Causal interventions with formal argumentation theory

Esteban Guerrero
Department of computing science
Umeå University, Sweden
esteban@cs.umu.se

Ailiana Santosa
School of Public Health and Community Medicine
University of Gothemburg, Sweden
ailiana.santosa@gu.se

Helena Lindgren
Department of computing science
Umeå University, Sweden
helena@cs.umu.se

ABSTRACT

Argumentation-based frameworks are used as a decision-making mechanism for software agents. This paper aims to investigate how a formal argumentation framework is affected when the underlying causal relationships of its theory is modified in counterfactual situations, the so-called "*what if*" scenarios. In contrast to previous approaches where *causality* relationships were derived from static probabilistic distributions, we address scenarios where causal models are intervened. Two novel contributions in the synergy between argumentation and causal theories are presented: 1) we characterize interventions and their consequences in causal argumentation frameworks; and 2) we introduce an account of the so-called *sequential interventions* that give a characterization of manipulations on time.

KEYWORDS

Argumentation theory, Causality, Counterfactuals, Health

ACM Reference Format:

Esteban Guerrero, Ailiana Santosa, and Helena Lindgren. 2021. Causal interventions with formal argumentation theory. In *Proc. of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2021), London, UK, May 3–7, 2021*, IFAAMAS, 3 pages.

1 MOTIVATION

Causal models are mathematical objects that provide interpretations of queries about a specific domain [7]. *Submodels* intended as the realization of a particular variable are useful for representing the effect of specific local actions, the so-called *interventions*, which answer questions such as "*what would happen if physical activities are increased in a person with pre-diabetes?*".

The generation of consistent scenarios given a probabilistic distribution from a causal model, has substantial research in the *argumentation theory* literature (see for example [10, 11]). However, during an intervention the underlying probabilistic distribution changes, and individual probabilities for those intervened variables are non-identifiable, that is, non-estimable from frequency data alone [17]. The aim of this paper is to provide a first characterization of the effects of causal models interventions on argumentation frameworks. We depart from general axiomatizations of causal models [6, 7, 9] and from an argumentation theory perspective of *deductive systems* (see [2, 5, 8]).

Our contributions are summarized as follows:

- We introduce a characterization of interventions and how those affect causal argumentation frameworks.
- We provide an account of sequential interventions, which provides a characterization of manipulations on time.

Additionally, we exemplify our theoretical findings using statistical data from a large database (more than 12 million of registers) of human activity behavior in northern Sweden [12].

2 FRAMEWORK FOR BUILDING CONSISTENT CAUSAL SCENARIOS

This section presents our framework for building consistent causal scenarios.

A *causal model* is a triple $M = \langle U, V, F \rangle$ where U and V are sets of *exogenous* and *endogenous* variables respectively, and F is a set of functions from $U \cup V$. See additional syntax details in [7]. A mapping from causal model's variables (structural equations) to a propositional language $\mathcal{L}$, allows us to form a theory Σ where atoms of the form x_1, y_1 can build complex structures (statements) about a particular causal hypothesis using basic operations: $\wedge$ (conjunction), $\neg$ (negation) and $\vdash$ (logical inference). We denote a mapped variable with the character (*). Additionally, we use the symbol $\vdash_c$ to represent a causal entailment that fulfills a set of axioms [7].

We introduce a framework called CLAIM: Causal argumentation framework, to generate sets of consistent hypotheses from a causal model that is intervened.

Definition 1 (CLAIM). *A CLAIM framework is a tuple* $\langle \Sigma, P \rangle$ *where* $\Sigma \subseteq U^* \cup V^*$ *and* U^*, V^* *are mappings from variables of a causal model* $M = \langle U, V, Fx \rangle$ *to atoms in* $\mathcal{L}$.

We can use argumentation theory in a CLAIM to build consistent scenarios from individual argument-based structures that we call *causal hypothesis chyp*s that are tuples $chyp = \langle S, \sigma, P_{\sigma'} \rangle$, fulfilling next conditions: 1) $S \subseteq A$; 2) S is consistent; 3) $S \vdash \sigma$; 4) $S \vdash_c \sigma$; and 5) $\nexists S' \subset S$ such that $S' \vdash \sigma$. Where S is called *support* and the tuple $(\sigma, P_{\sigma'})$ is the conclusion. Condition 3 and 4 ensure that σ is caused and logically inferred from a support S. A violation of these conditions may lead to a deductive correlation without causation.

Example (Causal hypotheses of PRED). Figure 1 presents a subset of causal variables for prediabetes. Two additional exogenous variables U_1 and U_2 are linked to the initial graph; these exogenous variables can be seen background (not measured) information, *e.g.* social or environmental aspects. From these structural equations, seven causal hypotheses can be built:

$$chyp_1 = \langle \{u_1\}, (u_1, P_{u_1}) \rangle$$

$$\ldots$$

$$chyp_6 = \langle \{Age \wedge u_2 \wedge u_{Age}\}, (Age, P_{Age}) \rangle$$

$$chyp_7 = \langle \underbrace{\{u_{PRED} \wedge Obe\}}_{support}, \underbrace{(PRED, P_{PRED})}_{conclusion} \rangle$$

Proc. of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2021), U. Endriss, A. Nowé, F. Dignum, A. Lomuscio (eds.), May 3–7, 2021, London, UK.

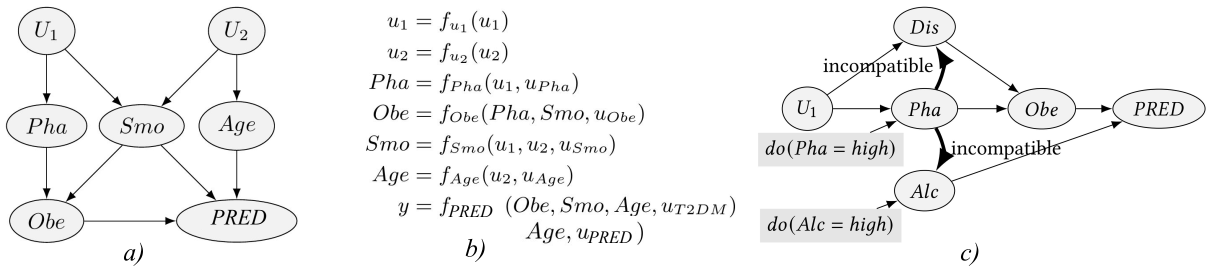

Figure 1: a) Causal graph of prediabetes (adapted from VIP data [12]); b) structural equations for prediabetes risk factors; c) double intervention of physical activity and alcohol variables.

A $chyp_7 = \langle\{u_{PRED} \wedge Obe\}, (PRED, P_{PRED})\rangle$ has an intuitive reading: "*There is a causal and deductive relationship between obesity and PRED with a probability of P_{PRED}*". Let us denote $Ch(\Sigma)$, the set of all causal hypotheses built from Σ.When a variable is manipulated, logical *incompatibilities* may emerge in the submodel [7, 15], *e.g.* a specific *physical disability* preventing a person from being active, could be considered as incompatible with an intervention $do(Pha = high)$, *i.e.* "*what if a person increases her/his physical activity level?*" (see Figure 1). This type of specific conflicts leads us to the notion of *hypothesis attack*, *i.e.* having $chyp_1, chyp_2 \in Ch$, we say that $chyp_1$ *attacks* $chyp_2$ *iff. i)* $\exists\ \sigma \in Supp(chyp_2)$ s.t. $\sigma \equiv \neg Conc(chyp_1)$, and/or *ii)* $Conc(chyp_2) \equiv \neg Conc(chyp_1)$. We use a function $att(chyp_a, chyp_b)$ that represents any type (*i* or *ii*) of attack from $chyp_a$ to $chyp_b$. A graph where causal hypotheses (*chyps*) are nodes, and attack relationships are edges will be called a *causal argumentation framework* $CAF = (Ch, \rightarrow, \mathcal{P})$, where the arrow $\rightarrow \subseteq Ch \times Ch$ represents all the attack relationships in a CAF, and $\mathcal{P}$ is the underlying probabilistic evidence sporting the causal model.

3 EFFECTS OF CAUSAL INTERVENTIONS IN ARGUMENTATION FRAMEWORKS

A causal intervention impacts a CAF in two ways: 1) it may introduce new relations of (deductive or causal) attacks between *chyps*, and 2) the underlying probability distribution changes given the intervention (variables affected by a counterfactual manipulation are those *descendants* of such manipulated variable). We denote an *intervened* CAF as CAF_x *w.r.t.* a $x \in M_x$. For example, action $do(Pha = high)$ will affect the conditional probabilities of *Pha* descendants, but the effect of *Age* (its *parent*) will remain invariant.

Proposition 1 (Probability independent argument structures). *During an intervention of a causal model, individual probabilities of the intervened variable's parents are not affected, therefore probabilities of argument-based structures linked to those parents are invariant and definable.*

Even when a causal model follows a Markov characteristic [13], individual probabilities of *chyps* are non-identifiable, then maximum and minimum boundaries ($\bar{P} = [P_{min}, P_{max}]$) of the causal inference can be defined for those argument-based children affected by the intervention (see [14, 17] for more details).

Proposition 2. *Let CAF_x be an intervened causal argumentation framework, the individual probability of every chyp $\in CAF_x$, may not be estimable from frequency data alone.*

A straightforward consequence for argument-based structures during an intervention, is that probability of those structures associated to the intervened variables' parents are not affected.

Corollary 1. *In a CAF_x the probability of argument-based structures associated to the intervened variable's parents is not affected and it is identifiable.*

Proposition 3 (Change in probability distribution of argumentation frameworks). *Let CAF and CAF_x be a causal argumentation framework and its intervened version when X is manipulated. The probability distributions associated to CAF and CAF_x are different, then for every causal argument structure except its parents, a re-computation of individual probabilities are necessary.*

In *sequential interventions* [16] (e.g. $do(Pha = high)$ then $do(Alc = high)$), we can prove that the joint effect of sequential interventions in a CAF is the same as if we perform individual interventions successively.

Proposition 4 (Sequential interventions in CAFs). *Let $X, Y \in M$ be two variables in the causal model M, and let CAF_{X^*} and CAF_{Y^*} two interventions in X, Y respectively. The set of extensions in a sequential intervention CAF_{X^*,Y^*} is the same as a joint effect of separated interventions CAF_{X^*} then CAF_{Y^*}, or conversely, iff the set of causal hypotheses of pre-intervention CAF remains invariable.*

The relevance of Proposition 4 can be seen when different interventions are applied in distinct time points in the same CAF.

4 CONCLUSIONS

We present a novel characterization of causal interventions in argumentation frameworks that opens the path to revisit some formal concepts of deductive systems from the perspective of causal theory. We made a first step to make a general conceptualization of counterfactuals in formal argumentation theory. We also report the current design and test of a tool implementing these theoretical approaches.

Our future work will be focused on three aspects: 1) the fulfillment of abstract axioms from deductive systems [1, 3, 4, 8] and causal inference [7]; and 2) the exploration of sequential and time-related interventions considering probabilistic argumentation frameworks.

Acknowledgment The authors would like to express their gratitude to the reviewers. Research was supported by Forte, the Swedish Research Council for Health, Working Life and Welfare, which supports the STAR-C project during 2019–2024 (Dnr. 2018-01461). STAR-C is a Swedish research collaboration between the Region Västerbotten and four faculties at Umeå University, including the Department of Epidemiology and Global Health at the Faculty of Medicine, the Department of Computing Science at the Faculty of Science and Technology, the Department of Social Work at the Faculty of Social Science and the Department of Culture and Media Studies at the Faculty of Arts and Humanities.

REFERENCES

[1] Leila Amgoud. 2014. Postulates for logic-based argumentation systems. *International Journal of Approximate Reasoning* 55, 9 (2014), 2028–2048.

[2] Philippe Besnard, Alejandro Garcia, Anthony Hunter, Sanjay Modgil, Henry Prakken, Guillermo Simari, and Francesca Toni. 2014. Introduction to structured argumentation. *Argument & Computation* 5, 1 (2014), 1–4.

[3] Martin Caminada and Leila Amgoud. 2007. On the evaluation of argumentation formalisms. *Artificial Intelligence* 171, 5 (2007), 286–310.

[4] Phan Minh Dung and Phan Minh Thang. 2014. Closure and Consistency In Logic-Associated Argumentation. *Journal of Artificial Intelligence Research* 49 (2014), 79–109.

[5] Vasiliki Efstathiou and Anthony Hunter. 2011. Algorithms for generating arguments and counterarguments in propositional logic. *International Journal of Approximate Reasoning* 52, 6 (2011), 672–704.

[6] David Galles and Judea Pearl. 1997. Axioms of causal relevance. *Artificial Intelligence* 97, 1-2 (1997), 9–43.

[7] David Galles and Judea Pearl. 1998. An axiomatic characterization of causal counterfactuals. *Foundations of Science* 3, 1 (1998), 151–182.

[8] Esteban Guerrero, Juan Carlos Nieves, and Helena Lindgren. 2015. Semantic-based construction of arguments: An answer set programming approach. *International Journal of Approximate Reasoning* 64 (2015), 54 – 74.

[9] Joseph Y Halpern. 2000. Axiomatizing causal reasoning. *Journal of Artificial Intelligence Research* 12 (2000), 317–337.

[10] Anthony Hunter. 2013. A probabilistic approach to modelling uncertain logical arguments. *International Journal of Approximate Reasoning* 54, 1 (2013), 47–81.

[11] Mariela Morveli-Espinoza, Juan Carlos Nieves, and Cesar Augusto Tacla. 2019. An Imprecise Probability Approach for Abstract Argumentation Based on Credal Sets. In *European Conference on Symbolic and Quantitative Approaches with Uncertainty*. Springer, 39–49.

[12] Nawi Ng, Malin Eriksson, Esteban Guerrero, Carina Gustafsson, John Kinsman, Jens Lindberg, Helena Lindgren, Kristina Lindvall, Anna Sofia Lundgren, Göran Lönnberg, Klas-Göran Sahlen, Ailiana Santosa, Linda Richter Sundberg, Lars Weinehall, and Patrik Wennberg. 2021. Sustainable behaviour change for health supported by person-Tailored, Adaptive, Risk-aware digital Coaching in a social context: Study protocol for the STAR-C research programme. *Frontiers in Public Health, section Digital Public Health* (2021). To appear.

[13] Judea Pearl. 1995. Causal diagrams for empirical research. *Biometrika* 82, 4 (1995), 669–688.

[14] Judea Pearl. 1999. Probabilities of causation: three counterfactual interpretations and their identification. *Synthese* 121, 1-2 (1999), 93–149.

[15] Judea Pearl. 2009. *Causality*. Cambridge university press.

[16] Judea Pearl. 2010. An introduction to causal inference. *The international journal of biostatistics* 6, 2 (2010).

[17] Jin Tian and Judea Pearl. 2000. Probabilities of causation: Bounds and identification. *Annals of Mathematics and Artificial Intelligence* 28, 1-4 (2000), 287–313.