Optimizing Causal Interventions in Hybrid Bayesian Networks

A discretization, knowledge compilation, and heuristic optimization approach

Maarten C. Vonk, Diederick Vermetten, Jacob de Nobel, Sebastiaan Brand,
Ninoslav Malekovic, Thomas Bäck, Alfons Laarman, and Anna V. Kononova

Leiden Institute of Advanced Computer Science
Leiden University, The Netherlands

Abstract. Causality is increasingly integrated into decision-making processes. Often, the goal is to optimize over causal interventions to achieve specific policy objectives. However, research into causal optimization has bifurcated into either the online optimization of interventions in causal models or the offline optimization of decision rules in causal influence diagrams. This paper introduces an approximate method for offline optimizing interventions in arbitrary hybrid Bayesian networks using observational data. The optimization problem is approached by compiling discretized Bayesian networks as binary decision diagrams, whereafter running interventional queries is very efficient. This efficiency is exploited by running heuristic optimization algorithms to optimize over the interventional queries. By running experiments on a variety of large hybrid Bayesian networks, we demonstrate the practical utility of our method and discuss policy relevance.

Keywords: Bayesian Networks · Knowledge Compilation · Causal Inference · Optimal Causal Interventions · Heuristic Optimization · Do-operator

1 Introduction

In the realm of decision-making challenges, policymakers draw increasingly on methods from causality to enhance policy analysis [21]. The underlying rationale is that by comprehending the causal dynamics inherent in a policy problem, one can craft more efficacious policy interventions. Consequently, optimizing causal interventions is crucial for informed decision-making in uncertain environments.

Computing the probabilistic effects of interventions, that involve forcing assignments of variables, relies on the do-calculus [33] to convert the resulting interventional distributions to observational ones. Often, the observational distributions can be computed using Bayesian network inference techniques [20]. Given the limitations of these techniques in the continuous setting [36], there has been a focused research effort on discretizing Bayesian networks (BNs) [9,31,29],
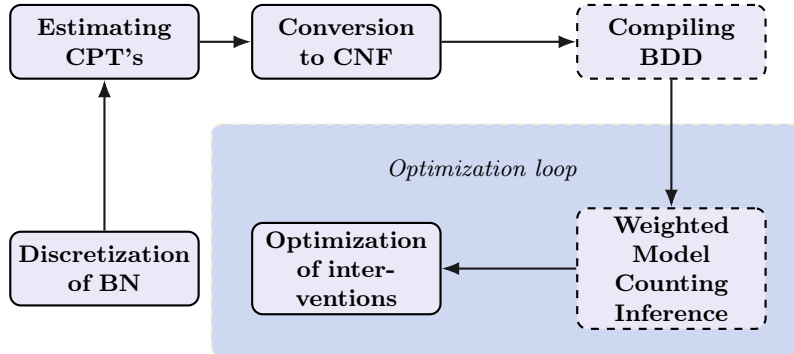
Fig. 1: The methodology used in this paper. After discretizing and estimating the conditional probability tables (CPTs) of the Bayesian network (BN), the network is encoded into conjunctive normal form (CNF). This encoding allows the compilation as a binary decision diagram (BDD) once, before the optimization starts. In the optimization process, heuristic optimization algorithms make use of the cheap weighted model counting operation on BDDs to compute the otherwise prohibitively costly interventional distributions.

which allows for the application of established discrete Bayesian network inference methods. However, the discrete Bayesian network inference techniques needed to compute the effect of interventions are computationally expensive, making optimization over multiple possible interventions even costlier. To tackle these computational challenges, we approach this problem with binary decision diagrams (BDDs). Inference with BDDs first requires performing a computationally expensive compilation step, after which computing inference queries can be done in time linear in the size of the BDD [14,44]. While compiling a BN into a BDD is a heuristic compression method without theoretical guarantees of reducing inference complexity, it often proves much faster in practice [44] and is particularly advantageous when optimizing over multiple interventional queries.

1.1   Approach and Contributions

In this paper, we propose a methodology (see Figure 1) to solve the *offline Causal Global Optimization problem* [2] on hybrid Bayesian networks by encoding discretized versions of the BNs as binary decision diagrams. We subsequently showcase the performance of the heuristic optimization algorithms that use these efficient encodings to optimize over interventional queries. A variety of synthetic and real-world causal Bayesian networks with different characteristics have been subjected to the proposed methodology to ensure generalizability of the results and application diversity. Our contribution is three-fold:

- We extend the implementation of the *do*-operator and interventional distributions to decision diagrams encoded from Bayesian networks without unobserved confounders.

- We compare and evaluate the performance of various optimization algorithms for optimizing causal interventions, thereby establishing their potential in causal contexts.
- Through practical demonstrations, we highlight the real-world value of our proposed methodology for decision-makers.

This paper is structured as follows. First, we discuss related research in Section 1.2. Then, the problem statement is formulated in Section 2, whereafter, our methodology is introduced in Section 3. The methodology is applied in Section 4. Finally, Section 5 reflects on the findings and identifies future research avenues.

1.2   Related Work

Research on causal decision-making is increasingly gaining traction. Earlier approaches have focused on online learning, with a particular focus on causal multi-armed-bandit problems. In these problems, an agent balances the exploration-exploitation trade-off through interventions in a causal system [8,24,23,22].

Extending bandits to the continuous domain and letting an agent carry out exploratory interventions prior to choosing the optimal intervention leads to the formulation of the *online Causal Global Optimization problem* [2], which has been further extended to the dynamic [1], constrained [3], stochastic [19], contextual [4] and adversarial settings [40].

While the aforementioned approaches all pertain to online scenarios, where agents progressively acquire knowledge through actions, performing interventions in real scenarios can be impractical or unethical. Therefore, the realm of causal decision-making also receives attention in the offline context. In this setting, (causal) Bayesian networks can be extended to (causal) influence diagrams, distinguishing clearly among nodes representing chance, decision, and utility [15]. Decision nodes consist of decision rules, which are sets of guidelines dictating actions in response to various assignments in parent nodes [20]. Effective methods for maximizing utility in influence diagrams by optimizing decision rules have been suggested [20]. These methods optimize the decision rules with respect to the (causal) structure of the diagram but do not consider causal interventions containing the *do*-operator.

Similarly, outside the online realm, there has been much research on optimizing individualized treatment [25,27,5]. These works go one step further in the causal hierarchy [7] by optimizing the counterfactual mean of the outcome with respect to the modified interventional distribution rather than the interventional mean. Nevertheless, their emphasis on one-dimensional treatments frees them from the computational complexities of the multi-dimensional interventions discussed in this paper.

2   Problem Statement

Throughout this paper, we denote random variables and their assignments using upper and lower case, respectively. Vectors are indicated in bold. The set of

random variables is denoted by $\boldsymbol{X} = \{X_1, \ldots, X_n\}$, where random variable X_i takes values x_i in corresponding state space Ω_{X_i}. A graph is denoted by $G = (\boldsymbol{V}, \boldsymbol{E})$ with nodes $\boldsymbol{V} = \{V_1, \ldots, V_n\}$ and edges $\boldsymbol{E} \subseteq \boldsymbol{V} \times \boldsymbol{V}$. The graph is called *directed* when every edge in the graph has a direction, and it is called *cyclic* when there exists a directed path from a node to itself. When the graph is directed and not cyclic it is called a *directed acyclic graph.*

A *Bayesian network* (BN) represents random variables by the nodes of a directed acyclic graph. The probabilistic dependencies of the random variables are captured by the graph's edges. Let $P(x_1, \ldots, x_n)$ be the joint probability distribution of random variables X_i associated to nodes $V_i \in \boldsymbol{V}$ in the corresponding directed acyclic graph $G = (\boldsymbol{V}, \boldsymbol{E})$. The joint probability can be factorized according to the structure of the Bayesian network [33]:

$$P(x_1, \ldots x_n) = \prod_{i=1}^{n} P(x_i \mid \boldsymbol{pa_i})$$

where $\boldsymbol{pa_i}$ represents the assignment of the random variables that correspond to parents of V_i. When $P(x_i \mid \boldsymbol{pa_i})$ is discrete, it can be expressed through a conditional probability table (CPT) [44].

In the case of *causal Bayesian networks*, the edges in BNs convey a causal meaning, and the *do*-operator can be used to model causal interventions. The behavior of the *do*-operator within a causal BN leads to a truncated factorization of the distribution [45]. Suppose $\boldsymbol{X_K}, \boldsymbol{X_S} \subset \boldsymbol{X}$ and $\boldsymbol{X_K} \cap \boldsymbol{X_S} = \varnothing$. Then:

$$P(\boldsymbol{x_S} \mid do(\boldsymbol{X_K} = \boldsymbol{x_K})) = \sum_{x_i \mid i \notin \boldsymbol{K} \cup \boldsymbol{S}} \prod_{i \notin \boldsymbol{K}} P(x_i \mid \boldsymbol{pa_i}). \tag{1}$$

More efficiently, the *do*-operator can be implemented by means of an adjustment set [41], that is, a set $\boldsymbol{X_J} \subset \boldsymbol{X}$ for which:

$$P(\boldsymbol{x_S} \mid do(\boldsymbol{X_K} = \boldsymbol{x_K})) = \begin{cases} P(\boldsymbol{x_S} \mid \boldsymbol{x_K}) & \text{if } \boldsymbol{X_J} = \varnothing, \\ \sum_{\boldsymbol{x_J}} P(\boldsymbol{x_S} \mid \boldsymbol{x_K}, \boldsymbol{x_J}) P(\boldsymbol{x_J}) & \text{otherwise.} \end{cases} \tag{2}$$

We focus on the *Causal Global Optimization* (CGO) problem [2] defined in Equation 3. The variables in the Bayesian network $\boldsymbol{X}$ can be further dissected into intervenable variables $\boldsymbol{Z}$, context variables $\boldsymbol{C}$, and outcome variable Y[1]. The *goal* is then to optimize the interventions among intervenable variables that minimize the expected value of the outcome variable in the associated interventional distribution:

$$\boldsymbol{X_K^*}, \boldsymbol{x_K^*} = \underset{\boldsymbol{X_K} \subset \boldsymbol{Z}, \boldsymbol{x_K} \in \Omega_{\boldsymbol{X_K}}}{\arg\min} \mathbb{E}[Y \mid do(\boldsymbol{X_K} = \boldsymbol{x_K}))]. \tag{3}$$

---

[1] In this case, we consider a one-dimensional outcome variable, although the proposed methodology is generalizable to multiple outcome variables.

3 Methodology

The highlighted methodology components in Figure 1, as discussed in this section, include the discretization of the hybrid Bayesian network, inference using the BDD encoding of this discretization, and the specifics of the optimization algorithms. More detailed information about inferring the CPTs, CNF formulas, and the BDD encodings can be found in [44].[2]

3.1 Discretization

The discretization process involves dividing the continuous state space of random variable X_i, denoted as Ω_{X_i} into distinct bins $B_j : j = 1, \ldots, m$ so that they form a *complete partition* of Ω_{X_i}. Each bin B_j is associated with the sample mean of the data points contained within that bin, $\frac{1}{|B_j|} \sum_{x_i \in B_j} x_i$, where $|B_j|$ indicates the number of data points x_i contained within bin B_j.

Two types of discretization have been considered. The equal width (EW) discretization technique involves dividing the state spaces Ω_{X_i} into bins of consistent width. On the other hand, the equal frequency (EF) discretization method separates the samples into quantiles. Because the EF showed better accuracy performance in the context of BN inference [44], discretizations in this paper involve this method. We infer the discretized conditional probability values with a Bayesian method with empirical Bayes priors as described in [44] to prevent *positivity violations* [45].

3.2 BDD Encoding and Inference

A binary decision diagram (BDD) [12] is a rooted directed acyclic graph which typically represents a Boolean function $f : \{0, 1\}^n \to \{0, 1\}$, although for our purposes it is used to represent a pseudo-Boolean function $f : \{0, 1\}^n \to \mathbb{R}$ (i.e. a discrete probability distribution). The two relevant properties of BDDs are their ability to compactly represent many functions by identifying redundancies, and their support for polynomial-time operations (e.g., computing marginal probabilities) based on BDD size.

A Bayesian network is encoded in a BDD by first encoding each CPT entry in a small Boolean expression. From these, a BDD can be built using primitive BDD operations for logical and ($\wedge$), or ($\vee$), not ($\neg$), etc. Consider, for example, the BN given in Figure 2a-2b. Boolean variables $\{a_0, b_0, b_1\}$ are introduced to encode the values of A and B, while Boolean variables ω_i are used to refer to specific probabilities. For example, the CPT entry $P(B = 2 \mid A = 0) = 0.2$ is encoded as $(\neg a_0 \wedge \neg b_1 \wedge b_0) \Rightarrow \omega_2$, where $\neg a_0$ corresponds to $A = 0$ and $\neg b_1 \wedge b_0$ corresponds to $B = 2_{\text{dec}} = 10_{\text{bin}}$. A mapping $\mathsf{val}(\cdot)$ from ω_i's to probabilities ($\mathsf{val}(\omega_2) = 0.2$ for this CPT entry) is stored separately.

Given a BDD that encodes a BN, computing marginal or conditional probabilities reduces to so-called *weighted model counting* [13], in which different paths
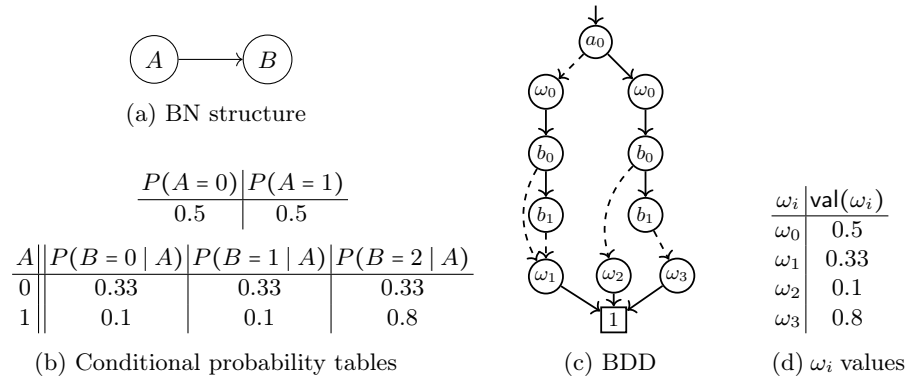
---

[2] The methodology is available on https://github.com/sebastiaanbrand/bn-dd.

(a) BN structure

$P(A = 0)$	$P(A = 1)$
0.5	0.5

A	$P(B = 0 \mid A)$	$P(B = 1 \mid A)$	$P(B = 2 \mid A)$
0	0.33	0.33	0.33
1	0.1	0.1	0.8

ω_i	$\mathsf{val}(\omega_i)$
ω_0	0.5
ω_1	0.33
ω_2	0.1
ω_3	0.8

(b) Conditional probability tables   (c) BDD   (d) ω_i values

Fig. 2: An example BN (a,b) and the corresponding BDD (c). The probabilities corresponding to the Boolean variables ω_i are stored separately (d). In the BDD solid (dashed) edges correspond to positive (negative) assignments to variables.

through the decision diagram are traversed and probabilities are multiplied and added along the way. During this traversal, every node in the BDD only needs to be visited (at most) once, and thus, weighted model counting can be done in time linear in the size of the BDD. In this paper, the *do*-operator has been implemented through the adjustment formula (Equation 2) that utilizes the efficiently computed marginal and conditional distribution. A more detailed description of the encoding and the weighted model counting procedure can be found in [44].

3.3   Optimization

Given the above discretization and encoding mechanisms, interventional queries can be computed and the CGO problem can be addressed heuristically. To illustrate this point, we select a variety of optimization algorithms from the *black-box optimization* literature [6], implemented in the Nevergrad [35] platform. Our set of algorithms includes: (a) RandomSearch as a baseline, (b) two evolutionary algorithms (Differential Evolution [39] (DE) and (1+1) Evolutionary Strategy [10] (OnePlusOne)), (c) local search methods (BFGS [11,16,18,38] and Powell's method [34]), (d) the algorithm wizard 'NGOpt39' [28], which automatically selects an optimizer based on problem characteristics such as the number of variables. To track the optimization performance, we use IOHexperimenter [30], a benchmarking module from the IOHprofiler environment [42], which allows us to fully track the optimization process.

For each optimization algorithm and each network, we perform 10 independent runs of 2000 evaluations of the objective function each, where an evaluation consists of calculating the expected value of the outcome variable given an intervention set (see Equation 2), which is to be minimized in all networks considered in this paper (see Equation 3). To represent the problem inside the optimization algorithms, we note that each node we can intervene on takes a value between

0 and the number of bins if it is intervened on; for convenience, we then use a negative value if no intervention is performed on this node. Within the Nevergrad library, such a representation gets translated to a real-valued one to allow continuous optimization algorithms to tackle this problem as well [35].

4   Experiments

This section details the experiments presented in the paper, considers their policy implications where relevant, and evaluates the performance of optimization algorithms. The discretizations were selected to maximize bin count without overfitting the sample data [44] or exceeding 64GB RAM during BDD compilation. Figure 4 presents the optimization results and Table 1 summarizes the Bayesian network features and discretizations.

Table 1: Characteristics of the used Bayesian networks.

Dataset	Origin	Samples	Network nodes	edges	Max parents[3]	Discretization bins	Intervention variables
Toy [2]	synthetic	1000	3	2	1	100	1
MC [17]	synthetic	4000	12	15	6	30	7
Climate [26]	real-world	293	8	11	3	20	3
Mehra [43]	real-world	5000	24	71	9	4	8
Arth [32]	real-world	5000	107	150	17	6	5

Toy This dataset [2] contains a three-node $X \rightarrow Z \rightarrow Y$ Bayesian network discretized into 100 bins. While not interesting from an optimization perspective, it benchmarks inference quality post-discretization, focusing on possibly-optimal minimal intervention set [24] $\{Z\}$. Most optimization algorithms quickly converge to the optimal solution in the discretization, $Y = -1.866$, which is near the known exact solution $Y = -1.856$.

Mixed Confounding (MC) This synthetic dataset, depicted in Figure 3a, is called mixed confounding as it contains many continuous as well as discrete variables that causally influence more than one variable in the graph, as further specified in Csuite benchmarking causal datasets [17]. After discretizing continuous variables into 30 bins, we conducted two experiments on outcome variables X_{10} and X_{11}, both with the minimum intervention set [24] $\{X_2, X_3, X_4, X_5, X_6, X_7, X_8\}$. Figures 4a and 4b illustrate that both DE and NGOpt19 outperform other optimization algorithms.

Climate The dataset comprises 294 samples, tapping into the interplay between climate and conflict as depicted in Figure 3b. It includes conflict, climate, environmental, and demographic data at the municipal level in southeastern Iraq. An

---

[3] The maximum number of parents is a proxy for computational complexity.

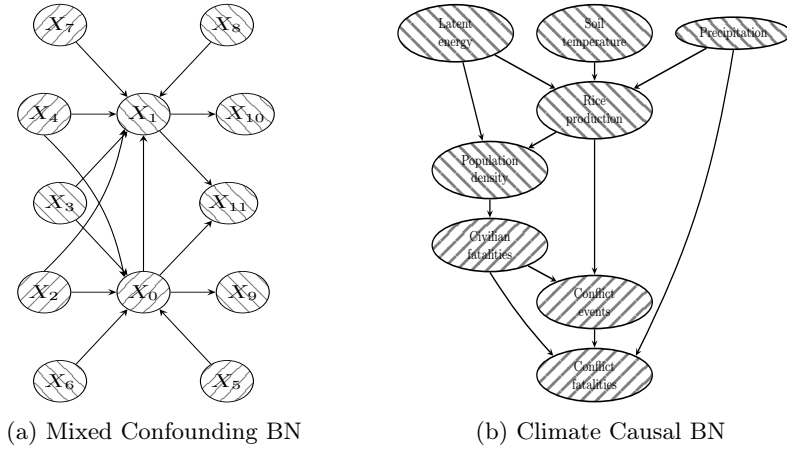(a) Mixed Confounding BN   (b) Climate Causal BN

Fig. 3: Bayesian network structures corresponding to the mixed confounding and climate experiments. The hybrid nodes are distinguished by interior line patterns: dashed lines oriented from the top-left to bottom-right denote nodes with *continuous* values, whereas those from the top-right to bottom-left signify nodes with *discrete* values.

understanding of the variables and the details of the empirically verified causal structure can be found in [26].

The outcome variable is the number of conflict fatalities. The interventions in consideration are precipitation, rice production, and population density, although direct intervention may be challenging. Indirect policy measures such as water management and development projects can be targeted to increase water availability and balance demographic distribution respectively.

Figure 4e shows that DE and NGOpt19 are the best-performing algorithms. The tables in Figure 4f indicate best-found objective values and intervention values compared to the sample means in the dataset. With interventions tailored to increasing rice supply/production, equitable population management leading to a more balanced demographic distribution, and increased precipitation or related water management interventions, the expected value of conflict fatalities *can be reduced* by 85.9% with respect to the mean.

Mehra This conditional linear Gaussian BN from bnlearn [37] explores the correlation between air pollution and health outcomes [43]. Due to its considerable size, the compilation of BDD is restricted to those segments of the network pertinent to the optimization task.[4] The results can be seen in Figure 4c, where the identified best-performing algorithm is Randomsearch closely followed by DE and NGOpt39 (same applies to the dataset below).

---

[4] Because of the nature of the *do*-operator and the adjustment formula of Equation 2, nodes that are not marginalized or conditioned upon in the optimization step can be eliminated from the compilation.

Arth This enormous Gaussian BN from bnlearn, originating from the GeneNet package, features plant expression data [32]. Like the Mehra dataset, the size necessitates limiting the BDD compilation to only those parts of the network that are essential for optimization. The results are available in Figure 4d.
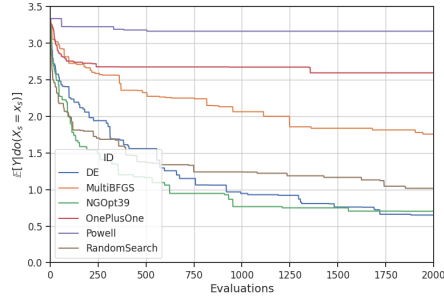
(a) Mixed Confounding: $Y = X_{10}$
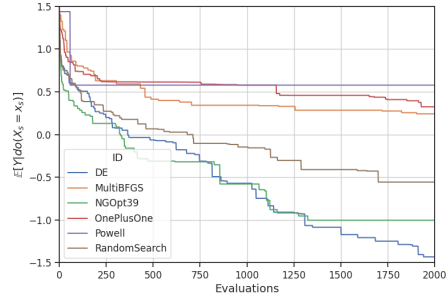
(b) Mixed Confounding: $Y = X_{11}$
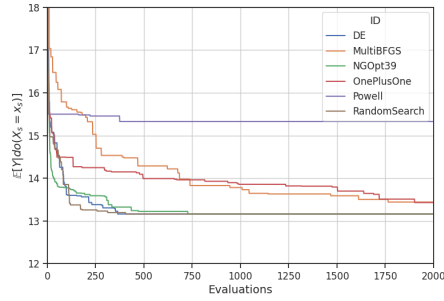
(c) Mehra
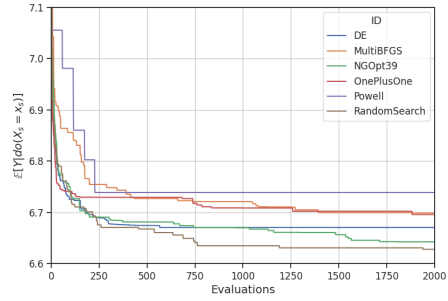
(d) Arth
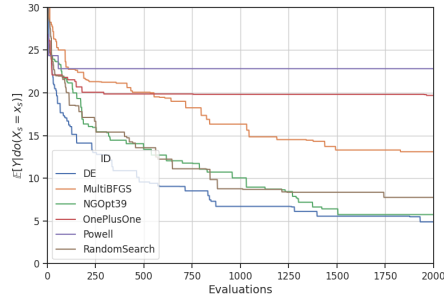
(e) Climate

	$\mathbf{y}^*$	$\bar{\mathbf{y}}$
Conflict Fatalities	4.89	34.6

	$\mathbf{x}^*$	$\bar{\mathbf{x}}$
Rice Production	36.9	13.0
Population Density	215	811
Precipitation	25.0	19.8

(f) Climate Dataset: Best-found solutions

Fig. 4: The best-found expected value of the interventional distributions of 6 optimization algorithms over 2000 evaluations, averaged over 10 runs for the 4 considered datasets (a-e). For the climate dataset, the table in (f) corresponds to found objective $\mathbf{y}^*$ and intervention values $\mathbf{x}^*$ compared to sample mean values $\bar{\mathbf{y}}$, $\bar{\mathbf{x}}$.

Performace of Optimization Algorithms We observe that the more local methods (BFGS, Powell, OnePlusOne) perform rather poorly. Only DE and NGOpt are able to outperform the Randomsearch baseline, suggesting that the underlying optimization problem might be multimodal. This might also be connected to the choice of internal problem representation selected from Nevergrad, as working directly on the discrete variables might be more suitable for these local search methods. While this points to a need for further examination of the specifics of the optimization procedure, the results nevertheless illustrate that in general these problems *contain sufficient structure* that heuristic black-box optimizers can improve over the performance of Randomsearch.

5   Conclusion

This paper proposes a methodology to solve the offline causal global optimization problem for large hybrid causal Bayesian networks using only observational data. The methodology consists of discretizing the hybrid Bayesian network and encoding to binary decision diagrams. Once the BDD is compiled, query costs become negligible, allowing the deployment of heuristic optimization algorithms that optimize for the best intervention.

Our contribution is three-fold. First, we have extended the use of interventional distributions to decision diagrams encoded from Bayesian networks in cases without unobserved confounders. Second, the encoding enables cheap evaluations of interventional queries that allow us to compare and evaluate various heuristic optimization algorithms aimed at optimizing causal interventions. Finally, we have demonstrated the practical utility of our methodology for decision-makers through its application to an empirically based causal Bayesian network.

Since collecting interventional data can be costly, impractical, or unethical, our research is limited to *observational data only*. Using strictly observational data brings as a limitation that optimal interventions may lie beyond the current observational domain of the variables.

We propose the following future research avenues: although some results suggest that discretization does not necessarily lead to a loss of information, future research could further specify for which data structure or distributions this is the case or even include the loss of information due to discretization in the optimization loop. The latter would lead to a multi-objective optimization problem. Alternatively, the proposed methodology could be benchmarked against methodologies incorporating different known approximate inference methods, such as sampling methods or variational inference [20].

As demonstrated, the encodings of Bayesian networks to decision diagrams bring computational advantages that pave the way for the *integration of optimization within the field of causality*.

References

1. Aglietti, V., Dhir, N., González, J., Damoulas, T.: Dynamic causal bayesian optimization. NeurIPS **34**, 10549–10560 (2021)

2. Aglietti, V., Lu, X., Paleyes, A., González, J.: Causal bayesian optimization. In: AISTATS. pp. 3155–3164. PMLR (2020)

3. Aglietti, V., Malek, A., Ktena, I., Chiappa, S.: Constrained causal bayesian optimization. arXiv preprint arXiv:2305.20011 (2023)

4. Arsenyan, V., Grosnit, A., Bou-Ammar, H.: Contextual causal bayesian optimisation. arXiv preprint arXiv:2301.12412 (2023)

5. Athey, S., Wager, S.: Policy learning with observational data. Econometrica **89**(1), 133–161 (2021)

6. Bäck, T., et al.: Evolutionary algorithms for parameter optimization—thirty years later. Evolutionary Computation **31**(2), 81–122 (2023)

7. Bareinboim, E., Correa, J.D., Ibeling, D., Icard, T.F.: On Pearl's hierarchy and the foundations of causal inference. Probabilistic and Causal Inference (2022)

8. Bareinboim, E., Forney, A., Pearl, J.: Bandits with unobserved confounders: A causal approach. NeurIPS **28** (2015)

9. Beuzen, T., Marshall, L., Splinter, K.D.: A comparison of methods for discretizing continuous variables in bayesian networks. Environmental modelling & software **108**, 61–66 (2018)

10. Beyer, H.G., Schwefel, H.P.: Evolution strategies–a comprehensive introduction. Natural computing **1**, 3–52 (2002)

11. Broyden, C.: The Convergence of a Class of Double-rank Minimization Algorithms: 2. The New Algorithm. IMA Journal of Applied Mathematics **6**(3) (09 1970)

12. Bryant, R.E.: Graph-based algorithms for Boolean function manipulation. Transactions on Computers **100**(8), 677–691 (1986)

13. Chavira, M., Darwiche, A.: On probabilistic inference by weighted model counting. Artificial Intelligence **172**(6-7), 772–799 (2008)

14. Dal, G., Lucas, P.: Weighted positive binary decision diagrams for exact probabilistic inference. Intl. Journal of Approximate Reasoning **90**, 411–432 (2017)

15. Everitt, T., Carey, R., Langlois, E.D., Ortega, P.A., Legg, S.: Agent incentives: A causal perspective. AAAI **35**(13), 11487–11495 (May 2021)

16. Fletcher, R.: A new approach to variable metric algorithms. The Computer Journal **13**(3), 317–322 (01 1970). https://doi.org/10.1093/comjnl/13.3.317

17. Geffner, T., et al.: Deep end-to-end causal inference. arXiv preprint arXiv:2202.02195 (2022)

18. Goldfarb, D.: A family of variable-metric methods derived by variational means. Mathematics of Computation **24**(109), 23–26 (1970)

19. Gultchin, L., Aglietti, V., Bellot, A., Chiappa, S.: Functional causal bayesian optimization. arXiv preprint arXiv:2306.06409 (2023)

20. Koller, D., Friedman, N.: Probabilistic graphical models: principles and techniques. MIT press (2009)

21. Kreif, N., DiazOrdaz, K.: Machine learning in policy evaluation: new tools for causal inference. arXiv (2019). https://doi.org/10.48550/arXiv.1903.00402

22. Lattimore, F., Lattimore, T., Reid, M.D.: Causal bandits: Learning good interventions via causal inference. NeurIPS **29** (2016)

23. Lee, S., Bareinboim, E.: Structural causal bandits: Where to intervene? Advances in neural information processing systems **31** (2018)

24. Lee, S., Bareinboim, E.: Structural causal bandits with non-manipulable variables. In: AAAI. vol. 33, pp. 4164–4172 (2019)

25. Li, H., Hubbard, A., van der Laan, M.: Targeted learning on variable importance measure for heterogeneous treatment effect. preprint arXiv:2309.13324 (2023)

26. Malekovic, N., Vonk, M., Birkman, L., Sweijs, T., Kononova, A., Bäck, T.: Angling for causality behind security: Natural causes of armed conflict in iraq. Preprints (February 2024). https://doi.org/10.20944/preprints202402.0556.v1
27. Malenica, I., Coyle, J., van der Laan, M.J.: tmle3mopttx: Targeted learning and variable importance with optimal individualized categorical treatment (2022), https://github.com/tlverse/tmle3mopttx, r package version 1.0.0
28. Meunier, L., et al.: Black-box optimization revisited: Improving algorithm selection wizards through massive benchmarking. IEEE Trans. Evol. Comput. **26**(3) (2022)
29. Neil, M., Tailor, M., Marquez, D.: Inference in hybrid bayesian networks using dynamic discretization. Statistics and Computing **17**(3), 219–233 (2007)
30. de Nobel, J., Ye, F., Vermetten, D., Wang, H., Doerr, C., Bäck, T.: IOHexperimenter: Benchmarking platform for iterative optimization heuristics. CoRR **abs/2111.04077** (2021), https://arxiv.org/abs/2111.04077
31. Nojavan, F., Qian, S.S., Stow, C.A.: Comparative analysis of discretization methods in bayesian networks. Environmental Modelling & Software **87**, 64–71 (2017)
32. Opgen-Rhein, R., Strimmer, K.: From correlation to causation networks: a simple approximate learning algorithm and its application to high-dimensional plant gene expression data. BMC systems biology **1**(1), 1–10 (2007)
33. Pearl, J.: Causality. Cambridge university press (2009)
34. Powell, M.: An efficient method for finding the minimum of a function of several variables without calculating derivatives. The Computer Journal **7**(2) (01 1964)
35. Rapin, J., Teytaud, O.: Nevergrad - A gradient-free optimization platform. https://GitHub.com/FacebookResearch/Nevergrad (2018)
36. Salmerón, A., Rumí, R., Langseth, H., Nielsen, T.D., Madsen, A.L.: A review of inference algorithms for hybrid bayesian networks. Journal of Artificial Intelligence Research **62**, 799–828 (2018). https://doi.org/10.1613/jair.1.11228
37. Scutari, M.: Learning bayesian networks with the bnlearn R package. Journal of Statistical Software **35**(3), 1–22 (2010)
38. Shanno, D.F.: Conditioning of quasi-newton methods for function minimization. Mathematics of Computation **24**(111), 647–656 (1970)
39. Storn, R., Price, K.: Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. Journal of Global Optimization **11**(4), 341–359 (1997)
40. Sussex, S., Sessa, P.G., Makarova, A., Krause, A.: Adversarial causal bayesian optimization (2023)
41. van der Zander, B., Liśkiewicz, M., Textor, J.: Separators and adjustment sets in causal graphs: Complete criteria and an algorithmic framework. Artificial Intelligence **270**, 1–40 (2019)
42. Vermetten, D., de Nobel, J., Ye, F., Wang, H., Doerr, C., Bäck, T.: IOHprofiler: Iterative Optimization Heuristics Profiler. https://github.com/IOHprofiler (2018), wiki available at https://iohprofiler.github.io/
43. Vitolo, C., Scutari, M., Ghalaieny, M., Tucker, A., Russell, A.: Modeling air pollution, climate, and health data using bayesian networks: A case study of the english regions. Earth and Space Science **5**(4), 76–88 (2018)
44. Vonk, M.C., Brand, S., Malekovic, N., Bäck, T., Laarman, A., Kononova, A.V.: Efficient inference in Bayesian networks with continuous variables a discretization and knowledge compilation approach. Available at SSRN 4620451 (2023)
45. Vonk, M.C., Malekovic, N., Bäck, T., Kononova, A.V.: Disentangling causality: assumptions in causal discovery and inference. Artificial Intelligence Review pp. 1–37 (2023)