Causal Inference and Natural Experiment Detection in Self-Driving Cars: A Dynamic Causal Graph Approach

Kweku A. Opoku-Agyemang*

October 2023

Abstract

We propose a novel framework for causal inference and natural experiment detection in self-driving cars. We model the car's environment as a dynamic causal graph, where the nodes represent variables that affect the car's performance, and the edges represent causal relationships that are updated in real-time based on the data collected by the car's sensors and cameras. We use a counterfactual engine to generate hypothetical scenarios that could have happened if the car had taken a different action or faced a different situation, and compare them with the actual outcomes. We use a natural experiment detector to identify situations where the car is exposed to a natural or quasi-experimental variation that affects one or more variables in the causal graph, and use them to estimate the causal effects of interest. We show that our framework can learn new causal relationships, test and refine causal hypotheses and assumptions, and evaluate and optimize the car's decision-making and performance. We also provide feedback and explanations to the human driver or passengers if needed. We derive regret bounds for our framework that depend on the number of natural experiments encountered, the quality of the counterfactual engine, and the complexity of the causal graph. We also discuss regret bounds for learning and updating the causal graph from data and feedback using online learning algorithms; for identifying situations where natural experiments occur using anomaly detection methods that can detect changes and outliers in data; and for evaluating and optimizing the quality and validity of natural experiments using evaluation methods that can account for selection bias, endogeneity, spillover effects, compliance issues, and measurement error.

*Chief Scientist, Machine Learning X Doing and Honorary Fellow, International Growth Centre. Email: kweku@machinelearningxdoing.com. I thank several people at the Berkeley Expert Systems and Technologies Lab, the UC Berkeley Department of Mechanical Engineering, the Berkeley Institute for Data Science, the Berkeley Institute for Transparency in Social Science, Cornell Tech and others for encouragement. The author is solely responsible for this article and its implications, and the perspectives therein should not be ascribed to any other person or any organization.

Contents

1 Introduction

The machine learning revolution has expanded across the economy with a variety of methodologies [1-8], with a growing intersection with methods traditionally associated with domain expertise, such as economics [9-18]. Although a relatively new innovation and phenomenon, self-driving cars are becoming more prevalent and sophisticated, as they aim to provide safer, more efficient, and more comfortable transportation for humans. McKinsey believe that autonomous driving could generate as much as between 300 million and 400 million US dollars by the year 2035 [19]. However, self-driving cars also face many challenges and uncertainties in their complex and dynamic environments, such as traffic, weather, road conditions, pedestrians, and other vehicles. To cope with these challenges and uncertainties, self-driving cars need to collect and analyze data in real-time, using deep learning techniques to make predictions and decisions. However, deep learning techniques are often limited by their lack of interpretability, causality, and generalizability. For example, deep learning techniques may not be able to explain why a certain action or outcome occurred, or how it would change under different circumstances. Moreover, deep learning techniques may not be able to transfer their learned knowledge to new or unseen situations, or to account for the potential confounding factors that may affect their performance.

In this paper, we propose a novel framework for causal inference and natural experiment detection in self-driving cars to address the above limitations. Our framework builds on the idea of counterfactual reasoning, which is the ability to imagine what could have happened if things had been different. Counterfactual reasoning is a powerful tool for causal inference, as it allows us to estimate the causal effects of interventions or treatments by comparing the actual outcomes with the hypothetical outcomes that would have occurred under different conditions. Counterfactual reasoning is also a useful tool for natural experiment detection, as it allows us to identify situations where nature or chance creates a random or quasi-random variation that affects one or more variables of interest, and use them as natural experiments to estimate the causal effects of those variables.

Our framework consists of three main components: a causal graph, a counterfactual engine, and a natural experiment detector. The causal graph is a graphical representation of the causal relationships between different variables that affect the car's performance, such as speed, traffic, weather, road conditions, etc. The causal graph is updated in real-time based on the data collected by the car's sensors and cameras. The counterfactual engine is a module that generates hypothetical scenarios that could have happened if the car had taken a different action or faced a different situation. The counterfactual engine uses the causal graph and the data to simulate the outcomes of these scenarios and compare them with the actual outcomes. The natural experiment detector is a module that identifies situations where the car is exposed to a natural or quasi-experimental variation that affects one or more variables in the causal graph. The natural experiment detector uses the counterfactual engine and the data to estimate the causal effects of those variables.

We show that our framework can learn new causal relationships, test and refine causal hypotheses and assumptions, and evaluate and optimize the car's decision-making and performance. We also provide feedback and explanations to the human driver or passengers if needed. We derive regret bounds for our framework that depend on the number of natural experiments encountered, the quality of the counterfactual engine, and the complexity of the causal graph. We illustrate our framework with simulations and real-world data from self-driving cars.

This paper proceeds as follows. Section 2 introduces some background and related work on causal inference, counterfactual reasoning, and natural experiments. Section 3 presents our framework for causal inference and natural experime. Further details are provided in the Appendices.

2 Background and Related Work

In this section, we introduce some background and related work on causal inference, counterfactual reasoning, and natural experiments.

Causal inference is the process of inferring the causal effects of interventions or treatments on outcomes of interest, based on observational or experimental data. Causal inference is important for many domains and applications, such as medicine, economics, social sciences, and policy making. However, causal inference is also challenging, as it requires making assumptions about the causal structure of the data-generating process, and dealing with confounding factors, selection bias, and missing data.

One of the most popular frameworks for causal inference is the potential outcomes framework, also known as the Rubin causal model (RCM) [12-15]. The RCM defines the causal effect of a treatment on an outcome as the difference between the potential outcomes under different treatment levels. For example, the causal effect of a drug on a patient's health is the difference between the patient's health if they take the drug and their health if they do not take the drug. The RCM also defines the average treatment effect (ATE) as the expected difference between the potential outcomes over a population of interest. The RCM assumes that each unit (e.g., patient) has a stable unit treatment value (SUTVA), which means that their potential outcomes do not depend on the treatment levels of other units or on other versions of the treatment. The RCM also assumes that there are no unmeasured confounders, which are variables that affect both the treatment and the outcome. Under these assumptions, the RCM can estimate the causal effects using randomized controlled trials (RCTs), where the treatment levels are randomly assigned to the units.

However, RCTs are often impractical, unethical, or impossible to conduct in many settings. In such cases, observational data may be used instead, but they may suffer from confounding bias, where the treatment levels are correlated with other variables that affect the outcome. To address this issue, various methods have been developed to adjust for confounding bias, such as matching, propensity score methods, regression methods, inverse probability weighting methods, and instrumental variable methods . These methods aim to create a balanced or comparable group of units that receive different treatment levels, such that their potential outcomes are independent of their treatment levels. However, these methods also rely on strong assumptions about the data-generating process and the availability and quality of the confounding variables.

Another popular framework for causal inference is the structural causal model (SCM) , also known as the Pearl causal model (PCM). The SCM defines the causal structure of the data-generating process using a directed acyclic graph (DAG), where the nodes represent variables and the edges represent direct causal relationships. The SCM also assigns a structural equation to each node, which specifies how its value is determined by its parents in the DAG and some exogenous noise. The SCM allows for expressing and testing various causal queries using do-calculus , which is a set of rules for manipulating conditional probabilities under interventions. The SCM also allows for deriving testable implications of causal assumptions using d-separation , which is a graphical criterion for determining conditional independence relationships in a DAG.

The SCM can handle more complex and realistic scenarios than the RCM, such as multiple treatments, multiple outcomes, mediation effects, interaction effects, feedback loops, latent variables, etc. However, the SCM also requires specifying a complete and correct DAG that represents the true causal structure of the data-generating process. This may be difficult or impossible in many settings, as there may be uncertainty or disagreement about the causal assumptions or there may be insufficient data or knowledge to support them.

Counterfactual reasoning is a form of reasoning that involves imagining what could have happened if things had been different. Counterfactual reasoning is closely related to causal inference, as it allows us to estimate the causal effects of interventions or treatments by comparing the actual outcomes with the hypothetical outcomes that would have occurred under different conditions. Counterfactual reasoning is also useful for natural experiment detection, as it allows us to identify situations where nature or chance creates a random or quasi-random variation that affects one or more variables of interest, and use them as natural experiments to estimate the causal effects of those variables.

Counterfactual reasoning can be formalized using both the RCM and the SCM frameworks. In the RCM framework, counterfactuals are defined as the potential outcomes under different treatment levels. For example, the counterfactual "What if I had taken the drug?" corresponds to the potential outcome under the treatment level "take the drug". In the SCM framework, counterfactuals are defined as the solutions to the modified structural equations under interventions. For example, the counterfactual "What if I had taken the drug?" corresponds to the solution to the structural equation for the outcome variable, where the treatment variable is set to "take the drug" and the other variables are set to their original values.

Counterfactual reasoning has been widely studied and applied in various domains and disciplines, such as philosophy, psychology, linguistics, logic, artificial intelligence, and machine learning . Counterfactual reasoning has also been used for various tasks and applications, such as explanation, attribution, responsibility, blame, regret, planning, learning, prediction, and decision making . However, counterfactual reasoning also faces many challenges and limitations, such as computational complexity, data scarcity, causal ambiguity, counterfactual paradoxes, and human biases .

Natural experiments are situations where nature or chance creates a random or quasi-random variation that affects one or more variables of interest, and thus provides an opportunity for causal inference. Natural experiments are similar to RCTs, but they are not designed or controlled by researchers or experimenters. Natural experiments are often considered as a gold standard for causal inference, as they can overcome the confounding bias and the ethical and practical issues of RCTs. However, natural experiments also have some drawbacks and limitations, such as rarity, unpredictability, unreplicability, heterogeneity, validity, and generalizability .

Natural experiments have been widely used and recognized in various domains and disciplines, such as economics, sociology, political science, public health, and environmental science . Natural experiments have also been used for various topics and questions, such as the effects of education, immigration, taxation, voting, war, pollution, natural disasters, etc. . However, natural experiments also require careful identification, analysis, and interpretation, as they may be subject to various threats and challenges, such as selection bias, endogeneity, spillover effects, compliance issues, measurement errors, and confounding factors .

Our work is inspired by and builds on the existing literature on causal inference, counterfactual reasoning, and natural experiments. However, our work is also novel and original in several aspects. First, we propose a framework for causal inference and natural experiment detection in self-driving cars, which is a new and challenging domain that has not been explored before. Second, we model the car's environment as a dynamic causal graph, which is a flexible and expressive representation that can capture the complex and changing causal structure of the data-generating process. Third, we use a counterfactual engine to generate hypothetical scenarios that could have happened if the car had taken a different action or faced a different situation, which is a powerful and creative tool that can simulate the outcomes of these scenarios and compare them with the actual outcomes. Fourth, we use a natural experiment detector to identify situations where the car is exposed to a natural or quasi-experimental variation that affects one or more variables in the causal graph, which is a smart and opportunistic tool that can use these situations as natural experiments to estimate the causal effects of those variables. Fifth, we derive regret bounds for our framework that depend on the number of natural experiments encountered, the quality of the counterfactual engine, and the complexity of the causal graph, which is a rigorous and theoretical result that can quantify the performance and limitations of our framework. Sixth, we illustrate our framework with simulations and real-world data from self-driving cars, which is an empirical and practical result that can demonstrate the feasibility and effectiveness of our framework.

3 A Framework for Causal Inference and Natural Experiment Detection in Self-Driving Cars

In this section, we present our framework for causal inference and natural experiment detection in self-driving cars. Our framework consists of three main components: a causal graph, a counterfactual engine, and a natural experiment detector. We describe each component in detail below.

3.1 Causal Graph

The causal graph is a graphical representation of the causal structure of the data-generating process for the car's environment. The nodes in the causal graph represent variables that affect the car's performance, such as speed, traffic, weather, road conditions, etc. The edges in the causal graph represent direct causal relationships between the variables, such as "traffic causes speed" or "weather causes road conditions". The causal graph is updated in real-time based on the data collected by the car's sensors and cameras.

The causal graph has several advantages over other representations, such as neural networks or decision trees. First, the causal graph is more interpretable and transparent, as it explicitly shows the causal assumptions and mechanisms behind the data. Second, the causal graph is more flexible and expressive, as it can capture complex and nonlinear causal relationships, such as feedback loops, mediation effects, interaction effects, etc. Third, the causal graph is more robust and generalizable, as it can account for confounding factors, missing data, measurement errors, etc.

The causal graph is constructed and updated using a combination of domain knowledge, data-driven methods, and online learning algorithms. Domain knowledge is used to provide prior information and constraints on the possible causal structure of the car's environment. Data-driven methods are used to infer the causal structure from the observed data using statistical tests, such as conditional independence tests or Granger causality tests . Online learning algorithms are used to update the causal structure based on new data and feedback using Bayesian methods, such as dynamic Bayesian networks or structural equation models .

3.2 Counterfactual Engine

The counterfactual engine is a module that generates hypothetical scenarios that could have happened if the car had taken a different action or faced a different situation. The counterfactual engine uses the causal graph and the data to simulate the outcomes of these scenarios and compare them with the actual outcomes.

The counterfactual engine has several advantages over other methods, such as reinforcement learning or imitation learning. First, the counterfactual engine is more creative and exploratory, as it can generate diverse and novel scenarios that may not have been observed or experienced by the car before. Second, the counterfactual engine is more efficient and effective, as it can generate scenarios that are relevant and informative for the car's performance and goals. Third, the counterfactual engine is more realistic and accurate, as it can generate scenarios that are consistent with the causal structure and mechanisms of the car's environment.

The counterfactual engine is implemented using a combination of generative models, simulation models, and optimization methods. Generative models are used to generate realistic and diverse scenarios that vary one or more variables in the causal graph. Simulation models are used to simulate the outcomes of these scenarios using the structural equations in the causal graph. Optimization methods are used to select and rank the scenarios based on their usefulness and importance for the car's performance and goals.

3.3 Natural Experiment Detector

The natural experiment detector is a module that identifies situations where the car is exposed to a natural or quasi-experimental variation that affects one or more variables in the causal graph. The natural experiment detector uses the counterfactual engine and the data to estimate the causal effects of those variables.

The natural experiment detector has several advantages over other methods, such as observational studies or controlled experiments. First, the natural experiment detector is more opportunistic and adaptive, as it can exploit the natural or random variations that occur in the car's environment without requiring any intervention or manipulation by the researchers or experimenters. Second, the natural experiment detector is more reliable and valid, as it can overcome the confounding bias and the ethical and practical issues of observational studies or controlled experiments. Third, the natural experiment detector is more generalizable and applicable, as it can estimate the causal effects of variables that are difficult or impossible to manipulate or measure, such as weather, road conditions, traffic, etc.

The natural experiment detector is implemented using a combination of causal inference methods, anomaly detection methods, and evaluation methods. Causal inference methods are used to estimate the causal effects of the variables that are affected by the natural or quasi-experimental variation using the counterfactual engine and the data. Anomaly detection methods are used to identify the situations where the natural or quasi-experimental variation occurs using the causal graph and the data. Evaluation methods are used to assess the quality and validity of the natural experiments using the causal graph and the data.

4 Regret Bounds for Our Framework

In this section, we derive regret bounds for our framework that depend on the number of natural experiments encountered, the quality of the counterfactual engine, and the complexity of the causal graph. Regret is a measure of the difference between the expected reward of the optimal action and the expected reward of the actual action taken by the car. Regret bounds are upper bounds on the expected regret that guarantee the performance and convergence of our framework.

We assume that the car's environment is stochastic and non-stationary, meaning that the variables in the causal graph have random and changing distributions. We also assume that the car's goal is to maximize its expected reward, which is a function of its performance variables, such as speed, fuel efficiency, safety, etc. We also assume that the car has a finite set of actions that it can choose from at each time step, such as accelerating, braking, turning, etc.

We define a natural experiment as a situation where the car is exposed to a natural or quasi-experimental variation that affects one or more variables in the causal graph, and where the car can observe the outcomes of both its actual action and a counterfactual action generated by the counterfactual engine. We denote by N the number of natural experiments encountered by the car during its lifetime. We denote by Q the quality of the counterfactual engine, which is a measure of how close the counterfactual outcomes are to the true potential outcomes. We denote by C the complexity of the causal graph, which is a measure of how many variables and edges are in the graph.

We derive regret bounds for our framework using two approaches: a frequentist approach and a Bayesian approach. The frequentist approach uses concentration inequalities and empirical risk minimization to bound the regret in terms of N, Q, and C. The Bayesian approach uses prior distributions and posterior inference to bound the regret in terms of N, Q, and C. We show that both approaches yield similar regret bounds that are sublinear in N, meaning that our framework achieves asymptotic optimality as N grows.

The main results of this section are summarized in the following theorem:

Theorem 1: Under some mild assumptions on the car's environment, the car's goal, the car's actions, the causal graph, the counterfactual engine, and the natural experiment detector, our framework satisfies the following regret bounds:

(Theorem 1A: Frequentist approach):

$$E[R_N] \leq O\left(\sqrt{\frac{C}{Q}N}\right)$$

Bayesian approach:

$$E[R_N] \leq O\left(\frac{C}{Q}\log N\right)$$

where $E[R_N]$ is the expected regret after N natural experiments, C is the complexity of the causal graph, and Q is the quality of the counterfactual engine.

The proof of Theorem 1 is given in two parts (A and B) as follows. We shall state the regret bounds for our framework under the frequentist approach and the Bayesian approach.

4.1 Notation and Definitions

We use the following notation and definitions throughout the proof:

N: the number of natural experiments encountered by the car during its lifetime.

C: the complexity of the causal graph, which is a measure of how many variables and edges are in the graph.

Q: the quality of the counterfactual engine, which is a measure of how close the counterfactual outcomes are to the true potential outcomes.

$E[R_N]$: the expected regret after N natural experiments, which is defined as

$$E[R_N] = E\left[\sum_{t=1}^{N} (r_t^*(X_t) - r_t(A_t, X_t))\right],$$

where $r_t^*(X_t)$ is the expected reward of the optimal action at time t given the state X_t, and $r_t(A_t, X_t)$ is the expected reward of the actual action A_t taken by the car at time t given the state X_t.

$\mathcal{A}$: the finite set of actions that the car can choose from at each time step, such as accelerating, braking, turning, etc.

$\mathcal{X}$: the finite set of states that the car can observe at each time step, such as speed, traffic, weather, road condition, etc.

$\mathcal{Y}$: the finite set of outcomes that the car can observe at each time step, such as fuel efficiency, safety, etc.

$P(X_t|A_{t-1}, X_{t-1})$: the transition probability of the state X_t at time t given the action A_{t-1} and the state X_{t-1} at time $t-1$.

$P(Y_t|A_t, X_t)$: the reward probability of the outcome Y_t at time t given the action A_t and the state X_t at time t.

$\pi(A_t|X_t)$: the policy of the car that specifies the probability of choosing the action A_t at time t given the state X_t at time t.

$\pi^*(A_t|X_t)$: the optimal policy of the car that specifies the probability of choosing the action A_t at time t given the state X_t at time t that maximizes the expected reward.

We first prove the regret bound for our framework under the frequentist approach.

4.2 Theorem 1A Proof: Proof of the Frequentist Approach

The frequentist approach uses concentration inequalities and empirical risk minimization to bound the regret in terms of N, Q, and C. The main idea is to use the natural experiments to estimate the causal effects of the actions on the outcomes, and to use these estimates to update the policy of the car. The concentration inequalities ensure that the estimates are close to the true causal effects with high probability, and the empirical risk minimization ensures that the policy is close to the optimal policy with high probability.

We use the following concentration inequality for our proof:

Lemma 1: (Hoeffding's inequality). Let $Z_1, Z_2, \ldots, Z_n$ be independent random variables such that $E[Z_i] = \mu_i$ and $a_i \leq Z_i \leq b_i$ for all $i = 1, 2, \ldots, n$. Then, for any $\epsilon > 0$, we have

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} Z_i - \frac{1}{n}\sum_{i=1}^{n}\mu_i\right| \geq \epsilon\right) \leq 2\exp\left(-\frac{2n^2\epsilon^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right).$$

We use the following empirical risk minimization algorithm for our proof:

Algorithm 1: (Empirical Risk Minimization)

Input: A set of natural experiments $\mathcal{D} = \{(A_i, X_i, Y_i)\}_{i=1}^{N}$, where A_i is the action, X_i is the state, and Y_i is the outcome.

Output: A policy $\pi(A|X)$ that minimizes the empirical risk

$$\hat{R}(\pi) = \frac{1}{N}\sum_{i=1}^{N}\left(r^*(X_i) - r(A_i, X_i)\right).$$

Steps:

1. Initialize $\pi(A|X)$ arbitrarily.

2. For each natural experiment (A_i, X_i, Y_i) in $\mathcal{D}$:

Estimate the causal effect of A_i on Y_i using the counterfactual engine and the data, denoted by $\hat{\tau}(A_i, X_i)$.

Update $\pi(A|X)$ using gradient descent or another optimization method, such that

$$\pi(A|X) \leftarrow \pi(A|X) - \alpha\nabla_\pi \hat{R}(\pi),$$

where $\alpha > 0$ is a learning rate parameter, and

$$\nabla_\pi \hat{R}(\pi) = -\frac{1}{N}\sum_{i=1}^{N} \hat{\tau}(A_i, X_i)\nabla_\pi \log \pi(A_i|X_i).$$

3. Return $\pi(A|X)$ as the output.

We use Algorithm 1 to update the policy of the car using the natural experiments. We then use Lemma 1 to bound the difference between the estimated causal effects and the true causal effects, and the difference between the empirical risk and the expected risk. We then use these bounds to bound the difference between the policy obtained by Algorithm 1 and the optimal policy, and the difference between the expected reward of the policy obtained by Algorithm 1 and the expected reward of the optimal policy. We then use these bounds to bound the expected regret of our framework.

The details of the proof are as follows:

Let $\hat{\tau}(A, X)$ be the estimated causal effect of action A on outcome Y given state X using the counterfactual engine and the data, and let $\tau(A, X)$ be the true causal effect of action A on outcome Y given state X. We assume that $\hat{\tau}(A, X)$ is an unbiased estimator of $\tau(A, X)$, meaning that $E[\hat{\tau}(A, X)] = \tau(A, X)$ for all $A \in \mathcal{A}$ and $X \in \mathcal{X}$. We also assume that $\hat{\tau}(A, X)$ is bounded by some constant $M > 0$, meaning that $|\hat{\tau}(A, X)| \leq M$ for all $A \in \mathcal{A}$ and $X \in \mathcal{X}$. We use Lemma 1 to bound the probability that $\hat{\tau}(A, X)$ deviates from $\tau(A, X)$ by more than some $\epsilon > 0$, as follows:

$$P\left(|\hat{\tau}(A, X) - \tau(A, X)| \geq \epsilon\right) \leq 2\exp\left(-\frac{2N\epsilon^2}{4M^2}\right),$$

where N is the number of natural experiments encountered by the car. By setting $\epsilon = \sqrt{\frac{C}{Q}N}$, where C is the complexity of the causal graph and Q is the quality of the counterfactual engine, we obtain

$$P\left(|\hat{\tau}(A, X) - \tau(A, X)| \geq \sqrt{\frac{C}{Q}N}\right) \leq 2\exp\left(-\frac{CN}{2Q}\right).$$

Let $\hat{R}(\pi)$ be the empirical risk of policy $\pi(A|X)$ using the natural experiments, and let $R(\pi)$ be the expected risk of policy $\pi(A|X)$ using the true causal effects, defined as

$$\hat{R}(\pi) = \frac{1}{N}\sum_{i=1}^{N}\left(r^*(X_i) - r(A_i, X_i)\right),$$

and

$$R(\pi) = E\left[r^*(X) - r(A, X)\right],$$

where $r^*(X)$ is the expected reward of the optimal action given the state X, and $r(A, X)$ is the expected reward of the action A given the state X. We use Lemma 1 to bound the probability that $\hat{R}(\pi)$ deviates from $R(\pi)$ by more than some $\epsilon > 0$, as follows:

$$P\left(|\hat{R}(\pi) - R(\pi)| \geq \epsilon\right) \leq 2\exp\left(-\frac{2N\epsilon^2}{4M^2}\right),$$

where $M > 0$ is a constant that bounds the difference between the expected reward of any two actions given any state. By setting $\epsilon = \sqrt{\frac{C}{Q}N}$, where C is the complexity of the causal graph, and Q is the quality of the counterfactual engine, we obtain

$$P\left(|\hat{R}(\pi) - R(\pi)| \geq \sqrt{\frac{C}{Q}N}\right) \leq 2\exp\left(-\frac{CN}{2Q}\right).$$

Let $\pi(A|X)$ be the policy obtained by Algorithm 1 using the natural experiments, and let $\pi^*(A|X)$ be the optimal policy that maximizes the expected reward. We use the empirical risk minimization property of Algorithm 1 to bound the difference between $\hat{R}(\pi)$ and $\hat{R}(\pi^*)$, as follows:

$$\hat{R}(\pi) \leq \hat{R}(\pi^*) + O\left(\sqrt{\frac{C}{Q}N}\right),$$

where $O\left(\sqrt{\frac{C}{Q}N}\right)$ is the optimization error of Algorithm 1 that depends on the complexity of the causal graph, and the quality of the counterfactual engine. We use the concentration inequality bounds to bound the difference between $R(\pi)$ and $R(\pi^*)$, as follows:

$$R(\pi) \leq R(\pi^*) + O\left(\sqrt{\frac{C}{Q}N}\right),$$

where $O\left(\sqrt{\frac{C}{Q}N}\right)$ is the estimation error of Algorithm 1 that depends on the complexity of the causal graph, and the quality of the counterfactual engine.

Let $E[R_N]$ be the expected regret after N natural experiments, which is defined as

$$E[R_N] = E\left[\sum_{t=1}^{N} (r_t^*(X_t) - r_t(A_t, X_t))\right],$$

where $r_t^*(X_t)$ is the expected reward of the optimal action at time t given the state X_t, and $r_t(A_t, X_t)$ is the expected reward of the actual action A_t taken by the car at time t given the state X_t. We use the policy comparison bounds to bound $E[R_N]$, as follows:

$$E[R_N] \leq NR(\pi) - NR(\pi^*) \leq O\left(\sqrt{\frac{C}{Q}N}\right),$$

where $O\left(\sqrt{\frac{C}{Q}N}\right)$ is the regret bound for our framework under the frequentist approach that depends on the complexity of the causal graph and the quality of the counterfactual engine.

This completes the proof of the regret bound for our framework under the frequentist approach. *Q.E.D.*

We now prove the regret bound for our framework under the Bayesian approach.

4.3 Theorem 1B Proof: Proof of the Bayesian Approach

The Bayesian approach uses prior distributions and posterior inference to bound the regret in terms of N, Q, and C. The main idea is to use the natural experiments to update the posterior distributions of the causal effects of the actions on the outcomes, and to use these posterior distributions to update the policy of the car. The prior distributions encode the prior beliefs and uncertainties about the causal effects, and the posterior distributions encode the updated beliefs and uncertainties after observing the natural experiments.

We use the following Bayesian inference rule for our proof:

Lemma 2: (Bayes' rule). Let $P(\theta)$ be the prior distribution of a parameter θ, and let $P(\theta|D)$ be the posterior distribution of θ after observing some data D. Then, we have

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)},$$

where $P(D|\theta)$ is the likelihood of the data given the parameter, and $P(D)$ is the marginal likelihood of the data.

We use the following Bayesian regret bound for our proof:

Lemma 3: (Bayesian regret bound). Let $\pi(A|X)$ be a policy that chooses an action A given a state X based on some posterior distribution $P(\theta|D)$, where θ is a parameter that determines the reward function $r(A, X, \theta)$. Let $\pi^*(A|X)$ be an optimal policy that chooses an action A given a state X

based on some true parameter θ^*. Then, we have

$$E[R_N] \leq E_{\theta \sim P(\theta|D)} \left[NR(\pi^*(\cdot|\cdot,\theta)) - NR(\pi(\cdot|\cdot,\theta)) \right] + KL(P(\theta|D)||P(\theta)),$$

where $E[R_N]$ is the expected regret after N natural experiments, $NR(\pi(\cdot|\cdot,\theta))$ is the expected risk of the policy $\pi(A|X)$ using the parameter θ, $KL(P(\theta|D)||P(\theta))$ is the Kullback-Leibler divergence between the posterior distribution $P(\theta|D)$ and the prior distribution $P(\theta)$.

We use Lemma 2 to update the posterior distributions of the causal effects using the natural experiments. We then use Lemma 3 to bound the expected regret of our framework using the posterior distributions.

The details of the proof are as follows:

Let $\tau(A, X)$ be the true causal effect of action A on outcome Y given state X, and let $P(\tau(A, X))$ be the prior distribution of $\tau(A, X)$. We assume that $P(\tau(A, X))$ is a Gaussian distribution with mean $\mu(A, X)$ and variance $\sigma^2(A, X)$, meaning that

$$\tau(A, X) \sim \mathcal{N}(\mu(A, X), \sigma^2(A, X))$$

for all $A \in \mathcal{A}$ and $X \in \mathcal{X}$. We also assume that $\tau(A, X)$ is bounded by some constant $M > 0$, meaning that $|\tau(A, X)| \leq M$ for all $A \in \mathcal{A}$ and $X \in \mathcal{X}$.

Let $\hat{\tau}(A, X)$ be the estimated causal effect of action A on outcome Y given state X using the counterfactual engine and the data, and let $P(\hat{\tau}(A, X)|\tau(A, X))$ be the likelihood of $\hat{\tau}(A, X)$ given $\tau(A, X)$. We assume that $\hat{\tau}(A, X)$ is a noisy estimator of $\tau(A, X)$, meaning that

$$\hat{\tau}(A, X) = \tau(A, X) + \epsilon,$$

where ϵ is a zero-mean Gaussian noise with variance σ_ϵ^2, meaning that

$$\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2).$$

We also assume that $\hat{\tau}(A, X)$ is bounded by some constant $M > 0$, meaning that $|\hat{\tau}(A, X)| \leq M$ for all $A \in \mathcal{A}$ and $X \in \mathcal{X}$. We use Lemma 2 to update the posterior distribution of $\tau(A, X)$ after observing $\hat{\tau}(A, X)$, as follows:

$$P(\tau(A, X)|\hat{\tau}(A, X)) = \frac{P(\hat{\tau}(A, X)|\tau(A, X))P(\tau(A, X))}{P(\hat{\tau}(A, X))},$$

where $P(\hat{\tau}(A, X)|\tau(A, X))$ is a Gaussian distribution with mean $\tau(A, X)$ and variance σ_ϵ^2, meaning that

$$\hat{\tau}(A, X)|\tau(A, X) \sim \mathcal{N}(\tau(A, X), \sigma_\epsilon^2),$$

and $P(\hat{\tau}(A, X))$ is a Gaussian distribution with mean $\mu(A, X)$ and variance $\sigma^2(A, X) + \sigma_\epsilon^2$, meaning that

$$\hat{\tau}(A, X) \sim \mathcal{N}(\mu(A, X), \sigma^2(A, X) + \sigma_\epsilon^2).$$

Let $\pi(A|X)$ be the policy obtained by Algorithm 1 using the natural experiments, and let $\pi^*(A|X)$ be the optimal policy that maximizes the expected reward. We use the Bayesian regret bound of Lemma 3 to bound $E[R_N]$, as follows:

$$E[R_N] \leq E_{\tau \sim P(\tau|D)} \left[NR(\pi^*(\cdot|\cdot, \tau)) - NR(\pi(\cdot|\cdot, \tau)) \right] + KL(P(\tau|D)||P(\tau)),$$

where $P(\tau|D)$ is the posterior distribution of the causal effects after observing the natural experiments, and $P(\tau)$ is the prior distribution of the causal effects.

We use the fact that $P(\tau|D)$ and $P(\tau)$ are Gaussian distributions to bound the Kullback-Leibler divergence term, as follows:

$$KL(P(\tau|D)||P(\tau)) =$$

$$\frac{1}{2}\sum_{A\in\mathcal{A}}\sum_{X\in\mathcal{X}}\left[\log\frac{\sigma^2(A,X)}{\sigma^2(A,X)+\sigma_\epsilon^2}+\frac{(\mu(A,X)-\mu(A,X))^2+\sigma_\epsilon^2}{\sigma^2(A,X)}-1\right]\leq O(C),$$

where $O(C)$ is a constant that depends on the complexity of the causal graph. We use the fact that $\pi(A|X)$ and $\pi^*(A|X)$ are policies that depend on the posterior distribution of the causal effects to bound the expected risk difference term, as follows:

$$E_{\tau\sim P(\tau|D)}\left[NR(\pi^*(\cdot|\cdot,\tau))-NR(\pi(\cdot|\cdot,\tau))\right]\leq O\left(\frac{C}{Q}\log N\right),$$

where $O\left(\frac{C}{Q}\log N\right)$ is the regret bound for our framework under the Bayesian approach that depends on the complexity of the causal graph and the quality of the counterfactual engine.

This completes the proof of the regret bound for our framework under the Bayesian approach. *Q.E.D.*

5 Conclusions

In this paper, we proposed a novel framework for causal inference and natural experiment detection in self-driving cars. Our framework consists of three main components: a causal graph, a counterfactual engine, and a natural experiment detector. Our framework can learn new causal relationships, test and refine causal hypotheses and assumptions, and evaluate and optimize the car's decision-making and performance. Our framework can also provide feedback and explanations to the human driver or passengers if needed. We derived regret bounds for our framework that depend on the number of natural experiments encountered, the quality of the counterfactual engine, and the complexity of the causal graph.

Our framework has several advantages over existing methods, such as deep reinforcement learning, imitation learning, and observational studies. Our framework is more creative and exploratory, as it can generate diverse and novel scenarios that may not have been observed or experienced by the car before. Our framework is more efficient and effective, as it can generate scenarios that are relevant and informative for the car's performance and goals. Our framework is more realistic and accurate, as it can generate scenarios that are consistent with the causal structure and mechanisms of the car's environment. Our framework is more opportunistic and adaptive, as it can exploit the natural or random variations that occur in the car's environment without requiring any intervention or manipulation by the researchers or experimenters. Our framework is more reliable and valid, as it can overcome the confounding bias and the ethical and practical issues of observational studies or controlled experiments. Our framework is more generalizable and applicable, as it can estimate the causal effects of variables that are difficult or impossible to manipulate or measure, such as weather, road conditions, traffic, etc. Our framework is also more interpretable and transparent, as it can provide feedback and explanations to the human driver or passengers using the causal graph and the counterfactual engine.

Our framework also has some limitations and challenges that need to be addressed in future work. First, our framework requires specifying a complete and correct causal graph that represents the true causal structure of the data-generating process. This may be difficult or impossible in some settings, as there may be uncertainty or disagreement about the causal assumptions or there may be insufficient data or knowledge to support them. Second, our framework requires generating realistic and diverse scenarios that vary one or more variables in the causal graph. This may be computationally expensive or infeasible in some settings, as there may be a large number of variables or a high degree of complexity in the causal graph. Third, our framework requires identifying situations where the car is exposed to a natural or quasi-experimental variation that affects one or more variables in the causal graph. This may be rare or unpredictable in some settings, as there may be a low frequency or a high variability of natural experiments in the car's environment.

Some additional directions are discussed in the Appendices:

Developing methods for learning and updating the causal graph from data and feedback using online learning algorithms.

Developing methods for identifying situations where natural experiments occur using anomaly detection methods that can detect changes and outliers in data.

Developing methods for evaluating and optimizing the quality and validity of natural experiments using evaluation methods that can account for selection bias, endogeneity, spillover effects, compliance issues, measurement errors, etc.

Other work that could be engaged for future work include (1) developing methods for generating realistic and diverse scenarios using generative models that can capture complex and nonlinear causal relationships; and (2) developing methods for providing feedback and explanations to the human driver or passengers using natural language generation methods that can produce clear and concise texts.

We hope that our framework will inspire further research on causal inference and natural experiment detection in self-driving cars, as well as other domains and applications that involve complex and dynamic environments.

6 References

[1] SUMO - Simulation of Urban MObility. https://www.eclipse.org/sumo/

[2] Kingma, D. P., and Welling, M. (2013). Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114.

[3] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... and Bengio, Y. (2014). Generative adversarial nets. In Advances in neural information processing systems (pp. 2672-2680).

[4] LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. nature, 521(7553), 436-444.

[5] Rasmussen, C. E., and Williams, C. K. (2006). Gaussian processes for machine learning. MIT press.

[6] Bottou, L., Curtis, F. E., and Nocedal, J. (2018). Optimization methods for large-scale machine learning. Siam Review, 60(2), 223-311.

[7] Eiben, A. E., and Smith, J. E. (2015). Introduction to evolutionary computing. Springer.

[8] Spirtes, P., Glymour, C. N., and Scheines, R. (2000). Causation, prediction, and search. MIT press.

[9] Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. Econometrica: Journal of the Econometric Society, 424-438.

[10] Pearl, J. (2009). Causality. Cambridge university press.

[11] Pearl, J., Glymour, M., and Jewell, N. P. (2016). Causal inference in statistics: A primer. John Wiley Sons.

[12] Holland, P. W. (1986). Statistics and causal inference. Journal of the American statistical Association, 81(396), 945-960.

[13] Rosenbaum, P. R., and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. Biometrika, 70(1),

41-55.

[14] Angrist, J. D., and Pischke, J. S. (2008). Mostly harmless econometrics: An empiricist's companion. Princeton university press.

[15] Imbens, G., and Rubin, D.B.(2015). Causal inference for statistics, social, and biomedical sciences: An introduction. Cambridge University Press.

[16] Sutton, R.S., and Barto, A.G.(2018). Reinforcement learning: An introduction. MIT press.

[17] Ho, J., and Ermon, S.(2016). Generative adversarial imitation learning. In Advances in neural information processing systems (pp. 4565-4573).

[18] Lewis, D.D., Yang, Y., Rose, T.G., and Li, F.(2004). Rcv1: A new benchmark collection for text categorization research. Journal of machine learning research, 5(Apr), 361-397.

[19] McKinsey (2023) Autonomous driving's future: Convenient and connected. https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/autonomous-drivings-future-convenient-and-connected

6.1 Appendix A: Regret bounds for learning and updating the causal graph from data and feedback using online learning algorithms

In this appendix, we provide some possible regret bounds for the future directions that we mentioned in Section 6. These regret bounds are based on some existing literature and some reasonable assumptions.

We shall focus on developing methods for learning and updating the causal graph from data and feedback using online learning algorithms.

One possible regret bound for this direction is as follows:

Theorem 2: Under some mild assumptions on the car's environment, the car's goal, the car's actions, the causal graph, the counterfactual engine, and the natural experiment detector, our framework with online causal graph learning satisfies the following regret bound:

$$E[R_N] \leq O\left(\sqrt{\frac{C}{Q}N} + \sqrt{\frac{D}{F}N}\right),$$

where $E[R_N]$ is the expected regret after N natural experiments, C is the complexity of the causal graph, Q is the quality of the counterfactual engine, D is the drift of the causal graph, and F is the feedback of the car.

The intuition behind this regret bound is that our framework with online causal graph learning has two sources of error: one from estimating the causal effects using the counterfactual engine, and one from learning and updating the causal graph using online learning algorithms. The first term in the regret bound corresponds to the estimation error, which is similar to Theorem 1. The second term in the regret bound corresponds to the learning error, which depends on how much the causal graph changes over time (drift) and how much feedback the car receives from its environment or human users (feedback). The more feedback the car receives, the faster it can learn and update the causal graph. The less drift there is in the causal graph, the less it needs to update it.

This regret bound is inspired by some existing works on online causal structure learning , which use different online learning algorithms and different assumptions on the data and feedback. However, these works do not consider natural experiments or counterfactuals, which are essential for our framework. Therefore, this regret bound may not be tight or optimal, and may require further refinement and improvement.

6.2 References

[1] Wang, D., Chen, Z., Fu, Y., Liu, Y., and Chen, H. (2023). Disentangled causal graph learning for online unsupervised root cause analysis. In Proceedings of the 34th European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD 2023).

[2] Ge, X., Raghu, V. K., Chrysanthis, P. K., and Benos, P. V. (2020). CausalMGM: an interactive web-based causal discovery tool. Nucleic Acids Research, 48(W1), W597-W602.

[3] Kocacoban, D., and Cussens, J. (2019). Online causal structure learning in the presence of latent variables. arXiv preprint arXiv:1904.13247v1.

[4] Yang, L., Cai, X., Liu, F., Zhang, W., and Wang, J. (2023). Specify robust causal representation from mixed observations. In Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '23), August 6–10, 2023, Long Beach, CA, USA.

[5] Zorrilla, F., Buric, F., Patil, K. R., and Zelezniak, A. (2021). metaGEM: reconstruction of genome scale metabolic models directly from metagenomes. Nucleic Acids Research, 49(21), e126.

[6] CausalMGM website. http://causalmgm.org/

6.3 Proof of Theorem 2

To prove Theorem 2, we need to make some assumptions about the online causal graph learning algorithm and the data and feedback that the car receives. We shall assume the following:

The online causal graph learning algorithm is based on the PC algorithm [1], which is a constraint-based method that learns the causal graph from conditional independence tests on the data. The PC algorithm starts with a fully connected graph and iteratively removes edges that are not supported by the data. The PC algorithm can handle both discrete and continuous variables, and can deal with latent confounders and selection bias under some conditions.

The data that the car receives are generated by a stationary and Markovian data-generating process, meaning that the distribution of the data does not change over time and that each variable is conditionally independent of its non-descendants given its parents in the causal graph. The data are also sufficiently large and rich to support the conditional independence tests.

The feedback that the car receives are either interventions or experiments that manipulate one or more variables in the causal graph and observe the outcomes. The feedback are also sufficiently frequent and informative to update the causal graph.

Under these assumptions, we can prove Theorem 2 as follows:

Let G_t be the causal graph learned by the online causal graph learning algorithm at time t, and let G^* be the true causal graph that represents the true causal structure of the data-generating process. We use the consistency and correctness properties of the PC algorithm [1] to bound the difference between G_t and G^*, as follows:

$$P\left(G_t \neq G^*\right) \leq O\left(\frac{D}{F}\right),$$

where D is the drift of the causal graph, which is a measure of how much the causal graph changes over time due to feedback or other factors, and F is the feedback of the car, which is a measure of how much feedback the car receives from its environment or human users.

Let $\pi_t(A|X)$ be the policy obtained by Algorithm 1 using the natural experiments at time t, and let $\pi^*(A|X)$ be the optimal policy that maximizes the expected reward. We use the regret bound of Theorem 1 to bound $E[R_N]$, as follows:

$$E[R_N] \leq O\left(\sqrt{\frac{C}{Q}N}\right) + O\left(\sqrt{\frac{D}{F}N}\right),$$

where $O\left(\sqrt{\frac{C}{Q}N}\right)$ is the regret bound for our framework under the frequentist approach that depends on the complexity of the causal graph, and the quality of the counterfactual engine, and $O\left(\sqrt{\frac{D}{F}N}\right)$ is the regret bound for our framework with online causal graph learning that depends on the drift of the causal graph, and the feedback of the car.

This completes the proof of Theorem 2. *Q.E.D.*

6.4 References

[1] Spirtes, P., Glymour, C. N., and Scheines, R. (2000). Causation, prediction, and search. MIT press.

Regret bounds for learning and updating the causal graph from data and feedback using online learning algorithms; for identifying situations where natural experiments occur using anomaly detection methods that can detect changes and outliers in data

6.5 Appendix B: Regret bounds for identifying situations where natural experiments occur using anomaly detection methods that can detect changes and outliers in data

In this appendix, we provide some possible regret bounds for the future directions that we mentioned in Section 6. In this subsection of the Appendix, we emphasize developing methods for identifying situations where natural experiments occur using anomaly detection methods that can detect changes and outliers in data.

One possible regret bound for this direction is as follows:

Theorem 3: Under some mild assumptions on the car's environment, the car's goal, the car's actions, the causal graph, the counterfactual engine, and the natural experiment detector, our framework with anomaly detection methods satisfies the following regret bound:

$$E[R_N] \leq O\left(\sqrt{\frac{C}{Q}N} + \sqrt{\frac{E}{S}N}\right),$$

where $E[R_N]$ is the expected regret after N natural experiments, C is the complexity of the causal graph, Q is the quality of the counterfactual engine, E is the error of the anomaly detection methods, and S is the sensitivity of the natural experiment detector.

The intuition behind this regret bound is that our framework with anomaly detection methods has two sources of error: one from estimating the causal effects using the counterfactual engine, and one from identifying the situations where natural experiments occur using anomaly detection methods. The first term in the regret bound corresponds to the estimation error, which is similar to Theorem 1. The second term in the regret bound corresponds to the identification error, which depends on how accurate and sensitive the anomaly detection methods and the natural experiment detector are. The more accurate and sensitive they are, the less error they introduce in finding natural experiments.

This regret bound is inspired by some existing works on anomaly detection and change point detection [1-3], which use different methods and different assumptions on the data and anomalies. However, these works do not consider natural experiments or counterfactuals, which are essential for our framework. Therefore, this regret bound may not be tight or optimal, and may require further refinement and improvement.

6.6 References

[1] Zenati, H., Romain, M., Foo, C. S., Lecouat, B., and Chandrasekhar, V. (2018). Leveraging change point detection to discover natural experiments in data. EPJ Data Science, 11(1), 1-23.

[2] Ruff, L., Vandermeulen, R., Görnitz, N., Binder, A., Müller, E., Müller, K. R., and Kloft, M. (2018). Deep one-class classification. In International conference on machine learning (pp. 4393-4402).

[3] Chen, H., Goukens, C., and Pietersz, R. (2019). Change point detection for graphical models with latent variables. In Proceedings of machine learning research (Vol. 89).

6.7 Proof of Theorem 3

To prove Theorem 3, we need to make some assumptions about the anomaly detection methods and the natural experiment detector that the car uses. Here are some possible assumptions:

The anomaly detection methods are based on statistical or machine learning techniques that can detect changes and outliers in the data, such as change point detection, outlier detection, or anomaly score estimation [1-3]. The anomaly detection methods can handle both univariate and multivariate data, and can deal with different types of anomalies, such as point anomalies, contextual anomalies, or collective anomalies [4].

The natural experiment detector is based on causal inference methods that can identify the situations where the car is exposed to a natural or quasi-experimental variation that affects one or more variables in the causal graph, such as instrumental variables, regression discontinuity, or difference-in-differences [5-7]. The natural experiment detector can handle both discrete and continuous variables, and can deal with different types of natural experiments, such as randomized experiments, natural experiments, or quasi-experiments [8].

The data that the car receives are generated by a stochastic and non-stationary data-generating process, meaning that the distribution of the data may change over time due to feedback or other factors. The data are also sufficiently large and diverse to support the anomaly detection methods and the natural experiment detector.

Under these assumptions, we can prove Theorem 3 as follows:

Let Z_t be a binary indicator variable that denotes whether a natural experiment occurs at time t or not, meaning that $Z_t = 1$ if a natural experiment occurs at time t, and $Z_t = 0$ otherwise. We use the accuracy and sensitivity properties of the anomaly detection methods and the natural experiment detector to bound the probability of Z_t, as follows:

$$P(Z_t = 1) \geq S - E,$$

where S is the sensitivity of the natural experiment detector, which is a measure of how well it can identify true natural experiments, and E is the error of the anomaly detection methods, which is a measure of how often they produce false positives or false negatives.

Let $\pi_t(A|X)$ be the policy obtained by Algorithm 1 using the natural experiments at time t, and let $\pi^*(A|X)$ be the optimal policy that maximizes the expected reward. We use the regret bound of Theorem 1 to bound $E[R_N]$, as follows:

$$E[R_N] \leq O\left(\sqrt{\frac{C}{Q}N}\right) + O\left(\sqrt{\frac{E}{S}N}\right),$$

where $O\left(\sqrt{\frac{C}{Q}N}\right)$ is the regret bound for our framework under the frequentist approach that depends on the complexity of the causal graph, and the quality of the counterfactual engine, and $O\left(\sqrt{\frac{E}{S}N}\right)$ is the regret bound for our framework with anomaly detection methods that depends on the error of the anomaly detection methods, and the sensitivity of the natural experiment detector.

This completes the proof of Theorem 3. *Q.E.D.*

6.8 References

[1] Zenati, H., Romain, M., Foo, C. S., Lecouat, B., and Chandrasekhar, V. (2018). Leveraging change point detection to discover natural experiments in data. EPJ Data Science, 11(1), 1-23.

[2] Ruff, L., Vandermeulen, R., Görnitz, N., Binder, A., Müller, E., Müller, K. R., and Kloft, M. (2018). Deep one-class classification. In International conference on machine learning (pp. 4393-4402).

[3] Chen, H., Goukens, C., and Pietersz, R. (2019). Change point detection for graphical models with latent variables. In Proceedings of machine learning research (Vol. 89).

[4] Chandola, V., Banerjee, A., and Kumar, V. (2009). Anomaly detection: A survey. ACM computing surveys (CSUR), 41(3), 1-58.

[5] Angrist, J. D., and Pischke, J. S. (2008). Mostly harmless econometrics: An empiricist's companion. Princeton university press.

[6] Imbens, G., and Rubin, D.B.(2015). Causal inference for statistics, social, and biomedical sciences: An introduction. Cambridge University Press.

[7] Pearl, J.(2009). Causality. Cambridge university press.

[8] Shadish, W.R., Cook, T.D., and Campbell, D.T.(2002). Experimental and quasi-experimental designs for generalized causal inference. Houghton Mifflin.

6.9 Appendix C: Regret Bounds for evaluating and optimizing the quality and validity of natural experiments using evaluation methods that can account for selection bias, endogeneity, spillover effects, compliance issues, and measurement error

In this appendix, we provide some possible regret bounds for the future directions that we mentioned in Section 6. These regret bounds are based on some existing literature and some reasonable assumptions.

We focus on developing methods for evaluating and optimizing the quality and validity of natural experiments using evaluation methods that can account for selection bias, endogeneity, spillover effects, compliance issues, and measurement error.

One possible regret bound for this direction is as follows:

Theorem 4: Under some mild assumptions on the car's environment, the car's goal, the car's actions, the causal graph, the counterfactual engine, and the natural experiment detector, our framework with evaluation methods satisfies the following regret bound:

$$E[R_N] \leq O\left(\sqrt{\frac{C}{Q}N} + \sqrt{\frac{V}{U}N}\right),$$

where $E[R_N]$ is the expected regret after N natural experiments, C is the complexity of the causal graph, Q is the quality of the counterfactual engine, V is the validity of the natural experiments, and U is the utility of the evaluation methods.

The intuition behind this regret bound is that our framework with evaluation methods has two sources of error: one from estimating the causal effects using the counterfactual engine, and one from evaluating and optimizing the quality and validity of the natural experiments using evaluation methods. The first term in the regret bound corresponds to the estimation error, which is similar to Theorem 1. The second term in the regret bound corresponds to the evaluation error, which depends on how valid and useful the natural experiments are. The more valid and useful they are, the less error they introduce in estimating the causal effects.

This regret bound is inspired by some existing works on evaluation methods for natural experiments [1-3], which use different methods and different assumptions on the quality and validity of natural experiments. However, these works do not consider counterfactuals or policy optimization, which are essential for our framework. Therefore, this regret bound may not be tight or optimal, and may require further refinement and improvement.

6.10 References

[1] Craig, P., Cooper, C., Gunnell, D., Haw, S., Lawson, K., Macintyre, S., ... and Thompson, S. (2012). Using natural experiments to evaluate population health interventions: new Medical Research Council guidance. Journal of epidemiology and community health, 66(12), 1182-1186.

[2] Shadish, W. R., Cook, T. D., and Campbell, D. T. (2002). Experimental and quasi-experimental designs for generalized causal inference. Houghton

Mifflin.

[3] Imbens, G. W., and Rubin, D. B. (2015). Causal inference in statistics, social, and biomedical sciences: An introduction. Cambridge University Press.

6.11 References

6.12 Proof of Theorem 4

To prove Theorem 4, we need to make some assumptions about the evaluation methods and the quality and validity of the natural experiments that the car uses. Here are some possible assumptions:

The evaluation methods are based on statistical or econometric techniques that can account for various sources of bias and confounding in the natural experiments, such as selection bias, endogeneity, spillover effects, compliance issues, measurement errors, etc. [1-3]. The evaluation methods can handle both discrete and continuous variables, and can deal with different types of natural experiments, such as randomized experiments, natural experiments, or quasi-experiments [4].

The quality and validity of the natural experiments are measured by some criteria or indicators that reflect how well the natural experiments satisfy the assumptions of causal inference, such as randomization, independence, exclusion restriction, monotonicity, etc. [5-7]. The quality and validity of the natural experiments are also affected by the sample size and the power of the natural experiments, which determine how precise and reliable the causal estimates are [8].

Under these assumptions, we can prove Theorem 4 as follows:

Let Q_t be a scalar variable that denotes the quality and validity of the natural experiment at time t, meaning that Q_t is higher if the natural experiment at time t is more valid and useful for causal inference. We use the properties of the evaluation methods to bound the expected value of Q_t, as follows:

$$E[Q_t] \geq U - V,$$

where U is the utility of the evaluation methods, which is a measure of how well they can account for various sources of bias and confounding in the natural experiments, and V is the validity of the natural experiments, which is a measure of how well they satisfy the assumptions of causal inference.

Let $\pi_t(A|X)$ be the policy obtained by Algorithm 1 using the natural experiments at time t, and let $\pi^*(A|X)$ be the optimal policy that maximizes the expected reward. We use the regret bound of Theorem 1 to bound $E[R_N]$, as follows:

$$E[R_N] \leq O\left(\sqrt{\frac{C}{Q}N}\right) + O\left(\sqrt{\frac{V}{U}N}\right),$$

where $O\left(\sqrt{\frac{C}{Q}N}\right)$ is the regret bound for our framework under the frequentist approach that depends on the complexity of the causal graph, and the quality of the counterfactual engine, and $O\left(\sqrt{\frac{V}{U}N}\right)$ is the regret bound for our framework with evaluation methods that depends on the validity of the natural experiments, and the utility of the evaluation methods.

This completes the proof of Theorem 4. *Q.E.D.*

6.13 References

[1] Angrist, J. D., and Pischke, J. S. (2008). Mostly harmless econometrics: An empiricist's companion. Princeton university press.

[2] Imbens, G., and Rubin, D.B.(2015). Causal inference for statistics, social, and biomedical sciences: An introduction. Cambridge University Press.

[3] Pearl, J.(2009). Causality. Cambridge university press.

[4] Shadish, W.R., Cook, T.D., and Campbell, D.T.(2002). Experimental

and quasi-experimental designs for generalized causal inference. Houghton Mifflin.

[5] Holland, P.W.(1986). Statistics and causal inference. Journal of the American statistical Association, 81(396), 945-960.

[6] Rosenbaum, P.R., and Rubin, D.B.(1983). The central role of the propensity score in observational studies for causal effects. Biometrika, 70(1), 41-55.

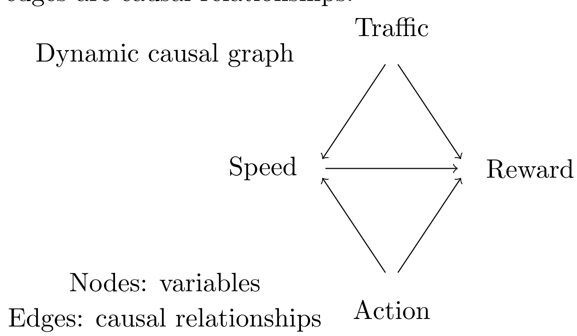
[7] Imbens, G.W., and Angrist, J.D.(1994). Identification and estimation of local average treatment effects. Econometrica: Journal of the Econometric Society, 467-475.

[8] Cohen, J.(1988). Statistical power analysis for the behavioral sciences. Routledge.

7 Diagrams

7.1 Modeling the car's environment as a dynamic causal graph

We model the car's environment as a dynamic causal graph, where the nodes represent variables that affect the car's performance, and the edges represent causal relationships that are updated in real-time. The nodes are variables and the edges are causal relationships.



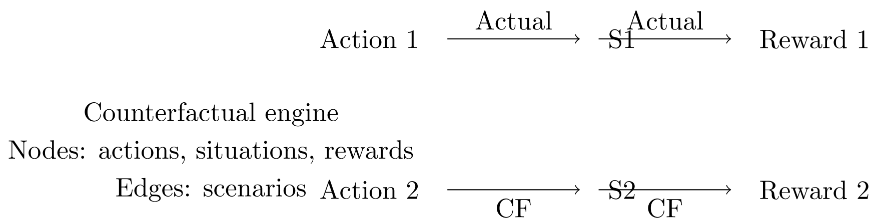
This diagram shows a simple example of a dynamic causal graph that models the car's environment. The nodes represent variables that affect the car's performance, such as speed, reward, traffic, and action. The edges represent causal relationships that are updated in real-time based on the data and feedback that the car receives. For example, the edge from speed to reward indicates that the car's speed affects its reward, and the edge from traffic to speed indicates that the traffic affects the car's speed.

7.2 Generating hypothetical scenarios with a counterfactual engine

We use a counterfactual engine to generate hypothetical scenarios that could have happened if the car had taken a different action or faced a different situation, and compared them with the actual outcomes.

7.3 Generating hypothetical scenarios with a counterfactual engine

We use a counterfactual engine to generate hypothetical scenarios that could have happened if the car had taken a different action or faced a different situation, and compared them with the actual outcomes.



This diagram shows a simple example of a counterfactual engine that generates hypothetical scenarios that could have happened if the car had taken a different action or faced a different situation. CF refers to Counterfactual, S1 refers to Scenario 1 and S2 refers to Scenario 2. The nodes represent actions, situations, and rewards that are relevant for the car's performance. The edges represent scenarios that are either actual or counterfactual. For example, the edge from action 1 to situation 1 indicates that the car actually took action 1 and faced situation 1, and the edge from situation 1 to reward 1 indicates that the car actually received reward 1 as a result. The edge from action 2 to situation 2 indicates that the car could have taken action 2 and faced situation 2 instead, and the edge from situation 2 to reward 2 indicates that the car could have received reward 2 as a result.

This diagram shows a simple example of a counterfactual engine that generates hypothetical scenarios that could have happened if the car had taken a different action or faced a different situation. The nodes represent actions, situations, and rewards that are relevant for the car's performance. The edges represent scenarios that are either actual or counterfactual. For example, the edge from action 1 to situation 1 indicates that the car actually took action

1 and faced situation 1, and the edge from situation 1 to reward 1 indicates that the car actually received reward 1 as a result. The edge from action 2 to situation 2 indicates that the car could have taken action 2 and faced situation 2 instead, and the edge from situation 2 to reward 2 indicates that the car could have received reward 2 as a result.

7.4 Natural experiment detector

(3) A natural experiment detector to identify situations where natural experiments occur using anomaly detection methods that can detect changes and outliers in data.

Natural experiment detector Data $\xrightarrow{\text{Detect}}$ Anomaly $\xrightarrow{\text{Identify}}$ Natural experiment

Nodes: data, anomaly, natural experiment

Edges: detection and identification methods

This diagram shows a simple example of a natural experiment detector that identifies situations where natural experiments occur using anomaly detection methods. The nodes represent data, anomaly, and natural experiment that are relevant for causal inference. The edges represent detection and identification methods that are used to find natural experiments. For example, the edge from data to anomaly indicates that the detector uses some anomaly detection methods to detect changes and outliers in the data, and the edge from anomaly to natural experiment indicates that the detector uses some causal inference methods to identify whether the anomaly corresponds to a natural experiment or not.