Parallel Attribute Exploration

Francesco Kriegel

Institute of Theoretical Computer Science, TU Dresden, Germany
`francesco.kriegel@tu-dresden.de`

Abstract. The canonical base of a formal context is a minimal set of implications that is sound and complete. A recent paper has provided a new algorithm for the parallel computation of canonical bases. An important extension is the integration of expert interaction for *Attribute Exploration* in order to explore implicational bases of inaccessible formal contexts. This paper presents and analyzes an algorithm that allows for *Parallel Attribute Exploration*.

Keywords: Formal Concept Analysis · Attribute Exploration · Canonical Base · Implication · Parallel Algorithm · Expert Interaction · Supervised Learning

1 Introduction

Implications provide an easily understandable means of logical knowledge representation. When learning terminological knowledge, it is hence straight-forward to extract valid implications from a data-set. If the data-set is complete, i.e., fully describes all individuals in the domain of interest, and if furthermore the data-set has a propositional structure or is represented as a formal context, then it suffices to compute the canonical base which is sound and complete for the set of implications holding in the data-set. This result has been found by Guigues and Duquenne [8] in the field of *Formal Concept Analysis*, where the authors utilize the notion of pseudo-intents to construct such canonical bases. A strong advantage of a canonical base is its minimality, i.e., there is no smaller set of implications which is sound and complete for the underlying data-set. Ganter [5] introduced the algorithm `NextClosure` for the computation of canonical bases, and proved its correctness. Previously, in the field of database theory, the deduction of dependencies has been investigated in Maier [12], but however no explicit construction of a base of dependencies was provided.

For the case of incomplete data-sets, i.e., if there are further individuals in the domain of interest that are not described in the data-set, a technique called *Attribute Exploration* has been developed by Ganter [4–6] and Stumme [13]. It allows for interaction with an expert which is able to provide unknown individuals that are not contained in the data-set but represent counterexamples to otherwise valid implications. This algorithm is merely an extension of the algorithm `NextClosure` by Ganter [5]. Unfortunately, the algorithm uses the lectic order on the attribute set, which is linear, to compute the implications in the canonical base. As a consequence, it is not (obviously) possible to parallelize this default *Attribute Exploration*. However, in [10, 11] we have introduced the algorithm `NextClosures` that is also able to compute canonical bases, but in a non-linear order which makes it possible to enumerate the elements of the canonical base in a parallel manner. In particular, the canonical base is constructed w.r.t. increasing premise cardinality. Benchmarks have shown that there is an inverse linear correlation between the computation time and the number of available CPU cores, provided that the underlying formal context is large enough, and that its performance on one CPU core is comparable to `NextClosure`, more specifically the quotient of the computation times for the same formal context is between $\frac{1}{3}$ and 3. The benchmarks in [10, 11] should only be interpreted in a relative way – using more efficient data structures (e.g., `java.util.BitSet`), or faster programming languages (e.g., C++), the computation times can be decreased further. An implementation of `NextClosures` in the programming language Java 8 can be found in [9].

In this paper, `NextClosures` is extended with the possibility of expert interaction. More specifically, we assume that there is a formal context that describes the domain of interest, but it is inaccessible and there is an expert (or a set of experts) that can correctly decide whether an implication holds in this context, and if she refutes then also provides a counterexample. Additionally, there may be an observed subcontext of the full domain context, which is used to decrease the number of questions posed to the expert. Using the technique of *Attribute Exploration*, it is possible to construct a minimal implicational base of the domain context. The algorithm `ParallelAttributeExploration` that will be described in the following sections implements this technique and furthermore allows for a parallel execution.

Angluin, Frazier, and Pitt [1] have also investigated the problem of learning propositional Horn-theories by means of oracles. In particular, they assume that there are two experts: a *membership oracle* and an *equivalence oracle*. While the first expert decides whether a certain object satisfies the (unknown) target theory, the second expert decides whether the current theory is equivalent to the target theory (and if not, returns a counterexample). Later, Arias and Balcázar [2] have proven that this learning approach always constructs the canonical base [8] for the target theory.

This document is structured as follows. Section 2 introduces the basic notions of *Formal Concept Analysis*, and Section 3 defines the notion of an expert as well as provides some important statements on the interplay of formal contexts and experts. Section 4 presents the algorithm `ParallelAttributeExploration`, and furthermore proves its soundness and completeness. Section 5 draws a comparison with the default algorithm for *Attribute Exploration*, as well as discusses some possibilities for the integration of several experts.

2 Formal Concept Analysis

In this section we shall introduce the basic notions of *Formal Concept Analysis*, cf. [7]. A *formal context* $\mathbb{K} = (G, M, I)$ consists of a set G of *objects*, a set M of *attributes*, and an *incidence relation* $I \subseteq G \times M$ such that $g\ I\ m$ indicates that object g *has* attribute m. Furthermore, for subsets $A \subseteq G$ and $B \subseteq M$, their *derivations* are defined as follows:

$$A^I \coloneqq \{\, m \in M \mid \forall g \in A \colon g\ I\ m \,\} \text{ and } B^I \coloneqq \{\, g \in G \mid \forall m \in B \colon g\ I\ m \,\}.$$

It is well-known [7] that both derivation operators form a so-called *Galois connection* between the powersets $\wp(G)$ and $\wp(M)$, i.e., the following statements hold true for all subsets $A, A_1, A_2 \subseteq G$ and $B, B_1, B_2 \subseteq M$:

1. $A \subseteq B^I \Leftrightarrow B \subseteq A^I$.
2. $A \subseteq A^{II}$.
3. $A^I = A^{III}$.
4. $A_1 \subseteq A_2 \Rightarrow A_2^I \subseteq A_1^I$.
5. $B \subseteq B^{II}$.
6. $B^I = B^{III}$.
7. $B_1 \subseteq B_2 \Rightarrow B_2^I \subseteq B_1^I$.

For subsets $A \subseteq G$ and $B \subseteq M$, the pair (A, B) is a *formal concept* of $\mathbb{K}$ if $A^I = B$ and $B^I = A$. Then we refer to A as the *extent*, and to B as the *intent* of (A, B). The set of all formal concepts is denoted as $\mathfrak{B}(\mathbb{K})$, and $\mathsf{Int}(\mathbb{K})$ denotes the set of all intents of $\mathbb{K}$. $\mathfrak{B}(\mathbb{K})$ can be ordered by $(A, B) \leq (C, D)$ if $A \subseteq C$ (or dually if $B \supseteq D$), and indeed then $(\mathfrak{B}(\mathbb{K}), \leq)$ is a complete lattice where infima and suprema are given as follows:

$$\bigwedge_{t\in T}(A_t, B_t) = (\bigcap_{t\in T} A_t, (\bigcup_{t\in T} B_t)^{II}) \text{ and } \bigvee_{t\in T}(A_t, B_t) = ((\bigcup_{t\in T} A_t)^{II}, \bigcap_{t\in T} B_t).$$

An *implication* over M is an expression of the form $X \to Y$ where $X, Y \subseteq M$. The set of all implications over M is denoted by $\mathsf{Imp}(M)$. We say that $X \to Y$ is *valid* in $\mathbb{K}$, denoted as $\mathbb{K} \models X \to Y$, if $X^I \subseteq Y^I$. $\mathsf{Imp}(\mathbb{K})$ is the set of all valid implications of $\mathbb{K}$. A subset $Z \subseteq M$ is a *model* of $X \to Y$ if $X \subseteq Z$ implies $Y \subseteq Z$, denoted by $Z \models X \to Y$, and $\mathsf{Mod}(X \to Y)$ is the set of all models of $X \to Y$. Furthermore, $\mathsf{Mod}(\mathcal{L}) \coloneqq \bigcap\{\, \mathsf{Mod}(X \to Y) \mid X \to Y \in \mathcal{L} \,\}$ is the set of all models of an implication set $\mathcal{L}$. It is well-known that the following statements are equivalent:

1. $X \to Y$ is valid in $\mathbb{K}$.
2. Each intent of $\mathbb{K}$ models $X \to Y$.
3. Each object intent of $\mathbb{K}$ models $X \to Y$.
4. $Y \subseteq X^{II}$.

Furthermore, the relation $\models$ may be lifted to implication sets as follows: Let $\mathcal{L} \cup \{X \to Y\} \subseteq \mathsf{Imp}(M)$, then $\mathcal{L}$ *entails* $X \to Y$, symbolized by $\mathcal{L} \models X \to Y$, if every model of $\mathcal{L}$ is a model of $X \to Y$. Then $\mathsf{Imp}(\mathcal{L})$ is the set of all implications that are entailed by $\mathcal{L}$. For each subset $X \subseteq M$, there is a smallest superset $X^{\mathcal{L}}$ of X that is a model of $\mathcal{L}$, since $\mathsf{Mod}(\mathcal{L})$ is closed under intersection. It is well-known that this set can be computed as follows:

$$X^{\mathcal{L}} = \textstyle\bigcup_{n\geq 1} X^{\mathcal{L}_n} \text{ where } X^{\mathcal{L}_{n+1}} \coloneqq (X^{\mathcal{L}_1})^{\mathcal{L}_n} \text{ for all } n \geq 1,$$
$$\text{and } X^{\mathcal{L}_1} \coloneqq X \cup \bigcup\{\, Z \mid Y \to Z \in \mathcal{L} \text{ and } Y \subseteq X \,\}.$$

It is easy to verify that the following statements are equivalent:

1. $\mathcal{L} \models X \to Y$.
2. For all $Z \subseteq M$, $Z \models \mathcal{L}$ implies $Z \models X \to Y$.
3. For all formal contexts $\mathbb{K}$ with attribute set M, $\mathbb{K} \models \mathcal{L}$ implies $\mathbb{K} \models X \to Y$.
4. $Y \subseteq X^{\mathcal{L}}$.

Note that a formal context is just another notion for a set of propositional models (where the attributes in M are considered as propositional variables). In particular, for a formal context $\mathbb{K} = (G, M, I)$ the set $\mathcal{P}_{\mathbb{K}} \coloneqq \{\, \chi^M_{g^I} \mid g \in G \,\}$, where χ^M_B is the characteristic function of B in M, is a set of propositional models such that for each implication $X \to Y$ over M, $X \to Y$ is valid in $\mathbb{K}$ if, and only if, $\bigwedge X \to \bigwedge Y$ is valid in $\mathcal{P}_{\mathbb{K}}$.

Analogously, if $\mathcal{P}$ is a set of propositional models over a set M of propositional variables, then the formal context $\mathbb{K}_{\mathcal{P}} := (\{ p^{-1}(1) \mid p \in \mathcal{P} \}, M, \ni)$ satisfies $\mathcal{P} \models \bigwedge X \to \bigwedge Y$ if, and only if, $\mathbb{K}_{\mathcal{P}} \models X \to Y$, for all implications $X \to Y$ over M.

An *implicational base* of a formal context $\mathbb{K}$ is an implication set that is *sound*, i.e., is valid in $\mathbb{K}$, and is *complete*, i.e., entails all valid implications of $\mathbb{K}$. An implicational base is *irredundant* if none of its implications follows from the others, and is *minimal* if it has minimal cardinality among all implicational bases for $\mathbb{K}$. It is straight-forward to show that the following statements are equivalent:

1. $\mathcal{L}$ is an implicational base for $\mathbb{K}$.
2. $\mathsf{Imp}(\mathcal{L}) = \mathsf{Imp}(\mathbb{K})$.
3. $\mathsf{Mod}(\mathcal{L}) = \mathsf{Int}(\mathbb{K})$.

A *pseudo-intent* of $\mathbb{K} = (G, M, I)$ is an attribute set $P \subseteq M$ such that $P \neq P^{II}$, and $Q^{II} \subseteq P$ for all pseudo-intents $Q \subsetneq P$. The set of all pseudo-intents of $\mathbb{K}$ is denoted by $\mathsf{PsInt}(\mathbb{K})$. The *canonical base* of a formal context $\mathbb{K}$ is defined as

$$\mathcal{B}_{\mathsf{can}}(\mathbb{K}) := \left\{ P \to P^{II} \;\middle|\; P \in \mathsf{PsInt}(\mathbb{K}) \right\},$$

and is a minimal implicational base for $\mathbb{K}$, cf. [5–8].

It has been shown that the set of all intents and pseudo-intents is a closure system. The corresponding closure operator $\cdot^{\mathbb{K}^*}$ is given by the following definition:

$$X^{\mathbb{K}^*} := \bigcup_{n \geq 1} X^{\mathbb{K}^*_n} \text{ where } X^{\mathbb{K}^*_{n+1}} := (X^{\mathbb{K}^*_1})^{\mathbb{K}^*_n} \text{ for all } n \geq 1,$$
$$\text{and } X^{\mathbb{K}^*_1} := X \cup \bigcup \left\{ P^{II} \;\middle|\; P \in \mathsf{PsInt}(\mathbb{K}) \text{ and } P \subsetneq X \right\}.$$

More specifically, then an attribute set $X \subseteq M$ is an intent or a pseudo-intent of $\mathbb{K}$ if, and only if, $X = X^{\mathbb{K}^*}$. Additionally, we may also define a pseudo-closure operator for implication sets $\mathcal{L} \subseteq \mathsf{Imp}(M)$:

$$X^{\mathcal{L}^*} := \bigcup_{n \geq 1} X^{\mathcal{L}^*_n} \text{ where } X^{\mathcal{L}^*_{n+1}} := (X^{\mathcal{L}^*_1})^{\mathcal{L}^*_n} \text{ for all } n \geq 1,$$
$$\text{and } X^{\mathcal{L}^*_1} := X \cup \bigcup \{ Z \mid Y \to Z \in \mathcal{L} \text{ and } Y \subsetneq X \}.$$

It is readily verified that both closure operators $\cdot^{\mathbb{K}^*}$ and $\cdot^{\mathcal{L}^*}$ coincide in case $\mathcal{L} = \mathcal{B}_{\mathsf{can}}(\mathbb{K})$.

3 Experts

An *expert* is an oracle that correctly answers questions in a certain domain of interest. For our purposes, the questions are expressed in form of implications, and an expert may either *accept* or *decline*. If the expert accepts an implication, then it must hold for all objects in the domain of interest, and otherwise she must return a refutation, i.e., an object that serves as a counterexample. In this section, we will formally define the notion of an expert, and provide some basic statements.

Definition 1 (Expert, [3, Definition 6.1.2]). *Let M be a set of attributes. An* expert *on M is a partial mapping $\chi\colon \mathsf{Imp}(M) \to_{\mathsf{p}} \wp(M)$ that satisfies the following properties:*

1. *If $\chi(X \to Y)$ is defined, then the value is not a model of $X \to Y$, i.e., $\chi(X \to Y) = C$ implies $X \subseteq C$ and $Y \not\subseteq C$. Furthermore, we then call C a* counterexample *against* $X \to Y$.
   (Experts return counterexamples for refuted implications.)
2. *If $\chi(X \to Y)$ is undefined, then every other counterexample given by χ must be a model of $X \to Y$, i.e., $\chi(U \to V) = C$ implies $X \not\subseteq C$ or $Y \subseteq C$.*
   (Counterexamples do not refute accepted implications.)

Furthermore, we say that χ accepts *$X \to Y$, and denote this as $\chi \models X \to Y$, if $\chi(X \to Y)$ is undefined, and that χ* refutes *$X \to Y$ otherwise. The set of all accepted implications of χ is denoted by $\mathsf{Imp}(\chi)$, and the set of all counterexamples of χ is denoted by $\mathsf{Cex}(\chi) \coloneqq \{\, C \mid \exists X, Y \subseteq M \colon \chi(X \to Y) = C \,\}$.*

There is a correspondence between formal contexts and experts as follows:

Definition 2 (Induced Expert). *An expert χ on M is* induced *by a formal context $\mathbb{K} = (G, M, I)$ if it accepts exactly those implications that are valid in $\mathbb{K}$, i.e., $\mathsf{Imp}(\mathbb{K}) = \mathsf{Imp}(\chi)$. If χ is an expert on M, then its* induced *formal context is $\mathbb{K}_\chi \coloneqq (\mathsf{Cex}(\chi), M, \ni)$.*

Lemma 3. *Let $\mathbb{K} = (G, M, I)$ be a formal context and χ an expert on M. Then χ is induced by $\mathbb{K}$ if, and only if, it accepts only valid implications of $\mathbb{K}$, and all counterexamples are intents of $\mathbb{K}$, i.e., $\mathsf{Imp}(\chi) \subseteq \mathsf{Imp}(\mathbb{K})$ as well as $\mathsf{Cex}(\chi) \subseteq \mathsf{Int}(\mathbb{K})$.*

Proof. The if-direction is trivial. For the converse direction, assume that χ accepts exactly those implications that are valid in $\mathbb{K}$. Of course, then each implication accepted by χ is valid in $\mathbb{K}$. Assume that χ refutes an implication with a counterexample C. Since the implication $C \to C^{II}$ is trivially valid in $\mathbb{K}$, χ accepts $C \to C^{II}$. Consequently, the counterexample C must be a model of $C \to C^{II}$, i.e., C is an intent of $\mathbb{K}$. □

Lemma 4. *If χ is an expert on M, then χ is an induced expert of $\mathbb{K}_\chi$.*

Proof. Let χ be an expert on M, and consider an implication $X \to Y$. If χ accepts $X \to Y$, then by Statement 2 of Definition 1 all counterexamples of χ are models of $X \to Y$. We conclude that all object intents of the $\mathbb{K}_\chi$ are models of $X \to Y$, i.e., $X \to Y$ is valid in $\mathbb{K}_\chi$.

Vice versa, if $X \to Y$ is valid in $\mathbb{K}_\chi$, then all intents of $\mathbb{K}_\chi$ are models of $X \to Y$. Hence, χ cannot refute $X \to Y$, as the counterexample would be an object intent of $\mathbb{K}_\chi$, but would not be a model of $X \to Y$. □

Corollary 5. *If $\mathbb{K}$ is a formal context with an induced expert χ, then an implication is valid in $\mathbb{K}$ if, and only if, it is valid in $\mathbb{K}_\chi$. Furthermore, then every (minimal) implicational base of $\mathbb{K}$ is a (minimal) implicational base of $\mathbb{K}_\chi$, and vice versa, i.e., the sets of intents of $\mathbb{K}$ and $\mathbb{K}_\chi$ coincide.*

Definition 6 (Optimal Expert). *An expert χ is* optimal *if for all implications $X \to Y$, it is true that χ accepts $X \to Y \cap C$ if χ refutes $X \to Y$ with counterexample C.*

Lemma 7. *Let $\mathbb{K}$ be a formal context. An induced expert χ of $\mathbb{K}$ is optimal if, and only if, for all implications $X \to Y$, $\chi(X \to Y) = C$ implies $Y \cap C \subseteq X^{II} \subseteq C$.*

Furthermore, the canonical *expert $\chi_{\mathbb{K}}$ for $\mathbb{K}$ is an optimal induced expert for $\mathbb{K}$, where*

$$\chi_{\mathbb{K}}(X \to Y) \coloneqq \begin{cases} \textit{undefined} & \textit{if } \mathbb{K} \models X \to Y, \\ X^{II} & \textit{otherwise.} \end{cases}$$

Proof. The if-direction is obvious. Vice versa, let χ be optimal for $\mathbb{K}$, and consider an implication $X \to Y$ that is refuted by χ with counterexample C, i.e., $X \subseteq C$ and $Y \nsubseteq C$. Since χ is induced by $\mathbb{K}$, C is an intent, and so $X^{II} \subseteq C$. Furthermore, as χ is optimal, $X \to Y \cap C$ is valid in $\mathbb{K}$, i.e., $Y \cap C \subseteq X^{II}$.

Eventually, $\chi_{\mathbb{K}}$ is an induced expert for $\mathbb{K}$, since it accepts all implications that are valid in $\mathbb{K}$, and all counterexamples are intents of $\mathbb{K}$. Furthermore, it is optimal, as implications $X \to Y \cap X^{II}$ are trivially valid in $\mathbb{K}$. □

Lemma 8. *Let χ be an expert. Then there is an optimal expert $\widehat{\chi}$ such that both accept the same implications, i.e.,* $\mathsf{Imp}(\chi) = \mathsf{Imp}(\widehat{\chi})$.

Proof. Consider an expert χ. We construct an equivalent optimal expert $\widehat{\chi}$ as follows. Let $X \to Y$ be an arbitrary implication. If χ accepts $X \to Y$, then $\widehat{\chi}$ accepts $X \to Y$, too. If χ rejects $X \to Y$, then proceed in the following way. Let $Y_0 \coloneqq Y$, and $n \coloneqq 0$. While χ rejects $X \to Y_n$ with the counterexample Z_n, set $Y_{n+1} \coloneqq Y_n \cap Z_n$ and increase n. Eventually, define $\widehat{\chi}(X \to Y) \coloneqq \bigcap_{k=0}^{n} Z_k$.

It remains to prove that $\widehat{\chi}$ is optimal and accepts the same implications as χ. Assume that $\widehat{\chi}$ refutes $X \to Y$ with counterexample Z. Then there exists a sequence $Z_0, \ldots, Z_n$ of counterexamples of χ as above such that Z equals their intersection, and χ accepts $X \to Y \cap Z$. By construction, then also $\widehat{\chi}$ accepts the adjusted implication $X \to Y \cap Z$.

By definition, we already know that $\mathsf{Imp}(\chi) \subseteq \mathsf{Imp}(\widehat{\chi})$. Vice versa, since χ rejects only if $\widehat{\chi}$ rejects, we conclude that χ accepts if $\widehat{\chi}$ accepts. □

Proposition 9. *Let $\mathbb{K} = (G, M, I)$ be a formal context and χ an expert on M. Then the following statements are equivalent:*

1. *χ is induced by $\mathbb{K}$.*
2. *χ accepts exactly those implications that are valid in $\mathbb{K}$, i.e.,* $\mathsf{Imp}(\chi) = \mathsf{Imp}(\mathbb{K})$.
3. *χ accepts only valid implications of $\mathbb{K}$, and all counterexamples are intents of $\mathbb{K}$, i.e.,* $\mathsf{Imp}(\chi) \subseteq \mathsf{Imp}(\mathbb{K})$ *and* $\mathsf{Cex}(\chi) \subseteq \mathsf{Int}(\mathbb{K})$.
4. *Each intent of $\mathbb{K}$ is an intersection of counterexamples of χ, and all counterexamples are intents of $\mathbb{K}$, i.e.,* $\langle \mathsf{Cex}(\chi) \rangle_{\bigcap} = \mathsf{Int}(\mathbb{K})$.

Proof. Statements 1 to 3 are equivalent by Definition 2 and Lemma 3.

3.⇔4. We consider the *optimization* $\widehat{\chi}$ from Lemma 8, then by construction it is true that every counterexample of $\widehat{\chi}$ is an intersection of counterexamples of χ. The maximal intent M is obtained as the empty intersection (of counterexamples). For each intent $B = B^{II}$ where $B \neq M$, the implication $B \to M$ is invalid in $\mathbb{K}$, and thus must be rejected by $\widehat{\chi}$ with a counterexample C. Then Lemma 7 yields $C = M \cap C \subseteq B^{II} \subseteq C$.

Vice versa, let every intent of $\mathbb{K}$ be an intersection of counterexamples of χ, and assume that all counterexamples are intents of $\mathbb{K}$. Consider an implication $X \to Y$ that is not valid in $\mathbb{K}$, i.e., $Y \nsubseteq X^{II}$. In particular, then X^{II} is an intersection of counterexamples $C_1, \ldots, C_n$ of χ, and the C_i are intents of $\mathbb{K}$. If χ accepts $X \to Y$, then Statement 2 of Definition 1 implies that all counterexamples of χ are models of $X \to Y$, and in particular each C_i is a model of $X \to Y$. Since the set of models of an implication is closed under intersection, X^{II} must be a model of $X \to Y$. Contradiction! □

Lemma 10. *Let χ be an induced expert of a formal context $\mathbb{K}$. If χ is optimal, then* $\mathsf{Cex}(\chi)$ *is closed under non-empty intersections.*

Proof. Let χ be optimal, and consider two counterexamples C_1 and C_2. The implication $C_1 \cap C_2 \to M$ must be rejected by χ with a counterexample C such that $M \cap C \subseteq (C_1 \cap C_2)^{II} \subseteq C$, i.e., $C = C_1 \cap C_2$, since both C_i are intents. □

However, the converse statement does not hold in general. To see this, consider a formal context $\mathbb{K}$ over $M \coloneqq \{a, b, c\}$ where $\mathsf{Int}(\mathbb{K}) = \{\{a\}, \{a, b\}, \{a, b, c\}\}$. Hence, for each induced expert χ, we have $\emptyset \neq \mathsf{Cex}(\chi) \subseteq \{\{a\}, \{a, b\}\}$ and it is easily verified that for each choice, the set of counterexamples is closed under non-empty intersections. However, an induced expert χ with $\chi(\{a\} \to \{b, c\}) \coloneqq \{a, b\}$ is not optimal.

4 Parallel Attribute Exploration

As the next step, we will introduce the algorithm for *Parallel Attribute Exploration*. Assume that we want to compute a minimal implicational base for an inaccessible formal context $\mathbb{D}$ with attribute set M, and we have observed an induced subcontext $\mathbb{K} = (G, M, I)$ of $\mathbb{D}$ as well as we know an expert χ that is induced by $\mathbb{D}$. Of course, it is not useful to simply compute an implicational base of $\mathbb{K}$, as wrong conclusions could be drawn. There are two naïve ways to accomplish the computation of a base. According to Corollary 5, we may construct the formal context $\mathbb{K}_\chi$ induced by χ, and compute its canonical base, e.g., by means of the algorithms in [5, 6, 11]. However, this is certainly no practical approach, as its puts a high workload on the expert by posing all possible implications as questions to her. A slight improvement would consist in first checking whether the implication in question is already refuted by the known subcontext $\mathbb{K}$, and only in case of validity ask the expert for acceptance. Of course, all implications holding in $\mathbb{D}$ are valid in $\mathbb{K}$, too. Unfortunately, this modification is still not efficient, since the number of questions posed to the expert will not be minimal in order to compute an implicational base. In general, calls to the expert are expensive, and it should be ensured that only a minimal amount of work is put on her. This is the starting point for an algorithm called *Attribute Exploration*, which is basically an extension of `NextClosure` with expert interaction as introduced by Ganter [5, 6] and Ganter and Wille [7]. It enumerates all pseudo-intents of the context $\mathbb{D}$ and only poses implicational questions to the expert whose premise is a pseudo-intent. This ensures the minimality on the number of questions w.r.t. both $\mathbb{K}$ and χ. If furthermore χ is optimal, then the number of questions is minimal w.r.t. $\mathbb{K}$, i.e., for each other expert χ' induced by $\mathbb{D}$, χ' must answer at least as many questions as χ.

While *Attribute Exploration* [5–7] constructs the canonical base of $\mathbb{D}$ in a lectic order, Algorithm 1 computes it w.r.t. increasing premise cardinality, which in turn allows to process all implications with the same premise cardinality in parallel. Note that Algorithm 1 is an extension of [11, Algorithm 1] with expert interaction.

Definition 11. *Let (G_1, M, I_1) and (G_2, M, I_2) be two formal contexts with disjoint object sets and the same attribute set. Their* subposition *is defined as the formal context*

$$\frac{(G_1, M, I_1)}{(G_2, M, I_2)} \coloneqq (G_1 \cup G_2, M, I_1 \cup I_2).$$

For a formal context $\mathbb{K} = (G, M, I)$ and an attribute set $X \subseteq M$, the formal context

$$\mathbb{K}[X] \coloneqq (G \cup \{g_X\}, M, I \cup \{g_X\} \times X)$$

is obtained by adding a new object $g_X \notin G$ that has all attributes from X. In particular, $\mathbb{K}[X]$ is a subposition of $\mathbb{K}$ and the row X. For a sequence $X_1, X_2, \dots, X_n \subseteq M$, we inductively define

$$\mathbb{K}[X_1, X_2, \dots, X_n] \coloneqq (\mathbb{K}[X_1])[X_2, \dots, X_n].$$

Furthermore, we write $\mathbb{K} \leq_M \mathbb{D}$ if there is a context $\mathbb{U}$ such that $\mathbb{D} = \frac{\mathbb{K}}{\mathbb{U}}$.

If $\mathcal{L}$ is an implication set, and $k \in \mathbb{N}$, then $\mathcal{L}\restriction_k \coloneqq \{\, X \to Y \in \mathcal{L} \mid |X| \leq k \,\}$ contains all implications from $\mathcal{L}$ whose premises have a cardinality of at most k. Furthermore, we define $\mathsf{PsInt}(\mathbb{K})\restriction_k \coloneqq \{\, P \in \mathsf{PsInt}(\mathbb{K}) \mid |P| \leq k \,\}$.

Lemma 12. *Let $\mathbb{K} = (G, M, I)$ be a formal context, and $X \subseteq M$ an attribute set. Furthermore, denote the incidence relation of $\mathbb{K}[X]$ by J. Then the following statements hold:*

1. *For all attribute sets $B \subseteq M$, it is true that*

$$B^{JJ} = \begin{cases} B^{II} \cap X & \text{if } B \subseteq X, \text{ and} \\ B^{II} & \text{otherwise.} \end{cases}$$

2. *If X is a model of the implication $Y \to Y^{II}$, then $Y^{II} = Y^{JJ}$.*
3. *If X is a model of all implications $P \to P^{II}$ where P is a pseudo-intent of $\mathbb{K}$ with $|P| \leq k$, then the pseudo-intents of $\mathbb{K}$ and $\mathbb{K}[X]$ with cardinality $\leq k$ coincide, i.e., $X \models \mathcal{B}_{\mathsf{can}}(\mathbb{K})\restriction_k$ implies $\mathsf{PsInt}(\mathbb{K})\restriction_k = \mathsf{PsInt}(\mathbb{K}[X])\restriction_k$.*

Proof. 1. Let $B \subseteq X$, then $B^J = B^I \cup \{g_X\}$, and hence $B^{JJ} = (B^I \cup \{g_X\})^J = B^{IJ} \cap g_X^J = B^{II} \cap X$. Otherwise, $B^J = B^I$, and thus $B^{JJ} = B^{IJ} = B^{II}$.

2. Assume that $X \models Y \to Y^{II}$, i.e., $Y \subseteq X$ implies $Y^{II} \subseteq X$. If $Y \subseteq X$, then $Y^{JJ} = Y^{II} \cap X = Y^{II}$. Otherwise, $Y^{JJ} = Y^{II}$ follows directly.

3. We prove the statement by induction on k. First let $k = 0$. Obviously, $\emptyset$ is the only set of cardinality 0. Since it has no strict subsets, it is a pseudo-intent if, and only if, it is no intent. If $\emptyset$ is a pseudo-intent of $\mathbb{K}$, then X is a model of $\emptyset \to \emptyset^{II}$. Statement 2 yields $\emptyset^{II} = \emptyset^{JJ}$, and thus $\emptyset \neq \emptyset^{JJ}$. If otherwise $\emptyset$ is an intent of $\mathbb{K}$, then it holds that $\emptyset = \emptyset^{II} \supseteq \emptyset^{II} \cap X \supseteq \emptyset$, i.e., $\emptyset$ must be an intent of $\mathbb{K}[X]$, too.

Now assume that the induction hypothesis holds for k. Consider a pseudo-intent P of $\mathbb{K}$ with $|P| = k+1$. Since X is a model of $P \to P^{II}$, Statement 2 yields $P^{II} = P^{JJ}$, and hence P is no intent of $\mathbb{K}[X]$. Now let $Q \subsetneq P$ be a pseudo-intent of $\mathbb{K}[X]$. Then $|Q| \leq k$, and hence Q is a pseudo-intent of $\mathbb{K}$ by induction hypothesis. Consequently, $Q^{II} \subseteq P$, and thus $Q^{JJ} \subseteq P$.

Vice versa, let P be a pseudo-intent of $\mathbb{K}[X]$ with $|P| = k+1$. Then P is no intent of $\mathbb{K}$, as $P \neq P^{JJ} \subseteq P^{II}$. Consider a pseudo-intent $Q \subsetneq P$ of $\mathbb{K}$. Then Q must be a $\mathbb{K}[X]$-pseudo-intent by induction hypothesis. Furthermore, $Q^{JJ} \subseteq P$. Since X is a model of $Q \to Q^{II}$, Statement 2 implies $Q^{JJ} = Q^{II}$, and thus $Q^{II} \subseteq P$. ◻

As an immediate consequence we deduce from the preceding lemma, more specifically from Statements 2 and 3, that the following corollary holds.

Corollary 13. *Let $\mathbb{K} = (G, M, I)$ be a formal context, and $X \subseteq M$ an attribute set. If X is a model of $\mathcal{B}_{\mathsf{can}}(\mathbb{K})\restriction_k$, then $\mathcal{B}_{\mathsf{can}}(\mathbb{K})\restriction_k = \mathcal{B}_{\mathsf{can}}(\mathbb{K}[X])\restriction_k$.*

Algorithm 1 `ParallelAttributeExploration`

`Require:` a formal context $\mathbb{K} = (G, M, I)$
`Require:` an expert χ on M
1 $\mathbf{C} := \{\emptyset\}$, $\mathcal{L} := \emptyset$
2 **for** $k = 0, 1, \ldots, |M| - 1$ **do**
3   **for all** $C \in \mathbf{C}$ **with** $|C| = k$ **do in parallel**
4   **if** $C = C^{\mathcal{L}^*}$ **then**
5   **while** $C \neq C^{II}$ **and** $\chi(C \to C^{II}) = X$ **do**
6         $\mathbb{K} := \mathbb{K}[X]$
7   **if** $C \neq C^{II}$ **then**
8         $\mathcal{L} := \mathcal{L} \cup \{C \to C^{II}\}$
9       $\mathbf{C} := \mathbf{C} \cup \{\, C^{II} \cup \{m\} \mid m \notin C^{II} \,\}$
10   **else**
11       $\mathbf{C} := \mathbf{C} \cup \{C^{\mathcal{L}^*}\}$
12   Wait for termination of all parallel processes.
13 **return** $(\mathbb{K}, \mathcal{L})$

By successive application of the previous corollary we get the following statement.

Lemma 14. *Let $\mathbb{K} = (G, M, I)$ be a formal context, and $X_1, \ldots, X_n \subseteq M$ attribute sets. If each X_i is a model of $\mathcal{B}_{\mathsf{can}}(\mathbb{K})\!\upharpoonright_k$, then $\mathcal{B}_{\mathsf{can}}(\mathbb{K})\!\upharpoonright_k = \mathcal{B}_{\mathsf{can}}(\mathbb{K}[X_1, \ldots, X_n])\!\upharpoonright_k$.*

Proof. We show by induction on $i \in \{1, \ldots, n\}$ that $\mathcal{B}_{\mathsf{can}}(\mathbb{K})\!\upharpoonright_k = \mathcal{B}_{\mathsf{can}}(\mathbb{K}[X_1, \ldots, X_i])\!\upharpoonright_k$. The induction base follows from Corollary 13. Now assume that $i < n$ and the induction hypothesis holds for i. Then $X_{i+1} \models \mathcal{B}_{\mathsf{can}}(\mathbb{K})\!\upharpoonright_k = \mathcal{B}_{\mathsf{can}}(\mathbb{K}[X_1, \ldots, X_i])\!\upharpoonright_k$, and again by Corollary 13 we conclude $\mathcal{B}_{\mathsf{can}}(\mathbb{K}[X_1, \ldots, X_i])\!\upharpoonright_k = \mathcal{B}_{\mathsf{can}}(\mathbb{K}[X_1, \ldots, X_i, X_{i+1}])\!\upharpoonright_k$. $\square$

In [11] we have shown that in order to correctly determine whether an attribute set with at most k elements is an intent or pseudo-intent of $\mathbb{K}$, it suffices to know the part of the canonical base that consists of all implications whose premise has a cardinality smaller than k. More specifically, we cite the following corollary.

Corollary 15 ([11, Corollary 3]). *If $\mathcal{L}$ contains all implications $P \to P^{II}$ where P is a pseudo-intent of $\mathbb{K}$ with $|P| < k$, and otherwise only implications with premise cardinality k, then for all attribute sets $X \subseteq M$ with $|X| \leq k$ the following statements are equivalent:*

1. *X is an intent or a pseudo-intent of $\mathbb{K}$.*
2. *X is $\mathcal{L}^*$-closed.*

Algorithm 1 describes *Parallel Attribute Exploration* in pseudo-code. If the expert χ is optimal, then the `while`-statement in Line 5 may be replaced with the analogous `if`-statement, since after χ refutes an implication $C \to C^{II}$ with a counterexample X we have that χ accepts $C \to C^{II} \cap X$, and $C^{JJ} = C^{II} \cap X$ where J is the incidence relation of $\mathbb{K}[X]$, i.e., in the second iteration, the condition of the `while`-statement always evaluates to `false`. In particular, the optimality of the expert is no restriction, since according to Lemma 8 we may always *optimize* an expert.

In the following text we will analyze Algorithm 1, and show its soundness, completeness, and termination. Beforehand, we define the following notions:

1. A run of `ParallelAttributeExploration` is *in state* k if all candidates of cardinality k have been processed, but none of cardinality $k+1$.
2. $\mathbf{C}_k$ denotes the set of candidates in state k.
3. $\mathcal{L}_k$ denotes the set of implications in state k.
4. $\mathbb{K}_k \coloneqq (G_k, M, I_k)$ denotes the formal context in state k.
5. $X_k^1, \ldots, X_k^{n_k}$ denote all counterexamples provided by the expert between states k and $k+1$, i.e., it is true that $\mathbb{K}_k[X_k^1, \ldots, X_k^{n_k}] = \mathbb{K}_{k+1}$.

Proposition 16. *Let $\mathbb{K} = (G, M, I)$ be a formal context, and χ an expert on M, such that all implications accepted by χ are valid in $\mathbb{K}$. Further assume that Algorithm 1 is started on $(\mathbb{K}, \chi)$ as input, and is currently in state k. Then the following statements are satisfied:*

1. *$\mathbf{C}_k$ contains all pseudo-intents of $\mathbb{K}_{k+1}$ with cardinality $k+1$.*
2. *$\mathcal{L}_k$ consists of all implications $P \to P^{I_k I_k}$ where P is a pseudo-intent of $\mathbb{K}_k$ with cardinality $\leq k$, i.e., $\mathcal{B}_{\mathsf{can}}(\mathbb{K}_k)\restriction_k = \mathcal{L}_k$.*
3. *Between the states k and $k+1$, every attribute set with cardinality $k+1$ is $\mathcal{L}^*$-closed if, and only if, it is either an intent or a pseudo-intent of $\mathbb{K}_{k+1}$.*

Proof. W.l.o.g. assume that the expert χ is optimal, and Line 5 of Algorithm 1 has been replaced with the analogous `if`-statement as discussed above.

We show the statements by induction on k. For the base case assume $k = -1$, as the initial state is -1. The candidate set is initialized as $\{\emptyset\}$, and thus $\mathbf{C}_{-1}$ indeed contains all pseudo-intents with 0 elements. As there are no pseudo-intents of $\mathbb{K}_{-1} = \mathbb{K}$ with at most -1 elements, the initial implication set $\mathcal{L}_{-1} = \emptyset$ satisfies Statement 2. Between the states -1 and 0 all candidates with 0 elements are processed, i.e., only $\emptyset$ is processed. Obviously, $\emptyset$ is either an intent or a pseudo-intent of $\mathbb{K}_0$, i.e., it is $(\mathbb{K}_0)^*$-closed. Furthermore, $\emptyset$ has no strict subsets, and hence it must be $\mathcal{L}^*$-closed for all implication sets $\mathcal{L}$ between states -1 and 0, i.e., $\mathcal{L}_{-1} \subseteq \mathcal{L} \subseteq \mathcal{L}_0$.

For the induction step assume that the statements hold for all states $\leq k$.

2. We will prove that $\mathcal{B}_{\mathsf{can}}(\mathbb{K}_{k+1})\restriction_{k+1} = \mathcal{L}_{k+1}$. Statement 2 of the induction hypothesis yields that $\mathcal{B}_{\mathsf{can}}(\mathbb{K}_k)\restriction_k = \mathcal{L}_k$. All counterexamples provided by the expert between states k and $k+1$ are models of $\mathcal{L}_k$, as the expert has accepted all implications in $\mathcal{L}_k$. As a consequence, Lemma 14 implies $\mathcal{B}_{\mathsf{can}}(\mathbb{K}_k)\restriction_k = \mathcal{B}_{\mathsf{can}}(\mathbb{K}_{k+1})\restriction_k$. Since Algorithm 1 does not remove or modify any implications in $\mathcal{L}$, and between the states k and $k+1$ only implications with a premise cardinality of $k+1$ are added to $\mathcal{L}$, it is true that $\mathcal{L}_{k+1}\restriction_k = \mathcal{L}_k$ and $\mathcal{L}_{k+1}$ cannot contain any implications with a premise cardinality $> k+1$. Hence, $\mathcal{L}_{k+1}$ contains $\mathcal{B}_{\mathsf{can}}(\mathbb{K}_{k+1})\restriction_k$, and it remains to show that $\mathcal{L}_{k+1}$ contains all implications of $\mathcal{B}_{\mathsf{can}}(\mathbb{K}_{k+1})$ with premise cardinality $k+1$.

By Statement 1 of the induction hypothesis, the candidate set $\mathbf{C}_k$ contains all pseudo-intents of $\mathbb{K}_{k+1}$ with $k+1$ elements. Of course, all these candidates are processed in Lines 4–12 of Algorithm 1 between the states k and $k+1$. Then Statement 3 of the induction hypothesis yields that for each candidate C between the states k and $k+1$, C is $\mathcal{L}^*$-closed if, and only if, C is an intent or a pseudo-intent of $\mathbb{K}_{k+1}$. Consequently, each pseudo-intent of $\mathbb{K}_{k+1}$ of cardinality $k+1$ is recognized in Line 4. Now consider one such recognized pseudo-intent C. If it were an intent of the current formal context $\mathbb{K}$, then also of $\mathbb{K}_{k+1}$. Thus, the test for non-closedness in Line 5 passes, and the question $C \to C^{II}$ is posed to the expert χ in Line 5. If the conclusion C^{II} is too large, i.e., χ rejects the implication, then the returned counterexample X is added as a new row to $\mathbb{K}$ in Line 6. After execution of Lines 5 and 6, the implication $C \to C^{II}$ is accepted, and hence is valid in $\mathbb{K}_{k+1}$. All other counterexamples provided by the expert between states k and $k+1$ must be models of $C \to C^{II}$, and by a repeated application of Statement 2 of Lemma 12 we conclude that $C^{II} = C^{I_{k+1}I_{k+1}}$. It follows that $C \to C^{II}$ is indeed an implication of the canonical base of $\mathbb{K}_{k+1}$, and it is contained in $\mathcal{L}_{k+1}$, since it has been added to $\mathcal{L}$ in Line 8.

Eventually, it remains to show that there are no other implications in $\mathcal{L}_{k+1}$ with premise cardinality $k+1$ which are not in the canonical base of $\mathbb{K}_{k+1}$. Consider any candidate C between states k and $k+1$ that is no pseudo-intent of $\mathbb{K}_{k+1}$. An implication with premise C could only have been added to $\mathcal{L}$ if C is recognized as $\mathcal{L}^*$-closed in Line 4, i.e., only if C is an intent of $\mathbb{K}_{k+1}$. If C is also an intent of the current context $\mathbb{K}$, then no implication with premise C is added to $\mathcal{L}$, cf. Lines 4–8. Otherwise, if C is no intent of the current context $\mathbb{K}$, then the question $C \to C^{II}$ must be rejected by χ with a counterexample X (that is trivially an intent of $\mathbb{K}[X]$). Furthermore, then it holds that $C^{JJ} = C^{II} \cap X$ where J is the incidence relation of $\mathbb{K}[X]$. It remains to prove that $C = C^{JJ}$. The adjusted implication $C \to C^{II} \cap X$ is valid in $\mathbb{K}_{k+1}$, as it trivially holds in the current context $\mathbb{K}$ and the expert must accept it due to optimality, i.e., all counterexamples provided by χ (between states k and $k+1$) are models of the implication. Consequently, $C^{II} \cap X$ is a subset of $C^{I_{k+1}I_{k+1}} = C$, and thus $C = C^{JJ}$. It follows that the check for non-closedness in Line 7 fails, and hence no implication with premise C is added to $\mathcal{L}$ in Line 8.

3. We have already shown that $\mathcal{B}_{\mathsf{can}}(\mathbb{K}_{k+1})\restriction_{k+1} = \mathcal{L}_{k+1}$. Lemma 14 states that $\mathcal{B}_{\mathsf{can}}(\mathbb{K}_{k+1})\restriction_{k+1} = \mathcal{B}_{\mathsf{can}}(\mathbb{K}_{k+2})\restriction_{k+1}$, since all counterexamples $X_{k+1}^1, \ldots, X_{k+1}^{n_{k+1}}$ are models of $\mathcal{L}_{k+1}$. Consequently, for each implication set $\mathcal{L}$ with $\mathcal{L}_{k+1} \subseteq \mathcal{L} \subseteq \mathcal{L}_{k+2}$, Corollary 15 yields that an attribute set of cardinality $k+2$ is an intent or a pseudo-intent of $\mathbb{K}_{k+2}$ if, and only if, it is $\mathcal{L}^*$-closed, since $\mathcal{L}$ is a superset of $\mathcal{B}_{\mathsf{can}}(\mathbb{K}_{k+2})\restriction_{k+1}$ and furthermore only contains implications with a premise cardinality $k+2$.

1. Let P be a pseudo-intent of $\mathbb{K}_{k+2}$ with cardinality $k+2$. We have to show that P occurs as a candidate in $\mathbf{C}_{k+1}$. Beforehand, we prove an auxiliary lemma:

Lemma 17. *If $\ell < k+2$, then for all h with $\ell \leq h \leq k+2$ it holds that* $\mathsf{PsInt}(\mathbb{K}_\ell)\restriction_\ell = \mathsf{PsInt}(\mathbb{K}_h)\restriction_\ell$ *and* $\mathcal{B}_{\mathsf{can}}(\mathbb{K}_\ell)\restriction_\ell = \mathcal{B}_{\mathsf{can}}(\mathbb{K}_h)\restriction_\ell$.

Proof. Assume that $\ell < k+2$. We prove the claim by induction on h. The base case $h = \ell$ is trivial. For the inductive step assume that the statement holds for h with $\ell \leq h < k+2$. In particular, then $\mathcal{L}_h = \mathcal{B}_{\mathsf{can}}(\mathbb{K}_h)\restriction_h$. We proceed by showing the inner induction: $\mathcal{B}_{\mathsf{can}}(\mathbb{K}_h)\restriction_\ell = \mathcal{B}_{\mathsf{can}}(\mathbb{K}_h[X_h^1, \ldots, X_h^i])\restriction_\ell$ for all $i \in \{1, \ldots, n_h\}$.

base case: $X_h^1 \models \mathcal{L}_h = \mathcal{B}_{\mathsf{can}}(\mathbb{K}_h)\restriction_h \supseteq \mathcal{B}_{\mathsf{can}}(\mathbb{K}_h)\restriction_\ell$ and thus Corollary 15 yields that $\mathcal{B}_{\mathsf{can}}(\mathbb{K}_h)\restriction_\ell = \mathcal{B}_{\mathsf{can}}(\mathbb{K}_h[X_h^1])\restriction_\ell$.

inductive step: $X_h^i \models \mathcal{L}_h = \mathcal{B}_{\mathsf{can}}(\mathbb{K}_h)\restriction_h \supseteq \mathcal{B}_{\mathsf{can}}(\mathbb{K}_h)\restriction_\ell = \mathcal{B}_{\mathsf{can}}(\mathbb{K}_h[X_h^1, \ldots, X_h^{i-1}])\restriction_\ell$ and hence Corollary 15 implies $\mathcal{B}_{\mathsf{can}}(\mathbb{K}_h)\restriction_\ell = \mathcal{B}_{\mathsf{can}}(\mathbb{K}_h[X_h^1, \ldots, X_h^i])\restriction_\ell$.

Since the statement holds in particular for $i = n_h$, we conclude that $\mathcal{B}_{\mathsf{can}}(\mathbb{K}_h)\restriction_\ell = \mathcal{B}_{\mathsf{can}}(\mathbb{K}_{h+1})\restriction_\ell$, since $\mathbb{K}_{h+1} = \mathbb{K}_h[X_h^1, \ldots, X_h^{n_h}]$. $\square$

Assume that there is a pseudo-intent Q of $\mathbb{K}_{k+2}$ that is maximal w.r.t. $Q \subsetneq P$. Then it holds that $Q \subsetneq Q^{I_{k+2}I_{k+2}} \subsetneq P$, and $Q^{I_{k+2}I_{k+2}}$ is the only intent of $\mathbb{K}_{k+2}$ between Q and P. Let $\ell \coloneqq |Q|$, i.e., $Q \in \mathsf{PsInt}(\mathbb{K}_{k+2})\restriction_\ell$. Then Q must be a pseudo-intent of $\mathbb{K}_\ell$, cf. Lemma 17. Consequently, $\mathcal{L}_\ell$ contains $Q \to Q^{I_\ell I_\ell}$, it is true that $Q^{I_\ell I_\ell} = Q^{I_{k+2}I_{k+2}}$, and the candidates $Q^{I_\ell I_\ell} \cup \{m\}$ for $m \in P \setminus Q^{I_\ell I_\ell}$ have been added to $\mathbf{C}$, cf. Line 9. Hence, define the sequence

$$C_0 \coloneqq Q^{I_\ell I_\ell} \cup \{m\} \quad \text{where } m \in P \setminus Q^{I_\ell I_\ell}, \text{ and}$$
$$C_{i+1} \coloneqq (C_i)^{\mathcal{L}^*} \qquad \text{where } \mathcal{L}_{|C_i|-1} \subseteq \mathcal{L} \subseteq \mathcal{L}_{|C_i|}.$$

The attribute m for the first element C_0 of the sequence may be chosen arbitrarily. Furthermore, all following elements are well-defined, since implications in $\mathcal{L}_{|C_i|} \setminus \mathcal{L}_{|C_i|-1}$ have no influence on the closure of C_i. It is obvious that each C_i occurs as a candidate during the algorithm's run, cf. Lines 9 and 11, and that the sequence increases, i.e., $C_i \subseteq C_{i+1}$ for all indices i. We now prove by induction on i that $C_i \subseteq P$. The base case for $i = 0$ is trivial. Assume that $C_i \subsetneq P$. Consider any implication set $\mathcal{L}$ where $\mathcal{L}_{|C_i|-1} \subseteq \mathcal{L} \subseteq \mathcal{L}_{|C_i|}$. Then $C_{i+1} = (C_i)^{\mathcal{L}^*}$. Furthermore, we have that $|C_i| \leq k+2$, and thus $\mathcal{L}_{|C_i|} \subseteq \mathcal{L}_{k+2}$. Consequently,

$$\begin{aligned} C_{i+1} = (C_i)^{\mathcal{L}^*} &= (C_i)^{(\mathcal{L}_{|C_i|})^*} \\ &\subseteq (C_i)^{(\mathcal{L}_{k+2})^*} \subseteq P^{(\mathcal{L}_{k+2})^*} = P. \end{aligned}$$

If there were an index i with $C_i = C_{i+1}$, i.e., C_i were $(\mathcal{L}_{|C_i|})^*$-closed, then C_i must be an intent or a pseudo-intent of $\mathbb{K}_{|C_i|}$. In particular, $Q^{I_{k+2} I_{k+2}} \subsetneq C_i \subseteq P$. If C_i were an intent, then also one of $\mathbb{K}_{k+2}$, which contradicts the maximality of Q. Hence C_i must be a pseudo-intent, and in particular one of $\mathbb{K}_{k+2}$ by Lemma 17. Due to the fact that Q is a maximal pseudo-intent below P, we may conclude that $C_i = P$. In summary, it follows that the sequence strictly converges to P (in finitely many steps if $P \setminus Q$ is finite), i.e., ends with P, and since each element is a candidate, P must occur as a candidate in $\mathbf{C}$.

Eventually, we have to consider the case where no pseudo-intent of $\mathbb{K}_{k+2}$ below P exists. In particular, then $\emptyset$ must be an intent of $\mathbb{K}_{k+2}$. As a consequence, $\emptyset$ is an intent of $\mathbb{K}_0$, too, as otherwise there would be an implication with premise $\emptyset$ in $\mathcal{L}$. Thus, the candidates $\{m\}$ where $m \in P$ have been inserted into $\mathbf{C}$. We may now define a sequence as above, but with $C_0 \coloneqq \{m\}$ for an $m \in P$, and argue similarly as above. However, we have to additionally take care of the case $C_i = C_{i+1}$, as we may not use the maximality argument. Instead, assume that i is a minimal such index, and then merely continue the sequence with $C_{i+1} \coloneqq C_i \cup \{m\}$ where $m \in P \setminus C_i$. This choice is suitable, since then C_{i+1} is a candidate, too, cf. Line 9. It follows that it is a sequence of candidates that ends with P, i.e., $P \in \mathbf{C}_{k+2}$. □

Theorem 18. *Let $\mathbb{D}$ be an (inaccessible) formal context with a finite attribute set, $\mathbb{K}$ be a finite subcontext of $\mathbb{D}$ such that $\mathbb{K}$ and $\mathbb{D}$ share the same attribute set, i.e., $\mathbb{K} \leq_M \mathbb{D}$, and χ an expert that is induced by $\mathbb{D}$ and answers questions in finite time. If Algorithm 1 is started on $(\mathbb{K}, \chi)$ as input, then it terminates, and returns a refinement $\mathbb{K}^\circ$ of $\mathbb{K}$ with $\mathsf{Int}(\mathbb{K}^\circ) = \mathsf{Int}(\mathbb{D})$ as well as a minimal implicational base of $\mathbb{D}$.*

Furthermore, there is no algorithm that computes a minimal implicational base of $\mathbb{D}$, but poses less questions to χ than Algorithm 1.

Proof. Termination is a consequence of finiteness of $\mathbb{K}$. If $\mathbb{K}$ is finite, then it has a finite attribute set M, and consequently there may only be finitely many candidates on each level. Furthermore, the computation of closures w.r.t. the operator $\mathcal{L}^*$ can always be obtained in finite time, since the implication set $\mathcal{L}$ consists of finitely many implications at any time during the algorithm's run. Obviously, also the intent closure $\cdot^{II}$ can be computed in finite time for finite contexts. Since each candidate is only used once to pose a question to the expert, it is not possible that the expert may return infinitely many counterexamples, and hence the adjusted context cannot grow to an infinite size.

The context $\mathbb{K}^\circ \coloneqq \mathbb{K}_{|M|}$ of the final state is returned by the algorithm. It is readily verified that it contains the initial context $\mathbb{K}$ as a subcontext. Furthermore, due to the fact that $\mathbb{K}$ is itself a subcontext of $\mathbb{D}$, and during the algorithm's run only intents of $\mathbb{D}$ are added as new rows to $\mathbb{K}$, we conclude that $\mathsf{Int}(\mathbb{K}^\circ) \subseteq \mathsf{Int}(\mathbb{D})$.

By Proposition 16, it follows that in the final state $|M|$, $\mathcal{L}^\circ \coloneqq \mathcal{L}_{|M|}$ is the canonical base of $\mathbb{K}^\circ$, i.e., $\mathsf{Imp}(\mathbb{K}^\circ) = \mathsf{Imp}(\mathcal{L}^\circ)$. Furthermore, $\mathsf{Imp}(\mathbb{D}) = \mathsf{Imp}(\chi)$ by Definition 2. Since all implications in $\mathcal{L}^\circ$ have been accepted by χ, we conclude $\mathcal{L}^\circ \subseteq \mathsf{Imp}(\chi)$, and hence $\mathsf{Imp}(\mathcal{L}^\circ) \subseteq \mathsf{Imp}(\chi)$. From $\mathsf{Int}(\mathbb{K}^\circ) \subseteq \mathsf{Int}(\mathbb{D})$ it follows that $\mathsf{Imp}(\mathbb{D}) \subseteq \mathsf{Imp}(\mathbb{K}^\circ)$. (If there were an implication that is valid in $\mathbb{D}$, but is not valid in $\mathbb{K}^\circ$, then a counterexample would exist which is an intent of $\mathbb{K}^\circ$, i.e., an intent of $\mathbb{D}$. Contradiction!) Consequently, the returned implication set $\mathcal{L}^\circ$ is indeed a minimal implicational base of $\mathbb{D}$. Since $\mathcal{L}^\circ$ is sound and complete for both $\mathbb{K}^\circ$ and $\mathbb{D}$, it follows that $\mathsf{Int}(\mathbb{K}^\circ) = \mathsf{Mod}(\mathcal{L}^\circ) = \mathsf{Int}(\mathbb{D})$.

The last claim is an immediate consequence of the fact, that $\mathcal{L}_{|M|}$ is a minimal implicational base. □

5 Discussion

Of course, it would be possible to utilize multiple experts in the default *Attribute Exploration* [4–7, 13], but however this would not give any performance boost (if we assume that all experts answer immediately), as only one question in form of an implication is constructed at a time. If we compare Algorithm 1 with default *Attribute Exploration*, the order of the questions is different. They are enumerated in the lectic order, while the `ParallelAttributeExploration` enumerates w.r.t. increasing set cardinality of the premises. This means that on the one hand between two states of Algorithm 1 several implications can be processed in parallel, and on the other hand the difficulty of the questions (when measured in premise size) increases during the algorithm's run. In the default *Attribute Exploration* the difficulty of the questions varies during the algorithm's run, as they are constructed in the lectic order that does not respect set cardinality. Furthermore, the default algorithm cannot continue before the last posed question has been answered, but in contrast the parallel algorithm may process all posed questions with same premise cardinality in parallel. However, both algorithms return the same result for the same experts.

For an integration of several experts, there are the following options:

1. Randomly choose an (idle) expert, and pose the question to her.
2. Pose the question to all experts, and return the first answer.
3. Pose the question to all experts, and accept if all experts accept.
4. Pose the question to all experts, and accept if at least one expert accepts.

However, if we assume that all available experts are indeed induced by the formal context describing the domain of interest, i.e., have the same knowledge, then each of the four possibilities above would yield the same result, and hence it suffices to equally distribute the questions to all available experts. The four choices would only create different behaviours if the knowledge of the experts is not equivalent, or if the answering delays vary.

For instance, assume that there are experts $\chi_1, \ldots, \chi_n$ such that each χ_i is induced by a formal context $\mathbb{D}_i \leq_M \mathbb{D}$, and $\bigcup_{i=1}^{n} \mathbb{D}_i = \mathbb{D}$, i.e., each expert knows a part of the domain of interest, and no part of the domain of interest is unknown. Then an implication is valid in $\mathbb{D}$ if, and only if, it is valid in each $\mathbb{D}_i$. An induced expert χ of $\mathbb{D}$ is then obtained with the following definition: For an implication $X \to Y$, let $\chi \models X \to Y$ if $\chi_i \models X \to Y$ for all indices $i \in \{1, \ldots, n\}$, and otherwise define $\chi(X \to Y)$ as an arbitrary element of $\{ C \mid \exists i \in \{1, \ldots, n\} \colon \chi_i(X \to Y) = C \}$, i.e., query all experts, and return the first counterexample, or accept otherwise.

6 Conclusion

We have considered the problem of *Parallel Attribute Exploration*, where a (minimal) implicational base for a domain of interest shall be computed in a parallel manner. The domain of interest is a formal context of which only some objects and their intents are known, and furthermore some experts are available that can correctly decide whether implications are valid. The introduced algorithm `ParallelAttributeExploration` is an extension of the algorithm `NextClosures` [10, 11], and a prototypical implementation is available [9]. It is planned to utilize it for a collaborative knowledge acquisition platform.

As a future step, the algorithm will be further extended to handle background knowledge, as this has been done by [4, 13] for the default *Attribute Exploration* with lectic order. Furthermore, the algorithm will be generalized to the case where the data-set is described in terms of a closure operator in a (graded) complete lattice.

Acknowledgements The author gratefully thanks the anonymous reviewers for their constructive hints and helpful remarks.

References

[1] Dana Angluin, Michael Frazier, and Leonard Pitt. "Learning Conjunctions of Horn Clauses". In: *Machine Learning* 9 (1992), pp. 147–164.

[2] Marta Arias and José L. Balcázar. "Canonical Horn Representations and Query Learning". In: *Algorithmic Learning Theory, 20th International Conference, ALT 2009, Porto, Portugal, October 3-5, 2009. Proceedings.* 2009, pp. 156–170.

[3] Daniel Borchmann. "Learning Terminological Knowledge with High Confidence from Erroneous Data". PhD thesis. Technische Universität Dresden, 2014.

[4] Bernhard Ganter. "Attribute Exploration with Background Knowledge". In: *Theor. Comput. Sci.* 217.2 (1999), pp. 215–233.

[5] Bernhard Ganter. *Two Basic Algorithms in Concept Analysis.* FB4-Preprint 831. Darmstadt, Germany: Technische Hochschule Darmstadt, 1984.

[6] Bernhard Ganter. "Two Basic Algorithms in Concept Analysis". In: *Formal Concept Analysis, 8th International Conference, ICFCA 2010, Agadir, Morocco, March 15-18, 2010. Proceedings.* 2010, pp. 312–340.

[7] Bernhard Ganter and Rudolf Wille. *Formal Concept Analysis: Mathematical Foundations.* Springer, 1999.

[8] Jean-Luc Guigues and Vincent Duquenne. "Famille minimale d'implications informatives résultant d'un tableau de données binaires". In: *Mathématiques et Sciences Humaines* 95 (1986), pp. 5–18.

[9] Francesco Kriegel. *Concept Explorer FX.* Software for Formal Concept Analysis. 2010–2016. URL: `https://github.com/francesco-kriegel/conexp-fx`.

[10] Francesco Kriegel. *NextClosures – Parallel Exploration of Constrained Closure Operators.* LTCS-Report 15-01. Dresden, Germany: Chair of Automata Theory, Institute of Theoretical Computer Science, Technische Universität Dresden, 2015.

[11] Francesco Kriegel and Daniel Borchmann. "NextClosures: Parallel Computation of the Canonical Base". In: *Proceedings of the 12th International Conference on Concept Lattices and Their Applications (CLA 2015), Clermont-Ferrand, France, October 13-16, 2015*. 2015, pp. 181–192.

[12] David Maier. *The Theory of Relational Databases*. Computer Science Press, 1983.

[13] Gerd Stumme. "Attribute Exploration with Background Implications and Exceptions". In: *Data Analysis and Information Systems*. Ed. by Hans-Hermann Bock and Wolfgang Polasek. Studies in Classification, Data Analysis, and Knowledge Organization. Springer, 1996, pp. 457–469.