Order **18:** 313–321, 2001.

On Computing the Size of a Lattice and Related Decision Problems

SERGEI O. KUZNETSOV
Institut für Algebra, Technische Universität Dresden, Germany, and
All-Russia Institute for Scientific and Technical Information (VINITI), Moscow, Russia.
E-mail: kuznetso@math.tu-dresden.de; serge@viniti.ru

(Received: 15 November 2000; accepted: 19 June 2001)

Abstract. The problem of determining the size of a finite concept lattice is shown to be #P-complete. Since any finite lattice can be represented as a concept lattice, the problem of determining the size of a lattice given by the ordered sets of its irreducibles is also #P-complete. Some results about NP-completeness or polynomial tractability of decision problems related to concepts with bounded extent, intent, and the sum of both are given. These problems can be reformulated as decision problems about lattice elements generated by a certain amount of irreducibles.

Mathematics Subject Classification (2000): 03G10.

Key words: lattices & related structures.

1. Introduction

Concept (Galois) lattices are known to be extensively used in various methods of data analysis, data mining, and machine learning. Some well-known batch polynomial-delay and incremental algorithms for computing the set of all concepts can be found in [2, 3, 5, 11, 12], and in a recent review [10]. The set of all concepts can be exponential in the size of the input, e.g., in the case where the resulting lattice is a Boolean one. Therefore, the knowledge of the number of concepts to be obtained could be helpful for efficient resourse allocation. In this paper we consider the problem of counting all formal concepts of a formal context and some decision problems related to concepts with size constraints that play the role of quality estimates in applications.

Since any complete lattice can be represented as a concept lattice [5] (see also [1] for the final case), the results of this paper apply to arbitrary finite lattices given by ordered sets of their irreducibles.

The following definition recalls some well-known notions from Formal Concept Analysis (FCA) [5, 15].

DEFINITION 1. Let G and M be two sets called the set of objects and the set of attributes, respectively, and I be a relation defined on $G \times M$: 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 *context*. If $A \subseteq G$, $B \subseteq M$ are arbitrary subsets, then the *Galois connection* is given as follows:

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

The pair (A, B), where $A \subseteq G$, $B \subseteq M$, $A' = B$, and $B' = A$, is called a *concept* (*of the context* K) with *extent* A and *intent* B (in this case we have also $A'' = A$ and $B'' = B$).

All concepts of a formal context K induce a lattice, called the concept lattice, which is usually denoted by $\underline{\mathfrak{B}}(K)$. By $|\underline{\mathfrak{B}}(K)|$ we denote the size of the concept lattice, i.e., the number of all concepts.

2. The Number of All Concepts

The problem of computing the number of all concepts is a long standing one. The best upper bound for this number proposed in [13] is $|\underline{\mathfrak{B}}| \leq \frac{3}{2} \cdot 2^{\sqrt{|I|+1}} - 1$ for $|I| > 2$.

The following theorem explains why it is hard to compute this number or even get a good estimate of it.

THEOREM 1. *The following problem "Number of all concepts" is #P-complete.*
INPUT *Context* $K = (G, M, I)$.
OUTPUT *The number of all concepts of the context* K, *i.e.,* $|\underline{\mathfrak{B}}(K)|$.

Proof. We shall reduce the following #P-complete problem to ours: "The number of binary vectors that satisfy monotone 2-CNF of the form $C = \bigwedge_{i=1}^{s}(x_{i_1} \vee x_{i_2})$" [14]:
INPUT Monotone (without negation) CNF with two variables in each conjunction $C = \bigwedge_{i=1}^{s}(x_{i_1} \vee x_{i_2})$, $x_{i_1}, x_{i_2} \in X = \{x_1, \ldots, x_n\}$ for all $i = \overline{1, s}$.
OUTPUT Number of binary n-vectors (corresponding to the values of variables) that satisfy CNF C.

First, we construct 2-DNF D, the negation of C: $D = \bigvee_{i=1}^{s}(\bar{x}_{i_1} \wedge \bar{x}_{i_2})$. We denote $D_i = (\bar{x}_{i_1} \vee \bar{x}_{i_2})$, $i = \overline{1, s}$. The set of binary vectors that satisfy D is a union of the sets of binary vectors that satisfy some conjunction D_i. Each disjunction is satisfied by every binary $(n-2)$-vector with zero i_1th and i_2th components.

We reduce this problem to that of the number of concepts by constructing the following context $K = (G, M, I)$. The set of attributes is $M = \{m_1, \ldots, m_n\}$, where each element of M biuniquely corresponds to a variable from X. For every conjunction D_i, $i = \overline{1, s}$ we construct a context $K_i = (G_i, M, I_i)$, where the set of attributes is $M_i = M \backslash \{m_{i1}, m_{i2}\} := \{m^{i_1}, \ldots, m^{i_{n-2}}\}$, the set of objects is $G_i = \{g_i^0, g_i^1, \ldots, g_i^{n-2}\}$, and the relation $I_i \subseteq M_i \times G_i$ is defined by object intents as follows: $\{g_i^0\}' = M_i$, $\{g_i^j\}' = M_i \backslash \{m^{i_j}\}$ for $j \in \overline{1, n-2}$. Now the context K is defined as $K = (\bigcup_{i=1}^{s} G_i, M, \bigcup_{i=1}^{s} I_i)$.

Now we show that every intent of K corresponds to an n-vector that satisfies D.

Every intent of K is an intent of K_i for some i (this i can be not unique). Note that for all $i = \overline{1, s}$ the closure system of intents of the context K_i form the power set of M_i (we denote it by $\mathcal{P}(M_i)$). An arbitrary element of this set of attributes corresponds biuniquely to a binary n-vector, where each component corresponds biuniquely to an element of M with the same number. This vector satisfies D_i, since it has zeroes at i_1th and i_2th places. Therefore, this vector satisfies D.

It remains to show that each binary n-vector that satisfies D corresponds biuniquely to an intent of K. In fact, each binary n-vector v that satisfies D, satisifies D_i for some i (this i should not be unique). Then this vector has zero i_1th and i_2th positions. Therefore, the corresponding set of attributes A belongs to $\mathcal{P}(M_i)$, where $M_i = M \backslash \{m_{i_1}, m_{i_2}\}$. Since $\mathcal{P}(M_i)$ is the closure system of intents of K_i for each i, there is a set of objects $\{g_1^i, \ldots, g_r^i\} \subseteq G_i$, $r \leq n-1$ such that $\{g_1^i, \ldots, g_r^i\}' = A$.

The one-to-one correspondence between the intents of K and binary n-vectors satisfying D is established. The intents are in one-to-one correspondence with concepts. Thus, if we figured out the number of all concepts of K, we obtain the number of all vectors satisfying D and, hence, that of the vectors satisfying C. The reduction is realized. The proof of its polynomiality in the input size is obvious, since the context K has $|M| = n$ attributes and $|K| = s(n-1)$ objects. □

COROLLARY. *The problem of determining the size of a finite lattice given by the ordered sets of its join- and meet-irreducibles is #P-complete.*

This follows directly from the Basic Theorem of the Formal Concept Analysis: any complete lattice L is isomorphic to the concept lattice of the context $K = (J(L), M(L), \leq)$, where $J(L)$, $M(L)$ are the sets of join- and meet-irreducibles of L, and $\leq$ is the order relation of L, respectively [5].

3. Decision Problems with Constraints on Concept Size

The following decision problems were motivated by applications of FCA and related methods in data analysis [4, 5], since the size of extent of a concept can be considered as "support" of a concept and the size of intent says how "detailed" is "description" of the concept.

The following classical NP-complete problem, called 3-Matching (or just 3-M) [6], is used throughout this section in the proofs of NP-completeness of decision problems concerning certain types of concepts.

INSTANCE Set $M \subseteq X \times Y \times Z$, where X, Y, Z are pairwise disjoint sets, $|X| = |Y| = |Z| = p$, $|M| = N$.
QUESTION Does there exist a set $M' \subseteq M$ such that $|M'| = p$ and no two element of M' have equal components.

We represent the 3-M problem by the following binary $(p \times N)$-matrix B. It consists of three submatrices X, Y, and Z, each submatrix has p columns and N rows.

The rows of the matrix B are in one-to-one correspondence with the elements of M: for each element $(x_i, y_j, z_k) \in M$ there is exactly one row in B such that ith entry of the row in submatrix X, jth entry in submatrix Y, and kth entry in submatrix Z are zero and the other entries in the row are ones. Thus, each row of matrix B has exactly three zeros and each submatrix has exactly one zero in a row:

X	Y	Z
$1\ldots101\ldots1$	$1\ldots101\ldots1$	$1\ldots101\ldots1$
$\ldots$	$\ldots$	$\ldots$
$\ldots$	$\ldots$	$\ldots$

A 3-matching corresponds to a set of p rows with exactly one 0 in each column in X, Y, and Z.

We define another matrix that will often be used in the proofs of completeness results. Matrix U_n is an $n \times n$-matrix with zeros in the diagonal and ones in all other entries. This is a matrix corresponding to the context $K = (A, A \neq)$ for some set A: $|A| = n$.

THEOREM 2. *The following "intent of exact size" problem is NP-complete.*
INSTANCE *Context* $K = (G, M, I)$, *parameter* k.
QUESTION *Does there exist a concept* (e, i) *from* $\underline{\mathfrak{B}}(K)$ *such that* $|i| = k$.

Proof. First, the problem obviously belongs to the class NP, since the test whether a concept $(e, i) \in \underline{\mathfrak{B}}(K)$ is a solution to the problem takes $O(|M|)$ operations.

To show the NP-completeness of our problem we reduce the 3-M problem to it.

For each 3-M problem given by matrix B we construct the "intent of exact size" problem for the context $K = (G, M, I)$, where $k = (N - p)(3p + 1)$, $|G| = N$, $|M| = N(3p + 1) + 3p$, and I is represented by the following matrix E with N rows and $N(3p + 1) + 3p$ columns. Its right submatrix E_2 is isomorphic to B and its left submatrix E_1 consists of N submatrices E_1^i. Each submatrix E_1^i has $3p + 1$ columns. The ith row of matrix E_1^i is filled with zeros, other rows of matrix E_1^i are filled with ones:

$$\begin{array}{cccccc} & E_1 & & & E_2 & \\ \overbrace{0\ldots0}^{3p+1} & \ldots & \overbrace{1\ldots1}^{3p+1} & \overbrace{1\ldots101\ldots1}^{p} & \overbrace{1\ldots101\ldots1}^{p} & \overbrace{1\ldots101\ldots1}^{p} \\ 1\ldots1 & \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & 1\ldots1 & \ldots & \ldots \\ \ldots & \ldots & \ldots & 0\ldots0 & \ldots & \ldots \end{array}$$

Thus, matrix E_1 is formed from U_N by copying each column exactly $3p + 1$ times. Now we show that the 3-M problem with the above specified parameters has a solution iff the "intent of exact size" problem for $k = (N - p)(3p + 1)$ and the context represented by the matrix above has a solution.

Let the 3-M problem have a solution, then the product of the right parts (corresponding to the submatrix E_2) of some p rows of the matrix E is a $3p$-row filled with zeros. The product of the left parts of these rows (corresponding to the submatrix E_1) is the row with $(3p + 1)p$ zeros and $(N - p)(3p + 1)$ units. This means that there is an intent of the context K with the same size and the "intent of exact size" problem has a solution. On the other hand, let the "intent of exact size" problem for the context $K = (G, M, I)$ given above and $k = (N - p)(3p + 1)$ have a solution. Then, there are r rows in the matrix E such that their product has $(N - p)(3p + 1)$ ones. Then, the number of zeros is $p(3p + 1) + 3p$ of which $p(3p + 1)$ zeros belong to the left side (submatrix E_1) and $3p$ zeros belong to the right side (submatrix E_2), since the right side is of width $3p$. The number r of rows cannot be less than p, since otherwise the number of ones in the left side of the product would not have exceeded $(p-1)(3p+1)$. The number of ones in the right side of the product does not exceed $3p$ by all means, so the total number of ones does not exceed $(p - 1)(3p + 1) + 3p < p(3p + 1) + 3p$.

At the same time, r cannot be greater than p, since otherwise the left side of the product would have had no less than $(p + 1)(3p + 1) > p(3p + 1) + 3p$ zeros, which contradicts our assumption. Therefore, $r = p$ and the problem 3-M has a solution.

The reduction is accomplished. Its polynomiality follows directly from the fact that the matrix E is polynomial with respect to p and N. □

COROLLARY. *The following "extent of exact size" problem is NP-complete:*

INSTANCE *Context* $K = (G, M, I)$, *parameter* k.

QUESTION *Does there exist a concept* (e, i) *from* $\underline{\mathfrak{B}}(K)$ *such that* $|e| = k$.

The proof follows directly from the duality of objects and attributes.

Note that the following "minimal (maximal) intent" problems are obviously solved in polynomial time:

INSTANCE Context $K = (G, M, I)$, parameter k.

QUESTION Does there exist a concept (e, i) from $\underline{\mathfrak{B}}(K)$ such that $|i| \leq k$ $(|i| \geq k)$.

Indeed, to get the minimal intents we look through all attribute concepts, i.e., concepts of the form (m', m'') for all $m \in M$.

If there are no intents with the desired property $|i| \leq k$, then other intents do not satisfy it a fortiori. The test takes $O(|M|^2|G|)$ operations. To get the maximal intents, we look through all object concepts, i.e., concepts of the form $(g'', g') \in M$. The test takes $O(|M||G|^2)$ operations.

Due to the duality between objects and attributes minimal intents correspond to maximal extents and maximal intents correspond to minimal extents. Thus, the following "minimal (maximal) extent" problem is also trivially polynomial:

INSTANCE Context $K = (G, M, I)$, parameter k
QUESTION Does there exist a concept (e, i) from $\mathfrak{B}(K)$ such that $|e| \leq k$ ($|e| \geq k$).

The situation is quite different when we consider problems related to the total size of concept, i.e., to $|e| + |i|$. If a context is represented by the binary matrix and (e, i) is a concept of this context, then the value $2(|e| + |i|)$ is the perimeter of the, maximal by inclusion, rectangular filled with ones that corresponds to the concept (e, i). We show that the size of a maximal concept is determined in polynomial time, whereas the problem of finding a minimal concept is intractable.

THEOREM 3. *The following "maximal concept" problem is solved in polynomial time.*
INSTANCE Context $K = (G, M, I)$, parameter k.
QUESTION Does there exist a concept (e, i) from $\mathfrak{B}(K)$ such that $|e| + |i| \geq k$.

Proof. Consider the bipartite graph B that corresponds to the context $K = (G, M, I)$. Each concept (e, i) from $\mathfrak{B}(K)$ corresponds to a maximal by inclusion complete bipartite subgraph of B and $|e| + |i|$ is the number of vertices of this subgraph. Now consider the context $K = (G, M, \bar{I})$ with relation $\bar{I} = G \times M \backslash I$, the complement of I, and the corresponding bipartite graph $\overline{B}$. A concept (e, i) from K corresponds to a maximal by inclusion independent set of vertices of $\overline{B}$, i.e., a set where where no pair of vertices is connected by an edge. A concept (e, i) of $\mathfrak{B}(K)$ with the largest $|e| + |i|$ corresponds to the largest (in the number of vertices) independent set of $\overline{B}$. According to the König theorem (see, e.g., [8]), the number of vertices in the largest independent set is $\overline{B} - |M|$, where $|\overline{B}|$ is the number of vertices in $\overline{B}$ and M is the number of edges in the maximal matching of $\overline{B}$. The size of a maximal matching can be found by polynomial-time algorithm, for example, by the algorithm from [9]. □

THEOREM 4. *The following "minimal concept" problem is NP-complete.*
INSTANCE Context $K = (G, M, I)$, parameter k.
QUESTION Does there exist a concept (e, i) from $\mathfrak{B}(K)$ such that $|e| + |i| \leq k$.

Proof. We prove the theorem by reducing the 3M problem to ours, however this time we need an intermediary problem, which is much similar to 3M. Let each column of the matrix B, which represents an individual 3M problem, be copied $2p^3 + p$ times and each row of the obtained matrix is copied twice. Thus, we obtained a binary matrix with $2N$ rows and $3p(2p^3 + p)$ columns. We call this matrix B_1.

It is obvious that a 3M problem has a solution iff there is a set of exactly p pairs of rows in the corresponding matrix B_1 such that rows in pairs are identical and for every column of B_1 there are exactly two rows from this set with zeros in this column. The binary product of all these $2p$ rows is a zero $3p(2p^3 + p)$-vector.

Consider the matrix B_2 formed from matrix U_N by copying twice each row. Thus, B_2 has $2N$ rows and N columns and for every $j = \overline{1, N}$ the rows $2j - 1$ and

$2j$ have jth zero component (other entries are ones). Now we adjoin matrix B_2 to matrix B_1 from the left. We denote the resulting matrix by B_3. Consider the context $K = (G, M, I)$ with $2N$ objects, $N + 3p(2p^3 + p)$ attributes, and relation I given by matrix B_3. Let $k = N + p$. We show that the initial 3M problem has a solution iff there is a concept (e, i) from $\underline{\mathfrak{B}}(K)$ such that $|e| + |i| \leq k$.

Suppose that there is a matching in the 3M problem given by matrix B, then there are p pairs of rows in B_3 such that for their binary product its part in B_1 is zero vector and its part in B_2 is a vector with p zeros (since the vectors in pairs are identical) and, hence, $N - p$ ones. These $2p$ pairs of rows make an intent of a concept corresponding to the matrix B_3, since any other row in matrix B_3 multiplied with those from the indicated p pairs will give another result. Thus, we have a concept (e, i) from $\underline{\mathfrak{B}}(K)$ such that $|e| = 2p$, $|i| = N - p$, and $|e| + |i| = 2p + N - p = N + p$.

Conversely, suppose that there is a concept (e, i) from $\underline{\mathfrak{B}}(K)$ such that $|e| + |i| \leq k = N + p$. Consider the rows of matrix B_3 that correspond to the extent of this concept. The part of these rows that correspond to matrix B_1 should be a zero vector, since otherwise (note that any nonzero product of rows from matrix B_1 contains at least $2p^3 + p$ ones and $N \leq p^3$) $|e| \geq 1$, $|i| \geq 2p^3 + p$, and $|e| + |i| \geq 2p^3 + p + 1 > N + p$, which violates the assumption. Now let us figure out the number of rows corresponding to the concept. Consider the part of the rows that correspond to matrix B_2. The number of rows cannot be less than $2p$, since otherwise there would be a matching for matrix B of size less than p. Suppose that this number is $r > 2p$, then $|e| + |i| = 2r + (N - r) = N + r > N + p$, which contradicts the assumption. Therefore, the number of rows is $2p$ and this gives us a matching for the 3M problem with matrix B.

The reduction is realized. Its polynomiality follows from the polynomiality of the number of rows ($2N$) and columns ($N + 3p(2p^3 + p)$) of matrix B_3. □

COROLLARY. *The following "exact size concept" is NP-complete.*

INSTANCE *Context* $K = (G, M, I)$, *parameter* k.

QUESTION *Does there exist a concept* (e, i) *from* $\underline{\underline{\mathfrak{B}}}(K)$ *such that* $|e| + |i| = k$.

Proof. Having an individual "minimal size" problem for $k = k_1$, we try to solve the "exact size concept" problem for $k = \overline{1, k_1}$. If for any k from this interval a solution exists, then the "minimal size" problem is solved, otherwise no solution exists. The reduction is realized and its polynomiality is obvious. □

4. Conclusion

We proved that the problem of determining the size of a finite lattice given by its irreducibles is #P-complete. It remains unknown whether this result holds for interesting classes, such as distributive or modular lattices. Note that the reduction that we used in Section 1 brings about a concept lattice, which is not necessarily distributive. Distributivity can be recognized in polynomial time if a lattice is given as an ordered set of irreducibles (e.g., by means of “arrow relations,” see [5]). The tree structure of distributive lattices helps to efficiently solve certain combinatorial problems [7]. So, we can hope that determining the size of a finite distributive lattice is polynomially tractable.

The results for the decision problems with constraints on the concept sizes can be briefly represented by the following table:

	$\leq$	$=$	$\geq$
$\|i\|$	P	NP	P
$\|e\|$	P	NP	P
$\|e\| + \|i\|$	NP	NP	P

Here P denotes that there exists a polynomial algorithm, and NP denotes NP-completeness of the problem. For instance, the upper left element of the table means that the problem “does there exist a concept such that $|i| \leq k$?” can be solved by a polynomial algorithm. It is obvious that all the mentioned counting and decision problems are solved in polynomial time when the number of ones in a column or/and in a row of the binary matrix representing the context is bound by a constant. As in the case of the counting problem it remains unknown whether these decision problems are tractable if the concept lattice is distributive.

5. Acknowledgements

This work was supported by the Alexander von Humboldt Foundation. The author thanks Bernhard Ganter for helpful comments.

References

1. Barbut, M. and Monjardet, B. (1970) *Ordre et Classification, Algèbre et Combinatoire*, 2 tomes, Hachette, Paris.
2. Bordat, J. P. (1986) Calcul pratique du treillis de Galois d’une correspondance, *Math. Sci. Hum.* **96**, 31–47.
3. Ganter, B. (1984) Two basic algorithms in concept analysis, Preprint 831, Technische Hochschule Darmstadt.
4. Ganter, B. and Kuznetsov, S. (2000) Formalizing hypotheses with concepts, in G. Mineau and B. Ganter (eds), *Proc. of the 8th Internat. Conf. on Conceptual Structures, ICCS’2000*, Lecture Notes in Artificial Intelligence 1867, pp. 342–356.
5. Ganter, B. and Wille, R. (1999) *Formal Concept Analysis: Mathematical Foundations*, Springer, Berlin.
6. Garey, M. and Johnson, D. (1979) *Computers and Intractability: A Guide to the Theory of NP-Completeness*, Freeman, New York.
7. Habib, M. and Nourine, L. (1996) Tree structure for distributive lattices and its applications, *Theoret. Comput. Sci.* **165**(2), 391–405.
8. Harary, F. (1969) *Graph Theory,* Addision-Wesley, Reading, MA.
9. Karp, R. and Hopcroft, G. A. (1973) 2.5-algorithm for maximum matching in bipartite graphs, *SIAM J. Comput.* **2**(2), 225–231.

10. Kuznetsov, S. O. and Objedkov, S. A. (2001) Comparing performance of algorithms for generating concept lattices, in E. Mephu Nguifo et al. (eds), *Proc. of the ICCS-2001 International Workshop on Concept Lattices-Based Theory, Methods and Tools for Knowledge Discovery in Databases* (CLKDD '01), Stanford University, Palo Alto, pp. 35–47.
11. Norris, E. M. (1978) An algorithm for computing the maximal rectangles in a binary relation, *Rev. Roumaine Math. Pures Appl.* **23**(2), 243–250.
12. Nourine, L. and Raynaud, O. (1999) A fast algorithm for building lattices, *Inform. Process. Lett.* **71**, 199–204.
13. Schütt, D. (1988) Abschätzungen für die Anzahl der Begriffe von Kontexten, Diplomarbeit TH Darmstadt, Darmstadt.
14. Valiant, L. G. (1979) The complexity of enumeration and reliability problems, *SIAM J. Comput.* **8**(3), 410–421.
15. Wille, R. (1982) Restructuring lattice theory: An approach based on hierarchies of concepts, in I. Rival (ed.), *Ordered Sets*, Reidel, Dordrecht/Boston, pp. 445–470.