Available online at www.sciencedirect.com

SCIENCE DIRECT

ELSEVIER

Discrete Applied Mathematics 142 (2004) 111–125

DISCRETE APPLIED MATHEMATICS

www.elsevier.com/locate/dam

Complexity of learning in concept lattices from positive and negative examples

Sergei O. Kuznetsov[a,b]

[a] *All-Russia Institute for Scientific and Technical Information (VINITI), Moscow 125219, Russia*
[b] *Institut für Algebra, Technische Universität Dresden, Germany*

Received 20 November 2000; received in revised form 19 November 2003; accepted 24 November 2003

Abstract

A model of learning from positive and negative examples in concept lattices is considered. Lattice- and graph-theoretic interpretations of learning concept-based classification rules (called hypotheses) and classification in this model are given. The problems of counting all formal concepts, all hypotheses, and all minimal hypotheses are shown to be #P-complete. NP-completeness of some decision problems related to learning and classification in this setting is demonstrated and several conditions of tractability of these problems are considered. Some useful particular cases where these problems can be solved in polynomial time are indicated.

Keywords: Concept lattices; Learning; Algorithmic complexity

1. Introduction

Many problems of data analysis are naturally formulated in terms of formal concept analysis (FCA) [7], e.g., the notion of an implication between sets of attributes was considered already in the first paper on FCA [25]. In this paper, we consider a model of learning from positive and negative examples from [4,5], which was recently described in terms of FCA [8,17]. Here, like in classical machine learning models [18], given a target property (attribute) and descriptions of some positive and negative examples, generalizations of positive examples that do not cover any negative examples are formed. These generalizations, called hypotheses, [1] can be used further for predicting the target attribute, i.e., classification.

To construct all possible hypotheses we can adapt an algorithm for constructing concepts, e.g., from [19,6,3,20]. Some algorithms for computing hypotheses and minimal hypotheses can be found in [14,8]. Thus, we do not present detailed descriptions of any algorithm, but rather discuss tractability of some problems of hypotheses generation and classification.

The paper is organized as follows. In Section 2, we recall the main definitions of FCA, give definitions of hypotheses and classifications, and consider an example. In Sections 3 and 4, we consider lattice- and graph-theoretic interpretations of hypotheses and classifications.

In Section 5, the algorithmic complexity of generating hypotheses and classifications is analyzed. We show that the problems of determining the number of all concepts and the number of all minimal hypotheses are #P-complete. We prove NP-completeness of some related decision problems for concepts and hypotheses with constraints on their sizes. The complexity of classification is discussed and its intractability in general case is shown. Some particular cases where the above problems can be solved in polynomial time are considered.

E-mail address: serge@viniti.ru (S.O. Kuznetsov).

[1] In [4] they were called JSM-hypotheses in honor of the English philosopher John Stuart Mill, who was one of the first to formalize inductive reasoning schemes.

0166-218X/$ - see front matter
doi:10.1016/j.dam.2003.11.002

2. Main definitions: concepts, hypotheses, and classifications

First, we recall some basic notions of FCA [25,7].

Definition 1. Let G and M be sets, called the set of objects and 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*. 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 (in this case we have also $A'' = A$ and $B'' = B$). The set of attributes B *is implied by the set of attributes* D, or *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 ($X \subseteq Y \Rightarrow X'' \subseteq Y''$). The set of all formal concepts of the context K, as any family of closed sets [1], forms a lattice, called a *concept lattice* and usually denoted by $\mathfrak{B}(K)$ in FCA literature. The meet and join of this lattice are given [7] as

$$\bigwedge_{j \in J} (A_j, B_j) = \left(\bigcap_{j \in J} A_j, \left(\bigcup_{j \in J} B_j \right)'' \right),$$

$$\bigvee_{j \in J} (A_j, B_j) = \left(\left(\bigcup_{j \in J} A_j \right)'', \bigcap_{j \in J} B_j \right).$$

Therefore, the order in the concept lattice is given as follows: if (A, B) and (C, D) are two concepts, $(A, B) \leqslant (C, D)$ iff $A \subseteq C$ (or, equivalently, $B \supseteq D$). Since an ordered set is naturally described by its (*Hasse*) *diagram*, called *line diagram* in FCA, it is natural to say that (A, B) lies *below* (C, D) in the diagram or (C, D) lies *above* (A, B) in the diagram, respectively.

Now, we present a learning model from [5] in terms of FCA. This model complies with the common paradigm of learning from positive and negative examples (see, e.g., [17]): given positive and negative examples of a target attribute, construct a generalization of the positive examples that would not cover any negative example.

Assume that there is a target property formally represented by attribute w different from attributes from the set M. For example, in some pharmacological applications the attributes from M may correspond to particular subgraphs of molecular graphs of chemical compounds and the target property corresponds to a biological activity of these compounds.

In most settings, the target attribute may have three values: positive, negative, and undetermined. Input data for learning can be represented by sets of positive, negative, and undetermined examples. *Positive examples* (or (+)-examples) are objects that are known to have the target property and *negative examples* (or (−)-examples) are objects that are known not to have this property. *Undetermined examples* (or (τ)-examples) are those that are neither known to have the property nor known not to have the property. The results of learning are supposed to be rules used for the classification of undetermined examples (or forecast of the target property for these examples).

In terms of FCA, this situation can be described by three contexts: a *positive context* $K_+ = (G_+, M, I_+)$, a *negative context* $K_- = (G_-, M, I_-)$, and an *undetermined context* $K_\tau = (G_\tau, M, I_\tau)$. Here G_+, G_-, and G_τ are sets of positive, negative, and undetermined examples, respectively; M is a set of attributes not containing the target attribute; $I_\varepsilon \subseteq G_\varepsilon \times M$, $\varepsilon \in \{+, -, \tau\}$ are relations that specify the structural attributes of positive, negative, and undetermined examples, respectively. The derivation operators in these three contexts are denoted by superscripts $+, -, \tau$, respectively, e.g., one can write A^+, A^{++}, A^-, $A^{\tau\tau}$, etc., where $A^{\varepsilon\varepsilon}$ denotes the double application of the derivation operator $(\cdot)^\varepsilon$.

Now, a positive hypothesis by Finn [4,5] (called "counterexample forbidding hypothesis" there) can be defined in the following way.

Definition 2. Consider a positive context $K_+ = (G_+, M, I_+)$, a negative context $K_- = (G_-, M, I_-)$, and an undetermined context $K_\tau = (G_\tau, M, I_\tau)$. The context $K_\pm = (G_+ \cup G_-, M \cup \{w\}, I_+ \cup I_- \cup G_+ \times \{w\})$ is called a *learning context*. The context $K_c = (G_+ \cup G_- \cup G_\tau, M \cup \{w\}, I_+ \cup I_- \cup I_\tau \cup G_+ \times \{w\})$ is called a *classification context*. The derivation operators in these two contexts are denoted by superscripts $\pm$ and c, respectively. If a pair (A_+, B_+) is a concept of the context K_+, then it is called a *positive* (or (+)-) *concept* and the sets A_+ and B_+ are called *positive* (or (+)-) *intent* and *extent*, respectively. If the intent B_+ of a positive concept (A_+, B_+) is not contained in the intent of any negative example (i.e., $\forall g_- \in G_-$, $B_+ \not\subseteq \{g_-\}^-$), then it is called a *positive* (or (+)−) *hypothesis with respect to the property* w.[2] A (+)-intent B_+ is called *falsified* if $B_+ \subseteq \{g_-\}^-$ for some negative example g_-. *Negative* (or (−)-) intents and hypotheses are defined similarly. The derivation operators in these contexts are denoted by superscripts $\pm, c$, respectively, e.g., one can write $A^{\pm}$, A^c, $A^{\pm\pm}$, A^{cc}, etc.

Obviously, if H_+ is a positive hypothesis with respect to the property w, then $H_+ \to \{w\}$ is an implication for the context $K_\pm$.

Hypotheses can be used for the classification of undetermined examples from G_τ (i.e., for forecasting whether they have the property w or not). If an undetermined example $g_\tau \in G_\tau$ contains a positive hypothesis H_+ (i.e., $\{g_\tau\}^\tau \supseteq H_+$), we say that H_+ is *for the positive classification of* g_τ. *A hypothesis for the negative classification of* g_τ is defined similarly. If there is a hypothesis for the positive classification of g_τ and no hypothesis for negative classification of g_τ, then g_τ is *classified positively*.[3] *Negative classifications* are defined dually. If $\{g_\tau\}^\tau$ does not contain any negative or positive hypothesis, then no classification is made. If $\{g_\tau\}^\tau$ contains both positive and negative hypotheses, then the classification is said to be contradictory.

We can distinguish a useful subset of hypotheses that is equivalent to the set of all hypotheses w.r.t. possible classifications. Formally, a positive hypothesis H_+ is a *minimal positive hypothesis* if no $H \subset H_+$ is a positive hypothesis. The set of all minimal hypotheses plays here a role similar to that of the basis of implications in FCA [10,7].

Example 1. In [8], we considered a context with winter wheel chains, which was adapted from the ADAC Magazine (1999, no. 11). Here, we analyze a somewhat simplified version of this context, with a subset of the set of initial examples (seven examples—four positive and three negative—instead of 17) and simpler scaling of attributes. The target attribute here is the high price of a chain (thus, positive examples correspond to expensive chains and negative examples correspond to cheap chains). The values of the attribute **system** give the type of a chain system: R—rope chain, S—steel ring chain, Q—quick mounting chain. The **mount** attribute takes the values F and B to denote that a chain of a particular type can be mounted either only on the front wheels or both on the front and rear wheels. Actually, to conform to the definitions above, we should have introduced the attributes **system R**, **system S**, **system Q** instead of **system**, but for the brevity sake we use attribute **system** as a shortcut, keeping in mind that this attribute is nominal, i.e., takes either the value **system R** or the value **system S** or the value **system Q**. The same for **mount**, which actually should be replaced by the attributes **mount B**, **mount F**. Various techniques of reducing many-valued attributes to binary ones, called *scalings*, are found in [7].

The original values of other attributes were numerical. The attribute **con** corresponds to the average expert assessment of the conveniency of a particular type of chain; the attributes **snow** and **ice** correspond to the average expert assessments of the maneuverability of a car, with a particular kind of chain, on snow and ice, respectively; the attribute **dur** corresponds to average expert assessments of the durability of a particular kind of chain; the attribute **grade** corresponds to average expert assessment of the general quality of a particular chain type. In the original setting, smaller values of attributes **con**, **snow**, **ice**, **dur**, and **grade** correspond to better assessments of the respective chain properties. In [8], we used some scales to turn numerical values into Boolean. Here we use a simpler scaling: these attributes take the Boolean value **true** (denoted by cross in the respective table entry) if the original values are less than some fixed thresholds (for details see [8]). The corresponding positive and negative contexts can be represented as follows:

Positive context:

chain	**system**	**mount**	**con**	**snow**	**ice**	**dur**	**grade**
2	S	B	×	×		×	×
5	Q	B	×		×	×	×
8	Q	B	×			×	×
14	Q	B	×			×	

[2] In [4,5], it is required that $|A_+| \geqslant 2$, however we omit this requirement here for the sake of uniformity.

[3] In [4,5] g_τ is called a "(+)-hypothesis of the second kind."

Negative context:

chain	system	mount	con	snow	ice	dur	grade
1	R	F	×	×	×	×	×
3	R	F	×	×	×		×
17	R	F				×	

Here, the minimal positive hypothesis is $\{\mathbf{B}, \mathbf{con}, \mathbf{dur}\}$. It is unique, since the intersection of all positive example intents is nonempty and is not contained in the intent of any negative example.

Other positive hypotheses are

$$\{\mathbf{Q}, \mathbf{B}, \mathbf{con}, \mathbf{dur}\}, \quad \{\mathbf{B}, \mathbf{con}, \mathbf{dur}, \mathbf{grade}\}.$$

These hypotheses can be useful if one is interested in the taxonomy of the class defined by the target attribute (i.e., the class of expensive chains in our case). For example, the first nonminimal hypothesis describes a class of chains of the same type, same mounting possibilities, convenient, but with possible bad behavior on snow.

The minimal negative hypothesis is $\{\mathbf{R}, \mathbf{F}\}$. It is also unique, since the intersection of all negative example intents is nonempty and is not contained in the intent of any positive example.

Another negative hypothesis is

$$\{\mathbf{R}, \mathbf{F}, \mathbf{con}, \mathbf{snow}, \mathbf{ice}, \mathbf{grade}\}.$$

Note that there are subsets of the minimal positive hypothesis (such as $\{\mathbf{B}, \mathbf{con}\}$, $\{\mathbf{B}, \mathbf{dur}\}$, and $\{\mathbf{B}\}$) that give smaller sufficient conditions for the occurrence of the target attribute, but the minimal hypothesis gives more detailed description of what is a "cheap chain" and what is an "expensive chain." Generally, hypotheses take account of all attributes that appear together with the target attribute to be meaningful and related to what can be considered as a "concept of the target attribute." Further discussion of the relation between minimal hypotheses and more general classifiers (such as implications based on pseudointents and minimal premises) can be found in [8].

3. Lattice-theoretic interpretation of hypotheses and classifications

In this section, we discuss hypotheses and classifications based on them from the lattice-theoretical viewpoint, using the representation of a concept lattice by its line (Hasse) diagram.

Line diagrams of concept lattices provide a data analyst with a useful visualization of data structure. Besides the hierarchy of concepts, which shows the trade-off between sizes of intents and extents, implication between sets of attributes also has a simple visual image [7]: if $B \to D$ is an implication of a context $K = (G, M, I)$ for $B, D \subseteq M$, then the meet (i.e., the result of applying the operation $\wedge$ from Definition 1) of all attribute concepts of the form $(\{b\}', \{b\}'')$, where $b \in B$, coincides with or lies below the meet of all attribute concepts of the form $(\{d\}', \{d\}'')$, where $d \in D$. In this section, we provide a means for similar visualization of hypotheses and classifications.

First, we consider positive hypotheses in terms of the lattice $\mathfrak{B}(K_+)$ of positive concepts. Each negative example cuts off some order filters from the lattice of positive concepts, these filters consisting of falsified $(+)$-intents. Thus, the set of all positive hypotheses, being the complementation of the set of falsified $(+)$-intents w.r.t. the set of all $(+)$-intents, is a set closed w.r.t. the meet operation of the lattice of positive concepts. Of course, the dual statement holds for $(-)$-hypotheses in the lattice $\mathfrak{B}(K_-)$.

The situation looks differently in the concept lattice $\mathfrak{B}(K_\pm)$. We can distinguish three types of concepts of this lattice: first, there are concepts of the form $(A, B \cup \{w\})$, where B is an intent of K_+, there are concepts of the form (A, B), where B is an intent of K_-, and there are concepts whose intents are intents of neither positive K_+ nor negative K_- contexts, i.e., concepts (A, B) such that $A = E_+ \cup E_-$, $E_+ \subseteq G_+$, $E_- \subseteq G_-$, $B \neq E_+^+$, $B \neq E_-^-$.

Proposition 1. *A $(+)$-hypothesis corresponds to a concept of $K_\pm$ of the form $(A, B \cup \{w\})$ such that there is no concept of $K_\pm$ with intent B.*

In fact, if such a concept exists, then there are examples that have all attributes from B, but do not have the attribute w, which means that the positive intent B is contained in intent of at least one negative example. A $(-)$-hypothesis corresponds to a concept of $K_\pm$ of the form (A, B), $w \notin B$ such that $(B \cup \{w\})^\pm = \emptyset$. This is equivalent to the fact that there is no concept of $K_\pm$ with nonempty extent and intent greater than or equal to $B \cup \{w\}$, i.e., lying below than or coinciding with $((B \cup \{w\})^{\pm}, (B \cup \{w\})^{\pm\pm})$.

In the concept lattice of $\underline{\mathfrak{B}}(K_\pm)$ the concepts corresponding to (+)- and (−)-hypotheses lie below concepts that are not hypotheses. In terms of this lattice the problem of classifying an undetermined example $g_\tau \in G_\tau$ looks as follows. Consider the order filter of $\underline{\mathfrak{B}}(K_\pm)$ given by the largest subsets of $\{g_\tau\}^\tau \cup \{w\}$ that are intents of $K_\pm$. If there is a concept $(A, B \cup \{w\})$ lying in this order filter such that $w \notin B$ and $(B^{\pm}, B)$ is not a concept of $\underline{\mathfrak{B}}(K_\pm)$, then B is a hypothesis for the positive classification of g_τ. The absence of a hypothesis for the negative classification of g_τ means that for any concept (A,B) of $\underline{\mathfrak{B}}(K_\pm)$ such that $w \notin B$ and (A,B) lies in the order filter of $\underline{\mathfrak{B}}(K_\pm)$ given by largest subsets of $\{g_\tau\}^\tau \cup \{w\}$ that are intents of $K_\pm$, one has $(B \cup \{w\})' \neq \emptyset$ or there is a concept with nonempty extent lying below or coinciding with $((B \cup \{w\})', (B \cup \{w\})'')$.

Things look different in the lattice of the classification context K_c.

Proposition 2. *Given a context K_c and an undetermined example g_τ, there is a positive hypothesis for the classification of g_τ iff the following conditions hold for some $A \subseteq G_\tau$, B, A_τ: $A, B \subseteq M$, $g_\tau \in A_\tau \subseteq G_\tau$:*

1. $(A, B \cup \{w\}) \in \underline{\mathfrak{B}}(K_c)$,
2. $(A \cup A_\tau, B) \in \underline{\mathfrak{B}}(K_c)$.

Proof. For an undetermined example g_τ to be classified positively, there should be a positive hypothesis B for this classification. This means, first, that there exists a set of positive examples (we denote it by A) with the intersection of intents in K_c equal to $B \cup \{w\}$. Since intents of undetermined and negative examples do not contain w, the extent of the concept of K_c with the intent $B \cup \{w\}$ is equal to A and the first statement is proved.

Second, B should be contained in $\{g_\tau\}^\tau$. Since g_τ does not have w, this means that B is an intent of K_c and there is some set A_τ: $g_\tau \in A_\tau \subseteq G_\tau$ such that the extent corresponding to the intent B should contain positive examples from A and undetermined examples from A_τ. It should not contain any negative examples, therefore, this extent is just the union $A \cup A_\tau$.

Conversely, if conditions 1 and 2 hold for some A_τ: $g_\tau \in A_\tau \subseteq G_\tau$, then B is a hypothesis for the positive classification of g_τ. □

Furthermore, there is no negative hypothesis for the negative classification of $g_\tau \in G_\tau$ iff for any concept $(C,D) \in \underline{\mathfrak{B}}(K_c)$ such that $w \notin D$ (the concept (C,D) does not lie in the order ideal of $\underline{\mathfrak{B}}(K_c)$ given by w) and $D \subseteq \{g_\tau\}^c$ (the concept (C,D) lies in the order filter of $\underline{\mathfrak{B}}(K_c)$ given by g_τ) either $C \subseteq G_\tau$ or $C \cap G_+ \neq \emptyset$ holds (C lies in an order filter of $\underline{\mathfrak{B}}(K_c)$ given by at least one positive example).

For contradictory classification this condition is violated, whenever conditions from Proposition 2 always hold. For an undetermined classification both this condition and that of Proposition 2 are violated.

Thus, we obtained complete lattice characterizations of hypotheses and classifications.

4. Graph-theoretic interpretation of hypotheses

First, we introduce some auxiliary constructions. Recall that a *vertex covering* of a graph $\Gamma = (V,E)$ is a subset of vertices $V_1 \subseteq V$ such that for every edge $(v_1, v_2) \in E$ we have $v_1 \in V_1$ and/or $v_2 \in V_1$.

Definition 3. For an arbitrary graph $\Gamma = (V,E)$ the *associated tripartite graph* is a graph T of the following form: $T = \langle W_1 \cup W_2 \cup W_3, E_1 \rangle$, $|W_1| = |W_2| = |V|$, $|W_3| = |E|$, $E_1 \subseteq (W_1 \times W_2) \cup (W_2 \times W_3)$. Pairs of vertices of the form (w_i^1, w_i^2), $w_i^1 \in W_1$, $w_i^2 \in W_2$ are in one-to-one correspondence to vertices of the form $v_i \in V$; $(w_i^1, w_j^2) \in E_1$ iff $i \neq j$. Vertices of the form $w_l^3 \in W_3$ are in one-to-one correspondence to edges of the form $e_l \in E$; $(w_j^2, w_l^3) \in E_1$ if the vertex $v_j \in V$ is incident to the edge $e_l \in E$.

We say that in a bipartite graph $B = (X \cup Y, Z)$ a set of nodes $X_1 \subseteq X$ *dominates* vertices from $Y_1 \subseteq Y$ if each vertex from Y_1 is adjacent to a vertex from X_1. The *common shadow* of a vertex set $X_1 \subseteq X$ is defined as the set $Y_2 \subseteq Y$ of all vertices adjacent to each vertex from the set X_1.

We also need the notion of a *biclique*, which is defined differently in the literature. Here we consider a biclique $D = (X_1 \cup Y_1, Z)$ in a bipartite graph B to be a complete bipartite subgraph of a graph (i.e., $Z = X_1 \times Y_1$) maximal by inclusion, i.e., for any X_2:$X_1 \subset X_2$ or any Y_2:$Y_1 \subset Y_2$ neither the bipartite subgraph of B induced by vertices (X_2, Y_1) nor the bipartite subgraph of B induced by vertices (X_1, Y_2) is a complete bipartite subgraph.

Lemma 3. *Let $\Gamma = (V, E)$ be a graph, and T be the associated tripartite graph. Γ has a vertex covering of size k if and only if in the tripartite graph T there is a triple (C, Z, W_3), where $C \subseteq W_1$, $Z \subseteq W_2$, Z is the common shadow of the vertex set C in the bipartite graph induced by vertices $W_1 \cup W_2$, and Z dominates vertices from W_3. Moreover, in this case $|C| = |W_1| - k = |V| - k$, and $|Z| = k$.*

Proof. The proof follows directly from the construction of the graph T. Indeed, the set of vertices $Z \subseteq W_2$ of size k dominates all vertices from W_3 if and only if it corresponds to a subset of vertices in graph Γ that makes a vertex covering of size k. In this case for the bipartite subgraph of T induced by vertices $W_1 \cup W_2$ the set of vertices Z is the common shadow of a set of nodes $C \subseteq W_1$ that corresponds to a subset of vertices of graph Γ complementary to the set of vertices that corresponds to the set Z. Therefore, $|C| = |W_1| - k = |V| - k$. □

Definition 4. Learning context corresponding to a tripartite graph $T{:}\langle W_1 \cup W_2 \cup W_3, E_1\rangle$, $E_1 \subseteq (W_1 \times W_2) \cup (W_2 \times W_3)$ is the context $K_\pm(T) = (G_+ \cup G_-, M \cup \{w\}, I_+ \cup I_- \cup G_+ \times \{w\})$, where $M = W_2$, positive examples of the form $g_i \in G_+$ are in one-to-one correspondence with the vertices of the form $w_i \in W_1$ and negative examples of the form $g_l \in G_-$ are in one-to-one correspondence with vertices of the form $w_l^3 \in W_3$. We define the relations I_+ and I_- by object intents. For a positive example $g_i \in G_+ \{g_i\}^+ := \{w^2 | w^2 \in W_2, (w_i^1, w^2) \in E_1\}$. For a negative example $g_l \in G_- \{g_l\}^- = \{w^2 | w^2 \in W_2, (w^2, w_l^3) \notin E_1$ for all $w_l^3 \in W_3\}$ (here superscripts $+$ and $-$ denote, as usual, derivation operators in the corresponding contexts).

Lemma 4. *Let T be a tripartite graph given by Definition 4 and $K_\pm(T)$ be the corresponding learning context. Then the following two statements are equivalent:*

1. *There is a triple (C, Z, W_3) of sets of vertices of graph T such that $C \subseteq W_1$ and $Z \subseteq W_2$ is a common shadow of vertices from C, Z dominates all vertices from W_3, and C is an inclusion-maximal set of the vertices whose common shadow is Z.*
2. *Z is a positive hypothesis obtained for the learning context $K_\pm(T)$ and C is the extent of the hypothesis Z.*

Proof. $\leftarrow$ If Z is a positive hypothesis, then it is the intent of a positive concept and, therefore, corresponds to a biclique of the bipartite graph $(W_1 \cup W_2, E_1 \cap (W_1 \times W_2))$. Therefore, the extent C of the hypothesis corresponds to the common shadow of the set of vertices that correspond to Z. Furthermore, since Z is not contained in intent of any negative example, by the definition of the relation I_- of the learning context corresponding to the tripartite graph T, the set of vertices corresponding to Z dominates the set of vertices W_3.

$\rightarrow$ The intersection of all sets of the form $\{w^2 | (w_i^1, w^2) \in E_1\}$ for $w_i^1 \in C$ is the set $\{w^2 | w^2 \in Z\}$, i.e., set Z, because Z is the common shadow of vertices from C. On the other hand, among vertices from W_1 there are no other vertices adjacent to all vertices from Z because C is inclusion-maximal by the conditions of the Lemma. Thus, the pair $\langle C, Z\rangle$ is a positive concept. Since Z dominates W_3, for all $w_l^3 \in W_3$ there is some $w^2 \in Z \subseteq W_2$ such that $(w^2, w_l^3) \in E_1$. Hence, by the definition of the intent of a negative example from the graph T, for every negative example $g_l \in G_-$ there is an element of Z that is not contained in the intent of $\{g_l\}^-$ and, therefore, Z is not contained in any negative example. Thus, Z is a positive hypothesis. □

Obviously, the converse is also feasible: given a learning context $K_\pm = (G_+ \cup G_-, M \cup \{w\}, I_+ \cup I_- \cup G_+ \times \{w\})$, one can build a tripartite graph $T = (W_1 \cup W_2 \cup W_3, E_1)$ such that vertices of the set W_1 are in one-to-one correspondence with $(+)$-examples, vertices of the set W_3 are in one-to-one correspondence with $(-)$-examples, vertices of the set W_2 are in one-to-one correspondence with elements of the set M, the subset of edges $E_1 \cap W_1 \times W_2$ is given by the relation I_+ and the subset of edges $E_1 \cap W_2 \times W_3$ is given by the complementation of the relation I_-, i.e., by $W_2 \times W_3 \setminus I_-$. A positive hypothesis H_+ of $K_\pm$ corresponds to a subset $V_2 \subseteq W_2$ that dominates all vertices of W_3 and induces a biclique on vertices $W_1 \cup W_2$ (i.e., the common shadow of V_2 onto W_1 is a set $V_1 \subseteq W_1$ and the common shadow of V_1 onto W_2 is $V_2 \subseteq W_2$).

Example 2. Consider graph T in Fig. 1. Here the corresponding learning context is given by that from Example 1.

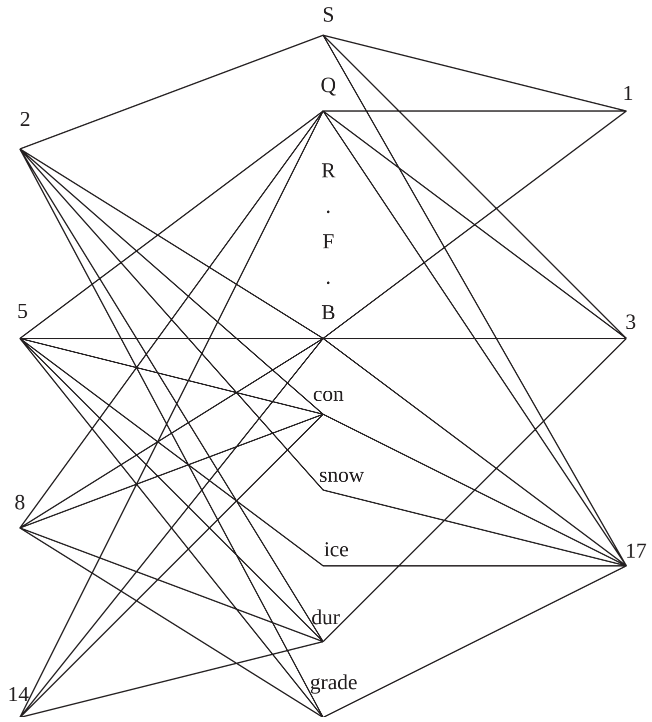

Fig. 1.

5. Complexity results about hypotheses

5.1. The number of hypotheses and minimal hypotheses

Several algorithms for computing the set of all concepts and the covering relation of its Hasse (line) diagram are known in the literature, e.g., [18,6] (see [7,3,14,19]. All these algorithms are polynomial-delay [11]. Recall that an algorithm listing a family of objects has a *delay* d if it satisfies the following conditions whenever it is run with an input of size p [11]:

1. It executes at most $d(p)$ computation steps before either outputting the first structure or halting;
2. After any output it executes at most $d(p)$ computation steps before either outputting the next structure or halting.

An algorithm whose delay is bounded from above by a polynomial in the length of the input is called a *polynomial delay algorithm* [11].

An experimental comparison of various algorithms can be found in a recent review [16]. To obtain an algorithm for computing hypotheses or minimal hypotheses from any of the above algorithms one needs to add an additional command for testing that the current intent of the positive context is not contained in the intent of any negative example.

The problem of counting the number of all concepts, given a formal context, is a long standing one. The knowledge of this number can be useful for effective resource allocation.

Obviously, for a context of the form $K=(M,M,\neq)$, which gives rise to a Boolean concept lattice, the number of concepts is exponential. The following upper bound for the size of the set of all concepts of a context $K=(G,M,I)$ was proposed in [22]: $|\underline{\mathfrak{B}}(K)| \leqslant \frac{3}{2} \times 2^{\sqrt{|I|+1}} - 1$ for $|I| > 2$.

The problem of counting all formal concepts of a context is equivalent to the problem of counting all bicliques of a bipartite graph. An upper bound on the number b of bicliques of a bipartite graph was proposed in [21]. For a bipartite graph $B=(X \cup Y, E)$ with n vertices, i.e., $|X|+|Y|=n$, $b \leqslant 2^{n/2}$. We can give the following very simple proof of this statement in terms of FCA. Consider a formal context $K=(G,M,I)$ (which is equivalent to the bipartite graph $(G \cup M, I)$). The number of concepts $|\underline{\mathfrak{B}}|$ can neither exceed the number of extents, which is not greater than $2^{|G|}$, no the number of intents, which is not greater than $2^{|M|}$. Thus, $|\underline{\mathfrak{B}}| \leqslant 2^{\min\{|G|,|M|\}}$, which is a better upper bound than that in [21], since $\min\{|G|,|M|\} \leqslant (|G|+|M|)/2 = n/2$, where $n=|G|+|M|$.

Now we show that the problem of counting all concepts of a formal context is intractable.

Theorem 5. *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 $C=\bigwedge_{i=1}^{s}(x_{i,1}\vee x_{i,2})$" [23]:

Input: Monotone (without negation) CNF with two variables in each disjunction $C=\bigwedge_{i=1}^{s}(x_{i,1}\vee x_{i,2})$, where $x_{i,1}$, $x_{i,2}\in X=\{x_1,\dots,x_n\}$ 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 $D_i=(\bar{x}_{i,1}\wedge\bar{x}_{i,2})$, $i=\overline{1,s}$. 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-vector with $n-2$ ones and two zeros in 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 $M=\{m_1,\dots,m_n\}$, where elements of M are in one-to-one correspondence with variables 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 $M_i=M\setminus\{m_{i,1},m_{i,2}\}:=\{m^{i,1},\dots,m^{i,n-2}\}$, the set of objects is $G_i=\{g_i^0,g_i^1,\dots,g_i^{n-2}\}$, and the relation $I_i\subseteq M_i\times G_i$ is defined by object intents as follows: $\{g_i^0\}'=M_i,\{g_i^j\}'=M_i\setminus\{m^{i_j}\}$ for $j\in\overline{1,n-2}$. Now the context K is defined as $K=(\dot\cup_{i=1}^{s}G_i,M,\cup_{i=1}^{s}I_i)$.

First, 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, which can be not unique. Recall that for a set M a set Q of subsets of M is called a *closure system* if for any X, $Y\in Q$ one has also $(X\cap Y)\in Q$ [1]. 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, denoted by $\mathscr{P}(M_i)$. Elements of this set of attributes are in one-to-one correspondence with binary n-vectors, where components are in one-to-one correspondence with elements of M with the same number. A vector of this form satisfies D_i, since it has zeros at i_1th and i_2th places. Therefore, this vector satisfies D.

It remains to show that binary n-vectors that satisfy D are in one-to-one correspondence with intents of K. In fact, each binary n-vector v that satisfies D, satisfies D_i for some i (this i may be not unique). Then this vector has zero i_1th and i_2th positions. Therefore, the corresponding set of attributes A belongs to $\mathscr{P}(M_i)$, where $M_i=M\setminus\{m_{i,1},m_{i,2}\}$. Since $\mathscr{P}(M_i)$ is the closure system of intents of K_i for each i, there is a set of objects $\{g_i^{j,1}\dots,g_i^{j,r}\}\subseteq G_i$, $r\leqslant n-2$ such that $\{g_i^{j,1}\dots,g_i^{j,r}\}'=A$.

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 counting all hypotheses of a learning context is #P-complete.*

Moreover, the problem of counting all minimal hypotheses is also intractable.

Theorem 6. *The following problem "the number of all hypotheses that are minimal by inclusion" is #P-complete*:

Input: Learning context $K_{\pm}$.
Output: The number of all minimal positive hypotheses, i.e., $\#\{H_+|H_+$ is a $(+)$-hypothesis of $K_{\pm}$ and any H such that $H\subset H_+$ is not a $(+)$-hypothesis.$\}$

Proof. We reduce the following problem of determining the number of inclusion-minimal vertex coverings (shown to be #P-complete in [23]) to our problem:

Input: Graph $\Gamma=(V,E)$.

Output: $\#\{W\subseteq V|[((u,v)\in E)\to(u\in A)\vee(v\in A)]$ holds for $A=W$ but not for any $A\subset W\}$.

By construction of Lemma 3, an inclusion-minimal vertex covering in graph Γ corresponds to a triple (C,Z,W_3) of subsets of the vertices of the tripartite graph T associated with Γ, such that $Z\subseteq W_2$ is the common shadow of $C\subseteq W_1$ and C is the common shadow of Z. Moreover, Z is an inclusion-minimal set of vertices from W_2 that dominates W_3. Conversely, each triple of this form corresponds to an inclusion-minimal vertex covering in graph Γ. By Lemma 4, triples of this form are in one-to-one correspondence with positive hypotheses of a learning context corresponding to the tripartite graph T. The inclusion-minimality of the set of vertices Z corresponds to the minimality of the hypothesis. □

The same result holds for negative hypotheses.

Theorem 6 also implies that problems of counting all hypotheses with inequality size constraints (such as $|H_+^+| \leqslant k$, $|H_+^+| + |H_+| \geqslant k$) are #P-complete. Indeed, by putting $k \geqslant |G| + |M|$, we obtain the reduction. The #P-completeness of counting problems with equality constraints, such as $|H_+^+| = k$, can be proved by summing over all possible values of $k : k = 1, \ldots, |G|$.

5.2. Decision problems related to hypotheses with size constraints

Since the number of all hypotheses can be exponential in the size of the learning context $K_\pm$, it is reasonable to generate only some "good" hypotheses. A natural quality estimate of a positive hypothesis H_+ is $|H_+^+|$, the size of the corresponding extent, or the number of examples that "support" H_+. This measure is typical in data analysis and machine learning, see [17]. Another aspect of quality of H_+ is the description size ("how detailed is a hypothesis"), i.e., $|H_+|$.

In this section, we consider complexity of some decision problems for hypotheses with size constraints.

Proposition 7. *The following "maximum-size hypothesis" problem can be solved in* $O(|M| \cdot |G_+| \cdot |G_-|)$ *time.*

Instance: Learning context $K_\pm$, positive integer k.
Question: Does there exist a $(+)$-hypothesis H_+ with extent H_+^+ such that $|H_+| \geqslant k$?

Proof. We consider intents of all positive examples and discard those that are contained in intent of a negative example. For the resulting set of positive intents we test whether there are intents larger than k. If there are no such intents, then there are no such intents at all, since we have considered the largest ones. The total worst-case time complexity in this case is $O(|M| \cdot |G_+| \cdot |G_-|)$. Since the largest intents correspond to the smallest extents, the same procedure gives an answer to the question about the "$|H_+^+| \leqslant k$" problem.

If we accept the restriction that a hypothesis H_+ should have support not less than two (i.e., $|H_+^+| \geqslant 2$, see comments to Definition 2), than we consider not the intents of positive examples, but all possible pairwise intersections of them. In this case the worst-case time complexity is $O(|M| \cdot |G_+|^2 \cdot |G_-|)$. □

Corollary. *The following "minimum-extent hypothesis" problem can be solved in* $O(|M| \cdot |G_+| \cdot |G_-|)$ *time.*

Instance: Learning context $K_\pm$, positive integer k.
Question: Does there exist a $(+)$-hypothesis H_+ with extent H_+^+ such that $|H_+^+| \leqslant k$?

Proof. The corollary follows from the fact that largest hypotheses have least extents.

Note that the algorithm in the proof of Theorem 7 also provides an answer to a more general question: "Does there exist at least one hypothesis for a given learning context?" If every object intent of the positive context is contained in a negative object intent, then every other intent is a fortiori contained in a negative object intent and, hence, there is no positive hypothesis.

In [15], we proved that the problem of finding a formal concept (A,B) with $|A|+|B| \leqslant k$ is NP-complete, which implies NP-completeness of the problem of finding a hypothesis H_+ with $|H_+| + |H_+^+| \leqslant k$.

Theorem 8. *The following "minimum-size hypothesis" problem is NP-complete*:

Instance: Learning context $K_\pm$, positive integer k.
Question: Does there exist a $(+)$-hypothesis H_+ such that $|H_+| \leqslant k$?

Proof. The problem obviously belongs to the NP class. For each potential solution, i.e., a subset of attributes $S \subseteq M$, the closure S^{++} is compared with S. In the case of coincidence, it is tested whether S is not contained in any negative intent and $|S|$ is compared with k. All these operations can be performed within $O(|M| \cdot (|G_+| + |G_-|))$ time. Now we reduce the problem of minimal vertex covering from [9] to that of ours.

Instance: Graph $\Gamma = (V,E)$, positive integer $k \leqslant |V|$.

Question: Does there exist a set $W \subseteq V$ such that $|W| \leqslant k$ and the relations "$v_i \in W$" or "$v_j \in W$" take place for an arbitrary $e = (v_i, v_j) \in E$.

Applying Definition 3, we construct a tripartite graph T associated with Γ. By Lemma 3, a vertex covering of size k of graph Γ corresponds, in graph T, to a triple (C,Z,W) such that $|C| = |V| - k$, $|Z| = k$, $|W| = |E|$; set Z is a common shadow of the set of vertices C and Z dominates the set W. By Lemma 4, this triple corresponds to a hypothesis with intent of size k formed by $|V|-k$ positive examples from the initial data corresponding to graph T by Definition 3. The reduction is realized within $O(|V|^2+|E|)$ time. □

Corollary. *The following "maximum-extent hypothesis" problem is NP-complete*:

Instance: Learning context $K_\pm$, positive integer k.
Question: Does there exist a (+)-hypothesis H_+ such that $|H_+^+| \geqslant k$?

Proof. The corollary follows from the fact that the largest extents correspond to the smallest intents. □

Theorem 9. *The following "largest hypothesis" problem is NP-complete*:

Instance: Learning context $K_\pm$, positive integer k.
Question: Does there exist a (+)-hypothesis H_+ such that $|H_+^+|+|H_+| \geqslant k$?

Proof. By Lemma 4, this problem is equivalent to the following one:
Instance: Tripartite graph $T_1=(V_1\cup V_2\cup V_3,E)$, $E\subseteq (V_1\times V_2)\cup(V_2\times V_3)$, natural number $\hat{k}\leqslant |V_1|+|V_2|$.
Question: Does there exist a biclique $B_1=(U_1\cup U_2, U_1\times U_2)$ of the graph T_1 such that $U_1\subseteq V_1$, $U_2\subseteq V_2$, $|U_1|+|U_2|\geqslant \hat{k}$, and U_2 dominates V_3.

We reduce to this problem the problem of "minimal vertex covering" (see the proof of Theorem 8). From graph Γ we construct an associated tripartite graph $T=(W_1\cup W_2\cup W_3,\ E)$. From T, we construct another tripartite graph $T_1=(V_1\cup V_2\cup V_3,E_1)$: $E_1\subseteq (V_1\times V_2)\cup(V_2\times V_3)$, $|V_1|=n\times|W_1|$, $|V_2|=|W_2|$, $|V_3|=|W_3|$, $V_1=V_1^1\cup\ldots\cup V_1^n$, where for any $i{:}1\leqslant i\leqslant n |V_1^i|=|W_1|$ and each subgraph induced by sets of vertices V_1^i, V_2, and V_3 is isomorphic to the graph T. Thus, each biclique of the graph T induced by subsets of vertices W_1 and W_2 corresponds to n isomorphic copies of it, which are bicliques of the graph T_1.

We will show that there exists a vertex covering in Γ not greater than k ($k\leqslant |V|=n$) if and only if the tripartite graph T_1 has a complete bipartite subgraph $B=(U_1\cup U_2, U_1\times U_2)$ such that $U_1\subseteq V_1$, $U_2\subseteq V_2$, $|U_1|+|U_2|\geqslant \hat{k}=n\cdot(n-k)+1$, and U_2 dominates V_3.

Indeed, suppose that the graph Γ has a vertex covering not greater than k. This implies that there is a biclique $(U_1\cup U_2, U_1\times U_2)$ of graph T_1 such that $U_1\subseteq V_1$, $U_2\subseteq V_2$, $1\leqslant |U_2|\leqslant k$, and U_2 dominates V_3. Here U_1 is not less than $n(n-k)$, and $|U_1|+|U_2|\geqslant n(n-k)+1$.

Conversely, suppose that graph T_1 contains biclique $(U_1\cup U_2,\ U_1\times U_2)$ such that $U_1\subseteq V_1$, $U_2\subseteq V_2$, U_2 dominates V_3, and $|U_1|+|U_2|\geqslant \hat{k}=n(n-k)+1$. Since $|U_2|\leqslant n$, we have $|U_1|\geqslant n(n-k)-n+1$. This biclique of graph T_1 corresponds in graph T to biclique $B=(Y_1\cup Y_2, Y_1\times Y_2)$ such that $Y_1\subseteq W_1$, $Y_2\subseteq W_2$ and $|Y_1|=|Y_2|/n$. Therefore, $|Y_1|\geqslant [(n\cdot(n-k)-n+1)/n]+1>n-k$. By definition of graph T (Definition 3) this means that $|Y_2|\leqslant k$ and, by Lemma 3, graph Γ has a vertex covering not greater than k. □

Note that in case where $G_-=\emptyset$, i.e., when there is no negative examples and the corresponding tripartite graph turns into a bipartite, the previous problem can be solved in polynomial time with the use of the following observation. This observation uses a well-known correspondence between formal contexts and bipartite graphs, see e.g., [3].

We represent the positive context $K_+=(G_+,M,I_+)$ as a bipartite graph Γ_B with sets of vertices G_+ and M and edges given by the relation I. Each concept (A,B) from $\mathfrak{B}(K_+)$ corresponds to a biclique of Γ_B, and $|A|+|B|$ is the number of vertices of this subgraph. To find a biclique of Γ_B with $|A|+|B|\geqslant k$ we use the following construction from [26]. Consider the context $\bar{K}_+=(G_+,M,\bar{I}_+)$ with relation $\bar{I}_+=G\times M\setminus I_+$, the complement of I_+, and the corresponding bipartite graph $\overline{\Gamma_B}$. A concept (A,B) of $\mathfrak{B}(K_+)$ corresponds to an inclusion-maximal independent set of vertices of $\overline{\Gamma_B}$, i.e., an inclusion-maximal set of vertices where no pair of vertices is connected by an edge. A concept (A,B) of $\mathfrak{B}(K_+)$ with the largest $|A|+|B|$ corresponds to the largest (in the number of vertices) independent set of $\bar{\Gamma}_B$. According to the König theorem (see, e.g., [24]), the number of vertices in the largest independent set is $|V(\bar{\Gamma}_B)|-|M|$, where $|V(\bar{\Gamma}_B)|$ is the number of vertices in $\bar{\Gamma}_B$ and M is the number of edges in the maximal matching of $\bar{\Gamma}_B$. The size of a maximal matching can be found by a polynomial-time algorithm, for example, by that from [12].

In the case where the intent size of positive examples is bounded by a constant k (i.e., $|\{g\}^+|\leqslant k$ for all $g\in G_+$) the above-mentioned #P-complete and NP-complete problems concerning hypotheses become tractable. Indeed, in this case the number of all positive concepts is $O(|M|^k)$ and any of the algorithms mentioned above can generate all hypotheses in time polynomial in $|G_+|$, $|G_-|$, and $|M|$. Due to the symmetry of objects and attributes, a similar condition $|\{m\}^+| \leqslant k$ for all $m \in M$ is also sufficient for the polynomial tractability of the problems considered in this chapter.

6. Graph-theoretic interpretation of classification

Since the number of all hypotheses and minimal hypotheses can be exponential in the size of the learning context, it is reasonable to raise a question of the possibility of fast classification of an object without generating all (minimal) hypotheses. Here we propose a graph-theoretic interpretation of the classification problem, which will be used further for the study of the classification complexity.

Definition 5. "Problem of domination by parts of complete bipartite subgraph" (DPCBS):

Instance: Quadripartite graph $Q = (V_1 \cup V_2 \cup V_3 \cup V_4, E)$, $E \subseteq (V_1 \times V_2) \cup (V_2 \times V_3) \cup (V_3 \times V_4)$. The graphs B_1, B_2, B_3 are subgraphs of the graph Q induced by the sets of vertices $(V_1 \cup V_2)$, $(V_2 \cup V_3)$, $(V_3 \cup V_4)$, respectively.

Question: Does there exist a complete bipartite subgraph $\langle W_2 \cup W_3, W_2 \times W_3 \rangle$ of the graph B_2 such that it is maximal by inclusion, $W_2 \subseteq V_2$, $W_3 \subseteq V_3$, the set of vertices W_2 dominates V_1, and the set of vertices W_3 dominates V_4?

Definition 6. The problem "hypothesis for a positive classification (HFPC)" corresponding to a DPCBS problem is defined as follows:

Instance: Classification context K_c with a single undetermined example $g_\tau \in G_\tau$, where $M = (V_1 \cup V_3)$, $G_+ = \{g_i | i = 1, \dots, |V_2|\}$, $G_- = \{f_l | l = 1, \dots, |V_4|\}$. The relation I_+ is given by object intents as follows: $\{g_i\}^+$ consists of all vertices of V_1 that are not adjacent to the vertex $v_i^2 \in V_2$ and all vertices from V_3 that are adjacent to the vertex $v_i^2 \in V_2$. The undetermined context is given by the one-element object set $G_\tau = \{g_\tau\}$ and relation I_τ defined by $\{g_\tau\}^\tau = V_3$. The relation I_- is defined by object intents as follows: $\{f_l\}^- = V_3 \setminus \{w_1^3, \dots, w_q^3\}$, where $\{w_1^3, \dots, w_q^3\}$ is the set of all vertices from V_3 adjacent to the vertex $v_l^4 \in V_4$.

Question: Does there exist a (+)-hypothesis H_+ such that $H_+ \subseteq \{g_\tau\}^\tau = V_3$, i.e., is a hypothesis for the positive classification of g_τ?

Lemma 10. *A DPCBS problem has a solution if and only if the corresponding HFPC problem has a solution.*

Proof. (1) Let H_+ be a (+)-hypothesis and $H_+ \subseteq \{g_\tau\}^\tau$. In this case, in graph Q, the subgraph induced by the vertices $v_{i_1}^2, \dots, v_{i_n}^2 \in V_2$ that correspond to the extent H_+^+ and vertices from V_3 that correspond to H_+ is a biclique. The set of vertices $v_{i_1}^2, \dots, v_{i_n}^2 \in V_2$ dominates V_1. Indeed, let a vertex $v_1 \in V_1$ be nonadjacent to any of the vertices from $\{v_{i_1}^2, \dots, v_{i_n}^2\}$. Then, by definition of (+)-examples, $v_1 \in \{g_1\}^+, \dots, v_1 \in \{g_n\}^+$ and $\{g_1\}^+ \cap \dots \cap \{g_n\}^+ \not\subset \{g_\tau\}^\tau$ (because $v_1 \notin \{g_\tau\}^\tau$). Let H_+ correspond to vertices $w_1^3, \dots, w_{|H_+|}^3$ in Q. Then, the set of vertices $\{w_1^3, \dots, w_{|H_+|}^3\}$ dominates the set V_4. Suppose that this is not the fact, and a vertex v_j^4 is not adjacent to any vertex from $\{w_1^3, \dots, w_{|H_+|}^3\}$. Then by definition of (−)-examples, for arbitrary (−)-example f_j we have $w_1^3 \in \{f_j\}^-, \dots, w_{|H_+|}^3 \in \{f_j\}^-$ and $H_+ \subseteq \{f_j\}^-$, which contradicts the fact that H_+ is a hypothesis.

(2) Let $W_2 \subseteq V_2$, $W_3 \subseteq V_3$ be sets of graph vertices such that a bipartite graph induced by these vertices is a biclique, the vertex set W_2 dominates V_1 and W_3 dominates V_4. By definition of (+)- and (−)-examples specified by graph Q and by Lemma 4, (W_2, W_3) corresponds to a (+)-hypothesis H_+. The bipartite graph B_w induced by vertices W_2 and W_3 is a biclique of the bipartite graph B_2. Due to the domination of W_2 over V_1 and by the definition of the bipartite graph B_1, the graph B_w is also a biclique of the graph $((V_1 \cup V_3) \cup V_2, E)$, where for $v \in V_1 \cup V_3$, $v_2 \in V_2$, one has $(v, v_2) \in E$ iff a positive example corresponding to v_2 has an attribute corresponding to v. Therefore, B_w corresponds to a positive concept. It remains to demonstrate that its intent lies in the set $\{g_\tau\}^\tau$ and is not contained in any negative object intent. The former is true by the definition of (+)-examples corresponding to graph Q and due to the fact that W_2 dominates V_1. Indeed, suppose that $\{g_1, \dots, g_n\}^+ = H_+ \not\subseteq \{g_\tau\}^\tau$. Then, by definition of g_i, it is possible to find a vertex $v \in V_1$ such that v is not adjacent to any vertex from W_2, which contradicts the fact that W_2 dominates V_1. The fact that the positive concept (W_2, W_3) is a positive hypothesis follows directly from the definition of (−)-examples from the graph Q and the fact that W_3 dominates V_4. □

Example 3. Consider the problem of classification of the τ-example with intent $\{\mathbf{R}, \mathbf{B}, \mathbf{con}, \mathbf{dur}, \mathbf{snow}, \mathbf{ice}\}$ for the learning context given in Example 1.

The undetermined example is classified positively, since the minimal positive hypothesis $\{\mathbf{B}, \mathbf{con}, \mathbf{dur}\}$ is contained in $\{g_\tau\}^\tau$ (note that the nonminimal hypotheses are not contained in it) and no negative hypothesis is contained in $\{g_\tau\}^\tau$. In

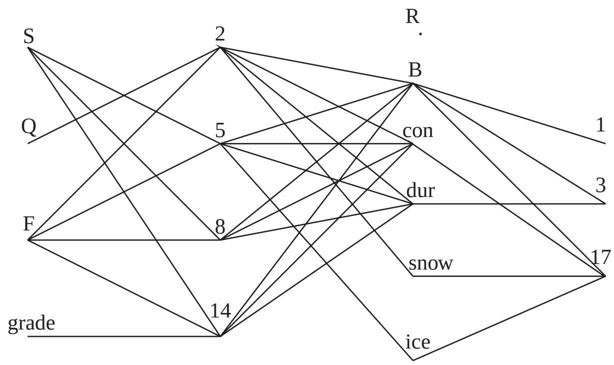

Fig. 2.

terms of the corresponding quadripartite graph (Fig. 2), the set of vertices $\{2,5,8,14\} \cup \{\mathbf{B},\mathbf{con},\mathbf{dur}\}$ induces a complete bipartite graph, the set of vertices $\{2,5,8,14\}$ dominates all the vertices from the first part and the set of vertices $\{\mathbf{B},\mathbf{con},\mathbf{dur}\}$ dominates all the vertices of the fourth part.

7. Complexity of classification

Now, we discuss algorithmic complexity of classification of objects from G_τ.

Recall that by Definition 2 an object $g_\tau \in G_\tau$ is classified positively if there exists a hypothesis for a positive classification of this object and there are no hypothesis for its negative classification. This definition can easily be implemented as an algorithm: first generate sets of (+)- and (−)-hypotheses, then test containment of the resulting hypotheses in the object intent $\{g_\tau\}^\tau$. However, this realization has an obvious drawback: if the number of hypotheses is exponential with respect to the input size, then time and memory required for classification of even a single object from G_τ is also exponential. Can a classification of a given undetermined example be realized without computing the set of all hypotheses or without computing the set of all minimal hypotheses? This question corresponds to the following problem:

Instance: A classification context K_c and an undetermined example $g_\tau \in G_\tau$.

Question: Is $g_\tau \in G_\tau$ classified positively with respect to K_c?

The motivation of studying this problem becomes even stronger when we consider the case where $G_- = \emptyset$. Here positive hypotheses are just positive concepts and minimal hypotheses are inclusion-minimal sets of the form $\{m\}^{++}$ for $m \in M$. Their generation is accomplished in $O(|G_+||M|^2)$ time. Testing the containment of attribute intents in $\{g_\tau\}^\tau$ is done in $|M|$ time. If there is $m \in M$ such that $\{m\}^{++} \subseteq \{g_\tau\}^\tau$, then classification is realized. If not, then other positive intents are also not contained in $\{g_\tau\}^\tau$ and positive classification is not possible. Thus, the classification problem is solved in $O(|G_+||M|^2)$ time even if the number of all hypotheses is exponential with respect to the input size.

We shall show that, for an arbitrary classification context and an arbitrary undetermined example $g_\tau \in G_\tau$, the decision problem of the existence of a "(+)-hypothesis for a positive classification of g_τ" is NP-complete. By the symmetry of (−)- and (+)-hypotheses, this implies that the problem "whether there is no (−)-hypothesis for the negative classification of g_τ" is coNP-complete.

The equivalence established in Lemma 9 allows us to reformulate the classification problem as a quadripartite graph problem.

Theorem 11. *The problem DPCBS is NP-complete.*

Proof. Consider the following special case of the DPCBS problem. Let $|V_2| = |V_3| = n$; $\forall i,j \ 1 < i,j < n$; for $v_j^2 \in V_2$, $v_i^3 \in V_3$ one has $(v_i^2, v_j^3) \in E$ if and only if $i \neq j$, and let biparite subgraphs induced by sets of vertices $V_2 \cup V_1$ and $V_3 \cup V_4$, i.e., B_1 and B_3, be isomorphic as unlabeled graphs (this isomorphism is given by the mapping that establishes one-to-one correspondence between sets of vertices V_2 and V_3, V_1 and V_4, respectively). In this case, any biclique of the bipartite graph B_2 consists of graphs of the form $(\{v_{i_1}^2,\dots,v_{i_k}^2\} \cup \{v_{j_1}^3,\dots,v_{j_m}^3\},\ \{v_{i_1}^2,\dots,v_{i_k}^2\} \times \{v_{j_1}^3,\dots,v_{j_m}^3\})$, where $\{j_1,\dots,j_m\} = \{1,\dots,n\} \setminus \{i_1,\dots,i_k\}$, i.e., the set of indices of vertices from V_2 is complementary to the set of indices of vertices from V_3. Considering that bipartite graphs induced by vertex sets V_2 and V_1, V_3 and V_4, respectively, are isomorphic, this special case of the DPCBS problem is equivalent to the following "domination by mutually complementary sets of vertices" (DMCSV) problem:

Instance: Bipartite graph $B = (V_1 \cup V_2, E)$, $E \subseteq V_1 \times V_2$.

Question: Does there exist a set $W_1 \subseteq V_1$ such that both sets W_1, $V_1 \setminus W_1$ dominate V_2?

Lemma 12. *The DMCSV problem is NP-complete.*

Proof. [13] We reduce the problem of CNF satisfiability to a DMCSV. The CNF satisfiability problem is stated as follows:

Instance: A CNF $C = D_1 \wedge \cdots \wedge D_n$, $D_i = (\neg)x_{i_1} \vee \cdots \vee (\neg)x_{i_k}$, where for arbitrary i, j $x_{i_j} \in X = \{x_1, \ldots, x_m\}$.

Question: Does there exist a Boolean vector $(a_1, \ldots, a_m)$ that satisfies C?

From the CNF C we construct the bipartite graph $B = (V_1 \cup V_2, E)$, $E \subseteq V_1 \times V_2$, $|V_1| = 2m + 1$, $|V_2| = n + m$.

In vertex set V_1 a vertex v_{i1} is in one-to-one correspondence with x_i and a vertex v_{i2} is in one-to-one correspondence with $\neg x_i$ and, besides these $2m$ vertices, there is an additional one, denoted by v_{2m+1}. In the set of vertices V_2 each disjunction D_j is assigned to a vertex v_j, $1 \leqslant j \leqslant n$. Each variable x_i is assigned to the vertex $v_{n+i} \in V_2$, $1 \leqslant i \leqslant m$. The pair of vertices (v_i^1, v_j^2), where $v_i^1 \in V_1$, $v_j^2 \in V_2$, is connected by an edge if and only if one of the following cases takes place:

(1) v_i^1 corresponds to the literal $(\neg)x_{i_1}$ that belongs to the disjunction D_{j_2}, which corresponds to the vertex v_j^2,
(2) v_i^1 corresponds to the literal $(\neg)x_{i_1}$, and the vertex v_j^2 corresponds to the variable x_{i_1} (i.e., $j = n + i_1$),
(3) $i = 2m + 1$, $1 \leqslant j \leqslant n$.

We shall show that a Boolean vector satisfying CNF C exists if and only if the graph B contains the vertex set $W_1 \subset V_1$ such that both sets W_1 and $V_1 \setminus W_1$ dominate V_2, i.e., the corresponding DMCSV problem has a solution.

Indeed, let C be satisfied by a vector $(a_{i_1}, \ldots, a_{i_n})$, where elements $a_{i_1}, \ldots, a_{i_k}$ are ones and the other elements are zeros. In this case all vertices from V_2 are dominated by vertices from $W_1 \subset V_1$ that correspond to literals satisfied by this vector. Since the vertex v_{2m+1} is connected to all vertices from $\{v_1^2, \ldots, v_n^2\}$ and vertices $v_{n+1}^2, \ldots, v_{n+m}^2$ are dominated by those vertices from $V_1 \setminus W_1$ which correspond to other literals, therefore, $(V_1 \setminus W_1)$ also dominates V_2.

Conversely, let a set $W_1 \subseteq V_1$ be such that W_1 and $V_1 \setminus W_1$ dominate V_2. Suppose that one of these sets (e.g. $V_1 \setminus W_1$) contains the vertex v_{2m+1}. The vertices corresponding to opposite literals can belong to either W_1 or $V_1 \setminus W_1$ (otherwise, vertices v_j, where $n + 1 \leqslant j \leqslant n + m$, that are connected with just a pair of vertices, are not dominated by W_1 or by $(V_1 \setminus W_1)$). Therefore, we can construct a Boolean vector by setting each literal that corresponds to a vertex belonging to W_1 be equal to one and assigning zero values to the remaining literals. The resulting vector satisfies CNF C. Indeed, suppose that this is not so. In that case, there should be a nonsatisfied disjunction D. However, the vertex corresponding to this disjunction is dominated by a certain vertex from W_1, and the literal corresponding to this vertex ought to be nonzero, i.e., should satisfy D. We have proved the reduction. Its polynomiality and the membership of the problem in NP are obvious. □

Note that in the degenerate cases where

- $V_1 = \emptyset$ $(M = \{g_\tau\}^\tau)$
- $V_4 = \emptyset$ ($G_- = \emptyset$, see the beginning of this subsection)

the quadripartite graph becomes tripartite and a polynomial algorithm for solving DPCBS problem exists. A polynomial algorithm is also possible when the size of $\{g_\tau\}^\tau$ is bounded from above by a constant. This assumption is well justified in various practical situations, for example, in the "structure—activity relationship" (SAR) problem (see, e.g., [2]), where the target attribute w is a biological activity and classification means forecasting the membership of a certain chemical compound (represented by a set of attributes) in the class of active or inactive compounds. The size of a compound description can be considered limited by a constant, at least when a sequence of classifications for a single compound with growing sets of examples and attributes (i.e., elements of M) is considered. Another case where the possibility of classification is tested in polynomial time is the situation when the example being classified is actually a (positive or negative) example from the initial sample. This test of internal consistency of data proposed in [5] under the name of "criterion of sufficient reason" is similar to cross-validation. This criterion requires that each positive example is classified positively by means of generated hypotheses and there is at least one positive hypothesis H_+ for its (+)-classification support greater than one (i.e., $|H_+^+| \geqslant 2$). For an arbitrary positive example g_+ this test can be realized as follows: First we look for the intersections of intents of positive examples with $\{g_+\}^+$. If there is an example $g_i \in G_+$ such that $\{g_+\}^+ \cap \{g_i\}^+ \not\subseteq \{g_-\}^-$ for all $g_- \in G_-$, then there is a hypothesis for the positive classification of g_+. If, on the

Table 1

	$\leqslant$	$\geqslant$
$\lvert H_+\rvert$	NP (Theorem 8)	P (Proposition 7)
$\lvert H_+^+\rvert$	P (Proposition 7)	NP (Theorem 8)
$\lvert H_+\rvert + \lvert H_+^+\rvert$	NP (Theorem 9)	NP (Theorem 9)

contrary, for every positive example $g_i \in G_+$ we have $\{g_+\}^+ \cap \{g_i\}^+ \subseteq \{g_-\}^-$ for some $g_- \in G_-$, then no positive intent contained in $\{g_+\}^+$ is a positive hypothesis. Hence, there is no hypothesis for the positive classification of g_+. The test is realized in $O(|G_+| \cdot |G_-| \cdot |M|)$ time.

An obvious algorithm for classification of an undetermined example g_τ in the general case can be based on any algorithm for constructing the set of all concepts. For testing if a current positive intent B_+ is a hypothesis for the positive classification, one should additionally test noncontainment of B_+ in any negative object intent and its containment in $\{g_\tau\}^\tau$ (i.e., the condition $B_+ \subseteq \{g_\tau\}^\tau$). If one of these conditions is not satisfied, B_+ is not a positive hypothesis for the classification of g_τ and the next intent is considered. In the same way it is tested whether there is no negative hypothesis against the positive classification of g_τ.

Let Int denote the set of all intents (both positive and negative) contained in $\{g_\tau\}^\tau$. Then the complexity of such an algorithm is either $O(|\mathrm{Int}|(|G_+|+|G_-|)^2 \cdot |M|)$ or $O(|\mathrm{Int}|(|G_+|+|G_-|) \cdot |M|^2)$, depending on the order in which intents are generated (from largest to smallest or from smallest to largest, respectively). When the condition $|\{g_\tau\}^\tau| \leqslant c$ is satisfied for a constant c, we have $|\mathrm{Int}| \leqslant 2^c$ and the algorithm runs in time polynomial in the input size.

8. Conclusion

We presented a model of learning from positive and negative examples and classification based on FCA. We showed that hypotheses correspond to certain subgraphs of tripartite graphs and hypotheses for classification of an object correspond to certain subgraphs of quadripartite graphs. At the same time hypotheses and classifications are naturally considered in terms of order filters of concept lattices.

We showed #P-completeness of the problem of counting all concepts and all minimal hypotheses. We showed intractability or polynomial solvability of some decision problems related to hypotheses with size constraints. These results are summarized in Table 1. We also considered cases where generally NP-complete problems are polynomially solvable. P denotes the existence of a polynomial algorithm for a particular problem, NP denotes NP-completeness of the problem. For example, the upper left entry of the table means that the problem “Does there exist a hypothesis with $|H_+| \leqslant k$” is NP-complete. Finally, we proved the intractability of the classification problem and considered cases where this problem can be solved in polynomial time.

Acknowledgements

This work was partially supported by the Alexander von Humboldt Foundation. The author thanks Bernhard Ganter, Sergei A. Obiedkov, and anonymous referees for helpful comments.

References

[1] G.D. Birkhoff, Lattice Theory, American Mathematical Society, Providence, RI, 1979.
[2] V.G. Blinova, D.A. Dobrynin, V.K. Finn, S.O. Kuznetsov, E.S. Pankratova, Toxicology analysis by means of the JSM-method, Bioinformatics 19 (2003) 1201–1207.
[3] J.P. Bordat, Calcul pratique du treillis de Galois d’une correspondence, Math. Sci. Hum. 96 (1986) 31–47.
[4] V.K. Finn, On machine-oriented formalization of plausible reasoning in the style of F. Backon–J.S. Mill, Semiotika Inform. 20 (1983) 35–101 (in Russian).
[5] V.K. Finn, Plausible reasoning in systems of JSM-type, Itogi Nauki Tekh., Inform. 15 (1991) 54–101 (in Russian).
[6] B. Ganter, Two basic algorithms in concept analysis, preprint 831, Technische Hochshule, Darmstadt, 1984.
[7] B. Ganter, R. Wille, Formal Concept Analysis: Mathematical Foundations, Springer, Berlin, 1999.

[8] B. Ganter, S.O. Kuznetsov, Formalizing Hypotheses with Concepts, in: G. Mineau, B. Ganter (Eds.), Proceedings of the Eigth International Conference on Conceptual Structures, ICCS'2000, Lecture Notes in Artificial Intelligence, Vol. 1867, Springer, Berlin, 2000, pp. 342–356.
[9] M. Garey, D. Johnson, Computers and Intractability: a Guide to the Theory of NP-Completeness, Freeman, San Francisco, 1979.
[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] D.S. Johnson, M. Yannakakis, C.H. Papadimitriou, On generating all maximal independent sets, Inform. Process. Lett. 27 (1988) 119–123.
[12] R. Karp, G.A. Hopcroft, 2.5-algorithm for maximum matching in bipartite graphs, SIAM J. Comput. 2 (2) (1973) 225–231.
[13] A.V. Karzanov, Private communication, Moscow, 1991.
[14] S.O. Kuznetsov, Learning of simple conceptual graphs from positive and negative examples, in: J. Zytkow, J. Rauch (Eds.), Proceedings Third European Conference on Principles of Data Mining and Knowledge Discovery, PKDD'99, Lecture Notes in Artificial Intelligence, Vol. 1704, Springer, Berlin, 1999, pp. 384–392.
[15] S.O. Kuznetsov, On computing the size of a lattice and related decision problems, Order 18 (4) (2001) 313–321.
[16] S.O. Kuznetsov, S.A. Obiedkov, Comparing performance of algorithms for generating concept lattices, J. Exp. Theoret. Artificial Intelligence 14 (2–3) (2002) 189–216.
[17] S.O. Kuznetsov, Machine learning and formal concept analysis, in: P. Eklund (Ed.), Proceedings Second International Conference on Formal Concept Analysis, Lecture Notes in Artificial Intelligence, 2961 (2004) 287–312.
[18] T. Mitchell, Machine Learning, McGraw-Hill, New York, 1997.
[19] E.M. Norris, An algorithm for computing the maximal rectangles in a binary relation, Rev. Roumaine Math. Pures Appl. 23 (2) (1978) 243–250.
[20] L. Nourine, O. Raynaud, A fast algorithm for building lattices, Inform Process. Lett. 71 (1999) 199–204.
[21] E. Prisner, Bicliques in graphs I: bounds on their number, Combinatorica 20 (1) (2000) 109–117.
[22] D. Schütt, Abschätzungen für die Anzahl der Begriffe von Kontexten, Diplomarbeit, TH Darmstadt, Darmstadt, 1988.
[23] L.G. Valiant, The complexity of enumeration and reliability problems, SIAM J. Comput. 8 (3) (1979) 410–421.
[24] D.B. West, Introduction to Graph Theory, Prentice-Hall, Upper Saddle River, 1996.
[25] 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.
[26] M. Yannakakis, Node-deletion problems of bipartite graphs, SIAM J. Comput. 10 (2) (1981) 310–327.