Transactions on Intelligent transportation systems* 21, 8 (2019), 3152–3168.
[38] Sandra Wachter, Brent Mittelstadt, and Chris Russell. 2018. Counterfactual explanations without opening the black box: Automated decisions and the GDPR. *Harv. JL & Tech.* 31 (2018), 841. Publisher: HeinOnline.
[39] Kui Yu, Xianjie Guo, Lin Liu, Jiuyong Li, Hao Wang, Zhaolong Ling, and Xindong Wu. 2020. Causality-based feature selection: Methods and evaluations. *ACM Computing Surveys (CSUR)* 53, 5 (2020), 1–36.