Noname manuscript No.
(will be inserted by the editor)

A common framework for learning causality

Eva Onaindia · Diego Aineto · Sergio Jiménez

Received: date / Accepted: date

Abstract Causality is a fundamental part of reasoning to model the physics of an application domain, to understand the behaviour of an agent or to identify the relationship between two entities. Causality occurs when an action is taken and may also occur when two happenings come undeniably together. The study of causal inference aims at uncovering causal dependencies amongst observed data and to come up with automated methods to find such dependencies. While there exist a broad range of principles and approaches involved in causal inference, in this position paper we argue that it is possible to unify different causality views under a common framework of symbolic learning.

Keywords Causal inference · Action Models · Behaviour Prediction

1 Introduction

Causality is a relationship between two events wherein one is identified as the *cause* and the other one as the consequence or *effect* caused by the former. When we say that event `A` causes event `B`, we can mean that `A` has been observed to probabilistically cause `B` (e.g. smoking causes lung cancer), that intervening on the value of `A` will affect the distribution of `B` (e.g. eating fast food for several weeks causes a weight gain independently of other causes?) or that `A` causally affects `B` via some action application (e.g. pouring water in a jug makes it become full).

In our exposition to learning causal inference, we will classify causality into two categories: (a) intrinsic causality and (b) extrinsic causality. In the former, causal inference is addressed from a statistical standpoint, typically using causal Bayesian networks [16]. In the *intrinsic causality*, the value of a variable is regarded as a consequence of the evolution of the other variables of the model. This has been the mainstream in the field of *causal discovery* whose principal aim is to study causal inference from independent and identically distributed random variables which rely on the conditional independence relationships in the data [18]. Ultimately, causal discovery is concerned with studying the mechanisms by which variables come to take on values, or with predicting the value of a variable after some other variable has been manipulated; i.e., analyzing the dynamics of beliefs under changing conditions. Philosophers have typically distinguished between the *type-level* or *general* causality, and the *token* or *actual* causality. The former is principally devoted to the investigation of generic and counterfactual relationships among variables that are applicable to any hypothetical scenario, while the latter focuses on particular events and it is mostly concerned with finding explanations or the causes of the observed events [8]. The difference between general and actual causality amounts to the difference between asking whether smoking causes lung cancer (the effect that results from the cause) and asking whether the years of smoking of an individual caused her to get cancer (the cause of an observed effect).

On the other hand, the *extrinsic causality* is determined by a given specific target and the actions taken by an agent towards the achievement of the desired target. Unlike intrinsic causality, which assumes that actions are external entities originating outside the theory and not as a mode of behaviour within the theory [15], in extrinsic causality actions are part of the set of variables that define the model theory.

Universitat Politècnica de València
Camino de Vera s/n, 46022 Valencia
E-mail: onaindia,dieaigar,serjice@dsic.upv.es

A central problem in extrinsic causality is to describe the changes or effects caused by the execution of actions. A large body of research has been devoted to the study of the frame problem (the problem of describing what does not change when actions are performed) [14], to the qualification problem (the problem of specifying all the sufficient conditions for an action to be executable) [13] and the ramification problem (the problem of stating all possible effects of an action) [19]. Causal theories have been introduced in [12] as a nonmonotonic formalism that provides a natural solution for both the frame and ramification problem in reasoning about actions. Intuitively, there is a difference between knowing the cause of a fact and knowing the conditions under which facts are caused, a distinction which is commonly disregarded in natural sciences. A causal theory is a set of causal rules $A \Rightarrow B$ that express a kind of a causal relation among propositions, indicating that B is caused if A is true. Informally, every fact that is caused obtains and every fact that obtains is caused [7].

In goal-driven applications, goals are the actual causal initiators that drive behavioural sequences, and actions are regarded as consciously acting causes towards the achievement of the goals. In extrinsic causality, the analysis of causal factors between the observations prior to the occurrence of an event and the observations after the event are very relevant, as they provide a viable explanation of the undertaken actions and are informative in terms of the predictions of behavioural outcomes. Particularly, the token-level reasoning in extrinsic causality is concerned with the extraction of the sequence of actions (plan) that causally explains the agent's behaviour in a particular scenario, while the type-level reasoning is involved with obtaining the description of the causal rules; that is, discovering the functional relationships among observations that explain the dynamics of a domain.

In this position paper, we argue that learning any of the aforementioned types of causality can be approached as learning the action model that governs the dynamics or behaviour of a domain. While this statement seems more appropriate for goal-driven applications that explicitly exhibit an action-based or extrinsic causality, we hypothesize that symbolic qualitative learning is also exploitable to learning intrinsic causality; that is, we hypothesize that domains that exhibit an intrinsic causality are likewise governed by an implicitly action model. The rationale behind this claim is supported by the fact that an action model accounts for both the facts that hold when an event occurs as well as the cause of the event.

In the following section, we briefly summarize the main literature to causal discovery or intrinsic causality. In section 3 we sketch a model of symbolic learning for causal inference and we show how the model can also be used to infer causal relations in intrinsic causality. Section 4 outlines some practical domains in which the symbolic learning framework is applicable to extract causality. Finally section 5 concludes and points at further research lines.

2 Causal discovery

Causal discovery algorithms derive causal relations between the measured variables of the observational data of a phenomenon. This type of algorithms are principally applied to uncover causal relations in natural and social sciences, examining how a phenomenon would change when a variable is manipulated. Due to the delicate nature of some experiments in these disciplines, scientists have put the focus of causal discovery on observational domain data obtained without intervention instead of experimental data.

Causal inference from statistical data has attracted much interest in many application domains. Causal inference methods in machine learning are explicitly designed to generate hypotheses on causal directions automatically based upon statistical independence tests and the causal Markov assumption that relates causal relations to probability densities [16,17].

Discovery of causal directed acyclic graphs (DAG) models has been addressed with two major approaches: the Bayesian approach and the constrained-based approach. The probabilistic framework of the Bayesian approach computes the probability that the independencies associated with an entire causal structure are true and hence it enables to average a particular hypothesis of interest, such as `Does X cause Y?`, over all possible causal structures. Bayesian networks allow information from several models to be combined to make better inferences and to better account for modeling uncertainty [9]. In contrast, the constrained-based model identifies first several constraints that the underlying causal structure must satisfy, and then it looks for those sets of causal structures that are consistent with the constraints. Constraints may consist, for instance, of particular conditional-independence statements [18]. A number of algorithms based on Boolean satisfiability solvers, as constraint optimization techniques, have opened new opportunities to integrate general background knowledge and discover causal structures in the presence of both directed cycles and latent variables [20].

As a whole, algorithms that learn the causal structure from purely or mostly observational data, as well as experimental data, and that typically use graphical model representations, are recently spreading widely in statistics, machine learning, and the social and natural sciences [11]. However, despite many algorithms provably find the correct causal structure under certain ideal circumstances, they are not proven to be effective in practice. Learning a complete causal Bayesian network is very costly and only applicable to low-dimensional data, and as such they are greatly limited due to the high computational complexity (bad scalability). Although introducing constraints on the structure of the DAG enable to deal efficiently with high-dimensional data, constrained-based methods are generally incomplete and unable to identify combined cause factors (when a change in an individual variable does not cause a change in the response variable but combined changes of variables do).

On the other hand, association rule mining is an efficient means for discovering potentially causal relationships in data. Causal rule mining approaches first extract association rules and then, following different methodologies of hypothesis testing, validate whether or not they are causal rules. One of the challenges in discovering causality in large datasets of observational studies is that even using domain knowledge, it is difficult to foresee a combined cause of an outcome, and this is where data mining research comes into play. Causal rule mining relies upon the idea that associations are necessary for causality. In the work presented in [10], hypothesized cause-effect relationships are represented as association rules, and then an observational study is conducted to test if each of the hypotheses is a real cause; i.e., to identify if the association rule is a causal rule.

We have presented here three different methodologies to find causal relationships: via probability, via constraints or via associations. Ultimately, the aim is to come up with a mechanism that allows to identify the *cause* variables that affect the *effect* variables.

3 Symbolic learning for causal inference

The symbolic learning paradigm has its roots at the inception of the field of Artificial Intelligence (AI). The models manipulated in this paradigm contain rules and syntactic combinations of explicit variables (symbols). Although these models were originally designed for search and representation, there have been important developments in the learning of symbolic models. Symbolic learning consists in finding a set of rules that explain the examples given to the learner. The strength of this approach, as compared to the connectionist approach, lies in its capabilities to generalize from very small amounts of examples. Another interesting feature of symbolic learning is that the inferred models are easily interpreted and understood by humans. A good example of symbolic learning is Inductive Logic Programming, a collection of techniques that, given a set of positive and negative examples, learn a logic program that entails all the positive examples and none of the negative ones.

In recent years, an interest in the learning symbolic action models has emerged from the AI Planning community, and different approaches have been proposed to solve this problem. We will focus our attention on this task, *Action Model Learning*, and we will show how this task allows us to tackle all the different types of causality.

An action model describes how a domain changes, that is, the valid transitions in the space of states of a given environment. Each action in the action model is usually defined in terms of its preconditions and effects, where the preconditions restrict the applicability of the actions to states meeting certain criteria and the effects describe the changes undertaken by the state. Action models can be fed to automated planners to find the solution to a planning problem, defined as a tuple $P = \langle F, A, I, G \rangle$, where F is a set of variables, A is the action model, I is an initial state, i.e. an instantiation of all variable in F, and G is a goal condition defined as an instantiated subset of F. The solution to P is a sequence of instantiated actions that transits from the state I to a state in which the conditions of G hold.

There exists several approaches to learning action models [2,21,4,22,1], varying both in methodology and input data. These algorithms usually rely on two types of inputs: (1) plan traces representing a valid sequence of instantiated actions, and (2) state observations, understood as total or partial instantiations of the world variables. The aim of the learning algorithms is to determine the action schemes (their preconditions and effects) that explain the input data. This means that the plan traces should be correct instantiations of the learned model and that the state observations should belong to the state space generated by the model. Usually these algorithms impose additional constraints that limit the space of possible models to a finite space. Some constraints come in the form of hyper-parameters (like maximum number of actions) and some are imposed by the domain (since the number of variables also restrict the space of possible models). Under these restrictions, learning action models can be seen as a deterministic search problem in an hypothesized space. Given that these learning algorithms originate from the automated planning field, which aims to automatically find plans for any given goal, the learned models represent the *physics* of a domain and determine all valid transi-

	s_i			s_j		
	EH	EX	WL	EH	EX	WL
Bob	0	0	0	1	0	1
Lisa	0	1	0	1	1	1
Robert	0	0	0	0	0	0

Table 1 Dataset for the example. EH: eating healthy, EX: exercising, WL: weight loss .

tions between the set of possible states of the domain. For causal inference, we are particularly interested in approaches that learn from state observations because they are more flexible and enable to learn both *intrinsic* and *extrinsic causality* depending on the nature of the observations, as we will show hereafter.

Following the causal inference taxonomy described in Section 1, action model learning can be seen as discovering inference in *extrinsic causation*. This is better understood when learning from observations that encode the behaviour of a particular agent. Since the behaviour of an agent is motivated by a goal, the space of states is constrained to the ones the agent understands as fruitful to achieving his goals. This is reflected in the action model in the form of additional preconditions that restrict the state space. When the observations represent the behaviour of an agent, the learned action model does not represent the *physics* of the domain and instead it is interpreted as an agent strategy or policy.

However, we also claim that an action model is able to represent the causal relationships among variables as understood in *causal discovery*, what we refer to as *intrinsic causation*. Moreover, an instantiation of such action model (i.e. a plan) can also shed light on the cause of a particular event, usually referred to as *actual cause* or *token causation*.

Let us illustrate now with an example how a symbolic approach is extensible to *intrinsic causation* and is able to infer both *type* and *token causation*.

Example

In this example we want to identify the causes of a person losing weight. Let us assume we have identified that the possible causes of `weight loss` are `eating healthy` and `exercising`. Using these three variables we build a dataset where each observation is a tuple (s_i, s_j) with $i < j$, meaning that an observation is comprised of a pre-state and a post-state of the observed individual. Using this type of observations, the learning algorithm should ensure that there exists a correct sequence of actions that allow the transition from s_i to s_j. Table 1 shows the dataset we will use throughout this example.

Using an action model learning algorithm and introducing the constraint that *an action can only modify a single variable*, we would be able to infer the following action model.

Action 1: `eat healthy`
  Preconditions: ¬ `eating healthy`
  Effects: `eating healthy`
Action 2: `exercise`
  Preconditions: ¬ `exercising`
  Effects: `exercising`
Action 3: `lose weight 1`
  Preconditions: `eating healthy`, ¬ `weight loss`
  Effects: `weight loss`
Action 4: `lose weight 2`
  Preconditions: `exercising`, ¬ `weight loss`
  Effects: `weight loss`

Analyzing the resulting action model, we can observe that in actions 1 and 2 only the variable under change participates, meaning that these actions can take place with independence of the values of other variables. Actions 3 and 4, on the other hand, present an additional variable in their preconditions set, meaning that the value of `weight loss` will change if a certain condition is met. This action model gives rise to the following causal structure, with `weight loss` being a common effect of `eating healthy` and `exercising`.

$$\texttt{exercising} \rightarrow \texttt{weight loss} \leftarrow \texttt{eating healthy}$$

We have, thus far, learned the causal relation between the variables under study, what in *causal discovery* literature is known as *type causation*. Let us go a step beyond and assume that now we want to know the cause of Bob (see Table 1) losing weight. The answer to this query is the solution to the planning problem $P = \langle F, A, si, sj \rangle$, with $F = \{$`eating healthy`, `exercising`, `weight loss`$\}$ and A being the inferred action model. An automated planner would find the following solution to P:

$$\texttt{eat healthy} \rightarrow \texttt{lose weight 1}$$

With this example we have demonstrated that a symbolic approach is able to deal with causation in both the type-token dimension and the intrinsic-extrinsic dimension. Specifically, an action model captures how variables are related to one another, and the instantiation of such model explains the sequence of events that gave place to a particular event. With respect to the intrinsic or extrinsic nature of the learned causality, the result depends on the input observations of the action model learning algorithm. When the observations reflect the behaviour of an agent in a goal-driven environment, we learn *extrinsic causation*. On the other hand, if the environment is not goal-driven and the observation just reflects its evolution, the learned model represents *intrinsic causation*.

We would like to end this section by clarifying that, although we have used a deterministic example, probabilistic planning is a very active research subfield in automated planning. While it is true that action model learning is still in its early stages and most proposed methodologies only learn classical or deterministic models, the field is quickly evolving and we can expect soon the appearance of learning algorithms for probabilistic action models. This would open the door to not just causal structure inference but also to quantify the causal dependencies among variables. Moreover, recent successful examples in action model learning and planning in continuous spaces [3] have expanded the applicability of planning to a wider range of domains.

4 Application domains

In this section we describe two application domains where the symbolic learning approach can be used to discover the causality of the environment. First, we present a popular strategic AI game to show the acquisition of causal relationships in a setting that exhibits extrinsic causality. In this example, we will analyze the difference between learning the planning rules that model the physics (not the rules) of the game in contrast to learning the strategy of the player, which obviously follows the game rules.

Next, we present the benefits of using the learning approach in a climate science environment, a domain that exhibits an intrinsic causality in the form of climate phenomena.

4.1 Extrinsic causality in strategic games

In this section we present an example of how the learning framework can be applied to learn the behaviour of strategic games. Games represent an interesting environment that require players to engage with different situations where action must be taken in order to progress towards a target. In this type of deliberative games, unlike more reactive-like games as video games, a player typically follows some kind of strategy that we can learn via discovering the underlying action model.

A strategy is a set of rules which guides the sequence of actions of a player towards a particular goal. In contrast, a set of planning rules aims to find a solution for any valid goal. The difference between a strategy and a set of planning rules is similar to the difference between informed and uninformed search, in the sense that a strategy uses the player knowledge as a heuristic to prune the search space.

Fig. 1 Sokoban game

Let us illustrate this difference using Sokoban, a prime example of an AI game. The game starts with a grid (see Figure 1) in which four elements are present: (1) a player, (2) boxes (represented with a cross inside a square), (3) stores (represented with a small circle), and (4) obstacles (brick-like squares). The only actions allowed to the player are `move a box` and `push a box` and the goal is for all boxes to end up in a store cell. The difficulty of the game lies in that the actions are not reversible (one can `push` but not `pull`), so a bad action choice can lead to a situation from which a solution no longer exists, like a box placed in a corner. This type of situation is known as a *dead end*.

This game is usually modeled in planning with three actions or planning rules: `move`, `push_to_nongoal` and `push_to_goal`. The specification of the actions does not contain any notion of *good* or *bad* moves, the only criteria is whether it is possible or not to apply the action. This means that the only precondition needed to push a box is that the adjacent cell (any cell in the `up`, `down`, `right` or `left` direction) where the box is going to be put is `empty`.

Let us now put aside what we know about the game and assume that instead we observe an agent playing this game without knowing which game it is; and let us consider the situation in Figure 1. This figure represents our initial observation, the initial state of the game. For the sake of explanation, we will represent the position of the player and the boxes as `(player|box,row,column)`, being the initial situation `(player,3,2)` and the two boxes located at `(box1,3,1)` and `(box2,4,3)`, respectively. Let us assume we have the following observations:

1. `(player,3,4)`

2. (player,4,3) (box2,4,2)
3. (player,3,2) (box2,2,2)

What can we learn from here? The valid transitions of the learnt action model will describe how the domain changes, and from the preconditions of the transitions we can infer that box2 is always located in a cell where at least three of its adjacent cells are not occupied by an obstacle, that is, the box is never placed in a corner. This learnt statement may represent either a rule of the game or a strategy of the player, but since we know from the planning rules that the game does not impose such rule, this is clearly a strategy aimed to avoid dead ends.

4.2 Intrinsic causality in climate science

Climate science is an interesting field for causal discovery because despite basic equations governing the evolutions of the states of the atmosphere and ocean and other climate elements, there are still many factors that we do not understand [5]. Climate science temporal information typically plays a crucial role, especially when dealing with daily data. This is the reason that scientists use temporal models to identify strong, robust causal signals [6].

One interesting application in climate science is to find the relationships between the indices that represent patterns of low-frequency tropospheric height variability, that is, pattern change. Roughly speaking, atmospheric oscillation compound indices like North Atlantic Oscillation (NAO), East Pacific Oscillation (EPO) or Pacific/North America (PNA), amongst others, are used as a signal of climate changes accordingly to their positive/negative values and the slope of the trending. Hence, finding the potential causal connections between the readings or observations of these compound indices along various days can provide interesting information about climate changes.

In this particular scenario, readings of the compound indices would be accompanied with climate phenomena occurring in that day, like `snowstorm, storm south, storm east, warm conditions east, warm conditions south, cold blasts`, etc. Thus, learning the action model would lead us to causal relationships like:

- A negative NAO and negative EPO and positive PNA fosters a `storm east`
- A positive EPO favors `warm conditions east`
- A negative NAO and a positive PNA do not make a `snowstorm`

	intrinsic	extrinsic
type	How does a positive EPO affect climate conditions?	How does a corner tile affect a player's decisions?
token	What caused the snowstorm in the east yesterday?	What sequence of actions did the player follow to solve this Sokoban grid?

Table 2 Causal inference taxonomy

We must note that symbolic learning approaches typically work with discretized values. This is not though a limitation to discover causal relationships in numeric datasets.

Our aim is to give answer to the questions that raise when studying causality, regardless the particular type of causality of the application domain. Table 2 showcases the orthogonality of the causality dimensions presented in our taxonomy as well as an example of question for each combination. The application domains shown in this section demonstrate that both *intrinsic* and *extrinsic causality* are inferible with a symbolic learning approach. Additionally, the learnt action model embodies the causal structure of the domain (*type causation*), and a sequence of instantiated actions of the model reports the actual cause of a particular event (*token causation*). With these examples of application we have shown that the proposed approach is able to answer any question regarding causality and hence we can affirm it is a first step towards a common framework for causal inference.

5 Conclusions

In this paper, we argue that symbolic learning can be exploited to uncover causal relationships from observational tuples of data. This is done by acquiring the underlying action model that explains the physics of the domain or the behaviour of the agent. While it seems clear that the symbolic learning approach is suitable for extrinsic causality domains in which there exists a conscious or explicit action taking, we have shown that it is also possible to adapt the symbolic scheme to domains that exhibit intrinsic causality. Although these domains are not driven by a goal-oriented behaviour, underlying transitions between observable states are always extractable, and these transitions precisely constitute the underpinning of our proposed symbolic learning approach.

References

1. Diego Aineto, Sergio Jiménez, and Eva Onaindia. Learning STRIPS action models with classical planning. In *Intenational Conference on Automated Planning and Scheduling, ICAPS-18*, 2018.
2. Eyal Amir and Allen Chang. Learning partially observable deterministic action models. *Journal of Artificial Intelligence Research*, 33:349–402, 2008.
3. Masataro Asai and Alex Fukunaga. Classical Planning in Deep Latent Space: Bridging the Subsymbolic-Symbolic Boundary. In *National Conference on Artificial Intelligence, AAAI-18*, 2018.
4. Stephen N Cresswell, Thomas L McCluskey, and Margaret M West. Acquiring planning domain models using LOCM. *The Knowledge Engineering Review*, 28(02):195–213, 2013.
5. Imme Ebert-Uphoff. Two applications of causal discovery in climate science. In *Workshop Case Studies of Causal Discovery with Model Search*, 2013.
6. Imme Ebert-Uphoff and Yi Deng. Causal discovery from spatio-temporal data with applications to climate science. In *13th International Conference on Machine Learning and Applications, ICMLA 2014, Detroit, MI, USA, December 3-6, 2014*, pages 606–613, 2014.
7. Enrico Giunchiglia, Joohyung Lee, Vladimir Lifschitz, Norman McCain, and Hudson Turner. Nonmonotonic causal theories. *Artif. Intell.*, 153(1-2):49–104, 2004.
8. Joseph Y. Halpern and Judea Pearl. Causes and explanations: A structural-model approach. part i: Causes. *The British Journal for the Philosophy of Science*, 56(4), 2005.
9. David Heckerman, Christopher Meek, and Gregory Cooper. A bayesian approach to causal discovery. In Lakhmi C. Jain (Eds.) Dawn E. Holmes, editor, *Innovations in Machine Learning. Theory and Applications*, Studies in Fuzziness and Soft Computing, chapter 1, pages 1–28. Springer, Springer, 2006.
10. Jiuyong Li, Thuc Duy Le, Lin Liu, Jixue Liu, Zhou Jin, Bing-Yu Sun, and Saisai Ma. From observational studies to causal rule mining. *ACM TIST*, 7(2):14:1–14:27, 2016.
11. Daniel Malinsky and David Danks. Causal discovery algorithms: A practical guide. *Philosophy Compass*, 13(e12470), 2018.
12. Norman McCain and Hudson Turner. Causal theories of action and change. In *Proceedings of the Fourteenth National Conference on Artificial Intelligence and Ninth Innovative Applications of Artificial Intelligence Conference, AAAI 97, IAAI 97, July 27-31, 1997, Providence, Rhode Island.*, pages 460–465, 1997.
13. John McCarthy. Epistemological problems of artificial intelligence. In *Proceedings of the 5th International Joint Conference on Artificial Intelligence. Cambridge, MA, USA, August 22-25, 1977*, pages 1038–1044, 1977.
14. John McCarthy and Patrick Hayes. Some philosophical problems from the standpoint of artificial intelligence. *Machine Intelligence*, 4:463–502, 1969.
15. Judea Pearl. Reasoning with cause and effect. *AI Magazine*, 23(1):95–112, 2002.
16. Judea Pearl. *Causality: Models, Reasoning and Inference (2nd ed.).* Cambridge University Press, 2009.
17. Clark Glymour Peter Spirtes and Richard Scheines. *Causation, Prediction and Search (2nd ed.).* The MIT Press, 2001.
18. Peter Spirtes and Kun Zhang. Causal discovery and inference: concepts and recent methodological advances. *Applied Informatics*, 3(3), 2016.
19. Michael Thielscher. Ramification and causality. *Artif. Intell.*, 89(1-2):317–364, 1997.
20. Sofia Triantafillou and Ioannis Tsamardinos. Constraint-based causal discovery from multiple interventions over overlapping variable sets. *Journal of Machine Learning Research*, 16:2147–2205, 2015.
21. Qiang Yang, Kangheng Wu, and Yunfei Jiang. Learning action models from plan examples using weighted MAX-SAT. *Artificial Intelligence*, 171(2-3):107–143, 2007.
22. Hankz Hankui Zhuo and Subbarao Kambhampati. Action-model acquisition from noisy plan traces. In *International Joint Conference on Artificial Intelligence, IJCAI-13*, pages 2444–2450. AAAI Press, 2013.