Data Mining Challenges and Opportunities to Achieve Net Zero Carbon Emissions: Focus on Electrified Vehicles

Mingzhou Yang * Bharat Jayaprakash* Matthew Eagon* Hyeonjung (Tari) Jung*
William F. Northrop* Shashi Shekhar*

Abstract

Society must achieve net zero carbon emissions to mitigate anthropogenic climate change and preserve a livable planet. Reducing transportation emissions is an important component to achieve net zero because such emissions account for a quarter of global carbon released into the environment. Driven by increasingly available transportation big data and enhanced computational speed, data mining techniques have become powerful tools to achieve transportation decarbonization. This paper describes existing gaps in transportation decarbonization research where data mining can help address problems related to medium and heavy vehicle electrification, electric micromobility safety, and analysis of alternative fuel-powered and plug-in hybrid electric vehicles. Our recommendations encompass open research problems, opportunities for data mining applications, and examples of areas where advancements in data mining techniques are needed. We encourage the data mining community to explore these challenges and opportunities to help achieve net zero emissions goals.

1 Introduction

Climate change has come to the fore as grand challenge for the global research community to address. Environmental effects attributed to climate change (e.g. storms, fires, etc.) have become increasingly frequent. Bringing anthropogenic greenhouse gas (GHG) emissions to zero [14], sometimes referred to as "net zero", is seen as a key to averting the worst impacts of global warming and climate change. According to the Paris Agreement, net zero must be achieved by 2050 to preserve a livable planet. Fig. 1 shows a pathway to net zero for the United States proposed by the Biden Administration [14]. Compared to the trends of historic emissions, the proposed targets are urgent and ambitious.

Decarbonizing transportation plays a vital role in achieving net zero since the transportation sector accounts for a quarter of global GHG emissions [12]. However, techniques commonly used in transportation science like controlled laboratory experiments and driving cycle-based certifications do not adequately address unsolved transportation decarbonization problems under actual driving conditions. Interventions like eco-routing [8, 9] are extraordinarily computationally-intensive due to the massive data volumes required and the spatiotemporal nature of the data.

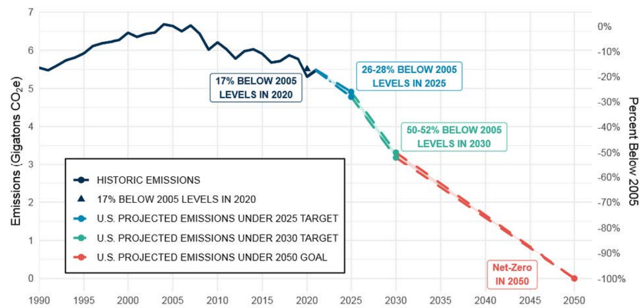

Figure 1: The pathway to net zero for the U.S. [14]

Data mining techniques, fed by emerging transportation datasets like vehicle on-board diagnostics (OBD) data, can be used to reduce transportation GHG emissions. Example applications include identifying patterns in vehicle powertrain operating conditions leading to high GHG emissions [2] and choosing the most environmental-friendly route from an origin to a destination [8, 9]. For example, Google Maps eco-routing uses a physics model and data mined from individuals to recommend the most fuel-efficient route, not just the fastest route [3]. This paper provides a vision of how innovative data mining techniques can be applied with highest impact to help achieve transportation decarbonization, focusing on electrified plug-in hybrid-electric (PHEVs) and battery-electric (BEVs) vehicles.

1.1 Emerging Datasets Newly available spatiotemporal datasets are opening up exciting opportunities in data mining research to address problems of reducing transportation emissions and achieving net zero. First, a tremendous quantity of trajectory data is collected daily by global positioning system (GPS) systems. Such data collection from vehicles, smartphones, ships, etc. has become ubiquitous. GPS data are annotated with

*University of Minnesota. {yang7492, jayap015, eagon012, jungx367, wnorthro, shekhar}@umn.edu

timestamps and corresponding location data. However, GPS datasets do not capture vehicle properties like battery state of charge or fuel use, making them less useful for determining their GHG emission impacts.

To improve the analysis of real-world driving, many modern vehicles are equipped with OBD systems. Vehicle OBD data includes physical parameters collected from the vehicle along with historical trajectories, each annotated with hundreds of attributes. Each multi-attribute trajectory records the vehicle status along a route at high frequency (e.g. 1Hz) as a sequence of spatial points, each of which is associated with a status such as location, GHG emission rate, speed, etc.

Finally, data collected from fixed services can help account for the emissions from electric vehicles (EVs). Although EVs have zero tailpipe emissions, fossil fuel power plants that generate carbon emissions still produce a significant portion of the power used to charge EVs globally. An example of a useful dataset is the electrical grid's marginal emissions rate data [1], which represents the emissions rate of electricity generators responding to changes in load on the local grid by location and time.

1.2 Vision To tackle climate change and achieve net zero emissions, we envision the use of innovative data mining techniques to enable an optimized transition to EVs as summarized in Table 1. Current work on reducing emissions of transportation mainly focuses on passenger cars [4, 10, 12]. For example, Google Maps uses a physics model of a mid-size sedan in their eco-routing application [3,8,9], and most of the EVs sold are passenger cars. However, relatively little attention has been paid to the electrification of medium- and heavy-duty vehicles (e.g. delivery trucks) and micromobility devices (e.g. electric bicycles). In addition, current data mining and machine learning techniques [12] cannot adequately address challenges raised by the spatiotemporal nature of transportation data and the physics of vehicle operation. Thus, advancements in data mining techniques like physics-aware prediction and spatiotemporal pattern mining are necessary to address these urgent problems in transportation decarbonization. In this paper, we focus on data mining opportunities in lowering the carbon footprint of transportation. In the sections that follow, we detail the challenges and opportunities around three research areas: (i) medium- and heavy-duty vehicle electrification, (ii) micromobility safety, and (iii) alternative fuels and PHEV analysis.

2 Medium and Heavy Vehicle Electrification

Medium- and heavy-duty vehicles are the second largest contributor to transportation emissions in the U.S. [7]. The large mass and high power requirement of these vehicles raise challenges for electrification since they cannot directly use public charging stations. Further, their mass varies significantly within a trip, and between trips depending on the cargo, which makes estimating their energy use more challenging than that of passenger cars, which have a largely consistent payload that is less significant compared to their curb weight.

2.1 Open Problems Eco-routing [8, 9], which aims to identify the most energy-efficient travel route between two locations, is important for optimizing EV operation. The previously cited Google Maps eco-routing application has limitations for heavy vehicles because it does not take vehicle type or mass into consideration. Energy consumption is influenced by vehicle mass [8,9], and the mass of heavy vehicles usually varies due to loading and unloading of goods or passengers. Thus, eco-routing for medium and heavy electric vehicles is still an open problem with several parts. First, vehicle energy use can be influenced by traffic patterns and congestion, so frequent patterns should be mined from historical trajectory data. Second, vehicle energy consumption estimation models should be developed for predicting the most energy-efficient paths. Identifying spatiotemporal emissions hotspots can also help reduce travel emissions by planning routes around hotspot locations [6].

While EVs have zero tailpipe emissions when operating only on electricity, energy production for battery charging may generate emissions. GHG emissions associated with EV charging are highly dependent on the energy mix of the local energy grid. Smart charge scheduling can reduce the carbon footprint of EVs by charging when the emissions from electricity production are low. If possible, vehicle-to-grid (V2G) technology can be used to return energy back to the grid from vehicle batteries when emissions are high [11].

Current work on charge scheduling ignores the impact of routing and charger placement its potential effectiveness. The placement of new charging stations could significantly impact driver behavior, EV ownership, and the power grid. Therefore, site selection becomes an important optimization problem that could be solved using data mining. Current work often uses integer programming or evolutionary algorithms to formulate and address the site selection problem [5], but these methods generally yield incomparable results.

2.2 Opportunities Advances in data mining are needed at both the theoretical and application levels to address prescient problems in transportation decarbonization.

Table 1: Research problems in net zero transportation and selected data mining techniques to address them.

Transportation decarbonization problems \ Data mining techniques		Prediction	(Frequent) Pattern mining	Clustering/ Anomaly detection	Optimization
Medium- and heavy- duty vehicle electrification	Eco-routing	Energy/range estimation	Traffic pattern	Emission hotspots, displacement	Routing methods
	Driver behaviors		Behaviors vs emissions		
	Charging station		EV ownership	Charging demand, grid capacity	Site selection
	Charging schedule	Emission estimation	Emission vs grid load/price		Optimal charging schedule
Micromobility safety				Battery fire	Ride sharing
Alternative fuels and plug-in hybrid electric vehicles analysis					Optimal carbon life cycle for vehicle mix and infrastructure

Theoretical: (i) *Physics-aware prediction methods:* Given the low adoption rate of heavy-duty electric vehicles, high-resolution OBD data for training energy estimation models for them are limited. Thus, purely data-driven machine learning prediction models have limited success in this task due to their large data requirements and have high estimation errors given the small amount of available training data [15]. Thus, research is needed to develop physics-informed prediction models that integrate physical information into machine learning models [9]. (ii) *Spatiotemporal clustering:* Emissions hotspots usually occur periodically on a route. This clustering problem is challenging because of the tremendous number of possible spatial and temporal partitions, making computational complexity very high. Also, spatiotemporal shifts can happen, which means that hotspots may not happen in exactly the same area at the same time interval. Furthermore, available statistically significant clustering and hotspot detection methods [16] on a network graph cannot identify a divergent sub-trip with unacceptable emissions if they do not represent the shortest path. (iii) *Spatiotemporal pattern mining:* Advancements in pattern mining methods are needed to identify periodical traffic patterns. Some patterns can be locally prominent but not prominent globally. In such cases, traditional pattern mining methods that focus on global frequent patterns may miss these locally prominent patterns. (iv) *Routing on a dynamic network:* The energy consumption of BEVs or PHEVs operating on batteries alone is influenced by dynamic traffic conditions and can be negative from regenerative braking. Thus, eco-routing algorithms [8, 9] should be performed in dynamic networks (e.g. a time-expanded/aggregated graph) with possible negative costs. This violates Dijkstra's assumption that the graph is static with non-negative cost.

Applications: (i) *Charging station site selection:* Standard benchmarks are needed to investigate the relative performance of charger placement techniques. Comparing these methods is challenging because of the wide variety of available data sources and objectives (e.g. costs, demand, etc.). One direction is measuring the benefits of reducing GHG emissions, either indirectly by increasing EV adoption or directly by siting chargers in areas with more renewable energy sources. Current methods generally do not consider whether existing disparities in EV ownership could be exacerbated by the site selection method [4], which could lead to growing inequalities in tailpipe emissions exposure. (ii) *Implementation of V2G:* Sparse attention has been given to chargers that incorporate V2G technology. Studies investigating the potential impact of bi-directional charging on GHG emissions and grid stability could help support efforts to improve and deploy V2G-capable charging infrastructure.

3 Micromobility Safety

Open Problems: The number of micromobility devices on the road, including electric bicycles, scooters, and motorcycles, has increased in recent years driven by advancements in battery technology [13]. These products are convenient for short-distance travel and consume less energy than passenger vehicles. However, battery fires associated with these products are a serious concern. From 2015 through early 2019, more than 330 micromobility fire incidents were reported in the U.S., resulting in more than \$9 million in property damage [13]. Many incidents are caused by unsafe home charging operations. Thus, building public charging stations and promoting micromobility device sharing may reduce battery fires.

Opportunities: Detecting anomalies in electricity burden within residential areas is a potential research direction for identifying high-risk areas for battery fire. Reporting spurious results leads to inefficient plans for preventing battery fires, thus reducing false positives should be an objective of this research. Further, the problem of site selection for micromobility charging stations can be defined as an optimization problem with costs as the objective and the safety distance between micromobility devices and electricity supply as constraints.

4 Alternative Fuels and PHEV Analysis

Open Problems: While EVs have zero tailpipe emissions, life cycle GHGs are emitted from manufacturing the battery and generating the electricity used for charging. Emissions displacement can occur because EV emissions mostly originate from power plants located in disadvantaged communities, raising equity concerns. Further, BEVs face resilience problems from power outages. Thus, charting the optimal vehicle mix, defined as identifying the optimal distribution of BEVs, PHEVs, and alternative fuel vehicles in a given geographic area, with emissions, resilience, and equity as objectives can be an open research area for net zero transportation data mining.

In the case of PHEVs and BEVs, detecting behavior patterns can provide inputs to range prediction algorithms with the aim of increasing prediction accuracy. On a broader scale, analyzing the spatiotemporal variations in driver behavior in a given region and detecting behavioral hotspots can be used to target interventions or controls aimed at reducing emissions.

Opportunities: Data mining has a key role to play in charting optimal vehicle mix for next-generation vehicles. One challenge of this problem is the spatiotemporal variance within emissions data: the location and hours of the vehicle operation can influence the level of GHGs emitted during fuel combustion or electricity production. Another challenge is raised by the integration of several spatiotemporal datasets for fuel optimization and PHEV operation. For example, understanding the emissions profile for PHEV operation will require the integration of roadside and OBD datasets.

5 Conclusions

In this paper, we envision innovative data mining techniques to achieve net zero transportation with next-generation EVs. Net zero is both important and urgent as the effects of climate change increase in frequency, scale, and severity. This vision cannot be fully realized without significant advancements in data mining techniques. We call on the data mining community to join the global effort towards net zero in transportation.

Acknowledgment

This material is based upon work supported by the National Science Foundation under Grants No. 1901099, and the USDOE Office of Energy Efficiency and Renewable Energy under FOA No. DE-FOA-0002044. We also thank Kim Koffolt and the Spatial Computing Research Group for valuable comments and refinements.

References

[1] *Marginal emissions methodology*, `https://www.watttime.org/marginal-emissions-methodology/`. Accessed: Jan. 2023.

[2] R. Y. Ali et al., *Discovering non-compliant window co-occurrence patterns*, GeoInformatica, 21 (2017), pp. 829–866.

[3] Google, *Google maps eco-friendly routing: How it works*, Nov. 2021, `https://www.gstatic.com/gumdrop/sustainability/google-maps-eco-friendly-routing.pdf`.

[4] H. T. Jung et al., *Revolutionizing electric vehicle management: Spatial computing challenges and opportunities*, in IWCTS '22, ACM, 2022.

[5] M. Kchaou-Boujelben, *Charging station location problem: A comprehensive review on models and solution approaches*, TR_C, 132 (2021), p. 103376.

[6] A. J. Kotz et al., *Lagrangian hotspots of in-use nox emissions from transit buses*, Environmental Science & Technology, 50 (2016), pp. 5750–5756.

[7] C. Ledna et al., *Decarbonizing medium- & heavy-duty on-road vehicles: Zero-emission vehicles cost analysis*, U.S. Department of Energy Office of Scientific and Technical Information, (2022).

[8] Y. Li et al., *Physics-guided energy-efficient path selection using on-board diagnostics data*, ACM/IMS Trans. Data Sci., 1 (2020).

[9] Y. Li, M. Yang, et al., *Eco-pinn: A physics-informed neural network for eco-toll estimation*, arXiv preprint arXiv:2301.05739, (2023).

[10] National Academies of Sciences Engineering and Medicine, *Assessment of Technologies for Improving Light-Duty Vehicle Fuel Economy—2025-2035*, The National Academies Press, Washington, DC, 2021.

[11] N. I. Nimalsiri et al., *A Survey of Algorithms for Distributed Charging Control of Electric Vehicles in Smart Grid*, IEEE TITS, 21 (2020), pp. 4497–4515.

[12] D. Rolnick et al., *Tackling climate change with machine learning*, ACM Comput. Surv., 55 (2022).

[13] U.S. Consumer Product Safety Commission., *Safety concerns associated with micromobility products.*, tech. report, 2020.

[14] U.S. Executive Office of the President, *The long-term strategy of the united states: Pathways to net-zero greenhouse gas emissions by 2050.*, (2021).

[15] J. Willard et al., *Integrating scientific knowledge with machine learning for engineering and environmental systems*, ACM Comput. Surv., (2022).

[16] Y. Xie et al., *Statistically-robust clustering techniques for mapping spatial hotspots: A survey*, ACM Computing Surveys (CSUR), 55 (2022), pp. 1–38.