On the Intractability of Computing the Duquenne-Guigues Base

Sergei O. Kuznetsov
(All-Russia Institute for Scientific and Technical Information (VINITI), Moscow, Russia, serge@viniti.ru)

Abstract: Implications of a formal context (G, M, I) obey Armstrong rules, which allows for definition of a minimal (in the number of implications) implication base, called Duquenne-Guigues or stem base in the literature. A long-standing problem was that of an upper bound for the size of a stem base in the size of the relation I. In this paper we give a simple example of a relation where this boundary is exponential. We also prove $\#P$-hardness of the problem of determining the size of the stem base (i.e., the number of pseudo-intents).
Key Words: implication base, computational complexity
Category: F.2, H.2

1 Main Definitions and Problem Statement

First we recall some basic notions of Formal Concept Analysis (FCA) [Wille 1982], [Ganter and Wille 1999].

Definition. 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 given 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*:

$$A' := \{m \in M \mid gIm \text{ for all } g \in A\},$$
$$B' := \{g \in G \mid gIm \text{ for all } m \in B\}.$$

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 [Ganter and Wille 1999], 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 [Birkhoff 1979], 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:

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

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 [Guigues and Duquenne 1986]. This subset is called Duquenne-Guigues or stem base in the literature. The premises of implications of the stem base can be given by pseudo-intents (see, e.g., [Ganter and Wille 1999]): 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 long standing 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 $|G| \times |M|$.

Now we recall some standard definitions. A **many-valued context** [Ganter and Wille 1999] 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 $w = v$. Thus, instead of $(g, m, w) \in I$ one can write $g(m) = w$. By definition, $\text{dom}(m) := \{g \in G \mid (g, m, w) \in I \text{ for some } w \in W\}$. An attribute m is **complete** if $\text{dom}(m) = G$. A many-valued context is complete if all its attributes are complete. $X \to Y$ is a **functional dependency** in a complete many-valued context (G, M, W, I) if the following holds for every pair of objects $g, h \in G$:

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

In [Ganter and Wille 1999] it was shown that having a complete many-valued context (G, M, W, I), one defines the context $K_N := (\mathcal{P}_2(G), M, I_N)$, where $\mathcal{P}_2(G)$ 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 $X \subseteq M$ iff the implication $X \to Y$ holds in the context K_N.

2 Counting pseudo-intents

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. [Kuznetsov 2001]). There are several polynomial-delay algorithms for computing the set of all concepts (see e.g. review [Kuznetsov and Obiedkov 2002]). However, neither an efficient (polynomial-delay) algorithm, nor a good upper bound for the size of stem base was known. It is easy to show that there can be a stem base exponential in the size 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 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 smallest base of functional dependencies can be exponential in the size of the relation [Mannila and Räihä 1992][1]. Although the reducibility of functional dependencies to implications implies similar statement for the implication base, a general form of a context that gives rise to exponentially large stem base was not clear. The reduction of a many-valued context (G, M, W, I) to a binary one $K_N = (\mathcal{P}_2(G), M, I_N)$ along the lines of [Ganter and Wille 1999] (see Section 1) results in contexts with $(2m + 3)^2$ objects for $m \geq 2$, so the smallest number of objects in such a context is 49. Here we propose simpler contexts with sizes of the stem base exponential in the relation size.

Consider a context $K_e = (G, M, I)$ given by the cross table in Figure 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 $G_1 = \{g_1, \ldots, g_n\}$, $G_2 = \{g_{n+1}, \ldots, g_{3n}\}$, $M_1 = \{m_1, \ldots, m_n\}$, $M_2 = \{m_{n+1}, \ldots, m_{2n}\}$. 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 $i \neq j - n$, $g_i I_3 m_j$ iff $i \neq j + n$ for g_i and m_j from corresponding sets of objects and attributes. For m_0 and $g \in G$ one has $m_0 I g$ iff $g \in G_2$.

[1] I am grateful to Lhouari Nourine for attracting my attention to this fact.

$G \setminus M$	m_0	$m_1, \ldots, m_n$	$m_{n+1}, \ldots, m_{2n}$
g_1 ⋮ ⋮ ⋮ g_n		I_1	I_2
g_{n+1} ⋮ ⋮ ⋮ ⋮ g_{3n}	× ⋮ ⋮ ⋮ ⋮ ×	I_3	

Figure 1

Theorem 1. *The number of pseudo-intents of the context K_e is 2^n.*

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

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

we have

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

and $B'' = B$, i.e., the set B is closed. 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. Since the set $\{m_1, \ldots, m_n\}$ is a pseudo-intent, if we replace $m_i \in \{m_1, \ldots, m_n\}$ with m_{n+i}, then the resulting set

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

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

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

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

and every subset

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

is closed by the same arguments as for $B \subset M_1$. We can replace each m_i with m_{n+i} obtaining another pseudo-intent. Since the replacement of m_i for m_{n+i} can be done independently for each i, we have 2^n pseudo-intents. $\diamond$

Note that in our example pseudo-intents are at the same time proper premises (see, e.g., [Ganter and Wille 1999]), which make the so-called direct base: all implications are deduced from this base by single application of Armstrong rules. Moreover, here all pseudointents are so-called minimal positive hypotheses (see, e.g., [Ganter and Kuznetsov 2000]) w.r.t. the target attribute m_0.

Besides the exponential boundary of the size of the stem base, the problem of counting pseudo-intents is also intractable by the following

Theorem 2. *The problem*

INPUT A formal context $K = (G, M, I)$

OUTPUT The number of pseudointents of K

is #P-hard.

Proof. Consider an abitrary graph (V, E) and three sets $M = \{m_1, \ldots, m_{|V|}\}$, $G_1 = \{g_1, \ldots, g_{|E|}\}$, and $G_2 = \{g_{|E|+1}, \ldots, g_{|E|+|V|}\}$ such that the elements of the set M are in one-to-one correpondence with the set of vertices V (so one can write, e.g., $v(m)$), the elements of the set G_1 are in one-to-one correspondence with the edges from E (so one can write, e.g., $e(g)$), the elements of the set G_2 are in one-to-one correspondence with vertices from V (so one can write, e.g., $v(g)$).

Now consider a context $K = (G_1 \cup G_2, M \cup \{m_0\}, I)$, where I is defined as follows: for $m \in M$ and $g \in G_1$ one has mIg iff $v(m) \notin e(g)$ (i.e., the vertex $v(m)$ is not incident to the edge $e(g)$). For $m \in M$ and $g \in G_2$ one has mIg iff $v(m) \neq v(g)$. For m_0 one has $m_0 Ig$ iff $g \in G_2$.

In terms of FCA, the context K is the subposition of two contexts, which can be reprsented by the cross table in Fig. 2. Here $\not I$ is the complement of the vertex-edge incidence relation of the graph (V, E): $v \not I e$ 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 no proper subset of it is a cover. The problem of counting all minimal covers was proved to be $\#P$-complete in [Valiant 1979]. 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 $\not I$, for each $g_i \in G_1$ there is an attribute $m_i \in W$ such that $g_j \not I m_i$ does not hold. Thus, the set W' will not contain any object from

G_1. Hence, W'' will contain m_0 and, thus W is not closed ($W'' \neq W$). However, for any subset $Q \subset W$ we have $Q'' = Q$ (because Q' contains an object from G_1). Thus, by definition, W is a pseudo-intent.

In the opposite direction, for each $m_i \in M$ consider $W\colon m \notin W$. Since $m_i \notin \{g_{|E|+i}\}'$, the implication $W \to \{m_i\}$ does not hold and there is no nontrivial implications with m_i in the right-hand side. The only possible nontrivial implications are of the form $W \to \{m_0\}$. Hence, if W is a pseudo-intent of the context, then W' should not contain any object from G_1. Thus, by the definition of $\not I$, the set W is a vertex cover. This cover is minimal, since otherwise there had existed a subset $Q \subset W$ which is not closed, $Q'' = Q \cup \{m_0\}$, which contradicts the fact that W is a pseudo-intent such that $W'' = W \cup \{m_0\}$.

$G \setminus M$	m_0	$m_1, \ldots\ldots\ldots\ldots, m_{\lvert V\rvert}$
g_1		
$\vdots$		
$\vdots$		$\not I$
$\vdots$		
$g_{\lvert E\rvert}$		
$g_{\lvert E\rvert+1}$	$\times$	
$\vdots$	$\vdots$	
$\vdots$	$\vdots$	$\neq$
$\vdots$	$\vdots$	
$g_{\lvert E\rvert+\lvert V\rvert}$	$\times$	

Figure 2

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

To show that the problem of counting pseudo-intents belongs to the class $\#P$ (and, thus is $\#P$-complete), one should prove that the following decision problem

INSTANCE A context $K = (G, M, I)$, $Q \subseteq M$
QUESTION Is Q a pseudo-intent?

is solvable in polynomial time. Note that the decision problem

INSTANCE A context $K = (G, M, I)$, a natural number $k \leq |M|$.
QUESTION Is there a pseudo-intent of the context K of size not greater than k?

is proved to be NP-hard with the same reduction as in the proof of Theorem 2 from the NP-complete problem of deciding the existence of a vertex cover of size no greater than k.

At the same time the problem

INSTANCE A context $K = (G, M, I)$

QUESTION Is there a pseudo-intent of the context K?

is solvable in polynomial time: The only situation when a context (G, M, I) does not have a pseudo-intent is the case where it has a "diagonal" subcontext of the form $(A, A, \neq)$. For an arbitrary context (G, M, I) one can test whether it has a diagonal subcontext in $(|G| \cdot |M|)$ time (by scanning once all rows of the cross-table of the context K).

Acknowledgments

I thank Lhouari Nourine for drawing my attention to upper bounds of covers of functional dependencies. I also thank Bernhard Ganter, Sergei A. Obiedkov and anonymous reviewers for helpful comments. This work was supported by the "Mathematical modeling and intelligent systems (2004)" fundamental research program of the Russian Academy of Sciences.

References

[Birkhoff 1979] G. D. Birkhoff, "Lattice Theory"; Amer. Math. Soc., Providence (1979).

[Ganter and Kuznetsov 2000] B. Ganter and S. Kuznetsov, "Formalizing Hypotheses with Concepts"; Proc. 8^th Int. Conf. on Conceptual Structures, ICCS'00, Lect. Notes Artif. Intel. 1867 (2000), 342-356.

[Ganter and Wille 1999] B. Ganter and R. Wille, "Formal Concept Analysis: Mathematical Foundations"; Springer, Berlin (1999).

[Garey and Johnson 1979] M. Garey and D. Johnson, "Computers and Intractability: a Guide to the Theory of NP-Completeness"; Freeman, San Francisco (1979).

[Guigues and Duquenne 1986] J.-L. Guigues and V. Duquenne, "Familles minimales d'implications informatives résultant d'un tableau de données binaires"; Math. Sci. Hum. 24, 95 (1986), 5–18.

[Kuznetsov 2001] S.O. Kuznetsov, "On Computing the Size of a Lattice and Related Decision Problems"; Order 18, 4 (2001), 313–321.

[Kuznetsov and Obiedkov 2002] 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.

[Mannila and Räihä 1992] H. Mannila and K.-J.Räihä, "Design of Relational Databases", Addison-Wesley (1992).

[Valiant 1979] L. G. Valiant, "The Complexity of Enumeration and Reliability Problems"; SIAM J. Comput. 8, 3 (1979), 410–421.

[Wille 1982] R. Wille, "Restructuring Lattice Theory: an Approach Based on Hierarchies of Concepts"; in: I. Rival, ed., Ordered Sets, Reidel, Dordrecht–Boston (1982), 445–470.