Available online at www.sciencedirect.com
ScienceDirect
Discrete Applied Mathematics 156 (2008) 1994–2003
DISCRETE APPLIED MATHEMATICS
www.elsevier.com/locate/dam

Some decision and counting problems of the Duquenne–Guigues basis of implications

Sergei O. Kuznetsov[a], Sergei Obiedkov[b,1]

[a]*Department of Applied Mathematics, State University Higher School of Economics, Moscow, Russia*
[b]*Department of Computer Science, University of Pretoria, Pretoria, South Africa*

Received 21 October 2005; received in revised form 25 December 2006; accepted 6 April 2007
Available online 8 May 2007

Abstract

Implications of a formal context obey Armstrong rules, which allows one to define a minimal (in the number of implications) implication basis, called Duquenne–Guigues basis or stem base in the literature. In this paper we show how implications are reduced to functional dependencies and prove that the problem of determining the size of the stem base is a #P-complete problem.

Keywords: Implication basis; Duquenne–Guigues basis; #P-completeness; Functional dependencies

1. Introduction

Formal concept analysis (FCA) is a mathematical theory oriented, in particular, at applications in knowledge representation, knowledge acquisition, and data analysis. It provides tools for understanding the structure of the data given as a set of objects described in terms of attributes they possess, which is done by representing the data as a hierarchy of concepts, or more exactly, as a concept lattice [7]. Every concept has extent (the set of objects that fall under the concept) and intent (the set of attributes that together are necessary and sufficient for an object to be an instance of the concept). Concepts are ordered in terms of being more general or less general (i.e., covering more objects or fewer objects).

The concept lattice, being a rather universal structure, provides a wealth of information about the relations among objects and attributes, which made possible applications in areas ranging from history and sociology to software engineering and machine learning to e-mail management and ontology building. Indeed, it can help in processing a wide class of data types (for example, any data represented as a table). Apart from structural representation of data, the concept lattice provides a framework in which various data analysis and knowledge acquisition techniques can be formulated.

An important notion in FCA is that of an attribute implication. An implication asserts a certain relationship between two attribute sets, called premise and conclusion: an implication is valid in the data set if every object that has all attributes from the premise of the implication also has all attributes from its conclusion. The set of all valid implications can be summarized by its (relatively small) subset, cover, from which all other implications follow semantically. For

[1] Present address: Department of Applied Mathematics, State University Higher School of Economics, Moscow, Russia.
E-mail address: skuznetsov@hse.ru (S.O. Kuznetsov).

0166-218X/$ - see front matter © 2007 Published by Elsevier B.V.
doi:10.1016/j.dam.2007.04.014

every data set, there is a particular implication cover, called the Duquenne–Guigues basis, which is proven to have the minimal size among all covers [10].

By capturing certain patterns in data, implications are quite relevant for data analysis, where they are known as exact association rules [16]. In fact, association rules can be regarded as (partial) implications that admit some exceptions, but do have a substantial support (i.e., there is a sufficiently large number of objects that have all the attributes of both the premise and the conclusion of the rule). There exist various bases of association rules [16,20]; typically, they consist of two parts: the Duquenne–Guigues basis of implications and some modification of the Luxenburger basis for partial implications [14]. Association rules with substantial support are easily constructed from (closed) sets of attributes that occur in a large number of objects ("frequent (closed) itemsets") [16], which gave rise to research in algorithmic problems related to generation of such attribute sets [2].

It is worth noting that there is a strong connection between implications in FCA and functional dependencies in database theory. For any relation in a database, it is possible to construct a context (binary object-attribute data set) such that the functional dependencies of the former are precisely the attribute implications of the latter [7] (in this paper we present the inverse reduction). Thus, the Duquenne–Guigues basis provides a compact representation of the functional dependencies of a relation.

The paper is organized as follows. In Section 2, we introduce basic notation and definitions. Section 3 studies the relationship between functional dependencies and implications. Next, we show that, in the worst case, the size of the Duquenne–Guigues basis exponentially depends on the size of the input data and that the problem of computing the size of the basis is #P-complete.

2. Main definitions and problem statement

First we recall some basic notions of FCA [7,19].

Definition 2.1. Let G and M be sets, called the set of objects and the set of attributes, respectively. Let I be a relation $I \subseteq G \times M$ between objects and attributes: for $g \in G,\ m \in M$, gIm holds iff the object g has the attribute m. The triple $K = (G, M, I)$ is called a (*formal*) *context*. Formal contexts are naturally represented by cross tables, where a cross for a pair (g, m) means that this pair belongs to the relation I. If $A \subseteq G$, $B \subseteq M$ are arbitrary subsets, then the *Galois connection* is given by the following *derivation operators*:

$$A' := \{m \in M \mid gIm \textit{ for all } g \in A\},$$

$$B' := \{g \in G \mid gIm \textit{ for all } m \in B\}.$$

The pair (A, B), where $A \subseteq G$, $B \subseteq M$, $A' = B$, and $B' = A$ is called a (*formal*) *concept* (*of the context* K) with *extent* A and *intent* B. For $g \in G$ and $m \in M$ the sets $\{g\}'$ and $\{m\}'$ are called *object intent* and *attribute extent*, respectively. The set of attributes B *is implied by the set of attributes* D, or an *implication* $D \to B$ holds, if all objects from G that have all attributes from the set D also have all attributes from the set B, i.e., $D' \subseteq B'$.

The operation $(\cdot)''$ is a closure operator [7], i.e., it is idempotent ($X'''' = X''$), extensive ($X \subseteq X''$), and monotone ($X \subseteq Y \Rightarrow X'' \subseteq Y''$). Sets $A \subseteq G$, $B \subseteq M$ are called *closed* if $A'' = A$ and $B'' = B$. Obviously, extents and intents are closed sets. Since the closed sets form a closure system or a Moore space [1], the set of all formal concepts of the context K forms a lattice, called a *concept lattice* and usually denoted by $\mathfrak{B}(K)$ in FCA literature.

Implications obey Armstrong rules:

$$\frac{}{A \to A}, \quad \frac{A \to B}{A \cup C \to B}, \quad \frac{A \to B,\ B \cup C \to D}{A \cup C \to D}.$$

A minimal (in the number of implications) subset of implications from which all other implications of a context can be deduced by means of Armstrong rules was characterized in [9,10]. This subset is called the Duquenne–Guigues basis or stem base in the literature. The premises of implications of the stem base can be given by pseudo-intents (see, e.g., [7]).

Definition 2.2. A set $P \subseteq M$ is a *pseudo-intent* if $P \neq P''$ and $Q'' \subset P$ for every pseudo-intent $Q \subset P$.

Since the introduction of the stem base, a longstanding problem was that concerning the upper bound of the size of the stem base: whether the stem base can be exponential in the size of the input, i.e., in $|G|$ and $|M|$.

3. Functional dependencies and implications

A *many-valued context* [7] is a tuple (G, M, W, I), where W is the set of attribute values, $I \subseteq G \times M \times W$, such that $(g, m, w) \in I$ and $(g, m, v) \in I$ implies $w = v$.[2] An attribute m is *complete* if for all $g \in G$ there is $w \in W$ such that $(g, m, w) \in I$. A many-valued context is complete if all its attributes are complete. For complete many-valued contexts, the value of the attribute m in the object g is denoted by $m(g)$; thus, $(g, m, m(g)) \in I$. *Functional dependency* $X \to Y$ is valid in a complete many-valued context (G, M, W, I) if the following holds for every pair of objects $g, h \in G$:

$$(\forall m \in X \; m(g) = m(h)) \Rightarrow (\forall n \in Y \; n(g) = n(h)).$$

In [7] it was shown that having a complete many-valued context $K = (G, M, W, I)$, one defines the context $K_N := (\mathscr{P}_2(G), M, I_N)$, where $\mathscr{P}_2(G)$ is the set of all pairs of different objects from G and I_N is defined by

$$\{g, h\} I_N m :\Leftrightarrow m(g) = m(h).$$

Then, a set $Y \subseteq M$ is functionally dependent on the set $X \subseteq M$ in K iff the implication $X \to Y$ holds in the context K_N. Theorem 3.1 introduces an inverse reduction.

Theorem 3.1. *For a context $K = (G, M, I)$ one can construct a many-valued context K_W such that an implication $X \to Y$ holds iff Y is functionally dependent on X in K_W.*

Proof. Assume that the objects from G are given by natural numbers from 1 to $|G|$. We construct the required many-valued context $K_W = (G \cup \{g_0\}, M, W, I_W)$. The attributes of M take values from the set $W = \{0, \ldots, |G|\}$. The relation I_W is defined as follows: $m(g_0) = 0$ for all $m \in M$ and for $g \in G$

$$m(g) = 0 \quad \text{if } (g, m) \in I \text{ and}$$

$$m(g) = g \quad \text{otherwise.}$$

Let $X \to Y$ be an implication of the context K. By the definition of the implication, we have $X' \subseteq Y'$. It means that $(g, x) \in I$ for all $x \in X$ implies $(g, y) \in I$ for all $y \in Y$. Then, in the context K_W, $x(g) = 0$ for all $x \in X$ implies $y(g) = 0$ for all $y \in Y$. Let $x(g_i) = x(g_j)$ for all $x \in X$. By the construction of the many-valued context K_W, all nonzero attribute values are different for different objects. Hence, $x(g_i) = x(g_j) = 0$ for all $x \in X$ and, by the above condition we have $y(g_i) = y(g_j) = 0$ for all $y \in Y$, which means that the dependency $X \to Y$ is valid in the context K_W.

In the other direction, let the dependency $X \to Y$ hold in the context K_W and let $x(g) = 0$ for some $g \in G$ and all $x \in X$. By the construction of K_W, $m(g_0) = 0$ for all $m \in M$. Hence, $x(g_0) = x(g)$ for all $x \in X$. By the definition of dependency $X \to Y$ for the context K_W, we have $y(g_0) = y(g)$ for all $y \in Y$, but since $m(g_0) = 0$ for all $m \in M$, we have $y(g) = 0$ for all $y \in Y$. Therefore, $(g, x) \in I$ for all $x \in X$ implies $(g, y) \in I$ for all $y \in Y$ in the context K, which means that $X' \subseteq Y'$ holds and $X \to Y$ is an implication in the context K. □

Example. Consider the context $K = (G, M, I)$ given by the following cross-table:

$G \setminus M$	a	b	c	d
1	×	×		
2	×		×	×
3		×		×
4	×		×	

[2] Many-valued contexts correspond to relations (tables) in databases.

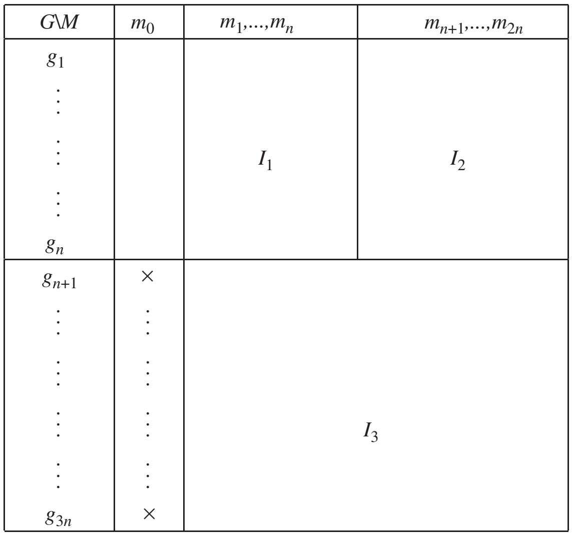

Fig. 1. A context with an exponential number of implications in the Duquenne–Guigues basis.

The many-valued context with the set of functional dependencies corresponding to implications in the above context is given by the following table:

$G\backslash M$	a	b	c	d
0	0	0	0	0
1	0	0	1	1
2	0	2	0	0
3	3	0	3	0
4	0	4	0	4

4. Counting pseudo-intents is hard

A concept lattice can be exponential in the size of the context (e.g., when it is a Boolean one). Moreover, the problem of determining the size of a concept lattice is #P-complete (see e.g., [11]). There are several polynomial-delay algorithms for computing the set of all concepts (see e.g., review [12]). However, neither an efficient (polynomial-delay) algorithm, nor a good upper bound for the size of the stem base was known. It is easy to show that there can be a stem base of size exponential with respect to $|M|$, for example when object intents are exactly all possible subsets of size $|M|/2$. However, in this case $|G|$, as well as $|I|$, are also exponential in $|M|$, and the number of pseudo-intents is polynomial in $|I|$.

A solution to the question whether a stem base can be exponential in the size of the context, i.e., in $|G| \times |M|$, is obtained by observing a fact about functional dependencies, namely, that the size of a minimum cover of functional dependencies can be exponential in the size of the relation [15]. Combined with the above-mentioned fact that having an $m \times n$ relation R, we can construct an $m(m-1)/2 \times n$ context K such that implications of K are exactly all functional dependencies of R, this immediately suggests that the size of the stem base can be exponential with respect to the context size. In [13] we proposed a very simple general form of contexts with the lattice of closed attribute sets isomorphic to that in [15]. For these contexts, the size of the stem base is exponential in the relation size.

Consider a context $K_{\exp,n} = (G, M, I)$ schematically represented by the table in Fig. 1, where $G = G_1 \cup G_2$, $M = M_1 \cup M_2 \cup \{m_0\}$, $I = I_1 \cup I_2 \cup I_3 \cup \{m_0\} \times G_2$, and subcontexts $K_1 = (G_1, M_1, I_1)$, $K_2 = (G_1, M_2, I_2)$, $K_3 = (G_2, M_1 \cup M_2, I_3)$ are of the form $(A, A, \neq)$. More formally, objects and attributes are $G_1 = \{g_1, \ldots, g_n\}$, $G_2 = \{g_{n+1}, \ldots, g_{3n}\}$, $M_1 = \{m_1, \ldots, m_n\}$, and $M_2 = \{m_{n+1}, \ldots, m_{2n}\}$. The relations I_1, I_2, and I_3 are defined as follows: $g_i I_1 m_j$ iff $i \neq j$, $g_i I_2 m_j$ iff $i \neq j - n$, and $g_i I_3 m_j$ iff $i \neq n + j$ for g_i and m_j from corresponding sets of objects and attributes. For m_0 and $g \in G$, one has gIm_0 iff $g \in G_2$.

Theorem 4.1. *The number of pseudo-intents of the context $K_{\exp,n}$ is 2^n.*

Proof. First note that the set of attributes $\{m_1, \ldots, m_n\}$ is a pseudo-intent. Indeed, for a subset

$$B = \{m_{j_1}, \ldots, m_{j_k}\} \subset \{m_1, \ldots, m_n\} = M_1$$

we have

$$B' = (G_1 \backslash \{g_{j_1}, \ldots, g_{j_k}\}) \cup (G_2 \backslash \{g_{n+j_1}, \ldots, g_{n+j_k}\})$$

and $B''=B$, i.e., the set B is closed. However, the set $\{m_1, \ldots, m_n\}$ is not closed, since $\{m_1, \ldots, m_n\}''=\{m_0, m_1, \ldots, m_n\}$. If a set is not closed and all its subsets are closed, then it is a pseudo-intent by definition. If we replace m_i in the pseudo-intent $\{m_1, \ldots, m_n\}$ with m_{n+i}, then the resulting set

$$\{m_1, \ldots, m_{i-1}, m_{n+i}, m_{i+1}, \ldots, m_n\}$$

is still a pseudo-intent, because it is not closed:

$$\{m_1, \ldots, m_{i-1}, m_{n+i}, m_{i+1}, \ldots, m_n\}'' = \{m_0, m_1, \ldots, m_{i-1}, m_{n+i}, m_{i+1}, \ldots, m_n\}$$

and every subset

$$C \subset \{m_1, \ldots, m_{i-1}, m_{n+i}, m_{i+1}, \ldots, m_n\}$$

is closed by the same argument as for $B \subset M_1$. We can replace each m_i with m_{n+i} obtaining another pseudo-intent. Since the substitution of m_{n+i} for m_i can be done independently for each i, we have 2^n pseudo-intents. It can be easily checked that there are no other pseudo-intents. □

Note that in our example pseudo-intents are at the same time proper premises (see, e.g., [7]), which are premises of the so-called direct basis [18], i.e., an irredundant set of implications such that, for any $A \subseteq M$ and any $m \in A'' \backslash A$, this set contains an implication $B \to C$, where $B \subseteq A$ and $m \in C$.[3] Moreover, here all pseudo-intents are so-called minimal positive hypotheses (see, e.g., [6]) w.r.t. the target attribute m_0.

It is worth considering the structure of the concept lattice of this context in more detail. Two important parts of the lattice can be distinguished: the one where intents contain attribute m_0 and the one with intents not containing m_0. The former consists of all subsets of $M_1 \cup M_2$ augmented with m_0. Thus, it is a Boolean sublattice of the lattice of $K_{\exp,n}$. The other part is generated by 2^n subcontexts of $K_{\exp,n}$ of the form $(G, M_{12}, (G \times M_{12}) \cap I)$, where $M_{12} = \{m_{i_1}, \ldots, m_{i_n}\}$ and $i_j \in \{j, j+n\}$. The concept lattices of these contexts are Boolean lattices, and, except for their bottom elements, they are (non-disjoint) subsets of the concept lattice of $K_{\exp,n}$. The intents of their bottom elements are exactly all the pseudo-intents of $K_{\exp,n}$, while their closures contain m_0 and belong to the first part.

Another important property of the context is that its nontrivial implications are only those with m_0 in the right-hand side.

Proposition 4.2. *In the context $K_{\exp,n} = (G, M, I)$, for any attribute implication of the form $A \to \{t\}$ where $A \subseteq M$, $t \in M$, $t \notin A$, it holds that $t = m_0$.*

Proof. Consider an implication $A \to \{t\}$, $A \subseteq M$, $t \in M$, $t \neq m_0$, $t \notin A$. Due to the structure of K_3, there is an object $g \in G_2$ such that $A \subseteq \{g\}'$ and $t \notin \{g\}'$. Therefore, $A \to \{t\}$ is not possible unless $t = m_0$. □

Apart from the exponential boundary on the size of the stem base, it should be noted that the problem of counting pseudo-intents is also intractable [13]. Recall the definition of the class of counting problems #P [17]. This is the class

[3] With a direct basis, to obtain A'' by adding to A the conclusions of implications whose premises are subsets of A, every implication must be considered at most once no matter in what order implications are presented. This is not the case in general, since the premise of an already considered implication may later become a subset of A as a result of enlarging A.

$G\backslash M$	v_0	$v_1, \ldots, v_{\lvert V\rvert}$
$e_1, \ldots, e_{\lvert E\rvert}$		$\overline{\mathscr{I}}$
$v_1, \ldots, v_{\lvert V\rvert}$	× … ×	$\neq$

Fig. 2. The context corresponding to an arbitrary graph from the proof of Theorem 4.3.

of problems of the form "compute $f(x)$", where $f(x)$ is the number of accepting paths of an NP machine on input x. In other words, a problem is in #P if there is a non-deterministic, polynomial-time Turing machine that, for each instance I of the problem, has a number of accepting computations that is exactly equal to the number of distinct solutions for instance I. A problem is #P-hard if any problem in #P can be reduced by Turing to it in polynomial time (whereas membership of the problem in #P may be not clear).

Theorem 4.3. *The problem*

INPUT A formal context $K = (G, M, I)$
OUTPUT The number of pseudo-intents of K
is #P-hard.

Proof. Consider an arbitrary graph (V, E) and the context $K = (E \cup V, V \cup \{v_0\}, I)$, where I is defined as follows. For $e \in E$ and $v, w \in V$, one has eIw iff $w \notin e$ (i.e., the vertex w is not incident to the edge e) and vIw iff $v \neq w$. For v_0, one has gIv_0 iff $g \in V$.

In terms of FCA, the context K is the subposition of two contexts, which is represented schematically in Fig. 2. Here, $\overline{\mathscr{I}}$ is the complement of the vertex–edge incidence relation of the graph (V, E): $e\overline{\mathscr{I}}v$ iff v is not incident to e (or $v \notin e$); $\neq$ denotes the "zero-diagonal" relation (only the diagonal pairs do not belong to it).

Recall that in a graph (V, E) a subset $W \subseteq V$ is a vertex cover if every edge $e \in E$ is incident to some $w \in W$. A cover is minimal if neither of its proper subsets is a cover. The problem of counting all minimal covers has been proved to be #P-complete in [17]. We show that for a graph (V, E) pseudo-intents of the context in Fig. 2 are in one-to-one correspondence with minimal vertex covers of (V, E).

Indeed, if a subset $W \subseteq V$ of vertices is a minimal cover, then by definition of $\overline{\mathscr{I}}$, for each $e \in E$ there is an attribute $w \in W$ such that $e\overline{\mathscr{I}}\, w$ does not hold. Thus, the set W' will not contain any object from e. Hence, W'' will contain v_0 and, thus, W is not closed ($W'' \neq W$). However, for any proper subset $Q \subset W$ we have $Q'' = Q$ (because Q' contains an object from E). Thus, by definition, W is a pseudo-intent.

In the opposite direction, for each $v \in V$ consider a set $W \subset V$ such that $v \notin W$. Since the object v has all attributes from M except for the attribute v, the implication $W \to \{v\}$ does not hold and there are no nontrivial implications with v on the right-hand side. The only possible nontrivial implications are of the form $W \to \{v_0\}$. Hence, if W is a pseudo-intent of the context, then W' should not contain any object from E. Thus, by the definition of $\overline{\mathscr{I}}$, the set W is a vertex cover. This cover is minimal, since otherwise there had existed a non-closed subset $Q \subset W$ with $Q'' = Q \cup \{v_0\}$, which contradicts the fact that W is a pseudo-intent such that $W'' = W \cup \{v_0\}$.

Thus, we have reduced the decision problem of finding a minimal vertex cover to the problem of finding a pseudo-intent. The reduction is obviously polynomial. □

5. Counting sets that are not pseudo-intents is in #P

In [4,9], the notions of (minimal) saturated gaps, called *noeuds* (*minimaux*) *de non-redondance* in [10], were formulated. The terms *quasi-closed* and *pseudo-closed* were introduced in [5]. The definitions of (minimal) saturated gaps and those of quasi- and pseudo-closed sets are different but equivalent (except that saturated gaps are not closed by definition). We use notation from [5].

Definition 5.1. A set $Q \subseteq M$ is *quasi-closed* if for any $R \subseteq Q$ one has $R'' \subseteq Q$ or $R'' = Q''$.

For example, closed sets are quasi-closed. Below we will use the following properties of quasi-closed sets from Propositions 5.2 and 5.3 considered in [3–5,9] in different terms. We present their detailed proofs to make the paper self-sufficient and to promote consistency of terminology.

Proposition 5.2. *A set $Q \subseteq M$ is quasi-closed iff $Q \cap C$ is closed for every closed set C with $Q \nsubseteq C$.*

Proof. Suppose that $Q \subseteq M$ is quasi-closed and let C be an arbitrary closed set with $Q \nsubseteq C$. Then $(Q \cap C)'' \subseteq C$, hence $(Q \cap C)'' \neq Q''$ and, by definition of a quasi-closed set, one has $(Q \cap C)'' \subseteq Q$. It follows that $(Q \cap C)'' \subseteq Q \cap C$ and $(Q \cap C)'' = Q \cap C$, i.e., $Q \cap C$ is closed.

In the other direction, suppose that $Q \cap C$ is closed for every closed set C with $Q \nsubseteq C$. Let $R \subseteq Q$ and $R'' \neq Q''$ (more precisely, $R'' \subset Q''$). Then $Q \cap R''$ is closed, which together with $R \subseteq Q$ implies $Q \cap R'' = R''$. Hence $R'' \subseteq Q$, which by definition means that Q is quasi-closed. □

Corollary. *$Q \subseteq M$ is quasi-closed iff adding Q to the set of concept intents results in a closure system.*

Proposition 5.3. *Intersection of quasi-closed sets is quasi-closed.*

Proof. Let Q and R be quasi-closed and C'' be an arbitrary closed subset of M. If $Q \subseteq C''$ or $R \subseteq C''$, then $(Q \cap R) \subseteq C''$. Otherwise, $(Q \cap R) \cap C'' = (Q \cap C'') \cap (R \cap C'')$ is closed. Thus, by Proposition 5.2 $Q \cap R$ is quasi-closed. □

Corollary. *Quasi-closed sets of a context make a closure system.*

Actually, we can weaken the condition over all subsets R of the set Q in Definition 5.1 to a condition over all quasi-closed subsets of Q.

Proposition 5.4. *Let $K = (G, M, I)$ be a context and $Q \subseteq M$. Then the following two statements are equivalent*:

(1) *Q is quasi-closed*;
(2) *For any quasi-closed $R \subseteq Q$ one has $R'' \subseteq Q$ or $R'' = Q''$.*

Proof. $1 \to 2$: Follows directly from Definition 5.1.

$2 \to 1$: The proof is by induction over the size of Q. For $|Q| = 0$ the proposition holds (the empty set is always quasi-closed). Let the proposition hold for any Q such that $|Q| = i < |M|$; let $|Q| = i + 1$. Consider an arbitrary set S such that $Q \nsubseteq S''$. We show that the set $Q \cap S''$ is quasi-closed. Since $|Q \cap S''| \leqslant i$, it suffices to show that for any quasi-closed set $R \subseteq Q \cap S''$, one has either $R'' \subseteq Q \cap S''$ or $R'' = (Q \cap S'')''$. By the condition, either $R'' \subseteq Q$ or $R'' = Q''$. In the first case, since $R \subseteq Q \cap S''$, one has $R \subseteq S''$ and, by the monotonicity and idempotence of the closure operator, $R'' \subseteq S''$, hence $R'' \subseteq Q \cap S''$. Since $R'' \subseteq S''$, in the second case we have $Q'' = R'' \subseteq S''$, which is impossible since Q, in contrast to R'', is not a subset of S'' by the assumption about the set S. Hence, only the first case is possible. The fact that $Q \cap S''$ is quasi-closed implies $(Q \cap S'')'' \subseteq Q$ by the condition of the proposition (the case when $Q'' = (Q \cap S'')'' \subseteq S''$ is impossible by the assumption that $Q \nsubseteq S''$). On the other hand, $(Q \cap S'')'' \subseteq S''$

and therefore, $(Q \cap S'')'' \subseteq Q \cap S''$, i.e., $Q \cap S'' = (Q \cap S'')''$. Thus, $Q \cap S''$ is closed for arbitrary S such that $Q \nsubseteq S''$, and therefore, Q is quasi-closed by Proposition 5.2. □

Definition 5.5. A quasi-closed set P is called *pseudo-closed* if $P'' \neq P$ and for every quasi-closed set $Q \subset P$ one has $Q'' \subset P$.

Proposition 5.6. *A set P is pseudo-closed if and only if $P \neq P''$ and $Q'' \subset P$ for every pseudo-closed $Q \subset P$.*

Proof. The necessity is obvious by the fact that pseudo-closed sets are quasi-closed. Now we prove sufficiency. Let $Q \subset P$ be a quasi-closed set. Either Q is pseudo-closed or there exists a pseudo-closed set R with $R'' = Q''$. In any case, $Q'' \subset P$ and therefore, P is pseudo-closed. □

Corollary. *Every pseudo-closed subset of M is a pseudo-intent and vice versa.*

Therefore, we use these terms interchangeably. By Proposition 5.6, a pseudo-intent is a minimal quasi-closed set in its closure class, i.e., among quasi-closed sets with the same closure. In some closure classes there can be several minimal quasi-closed elements.

Proposition 5.7. *A set S is quasi-closed iff for any object $g \in G$ either $S \cap \{g\}'$ is closed or $S \cap \{g\}' = S$.*

Proof. By Proposition 5.2, to test quasi-closedness of $S \subseteq M$, one should verify that for all $R \subseteq M$ the set $S \cap R''$ is closed or equal to S. Any closed set of attributes R'' can be represented as intersection of some object intents: $R'' = \{g_{i_1}\}' \cap \cdots \cap \{g_{i_k}\}'$ and $S \cap R'' = (S \cap \{g_{i_1}\}') \cap \cdots \cap (S \cap \{g_{i_k}\}')$. If $S \cap \{g_{i_j}\}' = S$ for all $j \leqslant k$, then $S \cap R'' = S$. Otherwise,

$$S \cap R'' = \bigcap_{j \in \{1,\ldots,k\}} \{S \cap \{g_{i_j}\}' \mid S \cap \{g_{i_j}\}' \neq S\}.$$

Thus, if the intersection of S with each of $|G|$ object intents is either closed or coincides with S, then this also holds for the intersection of S with any R''. If $S \cap \{g\}'$ is not closed and $S \cap \{g\}' \neq S$ for some g, then this suffices to say that S is not quasi-closed. □

Corollary. *Testing whether $S \subseteq M$ is quasi-closed in the context (G, M, I) can be done in* $\mathrm{O}(|G|^2 \cdot |M|)$ *time.*

Proof. By Proposition 5.7, to test whether S is quasi-closed, it suffices to compute intersection of S with intents of all objects from G and check whether these intersections are closed or equal to S. Testing closedness of intersection of S with an object intent takes $\mathrm{O}(|G| \cdot |M|)$ time, testing this for all $|G|$ objects takes $\mathrm{O}(|G|^2 \cdot |M|)$ time. □

Let $q : 2^M \to 2^M$ be a mapping taking a set to its minimal quasi-closed superset:

$$q(S) = \bigcap \{Q \mid S \subseteq Q \text{ and } Q \text{ is quasi-closed}\}.$$

It is easily seen that the mapping q is a closure operator. Indeed, the extensity, monotonicity, and idempotence of $q(S)$ are obvious.

Proposition 5.8. *Computing $q(S)$ for $S \subseteq M$ in the context (G, M, I) can be done in* $\mathrm{O}(|G|^2 \cdot |M|^2)$ *time.*

Proof. The quasi-closure of $S \subseteq M$, that is, $q(S)$, can be computed with the following simple algorithm:

$Q := S$
while there is $g \in G$ such that $(Q \cap \{g\}')'' \neq Q \cap \{g\}' \neq Q$
$Q := Q \cup (Q \cap \{g\}')''$

In more detail, let the set of objects G be linearly ordered (i.e., objects are numbered). Let us consider a set $Q \subseteq M$. First, let $Q := S$. By Proposition 5.7, the set Q is not quasi-closed iff there exists an object $g \in G$ such that $Q \cap \{g\}'$ is not closed and $Q \cap \{g\}' \neq Q$. Having found the first such g, we put $Q := Q \cup (Q \cap \{g\}')''$. This is necessary to attain quasi-closedness of Q: $Q \cap \{g\}' \subseteq Q$, but $(Q \cap \{g\}')'' \neq Q''$; therefore, $(Q \cap \{g\}')'' \subset Q$ should hold by Proposition 5.4. As a result, the intersection $Q \cap g'$ becomes closed. We find the next object $g \in G$ that violates quasi-closedness and continue in this fashion until the last element of G. We start again with the first element of G and iterate until we cannot find such an object. Obviously, $Q = q(S)$ at this point. At each step we add at least one attribute to Q. Therefore, there can be no more than $|M|$ updates of Q. However, in the worst case each update is preceded by $|G|$ intersections with object intents. Computing intersection $Q \cap \{g\}'$, testing $Q \cap \{g\}'$ for closedness, testing condition $Q \cap g' \neq Q$, and adding attributes take $O(|G| \cdot |M|)$ time. Therefore, computing $q(S)$ takes $O(|G|^2 \cdot |M|^2)$ time. □

This method of computing the smallest quasi-closed superset can be used in candidate selection when computing the implication basis by the algorithm from [5].

Proposition 5.9. *The problem*

INSTANCE A context $K = (G, M, I)$ and a set $S \subseteq M$.
QUESTION Is S not a pseudo-intent of $K = (G, M, I)$?
is in NP.

Proof. A nonclosed set S is pseudo-closed iff there is no quasi-closed $Q \subset S$ with $Q'' = S''$. First, we test if S is closed. If it is, then it is not pseudo-closed and the answer to our problem is positive. Otherwise, we obtain for S a suitable set Q and test whether Q is a quasi-closed subset of S such that $Q'' = S''$. By the corollary of Proposition 5.7, this test can be done in polynomial time. □

Corollary. *The problem*

INSTANCE A context $K = (G, M, I)$ and a set $S \subseteq M$.
QUESTION Is S a pseudo-intent of $K = (G, M, I)$?
is in coNP [8].

Now we consider the problem of counting the number of all pseudo-intents of a context. Since the problem of checking whether a set is non-pseudo-closed is in NP, the problem of determining the number of such sets is in #P. Since the problem of counting pseudo-intents (and thus, the problem of counting non-pseudo-intents) is #P-hard, we have

Proposition 5.10. *The problem*

INPUT A context $K = (G, M, I)$
OUTPUT The number of sets that are not pseudo-intents
is #P-complete.

Obviously, the latter problem is "trivially equivalent" to the problem of counting pseudo-intents, since the number of sets that are not pseudo-intents is k iff the number of pseudo-intents is $2^{|M|} - k$.

Acknowledgments

The first author would like to pay tribute to the brilliant lectures of Leonid G. Khachiyan on the theory of computational complexity held in the early 1980s at the Computer Center of the USSR Academy of Sciences (Moscow). The first author was supported by the DFG project Concepts and Models (COMO).

References

[1] G.D. Birkhoff, Lattice Theory, American Mathematical Society, 1979.
[2] E. Boros, V. Gurvich, L. Khachiyan, K. Makino, On the complexity of generating maximal frequent and minimal infrequent sets, in: Proceedings of the 19th International Symposium on Theoretical Aspects of Computer Science (STACS'02), Lecture Notes in Computer Science, vol. 2285, Springer, Berlin, 2002, pp. 133–141.

[3] V. Duquenne, Contextual implications between attributes and some representation properties for finite lattices, in: B. Ganter (Ed.), Beiträge zur Begriffsanalyse, Mannheim, 1987, pp. 213–239.
[4] V. Duquenne, J.-L. Guigues, Galois Lattice Revival, in: Communication to the Third Meeting of the Psychometric Society, Jouy en Josas, July 1983.
[5] B. Ganter, Two basic algorithms in concept analysis, Preprint No. 831, Technische Hochschule Darmstadt, 1984.
[6] B. Ganter, S. Kuznetsov, Formalizing hypotheses with concepts, in: Proceedings of the 8th International Conference on Conceptual Structures (ICCS'00), Lecture Notes in Artificial Intelligence, vol. 1867, 2000, pp. 342–356.
[7] B. Ganter, R. Wille, Formal Concept Analysis: Mathematical Foundations, Springer, Berlin, 1999.
[8] M. Garey, D. Johnson, Computers and Intractability: A Guide to the Theory of NP-Completeness, Freeman, San Francisco, 1979.
[9] J.-L. Guigues, V. Duquenne, Informative implications derived from a table of binary data. Preprint, Groupe Mathématiques et Psychologie, Université René Descartes, Paris, 1984.
[10] J.-L. Guigues, V. Duquenne, Familles minimales d'implications informatives résultant d'un tableau de données binaires, Math. Sci. Hum. 24 (95) (1986) 5–18.
[11] S.O. Kuznetsov, On computing the size of a lattice and related decision problems, Order 18 (4) (2001) 313–321.
[12] S.O. Kuznetsov, S.A. Obiedkov, Comparing performance of algorithms for generating concept lattices, J. Exp. Theor. Artif. Intell. 14 (2–3) (2002) 189–216.
[13] S.O. Kuznetsov, On the intractability of computing the Duquenne–Guigues base, J. Universal Comput. Sci. 10 (8) (2004) 927–933.
[14] M. Luxenburger, Implications partielles dans un contexte, Math. Inform. et Sci. Humaines 113 (29) (1991) 35–55.
[15] H. Mannila, K.-J. Räihä, Design of Relational Databases, Addison-Wesley, Reading, MA, 1992.
[16] N. Pasquier, Y. Bastide, R. Taouil, L. Lakhal, Closed set based discovery of small covers for association rules, in: Proceedings of the BDA Conference, 1999, pp. 361–381.
[17] L.G. Valiant, The complexity of enumeration and reliability problems, SIAM J. Comput. 8 (3) (1979) 410–421.
[18] M. Wild, Computations with finite closure systems and implications, Preprint No. 1708, Technische Hochschule Darmstadt, 1994.
[19] R. Wille, Restructuring lattice theory: an approach based on Hierarchies of concepts, in: I. Rival (Ed.), Ordered Sets, Reidel, Dordrecht–Boston, 1982, pp. 445–470.
[20] M.J. Zaki, M. Ogihara, Theoretical foundations of association rules, in: Proceedings of the 3rd SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery, 1998.