A Stratification of Possibilistic Partial Explanations

Sara Boutouhami[1] and Aicha Mokhtari[2]

[1] Institut d'Informatique, USTHB, BP 32, EL Alia, Alger, Algeria
s_boutouhami@yahoo.fr
[2] Institut d'Informatique, USTHB, BP 32, EL Alia, Alger, Algeria
mokhtari_aissani@yahoo.fr

Summary. Several problems are connected, in the literature, to causality: prediction, explanation, action, planning and natural language processing... In a recent paper, Halpern and Pearl introduced an elegant definition of causal (partial) explanation in the structural-model approach, which is based on their notions of weak and actual cause [5]. Our purpose in this paper is to partially modify this definition, rather than to use a probability (quantitative modelisation) we suggest to affect a degree of possibility (a more qualitative modelisation) which is nearer to the human way of reasoning, by using the possibilistic logic. A stratification of all possible partial explanations will be given to the agent for a given request, the explanations in the first strate are more possible than those belonging to the other strates. We compute the complexity of this strafication.

1 Introduction

Causation is a deeply intuitive and familiar relation, gripped powerfully by common sense, or so it seems. But as is typical in philosophy, deep intuitive familiarity has not led to any philosophical account of causation that is at once clean and precise [3]. A source of difficulties seems to be that the notion of causality is bound to other ideas like that of explanation. In a recent paper, Halpern and Pearl propose a new definition of explanation and partial explanation, using structural equations to model counterfactuals, the definition is based on the notion of actual cause. Essentially, an explanation is a fact that is not known for certain, but if found true, would constitute an actual cause of the fact to be explained, regardless of the agent's initial uncertainty [4, 5].

Our purpose in this paper is to partially modify this definition, i. e., rather than to use a probability (quantitative modelisation) we suggest to affect a degree of possibility (qualitative modelisation) which is nearer to the human reasoning [7]. A stratification of all possible partial explanations will be given to the agent for a given request (the explanations will be ordered in a set of strates), the explanations in the first strate are more possible than those belonging to the other strates. We compute the complexity of this stratification.

S. Boutouhami and A. Mokhtari: *A Stratification of Possibilistic Partial Explanations*, Advances in Soft Computing **6**, 291–298 (2006)
www.springerlink.com

The paper is organized as follows. We present in the section 2, the structural approach, the definition of actual cause and the definition of the explanation. In section 3 we suggest to affect a degree of possibility to the definition advocated by Halpern and Pearl and then we carry out a more qualitative reasoning. We propose a stratification of all possible partial explanations; this stratification reflects a hierarchy of priority between partial explanations.In Section 4, we analyze the complexity of the algorithm of stratification. Finally, in section 5, we conclude and we give some perspectives of this work.

2 Structural Approach

Halpern and Pearl propose a definition of cause ($actual\ cause$) within the framework of structural causal models. Specifically, they express stories as a structural causal model (or more accurately, a causal world), and then provide a definition for when one event causes another, given this model of the story [4, 5]. Structural models are a system of equations over a set of random variables. We can divide the variables into two sets: endogenous (each of which has exactly one structural equation that determines their value) and exogenous (whose values are determined by factors outside the model, and thus have no corresponding equation). Capital letters X, Y, etc. will denote variables and sets of variables, and the lower-case letters x, y, etc. denote values of the sets of variables X, Y. Formally, $a\ signature\ S$ is a tuple (U, V, R), where U is a set of exogenous variables, V is a set of endogenous variables, and R associates with every variable $Y \in U \cup V$ a nonempty set $R(Y)$ of possible values for Y (that is, the set of values over which Y ranges).

A $causal\ model$ (or $structural\ model$) over signature S is a tuple $M = (S, F)$, where F associates with each variables $X \in V$ a function denoted F_X such that $F_X : (\times_{u \in U} R(U)) \times (\times_{Y \in V - \{X\}} R(Y)) \rightarrow R(X)$.$F_X$ determines the values of X given the values of all the other variables in $U \in V$. Causal models can be depicted as a $causal\ diagram$: a directed graph whose nodes correspond to the variables in V with an edge from X to Y if F_Y depends on the value of X. Given a causal model $M = (S, F)$, a (possibly empty) vector X of variable in V, and vectors x and u of values for the variables in X and U, respectively, we can define a new causal model denoted $M_{X \leftarrow x}$ over the signature $S_X = (U, V - X, R_{|V-X})$. $M_{X \leftarrow x}$ is called $a\ submodel$ of M by [6], $R_{|V-X}$ is the restriction of R to the variables in $V - X$. Intuitively, this is the causal model that results when the variables in X are set to x by some external action that effects only the variables in X. Formally $M_{X \leftarrow x} = (S_X, F^{X \leftarrow x})$, where $F_Y^{X \leftarrow x}$ is obtained from F_Y by setting the values of the variables in X to x.

Given a signature $S = (U, V, R)$, a formula of the form $X = x$, for $X \in V$ and $x \in R(X)$, is called $primitive\ event$. A basic $causal\ formula$ (over S) is one of the form $[Y_1 \leftarrow y_1, ..., Y_k \leftarrow y_k]\varphi$ where : φ is a Boolean combination of primitive events, $Y_1, ..., Y_k$ are distinct variables in V, $y_i \in R(Y_i)$. Such formula is abbreviated as $[Y \leftarrow y]\varphi$. A basic causal formula is a boolean combination of basic formulas. A causal formula ψ is true or false in a causal model, given a context. We write $(M, u) \models \psi$ if ψ is true in the causal model M given the context u.

Definition 1. *Let $M = (U,V,F)$, be a causal model. Let $X \subseteq V$, $X = x$ is an actual cause of φ if the following three conditions hold:*

- $(AC1)$*:* $(M,u) \models X = x \wedge \varphi$ *(that is, both $X = x$ and φ are true in the actual world).*
- $(AC2)$*: There exists a partition (Z,W) of V with $X \subseteq V, W \subseteq V \backslash X$ and some setting (x',w') of the variables in (X,W) such that if $(M,u) \models Z = z^*$, then both of the following conditions hold :*
  a. $(M,u) \models [X \leftarrow x', W \leftarrow w']\neg\varphi$*. In worlds, changing (X,W) from (x,w) to (x',w') changes φ from true to false.*
  b. $(M,u) \models [X \leftarrow x, W \leftarrow w', Z' \leftarrow z^*]\varphi$*.for all subsets Z' of Z.*
- $(AC3)$*: X is minimal.*

2.1 Explanation

Essentially, an explanation is a fact that is not known to be certain but, if found to be true, would constitute an actual cause of the fact to be explained, regardless of the agent's initial uncertainty. An explanation is relative to the agent's epistemic state, in that case, one way of describing an agent's state is by simply describing the set of contexts the agent considers possible [4, 5].

Definition 2. *(Explanation) Given a structural model M, $X = x$ is an explanation of φ relative to a set K of contexts if the following conditions hold:*

- *EX1:* $(M,u) \models \varphi$ *for each $u \in K$. (that is, φ must hold in all contexts the agent considers possible. The agent considers what he is trying to explain as an established fact).*
- *EX2: $X = x$ is a weak cause (without the minimal condition AC3) of φ in (M,u) for each $u \in K$ such that $(M,u) \models X = x$.*
- *EX3: X is minimal; no subset of X satisfies EX2.*
- *EX4:* $(M,u) \models \neg(X = x)$ *for some $u \in K$ and $(M,u') \models (X = x)$ for some $u' \in K$.*

Halpern and Pearl propose a sophisticated definition for actual causality based on structural causal models, however although this definition works on many previously problematic examples, it still does not fit with intuition on all examples, moreover the explanation proposed in this approach is not qualitative. To handle this problem, we propose an improvement of this definition in the next section.

3 Possibilistic Explanation

Possibilistic logic offers a convenient tool for handling uncertain or prioritized formulas and coping with inconsistency [1]. Propositional logic formulas are thus associated with weight belonging to a linearly ordered scale. In this logic, at the semantic level, the basic notion is a possibility distribution denoted by π, which is a mapping from a set of informations Ω to the interval$[0,1]$. $\pi(\omega)$ represents the

possibility degree of the interpretation ω with the available beliefs. From a possibility distribution π, two measures defined on a set of propositional or first order formulas can be determined: one is the possibility degree of formula φ, denoted $\Pi(\varphi) = max\{\pi(\omega) : \omega \models \Omega\}$, the other is the necessity degree of formula φ is defined as $N(\varphi) = 1 - \Pi(\neg\varphi)$, for more details see [7, 8].

In order to give a more qualitative character to the previous explanation, we suggest to affect a degree of possibility rather than a degree of probability. A new definition of explanation using the possibilistic logic is proposed. It offers an ordering set of possible explanations. The agent's epistemic state will be represented by describing the set of the interpretations that the agent considers possible.

Definition 3. *(Possibilistic explanation) Let ω be an interpretation that the agent considers possible ($\omega \in \Omega$). Given a structural model M,$X = x$ is an explanation of φ relative to a set Ω of possible interpretations if the following conditions hold:*

- *$EX1'$: $(M,\omega) \models \varphi$ for each $\omega \in \Omega$. (that is, φ must be satisfied in all interpretation the agent considers possible).*
- *$EX2'$: $X = x$ is a weak cause of φ in (M,ω) for each $\omega \in \Omega$ such that $(M,\omega) \models X = x$.*
- *$EX3'$: X is minimal; no subset of X satisfies $EX2'$.*
- *$EX4'$: $(M,\omega) \models \neg(X = x)$ for some $\omega \in \Omega$ and $(M,\omega') \models X = x$ for some $\omega' \in \Omega$.*

Not all explanations are considered equally good. Some explanations are more plausible than others. We propose to define the goodness of an explanation by introducing a degree of possibility (by including priority levels between explanations). The measure of possibility of an explanation is given by:

$$\Pi(X = x) = max\{\pi(\omega) : \omega \models X = x, \omega \in \Omega\}$$

There is a situations where we can't find a complete explanation of an event (relative to Ω). But we can find a complete explanation relative to a sub-set Ω' of Ω. That explanation is a partial explanation relative Ω In the next section we give our definition a partial explanation and it's goodness.

Definition 4. *(partial explanation) Let π be a possibility distribution, i.e., a mapping from a set of interpretations Ω that the agent considers possible into the interval $[0,1]$. Let $\Omega_{X=x,\varphi}$ be the largest subset such that $X = x$ is an explanation of φ (it consists of all interpretations in Ω except those where $X = x$ is true but is not a weak cause of φ).*

$$\Omega_{X=x,\varphi} = \Omega - \{\omega : \omega \in \Omega | \omega \models X = x,\ \omega \models \varphi \text{ and } X = x \text{ is not a weak cause of } \varphi\}$$

- $X = x$ is a *partial explanation* of φ with the *goodness* $\Pi(\Omega_{X=x,\varphi}|X = x) = max\{\pi(\omega)\ :\ \omega \models X = x, \omega \in \Omega_{X=x,\varphi}\}$.
- $X = x$ is a $\alpha-$*partial explanation* of φ relative to π and Ω, if $\Omega_{X=x,\varphi}$ exists and $\Pi(\Omega_{X=x,\varphi}|X = x) \geq \alpha$.

- $X = x$ is an partial explanation of φ relative to π and Ω, iff $X = x$ is a $\alpha-$partial explanation of φ and $\alpha \geq 0$.

Partial explanations will be ordered, in a set of strates $S_{\alpha 1} \cup ... \cup S_{\alpha n}$ for a given request.

- The $S_{\alpha 1}$ will contain the complete explanations if there exists,
- $X = x$ is in the strate $S_{\alpha i}$, if $\Pi(\Omega_{X=x,\varphi}|X = x) = \alpha_i$,
- Let $X = x$ be a partial explanation in the strate $S_{\alpha i}$ and $Y = y$ a partial explanation in the strate $S_{\alpha j}$. $X = x$ is a partial explanation more plausible than the partial explanation $Y = y$, if $\Pi(\Omega_{X=x,\varphi}|X = x) = \alpha_i > \Pi(\Omega_{Y=y,\varphi}|X = x) = \alpha_j$.

Example 1. Suppose I see that Victoria is tanned and I seek an explanation. Suppose that the causal model includes variables for "Victoria took a vacation", "It is sunny in the Canary Islands", "Victoria went to a tanning". The set of Omega includes interpretations for all settings of these variables compatible with Victoria being tanned. Note that, in particular, there is an interpretation where Victoria both went to the Canaries (and didn't get tanned there, since it wasn't sunny) and to a tanning salon. Victoria taking a vacation is not an explanation (relative to Omega), since there is an interpretation where Victoria went to the Canary Islands but it was not sunny, and the actual cause of her tan is the tanning salon, not the vacation. However, intuitively it is "almost" satisfied, since it is satisfied by every interpretation in Omega, in which Victoria goes to the Canaries. "Victoria went to the Canary Islands" is a partial explanation of "Victoria being tanned". There is a situation where we can't find a complete explanation (it is inexplicable).

The usual definition of a conditional distribution of possibility is:

$$\pi(\omega|\varphi) = \begin{cases} 1 \text{ if } \Pi(\varphi) = \pi(\omega) \\ \pi^-(\varphi) \text{ if } = \pi^-(\omega) < \Pi(\varphi) \textit{ and } \neg(\omega \models \varphi) \\ 0 \textit{ else} \end{cases}$$

Conditioner with φ consists on a revision of degree of possibility associated to different interpretations, after having the certain information φ. (φ is a certain information, so interpretations that falsifie φ are impossibles).
We propose the measure of *explanatory power* of $X = x$ to be $\Pi^-(\Omega_{X=x,\varphi}|X = x) = max\{\pi^-(\omega) : \omega \models X = x, \omega \in \Omega_{X=x,\varphi}\}$.

3.1 Algorithm of Generation of Strates

The main idea of our algorithm is to provide a set of choices of ordered partial explanations for a given request of the agent.

Let φ be a request for which the agent seeks an explanation. Let V be the set of endogenous variables and let $X \subseteq V - \{Y_i\}$, $\forall Y_i \in \varphi$ be a set of possible variables that may formulate the explanation. For all subset X' of X, decide if there exists an attribution of values which makes it a partial explanation. If it is the case, then

compute $\Pi(\Omega_{X'=x',\varphi}|X'=x')$. Once that is done, add this partial explanation to the appropriate strate if it exists. If not, create a new strate which will contain this partial explanation. Finally, insert the new strate in its appropriate order according to the existing strates. This algorithm gives us all the partial explanations. This structure facilitates the search of a new explanation when we have a new consideration of the agent as an adaptation with the evolution of the agent believes.

Algorithm of Generation of strates

Input $\{S=\varphi, V, \varphi, \Omega, R(X)\}$
begin

a. $X = V - \{Y_j\}$, $\forall Y_j$, Y_j is a variable in φ
b. **for** all $X' \subseteq X$ **do**
   begin
   a) Decide if there exist $x' \in R(X')$, such that $X'=x'$ is an α-partial explanation of φ relative to Ω.
   b) **if** $X'=x'$ is an α-partial explanation **then**
      begin
      i. Compute $\Pi(\Omega_{X'=x',\varphi}|X'=x')$; Let $\alpha_i = \Pi(\Omega_{X'=x',\varphi}|X'=x')$
      ii. **If** the strate S_{α_i} exists **then** Add $\{X'=x'\}$ to the strate S_{α_i}
          else
          begin
          A. Create a new strate S_{α_i}
          B. Add $\{X'=x'\}$ to the strate S_{α_i}
          C. Insert the strate S_{α_i} in the good order
          D. $S = S \cup S_{\alpha_i}$
          end
      end
   end

end
Output $S = \cup S_{\alpha_i}$

---

4 Complexity of Stratification of Possibilistic Explanations

The complexity of our algorithm is driven from the results given by Eiter and Lukasiewicz [2]. An analysis of the computational complexity of Halpern and Pearl's (causal) explanation in the structural approach is given in a recent paper by Eiter and Lukasiewicz [2].

An explanation of an observed event φ is a minimal conjunction of primitive events that causes φ even when there is uncertainty about the actual situation at hand. The main idea of the stratification is to compute all the possible partial explanations. This problem can be reduced to that of computing the set of all partial explanations

which is equivalent to computing the set of all valid formulas among a Quantified Boolean Formulas $QBF = \exists A\ \forall C\ \exists D\ y$, where $\exists A\ \forall C\ \exists D\ y$ is a reduction of guessing some $X' \subseteq X$ and $x' \in R(x)$ and deciding whether $X' = x'$ is α-partial explanation. All complexity results from the two propositions:

Proposition 1. *For all $X, Y \in V$ and $x \in R(X)$, the values F_Y and $F_Y^{X \leftarrow x}$, given an interpretation $\omega \in \Omega$, are computable in polynomial time.*

Proposition 2. *Let $X \subseteq V$ and $x \in R(X)$. Given $\omega \in \Omega$ and an event φ, deciding whether $(M, \omega) \models \varphi$ and $(M, \omega) \models [X \leftarrow x]\varphi$ (given x) hold can be done in polynomial time.*

Given $M = (U, V, F)$, $X \subseteq V$, an event φ, a set of interpretations Ω such that $(M, \omega) \models \varphi$ for all interpretations $\omega \in \Omega$, for all $X' \subseteq X$ guessing an attribution of values x' of X' ($x' \in R(X')$)such that $X' = x'$ is a partial explanation of φ. After that we compute the explanatory power of the partial explanation $X' = x'$, once that done we insert it in the appropriate strate. Computing the set of strates is $FP_{\|}^{\Sigma_2^P}$-Complete. Recall that $X' = x'$ is a partial explanation of φ iff (a) $X' = x'$ is an explanation of φ relative to $\Omega^{\varphi}_{X'=x'}$ and (b) $\Pi(\Omega_{X'=x'} | X' = x') \geq 0$; To recognize partial explanation, we need to know the set of interpretations $\Omega^{\varphi}_{X'=x'}$. $\Omega^{\varphi}_{X'=x'}$ is the set of all $\omega \in \Omega$ such that either (i) $(M, \omega) \models \neg(X' = x')$ or (ii) $(M, \omega) \models (X' = x')$ and $X' = x'$ is a weak cause of φ under ω. Deciding (i) is polynomial, and deciding (ii) is in NP, $\Omega^{\varphi}_{X'=x'}$ can be computed efficiently with parallel calls to a $NP - oracle$, computing $\Omega^{\varphi}_{X=x}$ is in $P_{\|}^{NP}$. Once $\Omega^{\varphi}_{X'=x'}$ is given, deciding (a) is possible with two $NP - oracle$ calls and deciding (b) is polynomial. Hence, the problem is in $P_{\|}^{NP}$. Deciding whether $X' = x'$ is an α-partial explanation of φ is in $P_{\|}^{NP}$. Hence, guessing some $X' \subseteq X$ and $x' \in R(X')$ and deciding whether $X' = x'$ is an α-partial explanation of φ is in Σ_2^P.

The complexity of our algorithm is inherited from the complexity of guessing a partial explanation (is a Σ_2^P-complete) and of the complexity of the explanatory power ($P_{\|}^{NP}$), this complexity is lies to the problem of computing $\Omega_{X'=x',\varphi}$. The calculus of strate, is a problem of guessing all $X' \subseteq X$ and verifying the existence of partial explanation which is $FP_{\|}^{\Sigma_2^P}$-complete, and computing there explanatory power, so that the stratification problem is $FP_{\|}^{\Sigma_2^P}$-complete.

5 Conclusion and Perspectives

In this paper we have presented a partial modification of the notion of explanation related to the counterfactual idea. We have suggested the use of the possibilistic logic which provides a priority level between the explanations. We prefer the use of possibility instead of probability because possibility reflects better the human reasoning, which is rather qualitative than quantitative.

We have proposed a stratification of all partial explanation for a given request. This stratification facilitates the task of searching a new explanation when we have

a new consideration of the agent (an evolution of the agent beliefs). We gave an analysis of the computational complexity of this stratification. As perspectives, we plan to extend this work to deal with the problem of responsibility and blame.

Acknowledgments

This work was partially supported by CMEP project (06MDU687) entitled: *Developpement des reseaux causaux possibilistes : application a la securite informatique*. The authors are indebted to Daniel Kayser for his helpful remarks.

References

[1] S. Benferhat, S. Lagrue, O. Papini. A possibilistic handling of partially ordered information. In *Proceedings UAI'03*, pages 29–36, 2003

[2] T. Eiter, T. Lukasiewicz. Complexity results for explanations in the structural-model approach. *Artificial Intelligence*, 154(1-2):145–198, 2004

[3] N. Hall, L. A. Paul. Causation and counterefactuals. *Edited by John Collins, Ned Hall and L. A. Paul*, Cloth/june, 2004

[4] J.Y. Halpern, J. Pearl. Causes and Explanations: A Structural-model Approach. *British Journal for Philosophy of Science*, To appear

[5] J.Y. Halpern, J. Pearl. Causes and explanations: A structural-model approach, Part II: Explanations. In *Proceedings IJCAI'01*, pages 27–34, Seattle, WA, 2001

[6] J. Pearl. Causality Models, Reasoning, and Inference. Cambridge University Press, New York, 2000

[7] H. Prade, D. Dubois. Possibility theory: An approach to computerized, processing of uncertainty. Plenum Press, New York, 1988

[8] L. A. Zadeh. Fuzzy sets as a basic for a theory of possibility. In *Fuzzy Sets and Systems*, 1:3–28, 1978