

Avoiding Undesired Future with Sequential Decisions

Lue Tao, Tian-Zuo Wang, Yuan Jiang and Zhi-Hua Zhou

National Key Laboratory for Novel Software Technology, Nanjing University, China

School of Artificial Intelligence, Nanjing University, China

{taol, wangtz, jiangy, zhouzh}@lamda.nju.edu.cn

Abstract

Machine learning has advanced in predictive tasks, but practitioners often need to proactively avoid undesired outcomes rather than just predicting them. To this end, a framework called rehearsal has been introduced, which tackles the avoiding undesired future (AUF) problem by modeling how variables influence each other and searching for a decision that leads to desired results. In this paper, we propose a novel rehearsal approach for addressing the AUF problem by making a sequence of decisions, where each decision is dynamically informed by the latest observations via retrospective inference. Theoretically, we show that sequential decisions in our approach tend to achieve a higher success rate in avoiding undesired outcomes by more reliably inferring the outcome of actions compared with existing solutions. Perhaps surprisingly, our approach remains advantageous even under imprecise modeling of relations between variables, and we provide a sufficient condition under which the advantage holds. Finally, experimental results confirm the practical effectiveness of the proposed approach in both simulated and real-world tasks.

1 Introduction

“It is difficult to predict, especially the future,” the renowned physicist Niels Bohr once remarked. Decades later, developments in machine learning (ML) have significantly improved our ability to make accurate predictions [Scarpino and Petri, 2019; Brown *et al.*, 2020; Bi *et al.*, 2023]. However, predictions alone are not satisfactory, if the predicted results are unfavorable for us. It remains challenging to suggest proactive decisions to avoid the undesired results [Zhou, 2022b].

Machine learning techniques often fall short in addressing the problem of *avoiding undesired futures* (AUF), as they primarily focus on capturing statistical dependencies in observational data without understanding underlying mechanisms. For the AUF problem, modern methods leveraging causal relations [Pearl, 2009; Peters *et al.*, 2017] would be helpful. However, identifying causal relations is inherently difficult and not always a necessary prerequisite for decision-making.

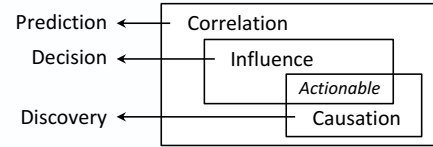


Figure 1: Relationship among correlation, causation, and influence.

After all, we humans make decisions everyday without entailing a complete understanding of the world around us. Moreover, even when identifiable, causal factors are useless for decision-making if unactionable, and factors influencing the future do not necessarily manifest causations. To this end, a new framework known as *rehearsal* has been developed [Zhou, 2022b], building on the concept of *influence* [Zhou, 2023], which was called rehearsal relation in Zhou [2022b] but then renamed to avoid confusion of the relation and the process. As illustrated in Fig. 1, influence serves as an intermediate concept between correlation and causation. Based on this, the AUF problem has been formulated as the search for a decision from hypothesized “rehearsals” of possible actions that leads to desired outcomes [Qin *et al.*, 2023].

It is noteworthy that many real-world tasks cannot be accomplished in one stroke, and thus searching for a single decision may not fully solve the AUF problem. For example, consider an online retailer who has decided to advertise a new product to avoid poor sales. Despite receiving a lot of clicks on the advertisement, not many people made a reservation. These observations allow us to retrospectively infer that the high click-through rate indicated customer interest, but the low reservation rate reflected that customers found the product too expensive. Subsequently, the retailer chose to offer a discount, which improved customer retention and ultimately led to the product reaching the desired sales level. This example highlights the need to make multiple decisions that sequentially adjust distinct variables based on available observations in order to fully address the AUF problem. While recent studies on rehearsal have proposed methods for handling multiple variables simultaneously [Du *et al.*, 2024; Qin *et al.*, 2025], these methods remain confined to composing a single decision. They cannot leverage past observations adaptively, thus leaving the more general and practically important case of sequential decision making untouched.

In this paper, we propose the first rehearsal approach for suggesting a sequence of decisions to avoid undesired out-

comes. A crucial ingredient of the proposed approach is its ability to integrate, at each decision stage, available observations into the structural rehearsal models [Qin *et al.*, 2023] through retrospective inference, i.e., inferring potential factors from present information. Hence, each decision in the sequence is informed by preceding actions and their consequences, constituting a collaborative effort to influence the outcomes. Theoretically, we show that the sequential decisions in our approach can lead to a higher success rate in avoiding undesired outcomes compared with existing solutions by more reliably inferring the outcome of actions. Notably, our approach maintains this advantage even under imprecise modeling of relations between variables, a common challenge in open and dynamic environments [Zhou, 2022a] where these relations might be changeable. A sufficient condition is established to ensure the superiority. Our main contributions are summarized as follows.

- We present a multi-stage formulation of the AUF problem, enabling a sequence of decisions to collaboratively prevent undesired outcomes.
- We propose the first rehearsal approach to make sequential decisions, where each decision is informed by the latest observations via retrospective inference.
- We provide a sufficient condition under which sequential decisions in our approach outperform single decisions, even when the modeling is imprecise in a linear setting.
- We corroborate the superiority of the proposed approach over existing solutions throughout the learning process for both simulated and real-world tasks.

2 Background

In this section, we present a brief overview of structural rehearsal models (SRMs) accommodating structural interactions between variables and the AUF problem.

2.1 Structural Rehearsal Models

An SRM is comprised of a set of potentially time-varying rehearsal graphs and structural equations [Qin *et al.*, 2023].

A rehearsal graph G consists of a set of vertices \mathbf{V} , representing variables of interest, and a set of edges \mathbf{E} connecting these variables. These edges can be either directional or bi-directional. $B \rightarrow C$ indicates that B causes C , while $B \leftrightarrow C$ denotes that B and C can affect each other.

For example, the graph in Fig. 2a illustrates a typical sales scenario for a store during a sales season. In this scenario, the number of salespersons B and the number of customers C mutually influence each other, collectively determining the level of discounts D , which in turn affects sales F . The rehearsal graph can change over time: in the next seasons, new marketing strategies can be implemented by the store manager. For example, the number of customers alone is used to determine the discount level in Fig. 2b, and a cashback (a form of discount) is offered based on the number of customers and the sales in Fig. 2c.

When sales are predicted to fall outside the desired range, the store manager can take actions, such as offering a 50% discount, to influence sales. This type of operation, which

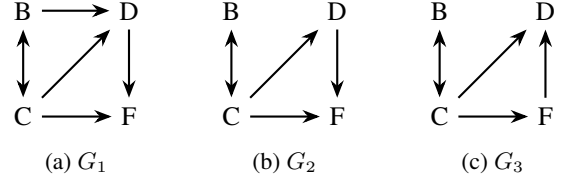


Figure 2: Illustration of rehearsal graphs.

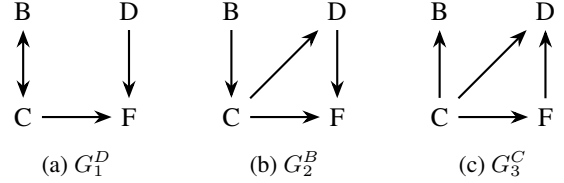


Figure 3: Illustration of alterations on rehearsal graphs.

alters variable A to a fixed value a , is called an *alteration* and denoted by $Rh(A = a)$. It indicates a rehearsal of altering A to a in realistic or hypothetical means. When an alteration is applied to a graph G , the incoming arrows of A are removed, resulting in an altered rehearsal graph G^A . For example, the action of altering D , B , or C in the rehearsal graph G_1 , G_2 , or G_3 as shown in Fig. 2, results in the altered graph G_1^D , G_2^B , or G_3^C in Fig. 3, respectively.

For a rehearsal graph G over variables $\{V_i\}_{i=1}^d$, the corresponding structural equations are defined as

$$V_i = f_i(pa_i, \epsilon_i), \quad (1)$$

where f_i determines the value of V_i by taking as argument $pa_i = \{V \mid V \rightarrow V_i \text{ in } G\}$, the parents of V_i , and ϵ_i , a noise term according to the probability distribution $p(\epsilon_1, \dots, \epsilon_d)$, abbreviated as $p(\epsilon)$. We use \mathbf{f} to denote the set of structural functions $\{f_1, \dots, f_d\}$ for a rehearsal graph G .

Together, the rehearsal graph G , the corresponding structural equations, and noises $p(\epsilon)$ constitute an SRM $\mathcal{M} = \langle G, \mathbf{f}, p(\epsilon) \rangle$. In this paper, the rehearsal graph, which encodes the structural connections among variables, is supposed to be provided based on background knowledge, while the structural equations are learned from data.

2.2 The AUF Problem

The AUF problem is formulated in a specific period of time, which we refer to as a *season*. During a season, an agent can observe variables \mathbf{X} , and the observation may lead to undesired predictions from a model. The agent can also perform alterations on actionable variables \mathbf{A} , aiming to ensure that outcome variables \mathbf{Y} fall within a desired region \mathcal{S} . Formally, the success of the outcome \mathbf{y} falling into \mathcal{S} is indicated by

$$\mathbf{1}(\mathbf{y} \in \mathcal{S}), \quad (2)$$

where the actual outcome \mathbf{y} depends on the observation \mathbf{x} and the alteration $Rh(A = a)$, in which A is one of the actionable variables in \mathbf{A} and a is in the set of feasible values $\Delta(A)$.

Unfortunately, it is infeasible to directly optimize Eq. (2) since the actual outcome \mathbf{y} cannot be altered after it has been observed. After all, no one can change what has already happened. One approach to addressing the AUF problem is to

predict the outcome in advance and minimize the probability of \mathbf{Y} falling out of \mathcal{S} :

$$\min_{A \in \mathbf{A}} \min_{a \in \Delta(A)} \mathbb{P}(\mathbf{Y} \notin \mathcal{S} \mid \mathbf{X} = \mathbf{x}, Rh(A = a)), \quad (3)$$

where the uncertainty is characterized by the noise terms of the structural equations in Eq. (1). It is noteworthy that while the decision by Eq. (3) involves a single actionable variable, the formulation can be easily adapted to utilize multiple actionable variables within a single decision.

A weakness of the approach of Eq. (3) is that all observable variables \mathbf{X} must be available before choosing the actionable variables \mathbf{A} . This requirement may not be realistic in practice; e.g., the observation of the click-through rate occurs after the advertisement action has been taken, as discussed in Section 1. The limitation persists in subsequent studies on the AUF problem [Du *et al.*, 2024; Qin *et al.*, 2025; Du *et al.*, 2025], so it is imperative to resolve this issue.

3 The Proposed Approach

In this section, we propose addressing the AUF problem with multi-stage decisions and introduce the first rehearsal approach that tackles AUF by suggesting a sequence of alterations, each informed by past alterations and observations.

3.1 Multi-Stage AUF

We treat AUF as a multi-stage decision-making problem. At the m -th stage, an agent might make an observation $\mathbf{o}^{(m)}$ on currently observable variables $\mathbf{O}^{(m)}$ and could perform an alteration $Rh(A^{(m)} = a^{(m)})$ on actionable variables in $\mathbf{A}^{(m)}$. After M stages, a season ends, and the outcome \mathbf{y} is revealed. The whole decision-making process is illustrated in Fig. 4. Formally, the objective is to minimize

$$\mathbb{1}(\mathbf{y} \notin \mathcal{S} \mid \mathbf{o}^{(1)}, Rh(a^{(1)}), \dots, \mathbf{o}^{(M)}, Rh(a^{(M)})), \quad (4)$$

where $Rh(a^{(m)})$ is an abbreviation of $Rh(A^{(m)} = a^{(m)})$, by finding a sequence of alterations $Rh(a^{(1)}), \dots, Rh(a^{(M)})$. Compared with the variables Eq. (3), we have $\mathbf{O}^{(m)} \subseteq \mathbf{X}$ and $\mathbf{A}^{(m)} \subseteq \mathbf{A}$ for all m . In addition, the sets $\mathbf{O}^{(m)}$ and $\mathbf{A}^{(m)}$ can be empty for some m . When only $\mathbf{O}^{(1)}$ and $\mathbf{A}^{(1)}$ are non-empty, Eq. (4) is degraded to the probability in Eq. (3).

Minimizing Eq. (4) is challenging for several reasons. First, outcomes \mathbf{y} are not known until the end of a season, similar to the case of Eq. (2). Second, the variables $\mathbf{O}^{(m+1)}$ are not observable at the m -th stage because they have not yet occurred. At this stage, the available information of $\mathbf{o}^{(1)}, \dots, \mathbf{o}^{(m)}$ and $Rh(a^{(1)}), \dots, Rh(a^{(m)})$ can be leveraged, but we still need to anticipate the variables that have not been observed or altered. Third, some variables in \mathbf{X} may remain unavailable even after they have occurred. For example, in the online retailer scenario, while the click-through rate and reservation rate can be obtained in real time, the price most customers are willing to pay may not be immediately observable. Gathering this information through a specialized customer survey is often costly and time-consuming, making it challenging for immediate decision-making use.

The above challenges are pervasive in multi-stage AUF. In Qin, Wang, and Zhou (2023), they focus solely on a single

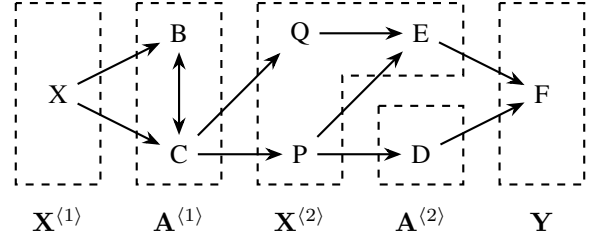


Figure 4: Illustration of multiple-stage AUF in a rehearsal graph. At each stage m , the variables $\mathbf{O}^{(m)}$ in $\mathbf{X}^{(m)}$ are observed, and some variables in $\mathbf{A}^{(m)}$ are altered to avoid undesired outcomes.

decision stage, so variables that are not observable before the decision stage are not represented and do not influence the decision. However, in multi-stage decision-making, the variables that are not immediately accessible at each stage are naturally modeled and can play an important role in inferring the outcome of alterations. In the next subsection, we attempt to address these challenges by proposing a novel multi-stage rehearsal approach, which minimizes the probability of failure computed based on the structural rehearsal models and integrates the observations at each stage into the structural rehearsal models through retrospective inference.

3.2 Multi-Stage Rehearsal

The multi-stage rehearsal (MSR) approach comprises three key components: probability estimation, alteration selection, and information integration. The primary component, alteration selection, is supported by the other two components. MSR operates by iteratively repeating the phases of observations and alterations at each stage, systematically integrating the information from both observations and alterations into the rehearsal model. The outline of MSR is shown in Algorithm 1. In what follows, we will first describe the components of probability estimation and alteration selection based on an informed rehearsal model, and then elaborate on the component of information integration, for which we employ retrospective inference to leverage available observations.

Probability Estimation. Estimating the probability of failure is straightforward by following Qin *et al.* [2023], provided that the information from existing observations and alterations have been incorporated into the rehearsal model. First, we sample N i.i.d. noises $\{\epsilon_i\}_{i=1}^N$ from the distribution $p(\epsilon)$. Each noise ϵ_i is then input into the structural equations \mathbf{f} to determine the outcome \mathbf{y}_i as specified in Eq. (1), vertex by vertex following the topological order of the graph G . Finally, the probability of failure is estimated as the proportion of outcome variables falling out of the desired region \mathcal{S} , i.e., $\sum_{i=1}^N \mathbb{1}(\mathbf{y}_i \notin \mathcal{S})/N$. This estimation procedure will be utilized in the component of alteration selection.

Alteration Selection. As shown in Line 7 of Algorithm 1, the goal here is to search for an alteration that minimizes the probability of failure given the current rehearsal model $\langle G, \mathbf{f}, p(\epsilon) \rangle$. Formally, we have the following objective

$$\min_{A \in \mathbf{A}^{(m)}} \min_{a \in \Delta(A)} \mathbb{P}(\mathbf{Y} \notin \mathcal{S} \mid \langle G^A, \mathbf{f}^A, p(\epsilon) \rangle), \quad (5)$$

Algorithm 1: Multi-Stage Rehearsal

Input: Number of stages M
Output: Sequence of alterations \mathcal{A}

```

1 Initialize the sequence of alterations  $\mathcal{A} = []$ .
2 for  $m \leftarrow 1$  to  $M$  do
3   Acquire rehearsal model  $\mathcal{M} \leftarrow \langle G, \mathbf{f}, p(\epsilon) \rangle$ .
4   Make a new observation  $\mathbf{o}$  on  $\mathbf{O}^{(m)}$ .
5   Obtain the updated noise  $p^o(\epsilon)$  by incorporating  $\mathbf{o}$ 
   into  $p(\epsilon)$  through retrospective inference.
6   Update rehearsal model  $\mathcal{M} \leftarrow \langle G, \mathbf{f}, p^o(\epsilon) \rangle$ .
7   Select an alteration  $Rh(A = a)$  from  $\mathbf{A}^{(m)}$  by
   minimizing the probability of failure.
8   Obtain the altered graph  $G^A$  from  $G$  by removing
   the incoming arrows of  $A$  in  $G$ .
9   Obtain the altered equations  $\mathbf{f}^a$  from  $\mathbf{f}$  by setting
   the equation of  $A$  to  $A = a$ .
10  Update rehearsal model  $\mathcal{M} \leftarrow \langle G^A, \mathbf{f}^a, p^o(\epsilon) \rangle$ .
11  Append the selected alteration  $Rh(A = a)$  to the
   sequence of alterations  $\mathcal{A}$ .
12 end
    
```

where G^A and \mathbf{f}^A are obtained as in Lines 8 and 9 of Algorithm 1, respectively. While the set of actionable variables $\mathbf{A}^{(m)}$ is finite, the feasible values $\Delta(A)$ can be continuous. In such cases, a direct method is to perform a grid search to determine its value, and a more advanced method is Bayesian optimization [Shahriari *et al.*, 2015], as adopted in Qin *et al.* [2023]. Besides, unlike those in Eqs. (3) and (4), the conditional probability employed here to derive the optimal solution are conditioned solely on the rehearsal model. We note that this is computationally sound, as information from both the given observations and alterations can be integrated into the rehearsal model, as detailed below.

Information Integration. Information integration is a key component of MSR, responsible for incorporating all available observations and alterations into the rehearsal model—a basic step for the other two components. When an alteration is applied, both the structure of the rehearsal graph and the corresponding structural equations are updated accordingly. Specifically, the rehearsal graph is modified by removing the incoming arrows of the altered variable, and the structural equations are updated by setting the equation of the altered variable to its new value. In contrast, observations are incorporated into the model by updating the noise distribution to ensure that the sampled variables during probability estimation align with the observed variables. While integrating alterations into the rehearsal model is straightforward, the integration of observations requires more effort and discussion.

If an observed variable O has no parents in the rehearsal graph, simply setting the variable to the observed value during probability estimation is enough. However, when O has parents, and the values of parent variables are not currently available, fixing the observed variable alone during probability estimation is inadequate. This is because the parent variables, which may affect future variables, also need to be consistent with the evidence provided by the observed variables. One straightforward method to align the parent variables with

the observed values is rejection sampling, which involves discarding samples that do not match the observed data during the probability estimation phase. However, this approach is notoriously inefficient. For example, when a standard normal random variable is observed, regardless of the observed value, its measure is always zero, making it nearly impossible to sample it by coincidence. In the next subsection, we address this issue by employing retrospective inference, which aims to efficiently integrate the information of any observations into the the noise distribution of rehearsal models.

3.3 Retrospective Inference

We start by illustrating the procedure of retrospective inference with a simple example. Consider we have a rehearsal model defined by the structural equations:

$$E = Q + P + \epsilon_e, \quad Q = \epsilon_q, \quad P = \epsilon_p, \quad (6)$$

where the noise terms ϵ_p and ϵ_e are independent and follow the normal distribution $\mathcal{N}(0, 1)$. This example represents an online retailer scenario in which the reservation rate E is affected by the click-through rate Q and the acceptable price of customers P , as depicted in Fig. 4.

Suppose that the reservation rate E and the click-through rate Q are observed to be e and q , respectively. To integrate these observations into the model, we would infer the noise posteriors, which ensures that the variables determined by the sampled noises during the probability estimation procedure are consistent with the observations. Concretely, given observations of $E = -1$ (indicating a low reservation rate) and $Q = 1$ (indicating a high click-through rate), we obtain $-1 = 1 + \epsilon_p + \epsilon_e$, from which the posterior of ϵ_p is inferred as $\mathcal{N}(-1, 0.5)$. This suggests that the acceptable price of customers $P = \epsilon_p$ is likely to be low, indicating that customers perceive the product too expensive. Consequently, the retailer should increase discounts rather than investing in additional advertisements to prevent poor sales.

The integration procedure described above is known as *retrospective inference*, as it updates the agent’s beliefs by retrospectively incorporating present information (e.g., q and e) with relevant past factors (e.g., ϵ_p). This ability emerges early in human development [Király *et al.*, 2018], suggesting its potential to facilitate machine decision-making. Thus, we incorporate this capability into the component of information integration. In scenarios where the posterior distribution is not analytically tractable, existing approximation techniques such as variational inference [Bishop, 2006] can be directly employed. Below, we present complete formulas for updating the noise distribution within a linear rehearsal model, accommodating arbitrary alterations and observations.

Formally, the model has structural equations of the form

$$\mathbf{v} = \Lambda \mathbf{v} + \boldsymbol{\epsilon}, \quad (7)$$

where $\mathbf{v} = [V_1, \dots, V_d]$ is a vector containing variables in a rehearsal graph G , Λ is a matrix containing the coefficients of the model with $\Lambda_{ij} = 0$ when there is no directional edge from V_i to V_j in G , and $\boldsymbol{\epsilon} = [\epsilon_1, \dots, \epsilon_d]$ is a vector of normal noise terms with mean $\boldsymbol{\eta}$ and covariance Ω . The mean and covariance of \mathbf{v} are given by

$$\begin{aligned} \boldsymbol{\mu}_{\mathbf{v}} &= (I - \Lambda)^{-1} \boldsymbol{\eta}, \\ \Sigma_{\mathbf{v}, \mathbf{v}} &= (I - \Lambda)^{-1} \Omega (I - \Lambda)^{-\top}. \end{aligned} \quad (8)$$

Before updating the noise distribution with current observations, we need to modify the coefficient matrix according to preceding alterations $\mathcal{A} = \{Rh(A^{(m)})\}_m$. Specifically, Λ is updated to a new matrix $\bar{\Lambda} = [\bar{\Lambda}_{ij}]$ defined by:

$$\bar{\Lambda}_{ij} = \begin{cases} 0 & \text{if } Rh(V_i) \text{ in } \mathcal{A} \\ \Lambda_{ij}^{new} & \text{if } Rh(V_j) \text{ in } \mathcal{A} \text{ and } V_j \leftrightarrow V_i \text{ in } G \\ \Lambda_{ij} & \text{otherwise} \end{cases} \quad (9)$$

where Λ_{ij}^{new} is the coefficient of a new parental relation activated by the alteration of one of the interrelated variables. The mean $\bar{\mu}_{\mathbf{v}}$ and covariance $\bar{\Sigma}_{\mathbf{v},\mathbf{v}}$ of the variables \mathbf{V} under alterations \mathcal{A} is thus updated to

$$\begin{aligned} \bar{\mu}_{\mathbf{v}} &= (I - \bar{\Lambda})^{-1} \eta, \\ \bar{\Sigma}_{\mathbf{v},\mathbf{v}} &= (I - \bar{\Lambda})^{-1} \Omega (I - \bar{\Lambda})^{-\top}. \end{aligned} \quad (10)$$

Then, we can obtain the conditional mean and covariance of \mathbf{V} under the observations \mathbf{o} [Bishop, 2006]:

$$\begin{aligned} \bar{\mu}_{\mathbf{v}|\mathbf{o}} &= \bar{\mu}_{\mathbf{v}} + \bar{\Sigma}_{\mathbf{v},\mathbf{o}} \bar{\Sigma}_{\mathbf{o},\mathbf{o}}^{-1} (\mathbf{o} - \bar{\mu}_{\mathbf{o}}), \\ \bar{\Sigma}_{\mathbf{v},\mathbf{v}|\mathbf{o}} &= \bar{\Sigma}_{\mathbf{v},\mathbf{v}} - \bar{\Sigma}_{\mathbf{v},\mathbf{o}} \bar{\Sigma}_{\mathbf{o},\mathbf{o}}^{-1} \bar{\Sigma}_{\mathbf{o},\mathbf{v}}. \end{aligned} \quad (11)$$

Finally, the distribution of ϵ under alterations and observations is updated from $\mathcal{N}(\eta, \Omega)$ to $\mathcal{N}(\bar{\eta}_{\epsilon|\mathbf{o}}, \bar{\Omega}_{\epsilon,\epsilon|\mathbf{o}})$, where

$$\begin{aligned} \bar{\eta}_{\epsilon|\mathbf{o}} &= \eta + (I - \bar{\Lambda}) \bar{\Sigma}_{\mathbf{v},\mathbf{o}} \bar{\Sigma}_{\mathbf{o},\mathbf{o}}^{-1} (\mathbf{o} - \bar{\mu}_{\mathbf{o}}), \\ \bar{\Omega}_{\epsilon,\epsilon|\mathbf{o}} &= \Omega - (I - \bar{\Lambda}) \bar{\Sigma}_{\mathbf{v},\mathbf{o}} \bar{\Sigma}_{\mathbf{o},\mathbf{o}}^{-1} \bar{\Sigma}_{\mathbf{o},\mathbf{v}} (I - \bar{\Lambda})^{\top}. \end{aligned} \quad (12)$$

4 Theoretical Analysis

In this section, we theoretically justify our approach by guaranteeing computational efficiency through structural information and demonstrating the advantage of sequential decisions over single decisions even with imprecise variable relations.

Efficiency Analysis. Let N be the number of samples used in probability estimation for alteration selection, and let C be the number of feasible values in Eq. (5). The complexity of the proposed approach is provided as follows.

Theorem 1. *Suppose that the structural equations in the rehearsal model are linear Gaussian. Then, the computational complexity of Algorithm 1 is $O(MNCd + Md^4)$.*

Theorem 1 states that the efficiency of MSR is governed by a polynomial term with respect to the number of variables d by leveraging the linearity of structural equations in SRMs. Note that the number of feasible values C naturally exists for discrete variables, and for continuous variables, it can be easily determined based on the granularity of grid search or Bayesian optimization, so our analysis is general and applicable across various scenarios. One feature of the modeling with SRMs is that it enables the utilization of structural information to facilitate both theoretical analysis and algorithmic design, such as the use of linear Gaussian equations considered in [Qin et al., 2023; Du et al., 2024]. Unlike previous studies that focus on single decisions, here we highlights the benefit of structural information in the context of sequential decision making.

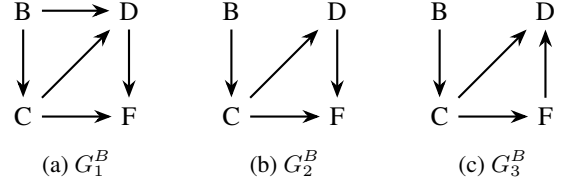


Figure 5: Illustration of the consequences of the alteration $Rh(B)$ on the rehearsal graphs in Fig. 2.

Effectiveness Analysis. In the following, we unveil the advantage of sequential decisions over a single decision. Specifically, we show how the outcome of alterations becomes more reliable after incorporating new observations even with imprecise variable relations. To this end, we provide a sufficient condition under which the inference of the outcome of future outcomes under alterations can be improved by new observations, even when the structural modeling is imprecise, a common case in practice when relations may change over time.

Consider the model given in Eq. (7) with coefficients updated according to a set of alterations \mathcal{A} . In sequential decisions, some alterations $\mathcal{A}^{(1:m-1)} \subseteq \mathcal{A}$ are supposed to be made before observing new variable O , while other alterations $\mathcal{A}^{(m:M)} \subseteq \mathcal{A}$ would be made after the observation. This means that the alterations $\mathcal{A}^{(m:M)}$ can be dynamically selected in a sequential manner by utilizing the latest observations, while single decisions select all alterations at once without the capability of incorporating new observations.

Formally, the distribution of an outcome variable under alterations is $p(Y|\mathcal{A}) = \mathcal{N}(\bar{\mu}_y, \bar{\Sigma}_{y,y})$, where $\bar{\mu}_y$ and $\bar{\Sigma}_{y,y}$ are derived from Eq. (10). Given the observation of variable O , the distribution of the future outcome is refined to $p(Y|\mathcal{A}, O) = \mathcal{N}(\bar{\mu}_{y|O}, \bar{\Sigma}_{y,y|O})$, where $\bar{\mu}_{y|O}$ and $\bar{\Sigma}_{y,y|O}$ are derived from Eq. (11). Evidently, when the rehearsal model in Eq. (7) is correct, the uncertainty of outcome Y will be reduced after observing O , i.e., $H(Y|\mathcal{A}, O) \leq H(Y|\mathcal{A})$. This indicates that $p(Y|\mathcal{A}, O)$ is more reliable at predicting the outcome of alterations. Based on this, sequential decisions produced in our approach by incorporating the latest observations can more reliably select alterations than single decisions for addressing the AUF problem. While this narrative requires that structural equations correctly reflect the underlying mechanisms, we will demonstrate the advantage of our rehearsal approach with incorrect structural equations.

Suppose that we are given an incorrect model $\mathbf{v} = \Lambda' \mathbf{v} + \epsilon$, where the coefficients Λ' are not consistent with the correct ones Λ , and the coefficients under alterations $\bar{\Lambda}'$ are similarly different from the correct ones under alterations $\bar{\Lambda}$. The corresponding outcome inferred using the incorrect model under alterations is denoted by $p(Y'|\mathcal{A})$. After observing O , the inferred distribution becomes $p(Y'|\mathcal{A}, O)$. The following theorem shows that $p(Y'|\mathcal{A}, O)$ is more accurate than $p(Y'|\mathcal{A})$ when $\|\bar{\Lambda}' - \bar{\Lambda}\|_2$ is bounded.

Theorem 2. *Consider any set of alterations \mathcal{A} and two models: the true but unknown model $\mathbf{v} = \Lambda \mathbf{v} + \epsilon$ and the incorrect but known model $\mathbf{v} = \Lambda' \mathbf{v} + \epsilon$. For any outcome variable Y , given any new observation of variable O , denote by $p(Y|\mathcal{A}, O)$ and $p(Y'|\mathcal{A}, O)$ the distributions of Y inferred using the two models under alterations, respectively. Then,*

$p(Y'|A, O)$ is closer to $p(Y|A, O)$ than $p(Y'|A)$ in the sense that both mean and covariance are closer, provided the following norm condition holds for arbitrary $0 < \rho < 1$:

$$\|\bar{\Lambda}' - \bar{\Lambda}\|_2 \leq \min \left\{ \rho s, \frac{s^6}{\sum_{i=0}^5 b_i s^i}, \frac{s^7}{\sum_{i=0}^6 c_i s^i} \right\},$$

where s is the smallest singular value of $I - \bar{\Lambda}'$, and b_i and c_i are constants uniquely determined by ρ , the values of O and A , and the parameters of the known model.

Theorem 2 indicates that it is plausible to more accurately infer the consequence of alterations based on new observations, even with an imprecise model. With a more accurate inference of future outcomes, we can make more reliable estimates for the probability of failure, thereby facilitating better decisions in the phase of alteration selection. Besides, the bound in Theorem 2 is related to the smallest singular value of $I - \bar{\Lambda}'$. The larger the value of s , the broader the bound.

We conclude this section by illustrating the significance of the norm condition in Theorem 2 with an example. Consider the rehearsal graph over four variables $\mathbf{v} = [B, C, D, F]$ as shown in Fig. 2a. Suppose that the corresponding structural equations are given by $\mathbf{v} = \bar{\Lambda}'\mathbf{v} + \epsilon$, where the mean of ϵ is a zero vector, and its covariance is a diagonal matrix with elements equal to one. After the alteration $Rh(B = 1)$, the rehearsal graph is changed to that shown in Fig. 5a, and the corresponding coefficient matrix is modified to

$$\bar{\Lambda}' = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0.1 & 0 & 0 & 0 \\ -0.002 & 0.1 & 0 & 0 \\ 0 & 1 & 0.001 & 0 \end{pmatrix}. \quad (13)$$

Then, given an observation $C = 1$, we obtain the norm condition $\|\bar{\Lambda}' - \bar{\Lambda}\|_2 \leq 0.005$ by applying Theorem 2. An intriguing implication is that the norm bound is large enough to cover variations in the graph structure. Specifically, by shifting $\bar{\Lambda}'_{3,1}$ from -0.002 to 0 , the edge $B \rightarrow D$ in Fig. 5a is removed, and the graph transforms to Fig. 5b. Moreover, by further shifting $\bar{\Lambda}'_{4,3}$ from 0.001 to -0.001 , the edge $D \rightarrow F$ in Fig. 5a is reversed, and the graph evolves to Fig. 5c. All these structural variations comply with the norm condition. As long as the underlying coefficients $\bar{\Lambda}$ are within the 2-norm ball around the given coefficients $\bar{\Lambda}'$, the observation on C will be helpful in inferring the impact of alterations on future outcomes, thereby facilitating subsequent decision-making. This result supports the applicability of the proposed approach in dynamic decision environments.

5 Experiments

In this section, we present experiments to validate the superiority of the proposed approach over existing solutions throughout the learning process of rehearsal models.

Tasks. We simulate a ride-hailing task by following Qin *et al.* [2023], where an SRM is abstracted for supporting a ride-hailing app in making decisions to improve user rating (RAT). The desired region of RAT is set to $[0.8, 1]$. Aside from RAT, this task involves six variables: three observable, two actionable, and one that is inaccessible at the time of decision.

For the Bermuda data [Aglietti *et al.*, 2020], which includes eleven variables, the goal is to maintain the net coral ecosystem calcification (NEC) within the desired range of $[0.5, 2]$. We consider four variables to be observable and five variables to be actionable, with a maximum of two variables being alterable per season. More detailed experimental settings are provided in the appendix.

Baselines. We compare multi-stage rehearsal (MSR) with the single-stage rehearsal (SSR) method. Specifically, MSR makes sequential decisions, where each alteration in the sequence is informed by the latest observations, while SSR makes a single decision, suggesting alterations without using retrospective inference to incorporate new observations. For a fair comparison, both MSR and SSR are allowed to alter the same number of variables. In addition, we consider a simple baseline named single-action decision (SAD), which can only alter one variable. These methods are provided with an SRM learned through Bayesian ridge regression over 100 seasons, as in Qin *et al.* [2023]. These experiments are repeated over 100 seasons 100 times. Furthermore, we compare our rehearsal approach with standard reinforcement learning (RL) methods, including DDPG [Lillicrap *et al.*, 2016], PPO [Schulman *et al.*, 2017], and SAC [Haarnoja *et al.*, 2018].

Comparison with Single-Stage Methods. Fig. 6 and Fig. 7 show the results for comparing with rehearsal approaches. Firstly, MSR consistently outperforms SSR and SAD in terms of the number of successful seasons and the probability of success across both tasks as the number of seasons increases. This demonstrates the superiority of the proposed MSR approach in making sequential decisions based on the latest observations. Secondly, the probability of success is observed to increase with the number of seasons as the SRM becomes more precise. Thirdly, MSR performs better than SSR even in early seasons when the SRM has not yet been well learned. This finding validates the advantage of sequential decisions over single decisions, even under imprecise modeling of relations between variables. We note that MSR shows a slightly higher standard deviation when the number of seasons is small in the Bermuda task. This is because its performance improves consistently with more seasons, while other methods that converge earlier; the deviation of MSR tends to decrease as its performance converges. Finally, the results of the last two columns, which report the number of successful seasons and the probability of success versus the length of the feasible alteration range for actionable variables, clearly demonstrate the superiority of our rehearsal approach when faced with a limited alteration range.

Comparison with Reinforcement Learning. The performance of rehearsal approaches is compared with standard reinforcement learning methods in Table 1. Rehearsal approaches consistently outperform DDPG, PPO, and SAC across both tasks. Notably, MSR achieves over 80 successes and 90 successes for the two tasks in 100 seasons, while the existing methods fail to do these with only 100 seasons. Although increasing the number of seasons may improve the performance of standard reinforcement learning, the number required to achieve the same level of performance as MSR is significantly higher. In the ride-hailing task, DDPG requires

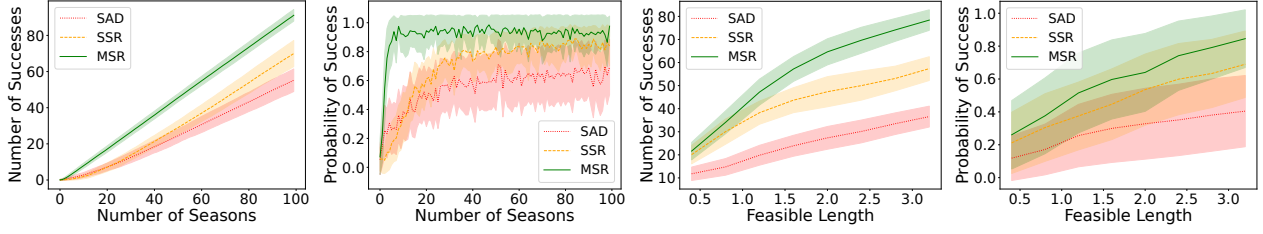


Figure 6: Results on the ride-hailing task. The bands depict standard deviations.

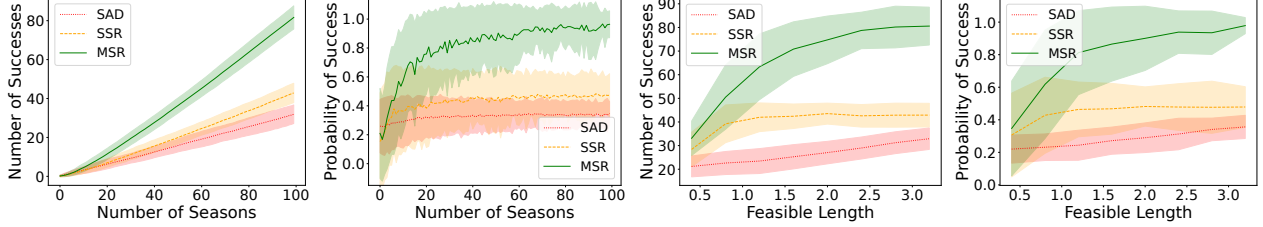


Figure 7: Results on the Bermuda task. The bands depict standard deviations.

| TASK | DDPG | PPO | SAC | SAD | SSR | MSR |
|--------------|------------------|------------------|------------------|------------------|------------------|------------------------------------|
| RIDE-HAILING | 11.99 ± 3.85 | 13.25 ± 4.77 | 10.31 ± 3.18 | 55.22 ± 6.23 | 69.96 ± 7.23 | 91.14 ± 3.20 |
| BERMUDA | 19.35 ± 4.33 | 18.60 ± 4.01 | 16.79 ± 3.82 | 31.93 ± 4.76 | 42.86 ± 4.98 | 81.67 ± 5.95 |

Table 1: Number of successes on the ride-hailing task and Bermuda task using six different methods.

over 1,000 seasons to achieve an average of 200 successful seasons, whereas MSR achieves this in just 218 seasons, highlighting the efficiency of rehearsal for AUF.

6 Related Work

Existing decision-making methods in RL [Sutton and Barto, 2018] are not effective for the AUF problem, primarily because of essential differences in the acquired information. Traditional RL methods typically rely on numerous cost-free interactions with *consistent* outcomes, such as a robot constantly hitting a wall during navigation regardless of whether it happens today or a week later. In real-world AUF scenarios, however, the acquired information is markedly different. Numerous interactions are not available due to the high cost, and more critically, acquiring “true” interaction results may be impossible, since real circumstances can vary over time, e.g., the result of an interaction today may differ from that of the same interaction a week later. These differences limit the applicability of existing RL methods to AUF tasks. It is important to note that gauging influence is the *objective* in solving the AUF problem, while RL is a *tool* that might be helpful to this objective with right adaptations in the future.

Much effort has also been invested in applying causal structures to decision-making problems [Bareinboim *et al.*, 2015; Lattimore *et al.*, 2016; Lee and Bareinboim, 2018; Zhang and Bareinboim, 2020; Majzoubi *et al.*, 2020; Lee *et al.*, 2021; Tsirtsis *et al.*, 2021; Aglietti *et al.*, 2021; Sussex *et al.*, 2023; Wang *et al.*, 2023a; Wang *et al.*, 2023b; Varici *et al.*, 2023; Joshi *et al.*, 2024; Sussex *et al.*, 2024]. Much of the existing research has assumed access to the true and static causal structure, and recent studies have tried to

relax this assumption [Pensar *et al.*, 2020; Toth *et al.*, 2022; Malek *et al.*, 2023; Branchini *et al.*, 2023]. Nevertheless, as discussed before, current causal modeling approaches could be too demanding and restrictive, making them unsuitable for the AUF problem. This has motivated the introduction of the concept of influence [Zhou, 2022b; Zhou, 2023]. Based on this concept, Qin *et al.* [2023] have developed structural rehearsal models, which capture structural interactions among variables and handle the challenges of open and dynamic environments [Zhou, 2022a]. In this paper, we build upon the rehearsal framework to enable sequential alterations dynamically informed by the latest observations through retrospective inference, demonstrating its advantages without requiring precise modeling—an essential capability for human-level decision-making [Zhou, 2022b].

7 Conclusion

In this paper, we propose the first rehearsal approach that enables a sequence of decisions to address the AUF problem. The proposed approach leverages retrospective inference to dynamically integrate new observations into the structural rehearsal models in the decision-making process. Thus, each decision in the sequence is made based on the latest observations, leading to more reliable alterations and better decisions. Theoretically, we show that the sequential decisions in our approach can lead to a higher success rate of avoiding undesired outcomes compared with existing solutions. Both theoretical and experimental results demonstrate the advantage of sequential decisions over single decisions even with imprecise modeling of relations between variables.

Acknowledgements

This research was supported by Jiangsu Science Foundation Leading-edge Technology Program (BK20232003), NSFC (62406137), the AI & AI for Science Project of Nanjing University. Tian-Zuo Wang was supported by National Postdoctoral Program for Innovative Talent and Xiaomi Foundation.

References

- [Aglietti *et al.*, 2020] Virginia Aglietti, Xiaoyu Lu, Andrei Paleyes, and Javier González. Causal bayesian optimization. In *AISTATS*, pages 3155–3164, 2020.
- [Aglietti *et al.*, 2021] Virginia Aglietti, Neil Dhir, Javier González, and Theodoros Damoulas. Dynamic causal bayesian optimization. In *NeurIPS*, pages 10549–10560, 2021.
- [Bareinboim *et al.*, 2015] Elias Bareinboim, Andrew Forney, and Judea Pearl. Bandits with unobserved confounders: A causal approach. In *NeurIPS*, pages 1342–1350, 2015.
- [Bi *et al.*, 2023] Kaifeng Bi, Lingxi Xie, Hengheng Zhang, Xin Chen, Xiaotao Gu, and Qi Tian. Accurate medium-range global weather forecasting with 3d neural networks. *Nature*, 619(7970):533–538, 2023.
- [Bishop, 2006] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer-Verlag, Berlin, Heidelberg, 2006.
- [Branchini *et al.*, 2023] Nicola Branchini, Virginia Aglietti, Neil Dhir, and Theodoros Damoulas. Causal entropy optimization. In *AISTATS*, pages 8586–8605, 2023.
- [Brown *et al.*, 2020] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. In *NeurIPS*, pages 1877–1901, 2020.
- [Du *et al.*, 2024] Wen-Bo Du, Tian Qin, Tian-Zuo Wang, and Zhi-Hua Zhou. Avoiding undesired future with minimal cost in non-stationary environments. In *NeurIPS*, pages 135741–135769, 2024.
- [Du *et al.*, 2025] Wen-Bo Du, Hao-Yi Lei, Lue Tao, Tian-Zuo Wang, and Zhi-Hua Zhou. Enabling optimal decisions in rehearsal learning under care condition. In *ICML*, 2025.
- [Haarnoja *et al.*, 2018] Tuomas Haarnoja, Aurick Zhou, Pieter Abbeel, and Sergey Levine. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In *ICML*, pages 1861–1870, 2018.
- [Joshi *et al.*, 2024] Shalmali Joshi, Junzhe Zhang, and Elias Bareinboim. Towards safe policy learning under partial identifiability: A causal approach. In *AAAI*, pages 13004–13012, 2024.
- [Király *et al.*, 2018] Ildikó Király, Katalin Oláh, Gergely Csibra, and Ágnes Melinda Kovács. Retrospective attribution of false beliefs in 3-year-old children. In *Proceedings of the National Academy of Sciences*, volume 115, pages 11477–11482, 2018.
- [Lattimore *et al.*, 2016] Finnian Lattimore, Tor Lattimore, and Mark D Reid. Causal bandits: Learning good interventions via causal inference. In *NeurIPS*, pages 1189–1197, 2016.
- [Lee and Bareinboim, 2018] Sanghack Lee and Elias Bareinboim. Structural causal bandits: Where to intervene? In *NeurIPS*, pages 2573–2583, 2018.
- [Lee *et al.*, 2021] Junkyu Lee, Radu Marinescu, and Rina Dechter. Submodel decomposition bounds for influence diagrams. In *AAAI*, pages 12147–12157, 2021.
- [Lillicrap *et al.*, 2016] Timothy P. Lillicrap, Jonathan J. Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. Continuous control with deep reinforcement learning. In *ICLR*, 2016.
- [Majzoubi *et al.*, 2020] Maryam Majzoubi, Chicheng Zhang, Rajan Chari, Akshay Krishnamurthy, John Langford, and Aleksandrs Slivkins. Efficient contextual bandits with continuous actions. In *NeurIPS*, pages 349–360, 2020.
- [Malek *et al.*, 2023] Alan Malek, Virginia Aglietti, and Silvia Chiappa. Additive causal bandits with unknown graph. In *ICML*, pages 23574–23589, 2023.
- [Pearl, 2009] Judea Pearl. *Causality*. Cambridge university press, 2009.
- [Pensar *et al.*, 2020] Johan Pensar, Topi Talvitie, Antti Hyttinen, and Mikko Koivisto. A bayesian approach for estimating causal effects from observational data. In *AAAI*, pages 5395–5402, 2020.
- [Peters *et al.*, 2017] Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. *Elements of causal inference: foundations and learning algorithms*. The MIT Press, 2017.
- [Qin *et al.*, 2023] Tian Qin, Tian-Zuo Wang, and Zhi-Hua Zhou. Rehearsal learning for avoiding undesired future. In *NeurIPS*, pages 80517–80542, 2023.
- [Qin *et al.*, 2025] Tian Qin, Tian-Zuo Wang, and Zhi-Hua Zhou. Gradient-based nonlinear rehearsal learning with multivariate alterations. In *AAAI*, 2025.
- [Scarpino and Petri, 2019] Samuel V Scarpino and Giovanni Petri. On the predictability of infectious disease outbreaks. *Nature communications*, 10(1):898, 2019.
- [Schulman *et al.*, 2017] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.
- [Shahriari *et al.*, 2015] Bobak Shahriari, Kevin Swersky, Ziyu Wang, Ryan P Adams, and Nando De Freitas. Taking the human out of the loop: A review of bayesian optimization. *Proceedings of the IEEE*, 104(1):148–175, 2015.
- [Sussex *et al.*, 2023] Scott Sussex, Anastasia Makarova, and Andreas Krause. Model-based causal bayesian optimization. In *ICLR*, 2023.
- [Sussex *et al.*, 2024] Scott Sussex, Pier Giuseppe Sessa, Anastasia Makarova, and Andreas Krause. Adversarial causal bayesian optimization. In *ICLR*, 2024.

- [Sutton and Barto, 2018] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.
- [Toth *et al.*, 2022] Christian Toth, Lars Lorch, Christian Knoll, Andreas Krause, Franz Pernkopf, Robert Peharz, and Julius Von Kügelgen. Active bayesian causal inference. In *NeurIPS*, pages 16261–16275, 2022.
- [Tsirtsis *et al.*, 2021] Stratis Tsirtsis, Abir De, and Manuel Rodriguez. Counterfactual explanations in sequential decision making under uncertainty. In *NeurIPS*, pages 30127–30139, 2021.
- [Varici *et al.*, 2023] Burak Varici, Karthikeyan Shanmugam, Prasanna Sattigeri, and Ali Tajer. Causal bandits for linear structural equation models. *Journal of Machine Learning Research*, 24(297):1–59, 2023.
- [Wang *et al.*, 2023a] Tian-Zuo Wang, Tian Qin, and Zhi-Hua Zhou. Estimating possible causal effects with latent variables via adjustment. In *ICML*, pages 36308–36335, 2023.
- [Wang *et al.*, 2023b] Tian-Zuo Wang, Tian Qin, and Zhi-Hua Zhou. Sound and complete causal identification with latent variables given local background knowledge. *Artificial Intelligence*, 322:103964, 2023.
- [Zhang and Bareinboim, 2020] Junzhe Zhang and Elias Bareinboim. Designing optimal dynamic treatment regimes: A causal reinforcement learning approach. In *ICML*, pages 11012–11022, 2020.
- [Zhou, 2022a] Zhi-Hua Zhou. Open-environment machine learning. *National Science Review*, 9(8):nwac123, 2022.
- [Zhou, 2022b] Zhi-Hua Zhou. Rehearsal: learning from prediction to decision. *Frontiers of Computer Science*, 16(4):164352, 2022.
- [Zhou, 2023] Zhi-Hua Zhou. Rehearsal: Learning from prediction to decision. Keynote at the CCF Conference on AI, 2023.