Counting Pseudo-intents and #P-completeness

Sergei O. Kuznetsov[1] and Sergei Obiedkov[2]

[1] VINITI Institute, ul. Usievicha 20, Moscow, 125190 Russia
serge@viniti.ru
[2] University of Pretoria, Pretoria 0002, South Africa
sergei.obj@gmail.com

Abstract. Implications of a formal context (G, M, I) have a minimal implication basis, called Duquenne-Guigues basis or stem base. It is shown that the problem of deciding whether a set of attributes is a premise of the stem base is in coNP and determining the size of the stem base is polynomially Turing equivalent to a #P-complete problem.

1 Introduction

Since the introduction of the Duquenne-Guigues basis of implications [4, 5] (called also the stem base in [2]), a long standing problem was that concerning the upper bound of its size: whether the size of the basis can be exponential in the size of the input. In [6] we proposed a general form of a context where the number of implications in the basis is exponential in the size of the context. Moreover, in [6] it was shown that the problem of counting pseudo-intents, which serve premises for the implications in the basis, is a #P-hard problem.

A closely related question is that posed by Bernhard Ganter at ICFCA 2005: what is the complexity class of the problem of determining if an attribute set is a pseudo-intent? There was also a conjecture that this problem is PSPACE-complete. This paper provides a proof that this problem is just in coNP. Then, the polynomial Turing equivalence to a #P-complete counting problem is a direct consequence of this fact and the previous #P-hardness result from [6].

2 Definitions and Main Results

We assume that the reader is familiar with basic definitions and notation of formal concept analysis [2]. Recall that, given a context (G, M, I) with derivation operator $(\cdot)'$ and $B, D \subseteq M$, an *implication* $D \rightarrow B$ holds if $D' \subseteq B'$.

A minimal (in the number of implications) subset of implications from which all other implications of a context follow semantically [2] was characterized in [4, 5]. This subset is called Duquenne-Guigues basis or stem base in the literature. The premises of implications of the stem base can be given by pseudo-intents [1, 2]: a set $P \subseteq M$ is a *pseudo-intent* if $P \neq P''$ and $Q'' \subsetneq P$ for every pseudo-intent $Q \subsetneq P$.

The notions of quasi-closed and pseudo-closed sets used below have first been formulated in [4] under the name of saturated gaps (*noeuds de non-redondance*

R. Missaoui and J. Schmid (Eds.): ICFCA 2006, LNAI 3874, pp. 306–308, 2006.

in [5]) and minimal saturated gaps (*noeuds minimaux* in [5]), respectively. The terms *quasi-closed* and *pseudo-closed* have been introduced in [1]. The corresponding definitions in [5] and [1] are different but equivalent (except that saturated gaps are not closed by definition). We use notation from [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:

Proposition 1. [1] *A set $Q \subseteq M$ is quasi-closed iff $Q \cap C$ is closed for every closed set C with $Q \nsubseteq C$. Intersection of quasi-closed sets is quasi-closed.*

A set P is called *pseudo-closed* if it is quasi-closed, not closed, and for any quasi-closed set $Q \subsetneq P$ one has $Q'' \subsetneq P$. It can be shown that a set P is pseudo-closed if and only if $P \neq P''$ and $Q'' \subsetneq P$ for every pseudo-closed $Q \subsetneq P$. Hence, a pseudo-closed subset of M is a pseudo-intent and vice versa, and we use these terms interchangeably. By the above, 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 2. *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 1, 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 coincides with S. Any closed set of attributes R'' can be represented as the intersection of some object intents:

$$R'' = \bigcap_{g \in R'} \{g\}' \text{ and } S \cap R'' = \bigcap_{g \in R'} (S \cap \{g\}').$$

If $S \cap \{g\}' = S$ for all $g \in R'$, then $S \cap R'' = S$. Thus, if intersection of S with each object intent 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 1. *Testing whether $S \subseteq M$ is quasi-closed in the context (G, M, I) may be performed in $O(|G|^2 \cdot |M|)$ time.*

Proof. By Proposition 2, 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 $O(|G| \cdot |M|)$ time, testing this for all $|G|$ objects takes $O(|G|^2 \cdot |M|)$ time. □

Proposition 3. *The following problem is in NP:*

INSTANCE: A context (G, M, I) and a set $S \subseteq M$
QUESTION: Is S not a pseudo-intent of (G, M, I)?

Proof. 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, note that a nonclosed set S is pseudo-closed if and only if there is no pseudo-closed set $P \subsetneq S$ with $P'' = S''$. However, such P exists if and only if there is a quasi-closed set $Q \subsetneq S$ with the same property. Therefore, we nondeterministically obtain for S such a set Q and verify if Q is indeed a quasi-closed subset of S such that $Q'' = S''$. By the corollary of Proposition 2, this test can be done in polynomial time. □

Corollary 2. *The following problem is in coNP:*

INSTANCE: A context (G, M, I) *and a set* $S \subseteq M$
QUESTION: Is S *a pseudo-intent of* (G, M, I)*?*

Consider the problem of counting the number of all pseudo-intents. #P [7] is the class of problems of the form "compute $f(x)$", where f is the number of accepting paths of an NP machine [3]. A problem is #P-hard if any problem in #P can be reduced by Turing to it in polynomial time. A problem is #P-complete if it is in #P and is #P-hard. #P-completeness of a problem in #P, can be proved by reducing a #P-complete problem to it in polynomial time.

Since the problem of checking whether a set is nonpseudo-closed is in NP, the problem of counting such sets is in #P. Since the number of pseudo-intents is $2^{|M|} - k$ if the number of sets that are not pseudo-intents is k, the #P-hardness of the problem of counting pseudo-intents [6] implies #P-hardness of the problem of counitng the sets that are not pseudo-intents. Hence, we proved

Proposition 4. *The following problem is #P-complete:*

INSTANCE: A context (G, M, I)
QUESTION: What is the number of sets that are not pseudo-intents?

Hence, the problem of counting pseudo-intents is polynomially Turing equivalent to a #P-complete problem. It remains still open if deciding that a set is a pseudo-intent can be done in polynomial time.

References

1. B. Ganter: Two Basic Algorithms in Concept Analysis, *Preprint Nr. 831, Technische Hochschule Darmstadt* (1984).
2. B. Ganter, R. Wille: *Formal Concept Analysis: Mathematical Foundations*; Springer, Berlin (1999).
3. M. Garey, D. Johnson: *Computers and Intractability: A Guide to the Theory of NP-Completeness*; Freeman, San Francisco (1979).
4. J.-L. Guigues, V. Duquenne: Informative implications derived from a table of binary data. *Preprint, Groupe Mathématiques et Psychologie, Université René Descartes*, Paris (1984).
5. 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.
6. S.O. Kuznetsov: On the Intractability of Computing the Duquenne-Guigues Base, *Journal of Universal Computer Science*, **10**, no. 8 (2004), 927-933.
7. L. G. Valiant: The Complexity of Enumeration and Reliability Problems; *SIAM J. Comput.* **8**, 3 (1979), 410–421.