Steps Towards Causal Formal Concept Analysis

Alexandre Bazin, Miguel Couceiro, Marie-Dominique Devignes, Amedeo Napoli

Steps Towards Causal Formal Concept Analysis

Alexandre Bazin, Miguel Couceiro, Marie-Dominique Devignes, Amedeo Napoli

LORIA, Université de Lorraine – CNRS – Inria, Nancy, France
`firstname.name@loria.fr`

Abstract. Efficiently discovering causal relations from data and representing them in a way that facilitates their use is an important problem in science that has received much attention. In this paper, we propose an adaptation of the Formal Concept Analysis formalism to the problem of discovering and representing causal relations. We show that Formal Concept Analysis structures and algorithms are well-suited to this problem.

1 Introduction

The study of causal relations is central in applied science [30]. Experimental protocols are often implemented in order to manipulate a potential cause (an object, state or process) in such a way that its effects can be inferred. Such a process produces what is called interventional data. However, when objects cannot be directly manipulated, one has to rely on purely observational data. Studying causality from data involves two related tasks:

- discovering the causal structure, i.e., is there a causal relation between two sets of variables?
- inferring causal effects, i.e., how does the cause lead to the effect?

In this paper, we are interested in the discovery of the causal structure in observational data.

Inferring causal relations from observational data is a challenging task. Most of the approaches in the literature are probabilistic [25, 28, 37, 39, 43]. Nonetheless, recent algorithmic approaches based on approximations of Kolmogorov complexity are becoming quite popular [9, 10, 8, 34, 36]. The set of causal relations forms the causal structure. It can be presented to humans to help them understand a situation described by data and support decision making [14] or it can be used to automatically select important variables in data [49]. Most works on causality focus on univariate *causal inference* where the task consists in determining whether a causal relation exists between two variables x and y, and deciding its direction, *i.e.*, the *cause* and the *effect*. The causal structure is then represented by a directed acyclic graph called the *causal diagram*.

In practice, however, effects often require the interaction of multiple causes to appear. We then talk of *multivariate* causal relations. In the multivariate case, we allow causal relations between sets X and Y of variables. We will use $X \xrightarrow{c} Y$

to denote the fact that X causes Y. The corresponding causal diagrams have yet to be properly formalised and studied. In this paper we propose an approach based on *Formal Concept Analysis* to address and tackle this issue.

Formal Concept Analysis (FCA) [17] is a lattice-theoretic mathematical framework that enables the representation of the underlying structure of data. Originally, FCA is applied to binary data, but several extensions have been proposed to manage other data types, *e.g.*, numerical [15], fuzzy [5], incomplete [33], multidimensional [47], relational [41] and sequential [11] data. FCA is particularly suited to the discovery and representation of rule patterns such as implications [42, 40], association and decision rules [38, 50], and link keys [1].

In this paper, we propose an FCA-based framework that relies on causal inference to represent the causal structure of the variables in datasets. In Section 2, we briefly survey existing causal inference methods and recall basic FCA notions needed throughout the paper. We discuss properties of causal relations in Section 3, and use them to define a closure operator that will be the key to the proposed framework. In Section 4, we show how FCA structures can be used to represent causal structures, and provide algorithms for computing them. In Section 5, we illustrate our approach on the well-known public dataset Iris. We end the paper with a discussion on the usefulness of the framework for computing and representing causal relations in Section 6.

2 Preliminaries

2.1 Notation

Throughout the paper, we use calligraphic letters ($\mathcal{A}, B, \ldots$) to denote important sets and structures, whereas (sub)sets and individual elements are denoted by upper case letters ($A, B, \ldots$) and by lower case letters ($a, b, \ldots$), respectively. We may specify a set extensively, *e.g.*, $\{a, b, c, d, e\}$, or in the simplified form *abcde*, especially to improve readability of figures.

2.2 Causal Inference

A causal relation between a single cause x and a single effect y is called univariate. The structure of a set of such relations is usually represented by a directed acyclic graph (DAG), called a *causal diagram* or *causal network*, in which the vertices are variables and the edges denote a causal relation between the cause and its effect. Recent efforts have been devoted to learning a causal diagram [20] from observational data, whether it be the complete diagram [44, 45] or just parts of it [18]. Some approaches try to learn the skeleton of the diagram (is there a causal relation between these two variables?) first, and then to orient the edges (which of these variables is the cause?).

Inferring the direction of a causal relation between x and y is a difficult task, and authors usually rely on some assumptions. The most common one is the absence of *confounders* [29, 46, 51], *i.e.*, variables z that cause both x and

y. Also, many current approaches are based on *additive noise models* [25, 26], which assume that, if x causes y, then the value of y is a function of the value of x plus some additive noise:

$$y = f(x) + n, \text{ for some function } f,$$

and where n is independent of x. Under this assumption, state of the art methodologies [25, 28, 37, 39, 43] assume that the marginal distribution $P(x)$ of the cause x and the conditional distribution $P(y|x)$ of the effect y given the cause x are independent.

In contrast, algorithmic approaches are based on the Markov condition [29] and the idea that it is algorithmically easier to compute the effect from the cause than the other way around. The Kolmogorov complexity $K(s)$ of a finite binary string s is the length of the shortest program for which a universal Turing machine outputs s and halts. In this setting, x is said to *cause* y when $K(P(x)) + K(P(y|x)) << K(P(y)) + K(P(x|y))$. As the Kolmogorov complexity is not computable, approximations are used to infer causal directions in Boolean [9, 10] and numerical [8, 34, 36] settings.

Probabilistic [12, 27, 52] and algorithmic [48] approaches have been also proposed for the multivariate case, when causal relations are between sets of variables X and Y. However, this has received much less attention than the univariate case. In the experimental example presented in Section 5, we use ERGO [48], an algorithmic approach for inferring causal directions in the multivariate case by approximating the Kolmogorov complexity using cumulative and Shannon entropies.

2.3 Formal Concept Analysis

Formal Concept Analysis (FCA) is a mathematical framework based on lattice theory that aims at representing the information lying in binary datasets in terms of concept hierarchies and rules. Datasets are formalised as *formal contexts* that are triples $(\mathcal{G}, \mathcal{M}, \mathcal{R})$ in which $\mathcal{G}$ is a set of *objects*, $\mathcal{M}$ is a set of *attributes* and $\mathcal{R} \subseteq \mathcal{G} \times \mathcal{M}$ is a binary relation between objects and attributes. An object g is said to be *described by* an attribute m when $(g, m) \in \mathcal{R}$.

	a	b	c	d	e
1	×	×			
2		×	×	×	
3		×		×	×
4	×		×		
5				×	×

Fig. 1. A crosstable representing a formal context with five objects ($\{1, 2, 3, 4, 5\}$) and five attributes ($\{a, b, c, d, e\}$).

Given a formal context $(\mathcal{G}, \mathcal{M}, \mathcal{R})$, we can define the two following *derivation operators*:

- $\cdot' : 2^{\mathcal{G}} \to 2^{\mathcal{M}}$, $G \mapsto G' = \{m \in \mathcal{M} \mid \forall g \in G, (g, m) \in \mathcal{R}\}$, and
- $\cdot' : 2^{\mathcal{M}} \to 2^{\mathcal{G}}$, $M \mapsto M' = \{g \in \mathcal{G} \mid \forall m \in M, (g, m) \in \mathcal{R}\}$.

This pair of derivation operators forms a *Galois connection*, and so both compositions $\cdot'' : 2^{\mathcal{G}} \to 2^{\mathcal{G}}$ and $\cdot'' : 2^{\mathcal{M}} \to 2^{\mathcal{M}}$ constitute *closure operators* on $\mathcal{G}$ and $\mathcal{M}$, respectively. We say that a formal context is *object-reduced* if, for each object o, there is no pair of objects p and q such that $\{o\}' = \{p\}' \cap \{q\}'$. We define *attribute-reduced* contexts analogously.

A pair $(G, M) \in 2^{\mathcal{G}} \times 2^{\mathcal{M}}$ is called a *formal concept* of $\mathcal{K}$ if $M = G'$ and $G = M'$. In this case, G is called the *extent* and M the *intent* of the formal concept (G, M). The set of all the formal concepts of a formal context together with the partial order of extents induced by the inclusion of sets of objects or, equivalently, by the reversed inclusion of intents, forms a complete lattice called the *concept lattice* of the formal context.

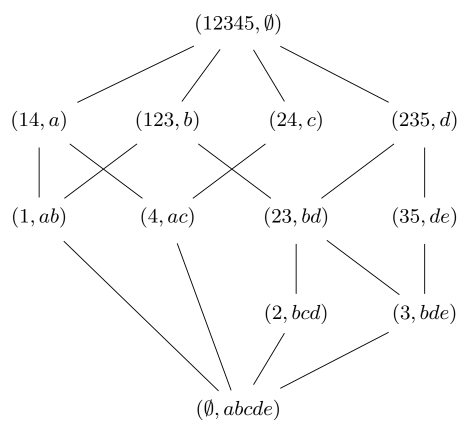

Fig. 2. Concept lattice of the formal context depicted in Fig. 1.

Definition 1. *(*IMPLICATIONS*) Let $\mathcal{K} = (\mathcal{G}, \mathcal{M}, \mathcal{R})$ be a formal context. An* implication *is a pair (X, Y) of attribute sets, often written in rule form $X \to Y$. An implication $X \to Y$* holds *in $\mathcal{K}$ when $Y' \subseteq X'$, i.e., the objects described by the attributes in X are also described by the attributes in Y. Let $\mathcal{I}_{\mathcal{K}}$ denote the set of all the implications that hold in $\mathcal{K}$.*

Implications capture certain regularities in the description of objects. For instance, in the formal context depicted in Figure 1, implications $\{b, c\} \to \{d\}$ and $\{e\} \to \{d\}$ hold, whereas $\{a\} \to \{b\}$ does not. The number of implications in a formal context $(\mathcal{G}, \mathcal{M}, \mathcal{R})$ can grow exponentially with $|\mathcal{G}|$, $|\mathcal{M}|$ and $|\mathcal{R}|$.

Hence, even relatively small formal contexts can give rise to an exponential number of implications [32]. Whether to present to a human analyst or to use in a computation, it is advantageous to avoid "redundant" implications. Two main approaches to reduce implication sets are usually considered, which may or may not entail loss of information. When loss of information is acceptable, one can use the many interestingness measures that have been proposed in the literature [19]. When information loss is to be avoided, we can make use of so-called *implication bases.*

Definition 2. *(*Implication bases*) Let $\mathcal{K}$ be a formal context. A subset $I \subseteq \mathcal{I}_{\mathcal{K}}$ is said to be an* implication base *of $\mathcal{K}$ if every implication in $\mathcal{I}_{\mathcal{K}}$ can be derived from those of I through Armstrong's axioms:*

- *If $Y \subseteq X$, then $X \to Y$.*
- *If $X \to Y$, then $X \cup Z \to Y \cup Z$ for all Z.*
- *If $X \to Y$ and $Y \to Z$, then $X \to Z$.*

As every valid implication of a formal context $\mathcal{K}$ can be derived from an implication basis of $\mathcal{K}$, implication bases contain the same information as the whole implication set $\mathcal{I}_{\mathcal{K}}$. Different implication bases, with different properties, have been studied [22, 42, 7]. Here, we present the two best-known.

Definition 3. *(*Logical closure*) Let I be an implication set. The* logical closure *$I(X)$ of an attribute set X by I is the smallest superset of X such that $(A \to B \in I$ and $A \subseteq I(X))$ implies $B \subseteq I(X)$.*

When I is an implication base, $I(X) = X''$ for all $X \subseteq \mathcal{M}$. If $\mathcal{K}$ is the context depicted in Figure 1, then $I(\{e\}) = \mathcal{I}_{\mathcal{K}}(\{e\}) = \{d, e\}$.

Definition 4. *(*Logical pseudo-closure*) Let I be an implication set. The* logical pseudo-closure *$I^{\Box}(X)$ of an attribute set X by I is the smallest superset of X such that $(A \to B \in I$ and $A \subset I^{\Box}(X))$ implies $B \subseteq I^{\Box}(X)$.*

Note that the only difference between the logical closure and pseudo-closure resides in the strictness of the inclusion of premises, and it is not difficult to verify that the logical pseudo-closure is a closure operator. For instance, if $\mathcal{K}$ is the context depicted in Figure 1, then $\mathcal{I}_{\mathcal{K}}^{\Box}(\{e\}) = \{e\}$ because no implications in which the premise is a proper subset of $\{e\}$ holds.

Definition 5. *(*Pseudo-intent*) Let $\mathcal{K} = (\mathcal{G}, \mathcal{M}, \mathcal{R})$ be a formal context. An attribute set $P \subseteq \mathcal{M}$ is called a* pseudo-intent *if $P \neq P''$ and $P = \mathcal{I}_{\mathcal{K}}^{\Box}(P)$.*

Note that the attribute sets closed under the logical pseudo-closure are either intents or pseudo-intents. Thus the set of pseudo-intents can be obtained by computing the logical pseudo-closure $\mathcal{I}_{\mathcal{K}}^{\Box}(.)$ of attribute sets [2, 4, 31, 13].

Definition 6. *(*Canonical base*) The* canonical base *of a formal context is*

$$\{P \to P'' \mid P \text{ is a pseudo-intent}\}$$

The notion of canonical base is also referred to as the *Duquenne-Guigues base* [22]. It is the smallest implication base of a formal context in terms of number of implications, and it constitutes an optimal compression of the information contained in the implications or, equivalently, in the formal context. The canonical base of the formal context depicted in Figure 1 is:

- $\{e\} \rightarrow \{d, e\}$
- $\{a, d\} \rightarrow \{a, b, c, d, e\}$
- $\{b, c\} \rightarrow \{b, c, d\}$
- $\{c, d\} \rightarrow \{b, c, d\}$
- $\{b, c, d, e\} \rightarrow \{a, b, c, d, e\}$

Definition 7. *(*Proper premise*) Let $\mathcal{K}$ be a formal context and a an attribute. An attribute set X is a* proper premise *of a if $X \rightarrow \{a\}$ holds, $X \neq \{a\}$, and, for all $Y \subset X$, $Y \rightarrow \{a\}$ does not hold.*

In words, a proper premise is a minimal attribute set that implies another attribute.

Definition 8. *(*Proper premise base*) The* base of proper premises *[42] of a formal context is defined by*

$$\{X \rightarrow X'' \mid X \textit{ is a proper premise}\}$$

The base of proper premises is the implication base with the smallest premises, and it constitutes the optimal compression of the information contained in the implications when the size of the premises is a concern. For example, the base of proper premises of Figure 1 is:

- $\{e\} \rightarrow \{d, e\}$
- $\{b, c\} \rightarrow \{b, c, d\}$
- $\{c, d\} \rightarrow \{b, c, d\}$
- $\{a, d\} \rightarrow \{a, b, c, d, e\}$
- $\{c, e\} \rightarrow \{a, b, c, d, e\}$
- $\{a, e\} \rightarrow \{a, b, c, d, e\}$
- $\{a, b, c\} \rightarrow \{a, b, c, d, e\}$

It should be noticed that minimal pseudo-intents are necessarily proper premises.

3 Causal Closure

Works on modelling or discovering causal relations often make a number of assumptions on the nature of the causal relation: univariate or multivariate, absence of confounders, linearity of the function in the Additive Noise Models... Following the same tracks, we assume that causal relations have some desirable (and generally accepted) properties that resemble those of implications. In the remainder of this paper, we assume that all variables belong to a set $\mathcal{V}$.

For every $X, Y, Z \subseteq \mathcal{V}$, suppose that the following five properties hold:

Property 1. *A set of variables causes itself*:

$$X \xrightarrow{c} X.$$

Property 2. *Causality is transitive*:

$$X \xrightarrow{c} Y \ \wedge \ Y \xrightarrow{c} Z \Rightarrow X \xrightarrow{c} Z.$$

In words, if X directly causes Y and Y directly causes Z, then X indirectly causes Z.

Property 3. *If X causes a set of variables Y, then it causes all of its subsets*:

$$X \xrightarrow{c} Y \Rightarrow X \xrightarrow{c} Z, \ \forall Z \subseteq Y.$$

In particular, X causes $\{y\}$ for all y in Y.

Property 4. *All supersets of X cause its effects*:

$$X \xrightarrow{c} Y \Rightarrow X \cup Z \xrightarrow{c} Y.$$

Note that this property is not incompatible with possible notions of *negation* as $\xrightarrow{c}$ only denotes the presence of a causal relation and not the causal effect itself, which can be different between $X \xrightarrow{c} Y$ and $X \cup Y \xrightarrow{c} Y$.

Property 5. *A variable causes the union of its effects*:

$$X \xrightarrow{c} Y \ \wedge \ X \xrightarrow{c} Z \Rightarrow X \xrightarrow{c} Y \cup Z.$$

Properties 1 and 3 imply that if $Y \subseteq X$, then $X \xrightarrow{c} Y$. Properties 1, 3 and 4 imply that if $X \xrightarrow{c} Y$, then $X \cup Z \xrightarrow{c} Y \cup Z$. Since Property 2 is transitivity, these five properties imply Armstrong's axioms (see Definition 2). Note that transitivity, despite intuitive, is disputed [35] and counterexamples have been presented. Halpern [23] identified conditions under which causality is transitive.

Proposition 1. *Consider the operator $\xi \colon 2^{\mathcal{V}} \to 2^{\mathcal{V}}$ defined by*

$$\xi(X) = \bigcup_{X \xrightarrow{c} Y} Y.$$

Then ξ is a closure operator.

Proof. By Property 1, we have that $X \subseteq \xi(X)$ so $\xi(.)$ is extensive. From Properties 1, 2 and 3, we have that $X \subseteq Y \Rightarrow \xi(X) \subseteq \xi(Y)$, for every $X, Y \subseteq \mathcal{V}$, and thus $\xi(.)$ is monotone. Furthermore, by Properties 1 and 2 we also have that $\xi(\xi(X)) = \xi(X)$, for every $X \subseteq \mathcal{V}$, and hence ξ is idempotent. This shows that ξ is indeed a closure operator.

We refer to $\xi(X)$ as the *causal closure* of X. The causal closure of X is made of X and all of its direct and indirect effects. When all causal relations are known, the causal closure can be computed by applying ξ successively until a fixpoint is reached. However, inferring causal relations from a dataset may be costly, which prevents us from testing whether $X \xrightarrow{c} Y$ for all possible Y. As we will now observe, Property 3 ensures that the computation of $\xi(X)$ can be simplified.

Proposition 2. *Let X be a variable set and y a variable. Then,*

$$\xi(X) = \{y \mid X \xrightarrow{c} \{y\}\}$$

Proof. This follows from the fact that $X \xrightarrow{c} Y \Rightarrow X \xrightarrow{c} Z,\ \forall Z \subseteq Y$.

Hence, computing $\xi(X)$ directly from a dataset requires $|\mathcal{V} \setminus X|$ causal tests.

4 Reconstructing the FCA Trinity with Causality

In the previous section we showed that causal relations give rise to a closure operator ξ. Let $\mathcal{V}$ be the set of variables in a dataset. Suppose that $\mathcal{K}_c$ is a formal context whose set of attributes is $\mathcal{V}$, and that an implication $X \rightarrow Y$ holds if and only if $X \xrightarrow{c} Y$. As causal relations respect Armstrong's axioms, $\mathcal{I}_{\mathcal{K}_c}(X) = \xi(X)$, for all $X \subseteq \mathcal{V}$, and any implication base of $\mathcal{K}_c$ enables the derivation of all the causal relations. Thus, by using the closure operator ξ as if it were the closure operator induced by the (unknown) formal context $\mathcal{K}_c$, it is possible to apply the FCA machinery to compute representations of the causal relations in the dataset. FCA offers many structures to represent the content of the closure operator and, thus, of the causal relations: formal contexts, concept lattices, AOC-posets [21], and implication bases. Each of them provides its unique perspective, but some are easier to handle when the closure operator is thought of as a *black box*.

In this section, we go through these structures and discuss both their usefulness as representations of causal relations and their ease of use. We illustrate the various notions and results that we present on the following set of causal relations from a dataset with five variables:

- v_1 causes v_2: $\{v_1\} \xrightarrow{c} \{v_2\}$
- v_3 causes v_4: $\{v_3\} \xrightarrow{c} \{v_4\}$
- v_1 and v_4 together cause v_5: $\{v_1, v_4\} \xrightarrow{c} \{v_5\}$

4.1 The Causal Canonical Base

The canonical base of $\mathcal{K}_c$ is the smallest implication base of $\mathcal{K}_c$, that is, it corresponds to the smallest set of causal relations that allow the derivation of all the causal relations in the dataset. We call it the *causal canonical base* of the dataset. It is a set of multivariate causal relations of the form $X \rightarrow \xi(X)$ (analogue to the implications $X \rightarrow X''$ in classical FCA) in which X contains all the variables it causes except for those specified by the rule $X \rightarrow \xi(X)$ itself. In our running example, the causal canonical base is

- $\{v_1\} \rightarrow \{v_2\}$
- $\{v_3\} \rightarrow \{v_4\}$
- $\{v_1, v_2, v_4\} \rightarrow \{v_5\}$

As the smallest rule-based representation of causal relations, the causal canonical base is most useful when the number of rules is a concern: when visualising the structure or when storing it for computer use. This base is also arguably the easiest representation to compute as existing algorithms for computing pseudo-closed sets make use of the closure operator as a black box. For instance, the NEXT CLOSURE [16] algorithm (Algorithm 1) can be used. This algorithm relies on the fact that the sets of variables closed under $\xi^{\Box}$ are either closed under ξ or premises of the causal relations in the causal canonical base. It enumerates the elements of this closure system and, for each one, checks whether it is closed under ξ or a premise. In order to avoid computing the same element twice, it enumerates the sets in the *lectic order*: given an arbitrary ordering of variables, $A \leq B$ if and only if the smallest element in $(A \cup B) \setminus (A \cap B)$ is in B.

Algorithm 2 uses the NEXT procedure (Algorithm 2) that takes a set of variables as input and returns the set immediately succeeding it in the lectic order. When applied to the running example, Algorithm 1 starts with $V = \emptyset$. Assuming that $v_1 < v_2 < v_3 < v_4 < v_5$, the algorithm then enumerates the sets $\{v_5\}$, $\{v_4\}$ and $\{v_4, v_5\}$. As all of them are closed under ξ, no causal relation is added to I. Then, the algorithm reaches $\{v_3\}$. As $\xi(\{v_3\}) = \{v_3, v_4\}$, the causal relation $\{v_3\} \rightarrow \{v_4\}$ is added to the implication set I. The algorithm continues until it reaches $\{v_1, v_2, v_3, v_4, v_5\}$, at which point it stops and outputs I, which contains the three causal relations in the causal canonical base.

Algorithm 1: Causal NEXT CLOSURE for implications

Input: Dataset with variables $\mathcal{V}$
Output: The causal canonical base of the dataset

begin
  $I = \emptyset$;
  $V = \emptyset$;
  while $V \neq \mathcal{V}$ **do**
    if $V \neq \xi(V)$ **then**
      $I = I \cup \{V \rightarrow \xi(V)\}$;
    V = NEXT(V);
return I

Different algorithms can be chosen to compute the logical closure of a set (see [3]). The time complexity of Algorithm 2 is at most linear with respect to the complexity of the chosen algorithm, since Algorithm 2 performs at most $|\mathcal{V}|$ logical closures. The naïve logical closure algorithm is quadratic in the number of rules and, in this case, the complexity of Algorithm 2 would be quadratic in

Algorithm 2: NEXT

Input: Variable set $V \in \mathcal{V}$, causal relation set I
Output: The next variable set closed under $\xi^{\Box}$ in the lectic order

```
for every variable x ∈ 𝒱 \ V in decreasing order do
    W = I({v ∈ V | v < x} ∪ {x});
    if min(W \ V) = x then
        return W;
```

the size of the causal canonical base. Algorithm 1 calls Algorithm 2 as many times as there are sets closed under $\xi^{\Box}$, which is $2^{|\mathcal{V}|}$ in the worst case.

4.2 Causal Proper Premises

The base of proper premises of $\mathcal{K}_c$ is the implication base with the smallest premises. It is made of rules of the form $X \rightarrow \xi(X)$ in which X is a *minimal* (called a *sufficient*) cause of the variables $y \in \xi(X) \setminus X$. We call this base the *causal sufficiency base*. In our running example, the causal sufficiency base is

- $\{v_1\} \rightarrow \{v_2\}$
- $\{v_3\} \rightarrow \{v_4\}$
- $\{v_1, v_3\} \rightarrow \{v_5\}$
- $\{v_1, v_4\} \rightarrow \{v_5\}$

Sufficient causes are particularly important because they allow us to pinpoint what can be acted on to predict or modify the future. For instance, if the association of high blood pressure and sedentary lifestyle is found to be a sufficient cause of the development of a disease, correcting either of those will certainly help preventing the disease. The causal sufficiency base is therefore interesting as it readily contains all the sufficient causes of all the variables. However, it is more difficult to compute than the causal canonical base. Indeed, computing proper premises is usually seen as computing the minimal transversals of hypergraphs constructed from the formal context [42]. In our case, we only have a closure operator and the formal context is unknown. Thus, computing the proper premises of the causal relation requires the use of a well-known FCA algorithm tailor-made for when the underlying formal context is unknown: ATTRIBUTE EXPLORATION [17].

ATTRIBUTE EXPLORATION is an algorithm for computing an implication base of a formal context that is only known to an expert. It works by repeatedly asking the expert whether an implication holds and, if not, to provide a counterexample in the form of an object whose description invalidates the implication. The algorithm thus produces an implication base and a formal context that contains the same information as the underlying, unknown formal context. In our case, we use the version proposed in [42] for proper premises and the expert is played by the causal closure operator ξ in Algorithm 3.

Algorithm 3: Causal ATTRIBUTE EXPLORATION.

Input: Dataset with variables $\mathcal{V}$

Output: The causal sufficiency base of the dataset

1 **begin**
2    $I = \emptyset$;
3    $\mathcal{G} = \{o_1\}$;
4    $\mathcal{R} = \emptyset$;
5    $K = (\mathcal{G}, \mathcal{V}, \mathcal{R})$;
6    $i = 2$;
7   **for** $a \in \mathcal{V}$ **do**
8      $\mathcal{E} = H^{\notin}_{K,a}$;
9      $T = \{\emptyset\}$;
10   **while** *there exists* $E \in \mathcal{E}$ **do**
11        $T = min(T \vee \{\{a\} \mid a \in E\})$;
12        $\mathcal{E} = \mathcal{E} \setminus \{E\}$;
13   **while** *there exists* $Q \in T$ *with* $I \nvDash \{Q \rightarrow Q''\}$ **do**
14   **if** $Q'' \subseteq \xi(Q)$ **then**
15            $I = I \cup \{Q \rightarrow Q''\}$;
16   **else**
17            $\mathcal{G} = \mathcal{G} \cup \{o_i\}$;
18            $\mathcal{R} = \mathcal{R} \cup \{(o_i, x) \mid x \in \xi(Q)\}$;
19   **if** $a \notin o_i'$ **then**
20              $\mathcal{E} = \mathcal{E} \cup \{\mathcal{V} \setminus o_i'\}$;
21            $i = i + 1$;
22   **return** I

Initially, we suppose that we have no information about the unknown formal context, *i.e.*, the causal relations. So we start with an empty implication set I and the trivial formal context $K = (\{o_1\}, \mathcal{V}, \emptyset)$ with a single object (lines 2–5). The algorithm then enumerates implications $Q \rightarrow Q''$ that hold in K and checks whether $Q \xrightarrow{c} Q''$ by computing $\xi(Q)$. If $Q \rightarrow Q''$ does not hold in the underlying context because Q does not cause Q'', the counterexample is a new object o such that $o' = \xi(Q)$. The proper premises of an attribute a in the context K are the minimal transversals of the hypergraph

$$H^{\notin}_{K,a} = \{\overline{o'} \setminus \{a\} \mid a \notin o'\}.$$

where $\overline{o'} = \mathcal{M} \setminus o'$. Thus, the implications are enumerated as follows.

For each attribute a (line 6), the algorithm computes $H^{\notin}_{K,a}$ (line 7) and then incrementally computes its set of minimal transversals T using Berge's multiplication algorithm [6] (lines 9–11). For each transversal, and thus each proper premise Q (line 12), the validity of $Q \rightarrow \xi(Q)$ is tested (line 13). If the causal relation holds, $Q \rightarrow Q''$ is added to the set of implications I (lines 13–14).

If it does not hold, a new object o_i whose description is $\xi(Q)$ is created (lines 15–19).

When applied on the running example, Algorithm 3 starts with the trivial context $K = (\{o_1\}, \mathcal{V}, \emptyset)$. Suppose that the first variable to be considered is $a = v_1$. As the context is empty but still contains an object, the hypergraph $H^{\not\subseteq}_{K,v_1}$ contains a single hyperedge, $\{v_2, v_3, v_4, v_5\}$ and its set of minimal transversals is $\{\{v_2\}, \{v_3\}, \{v_4\}, \{v_5\}\}$. We have that $\{v_3\}'' = \{v_1, v_2, v_3, v_4, v_5\}$ and $I \not\models \{v_3\} \rightarrow \{v_1, v_2, v_3, v_4, v_5\}$. Hence, the algorithm enters the second while loop (line 13) for $Q = v_3$. As $\{v_3\}'' \not\subseteq \xi(\{v_3\}) = \{v_3, v_4\}$, a new counterexample is created and added to the context. This counterexample o_i is such that $o_i' = \xi(\{v_3\})$. The algorithm then continues until all five variables have been considered, at which point it outputs the set I of all causal relations in the causal sufficiency base.

4.3 Causal Intent Lattice and Causal Intent Context

The lattice made of the sets of variables closed under ξ and ordered by inclusion, is another representation of the causal structure. Closed sets C are the maximal elements of their equivalence classes $C_{\equiv}$ that contain all the variable sets that have the same direct and indirect causes as C. We call them *causal intents*. The causal intent lattice corresponding to our running example is depicted in Fig. 3. The set $\{v_3, v_4\}$ is a causal intent as it is the maximal element of the equivalence class that contains $\{v_3\}$ and $\{v_3, v_4\}$. The fact that $\{v_3\}$ belongs to the same equivalence class as $\{v_3, v_4\}$ means that v_3 causes v_4. As such, the causal relations can be inferred from the causal intent lattice. However, despite the similarities between the graphical representations of causal intent lattices and causal diagrams [20], the causal relations do not directly correspond to edges in the lattice. For instance, the edge between $\{v_2, v_5\}$ and $\{v_2, v_4, v_5\}$ does not mean that $\{v_2, v_5\} \xrightarrow{c} \{v_4\}$. Similarly, the edge between $\{v_2\}$ and $\{v_1, v_2\}$ does not mean that $\{v_2\} \xrightarrow{c} \{v_1\}$. Instead, the two closed sets $\{v_2\}$ and $\{v_1, v_2\}$ together with the fact that $\{v_1\}$ is not closed mean that all the objects described by v_1 are also described by v_2 while some objects are described by v_2 and not by v_1. Hence, it means that $\{v_1\} \xrightarrow{c} \{v_2\}$.

The formal context induced by the closure operator ξ (and for which the causal intent lattice is the intent lattice) is called the *causal context*. The causal context whose objects are $\wedge$-irreducible is called the *reduced causal context*. The causal context of our running example is depicted in Fig 4.

5 A Concrete Example

The FCA based framework that we proposed to computing the causal structure requires the ξ operator. In practice, it is hard to infer all the causal relations from an observational dataset. Thus we have to rely on approximations provided by the various approaches presented in Subsection 2.2.

In our framework we suppose that a variable set X causes a variable y if and only if X and $\{y\}$ are sufficiently correlated and that $X \xrightarrow{c} \{y\}$ is inferred

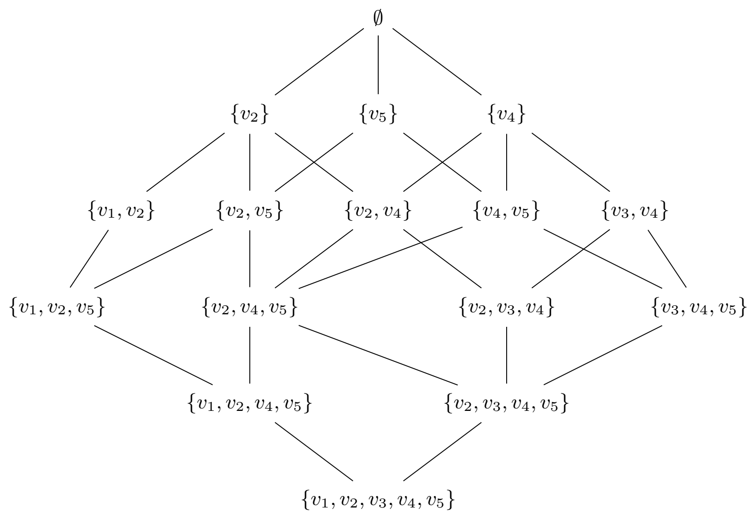

Fig. 3. Causal intents lattice of the running example.

	v_1	v_2	v_3	v_4	v_5
	×	×		×	×
		×	×	×	×
	×	×			×
		×	×	×	
			×	×	×
	×	×			

Fig. 4. The reduced causal context of the running example.

by ERGO, the multivariate causal relation inference approach proposed in [48]. The correlation between the variable sets X and $\{y\}$, denoted by $corr(X, \{y\})$, is measured using Linear Canonical Correlation Analysis [24] and we consider that two subsets X and Y are sufficiently correlated if $corr(X, \{y\}) > 0.8$.

The Iris dataset is well known in pattern recognition and machine learning. It describes 150 flowers of the *iris* genus with the values of five variables *sepal_length*, *sepal_width*, *petal_length*, *petal_width* and *class*. The first four are numerical variables while the class is nominal with three possible values.

In the Iris dataset, both the causal canonical base and the causal sufficiency base contain the same rules:

– $\{petal\_width\} \rightarrow \{petal\_length, class\}$
– $\{petal\_length\} \rightarrow \{class\}$
– $\{sepal\_length\} \rightarrow \{petal\_length, petal\_width, class\}$

We observe that the *class* is caused by *petal_width*, *petal_length* and *sepal_length* but not by *sepal_width*. Additionally, *petal_length* appears to be a direct cause of *class*, itself caused by *petal_width*, itself caused by *sepal_length*.

The corresponding reduced causal context is depicted in Fig. 5 and the causal intent lattice in Fig. 6. The causal relations can be read from the lattice, *e.g.*, $\{petal\_length\} \overset{c}{\rightarrow} \{class\}$ can be inferred from the fact that $\{petal\_length\}$ belongs to the equivalence class of $\{petal\_length, class\}$. Hence, from the Iris dataset, we infer that the classes of irises are directly "caused" by the petal length, itself caused by both the petal width and the sepal length.

	sepal_length	*sepal_width*	*petal_length*	*petal_width*	*class*
	×		×	×	×
		×	×	×	×
		×	×		×
		×			×
		×			

Fig. 5. The reduced causal context of the Iris dataset.

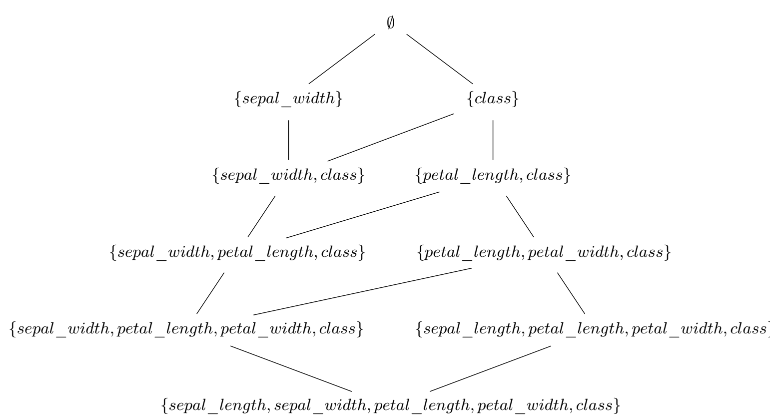

Fig. 6. Causal intent lattice of the Iris dataset.

6 Discussion and Conclusion

In this paper, we adapted the FCA framework to compute and represent (multivariate) causal relations in datasets. Inference of (multivariate) causal relations is a challenging problem, for which the current state of the art approaches only manage to obtain satisfactory results in very specific settings. Our framework uses causal inference as a black box, and it can integrate by design any existing (or future) approaches to causal inference. As FCA can be extended to different data types, causal inference can also be adapted to different data types. This not only includes Boolean or numerical data, but also more complex data such as sequences. However, one must be careful when extending the causal framework in this paper. The properties of causality discussed in Section 3 are necessary for defining the underlying closure operator, and some of them, *e.g.*, transitivity, do not necessarily hold.

Also, our empirical studies used a combination of ERGO and linear canonical correlation analysis for inferring causal relations. Using different approaches for measuring correlation and orienting the causal direction [12, 27, 52] could have produced different causal relations and thus a different lattice and a different causal context. The algorithms' performance for univariate causal inference was evaluated on real and synthetic datasets [37] for which the true causal relations are known. In the multivariate case, there are very few datasets with ground truth, and they only contain single cause and effect examples. This is clearly not suited to evaluate an approach conceived to representing multiple multivariate causal relations.

Furthermore, it is not clear how to generate synthetic data according to multivariate causal relations. Having in mind the comparison of algorithmic approaches, we think that the next milestone in the study of causal inference is the development of new approaches for generating synthetic datasets with multivariate causal relations. In general, structures with cycles such as $\{\{a\} \overset{c}{\to} \{b\}, \{b\} \overset{c}{\to} \{a, c\}\}$, should also be taken into account. However, the existing causal inference approaches cannot deal with such relations. Being able to generate datasets containing such structures would thus pave the way to new (multivariate) causal inference approaches.

Acknowledgements

This work was supported partly by the french PIA project "Lorraine Université d'Excellence", reference ANR-15-IDEX-04-LUE, by TAILOR, a project funded by EU Horizon2020 research and innovation programme under GA No 952215, and by the Inria ProjectLab "Hybrid Approaches for Interpretable AI" (HyAIAI).

References

1. Manuel Atencia, Jérôme David, Jérôme Euzenat, Amedeo Napoli, and Jérémy Vizzini. Link key candidate extraction with relational concept analysis. *Discrete Applied Mathematics*, 273:2–20, 2020.
2. Konstantin Bazhanov and Sergei Obiedkov. Optimizations in computing the Duquenne–Guigues basis of implications. *Annals of mathematics and artificial intelligence*, 70(1):5–24, 2014.
3. Konstantin Bazhanov and Sergei A. Obiedkov. Comparing performance of algorithms for generating the duquenne-guigues basis. In Amedeo Napoli and Vilém Vychodil, editors, *Proceedings of The Eighth International Conference on Concept Lattices and Their Applications, Nancy, France, October 17-20, 2011*, volume 959 of *CEUR Workshop Proceedings*, pages 43–57. CEUR-WS.org, 2011.

4. Alexandre Bazin and Jean-Gabriel Ganascia. Computing the Duquenne–Guigues basis: An algorithm for choosing the order. *International Journal of General Systems*, 45(2):57–85, 2016.
5. Radim Bělohlávek. Concept lattices and order in fuzzy logic. *Annals of pure and applied logic*, 128(1-3):277–298, 2004.
6. Claude Berge. Hypergraphs, North-Holland Mathematical Library. 1989.
7. Karell Bertet and Bernard Monjardet. The multiple facets of the canonical direct unit implicational basis. *Theoretical Computer Science*, 411(22-24):2155–2166, 2010.
8. Patrick Blöbaum, Dominik Janzing, Takashi Washio, Shohei Shimizu, and Bernhard Schölkopf. Cause-effect inference by comparing regression errors. In *International Conference on Artificial Intelligence and Statistics*, pages 900–909. PMLR, 2018.
9. Kailash Budhathoki and Jilles Vreeken. MDL for causal inference on discrete data. In *2017 IEEE International Conference on Data Mining (ICDM)*, pages 751–756. IEEE, 2017.
10. Kailash Budhathoki and Jilles Vreeken. Origo: causal inference by compression. *Knowledge and Information Systems*, 56(2):285–307, 2018.
11. Aleksey Buzmakov, Elias Egho, Nicolas Jay, Sergei O Kuznetsov, Amedeo Napoli, and Chedy Raïssi. On mining complex sequential data by means of fca and pattern structures. *International Journal of General Systems*, 45(2):135–159, 2016.
12. Zhitang Chen, Kun Zhang, and Laiwan Chan. Nonlinear causal discovery for high dimensional data: A kernelized trace method. In *2013 IEEE 13th International Conference on Data Mining*, pages 1003–1008. IEEE, 2013.
13. Felix Distel and Barış Sertkaya. On the complexity of enumerating pseudo-intents. *Discrete Applied Mathematics*, 159(6):450–466, 2011.
14. Ellery Eells. *Rational decision and causality.* Cambridge University Press, 2016.
15. Bernhard Ganter and Sergei O Kuznetsov. Pattern structures and their projections. In Harry S. Delugach and Gerd Stumme, editors, *Conceptual Structures: Broadening the Base, 9th International Conference on Conceptual Structures, ICCS 2001, Stanford, CA, USA, July 30-August 3, 2001, Proceedings*, volume 2120 of *Lecture Notes in Computer Science*, pages 129–142. Springer, 2001.
16. Bernhard Ganter and Klaus Reuter. Finding all closed sets: A general approach. *Order*, 8(3):283–290, 1991.
17. Bernhard Ganter and Rudolf Wille. *Formal Concept Analysis: Mathematical Foundations*. Springer - Verlag, 1999.
18. Tian Gao and Qiang Ji. Local causal discovery of direct causes and effects. *Advances in Neural Information Processing Systems*, 28:2512–2520, 2015.
19. Liqiang Geng and Howard J Hamilton. Interestingness measures for data mining: A survey. *ACM Computing Surveys (CSUR)*, 38(3):9–es, 2006.
20. Clark Glymour, Kun Zhang, and Peter Spirtes. Review of causal discovery methods based on graphical models. *Frontiers in genetics*, 10:524, 2019.
21. Robert Godin and Hafedh Mili. Building and maintaining analysis-level class hierarchies using galois lattices. In *Proceedings of the eighth annual conference on Object-oriented programming systems, languages, and applications*, pages 394–410, 1993.
22. Jean-Louis Guigues and Vincent Duquenne. Familles minimales d'implications informatives résultant d'un tableau de données binaires. *Mathématiques et Sciences humaines*, 95:5–18, 1986.
23. Joseph Y Halpern. Sufficient conditions for causality to be transitive. *Philosophy of Science*, 83(2):213–226, 2016.

24. Harold Hotelling. Relations between two sets of variates. *Biometrika*, 1936.
25. Patrik Hoyer, Dominik Janzing, Joris M Mooij, Jonas Peters, and Bernhard Schölkopf. Nonlinear causal discovery with additive noise models. *Advances in neural information processing systems*, 21:689–696, 2008.
26. Shoubo Hu, Zhitang Chen, Vahid Partovi Nia, Laiwan Chan, and Yanhui Geng. Causal inference and mechanism clustering of a mixture of additive noise models. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, pages 5212–5222, 2018.
27. D Janzing, P Hoyer, and B Schölkopf. Telling cause from effect based on high-dimensional observations. In *27th International Conference on Machine Learning (ICML 2010)*, pages 479–486. Omnipress, 2010.
28. Dominik Janzing, Joris Mooij, Kun Zhang, Jan Lemeire, Jakob Zscheischler, Povilas Daniušis, Bastian Steudel, and Bernhard Schölkopf. Information-geometric approach to inferring causal directions. *Artificial Intelligence*, 182:1–31, 2012.
29. Dominik Janzing and Bernhard Schölkopf. Causal inference using the algorithmic markov condition. *IEEE Transactions on Information Theory*, 56(10):5168–5194, 2010.
30. Samantha Kleinberg. *Why: A guide to finding and using causes.* "O'Reilly Media, Inc.", 2015.
31. Sergei O Kuznetsov. On the intractability of computing the Duquenne–Guigues base. *Journal of Universal Computer Science*, 10(8):927–933, 2004.
32. Sergei O Kuznetsov and Sergei Obiedkov. Counting pseudo-intents and #P-completeness. In *Formal Concept Analysis*, pages 306–308. Springer, 2006.
33. Jinhai Li, Changlin Mei, and Yuejin Lv. Incomplete decision contexts: approximate concept construction, rule acquisition and knowledge reduction. *International Journal of Approximate Reasoning*, 54(1):149–165, 2013.
34. Alexander Marx and Jilles Vreeken. Identifiability of cause and effect using regularized regression. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 852–861, 2019.
35. Neil McDonnell. Transitivity and proportionality in causation. *Synthese*, 195(3):1211–1229, 2018.
36. Osman Mian, Alexander Marx, and Jilles Vreeken. Discovering fully oriented causal networks. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 2021.
37. Joris M Mooij, Jonas Peters, Dominik Janzing, Jakob Zscheischler, and Bernhard Schölkopf. Distinguishing cause from effect using observational data: methods and benchmarks. *The Journal of Machine Learning Research*, 17(1):1103–1204, 2016.
38. Nicolas Pasquier. Mining association rules using formal concept analysis. In *ICCS'2000 International Conference on Conceptual Structures*, volume 1867, pages 259–264, 2000.
39. Jonas Peters, Joris M Mooij, Dominik Janzing, and Bernhard Schölkopf. Causal discovery with continuous additive noise models. *Journal of Machine Learning Research*, 15:2009–2053, 2014.
40. José Manuel Rodríguez-Jiménez, Pablo Cordero, Manuel Enciso, and Angel Mora. Negative attributes and implications in formal concept analysis. *Procedia Computer Science*, 31:758–765, 2014.
41. Mohamed Rouane-Hacene, Marianne Huchard, Amedeo Napoli, and Petko Valtchev. Relational concept analysis: mining concept lattices from multi-relational data. *Annals of Mathematics and Artificial Intelligence*, 67(1):81–108, 2013.

42. Uwe Ryssel, Felix Distel, and Daniel Borchmann. Fast algorithms for implication bases and attribute exploration using proper premises. *Annals of Mathematics and Artificial Intelligence*, 70(1):25–53, 2014.
43. Eleni Sgouritsa, Dominik Janzing, Philipp Hennig, and Bernhard Schölkopf. Inference of cause and effect with unsupervised inverse regression. In *Artificial intelligence and statistics*, pages 847–855. PMLR, 2015.
44. Peter Spirtes, Clark Glymour, Richard Scheines, Stuart Kauffman, Valerio Aimale, and Frank Wimberly. Constructing bayesian network models of gene expression networks from microarray data. 2000.
45. Peter Spirtes, Clark N Glymour, Richard Scheines, and David Heckerman. *Causation, prediction, and search*. MIT press, 2000.
46. Xiaohai Sun, Dominik Janzing, and Bernhard Schölkopf. Causal reasoning by evaluating the complexity of conditional densities with kernel methods. *Neurocomputing*, 71(7-9):1248–1256, 2008.
47. George Voutsadakis. Polyadic concept analysis. *Order*, 19(3):295–304, 2002.
48. Jilles Vreeken. Causal inference by direction of information. In *Proceedings of the 2015 SIAM International Conference on Data Mining*, pages 909–917. SIAM, 2015.
49. Kui Yu, Xianjie Guo, Lin Liu, Jiuyong Li, Hao Wang, Zhaolong Ling, and Xindong Wu. Causality-based feature selection: Methods and evaluations. *ACM Computing Surveys (CSUR)*, 53(5):1–36, 2020.
50. Yanhui Zhai, Deyu Li, and Kaishe Qu. Decision implication canonical basis: a logical perspective. *Journal of computer and system sciences*, 81(1):208–218, 2015.
51. K Zhang and A Hyvärinen. On the identifiability of the post-nonlinear causal model. In *25th Conference on Uncertainty in Artificial Intelligence (UAI 2009)*, pages 647–655. AUAI Press, 2009.
52. Jakob Zscheischler, Dominik Janzing, and Kun Zhang. Testing whether linear equations are causal: a free probability theory approach. In *Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence*, pages 839–848, 2011.