ISSN 0361-7688, Programming and Computer Software, 2012, Vol. 38, No. 5, pp. 219–230. © Pleiades Publishing, Ltd., 2012.
Original Russian Text © E.E. Vityaev, A.V. Demin, D.K. Ponomaryov, 2012, published in Programmirovanie, 2012, Vol. 38, No. 5.

Probabilistic Generalization of Formal Concepts

E. E. Vityaev*, A. V. Demin, and D. K. Ponomaryov****

**S.L. Sobolev Institute of Mathematics, Siberian Branch, Russian Academy of Sciences, pr. Akademika Koptyuga 4, Novosibirsk, 6300090 Russia*

***A.P. Ershov Institute of Information Systems, Siberian Branch, Russian Academy of Sciences, pr. Akademika Lavrent'eva 6, Novosibirsk, 630090 Russia*

e-mail: vityaev@math.nsc.ru, alexandredemin@yandex.ru, ponom@iis.nsk.su

Received December 14, 2011

Abstract—An inductive probabilistic approach to formal concept analysis (FCA) is proposed in which probability on formal contexts is considered; probabilistic formal concepts that have predictive force are defined: nonclassified objects can be assigned to earlier found probabilistic formal concepts; random attributes are eliminated from probabilistic formal concepts; probabilistic formal concepts are robust with respect to data noise. A result of experiment is presented in which formal concepts (in their standard definition in FCA) are first distorted by random noise and then recovered by detecting probabilistic formal concepts.

DOI: 10.1134/S0361768812050076

1. INTRODUCTION

In formal concept analysis (FCA) [1, 2], formal concepts are used as classification units that arise in relational data analysis. These kinds of data are presented as tables (formal context) over sets of objects G and attributes M in which the rows are labeled by the names of objects from G and the columns, by the names of attributes from M. In this case each cell (i, j) of a table contains the value 1 if and only if the ith object has attribute j. During the analysis, the objects are grouped into classes in such a way that objects that have some common set of attributes fall into one class and this set of objects is maximal: i.e., no other object outside this class has precisely this set of attributes. It is well known that the pairs <object group, attribute set> thus obtained can be ordered in a natural way and represented as a complete lattice. FCA is closely related to the studies on association rules, which have been the subject of intense investigation in the field of data mining. For example, it is known [3] that, having the set of all formal concepts of a given context, one can construct a basis for finding association rules in this context.

The goal of association rules is to find all the relationships between attributes in relational data. Originally this problem arose from the analysis of the basket of goods and from the problem of finding relationships between the sales of different groups of commodities. The dimension of real databases may be very large; therefore, the number of detectable rules may be very large, and is really such in practice. The selection of significant association rules is quite a nontrivial problem and is related to the problem of assessing the quality of rules [4]. One of approaches to solving this problem consists in determining the statistical estimates of association rules [4, 5].

If, in addition, we want that the rules have predictive force, then we fall into another paradigm of data mining methods—the paradigm of inductive inference of rules. As mentioned in [6], these two approaches (finding association rules and inductive inference of rules) have different goals. The problem of inductive inference of rules consists in providing predictions, while the problem of association rules consists in providing a data overview to a user.

The paradigm of FCA is related to association rules and consists in detecting a complete set of classification units and their relationships on relational data. Here the following questions that may arise in the inductive paradigm are skipped:

1. The predictive force of these concepts: the possibility of assigning a new object to a given concept.

2. The stability of a concept with respect to possible data errors.

3. Minimality of the description of concepts: elimination of random attributes from the description of a concept.

Today, there is no inductive paradigm for FCA. In the related paper [7], the main objects of FCA are reformulated in terms of probabilistic logic and are used for formulating new patterns. However, the definition of these objects within FCA remains unchanged.

The goal of the present study is to propose an inductive generalization of FCA and answer questions 1–3 formulated for the inductive paradigm. To this end, we consider two different types of proba-

I	m_1	m_2	m_3	m_4
g_1	×	×	×	
g_2	×	×		
g_3			×	
g_4				×

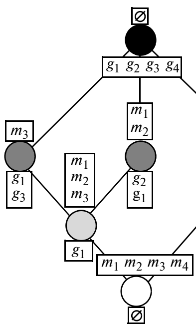

Fig. 1. Context and the corresponding lattice of concepts.

bility: probability on the families of formal contexts as sets of possible worlds (Section 3), and probability on the set of objects of the same context as a general population (Section 4). It is well known that, within formal logic, probability can be introduced by these two methods: on the set of possible worlds (models), and on the domain (the underlying sets of a specific model) [8]. To pass to the predictive definition of a formal concept, we take, as a basis, the definition of a formal concept in FCA in terms of fixed points of implications that are true on data. We generalize the concept of the truth of an implication on data, which is standard in FCA, by a certain truth estimate based on a probability measure in such a way that, first, consistency with the original definition of a formal concept is preserved and, second, these implications follow the idea of inductive rules—the minimization of the intent of concepts and elimination of random attributes—rather than the idea of association rules (finding all possible associations). After that, by analogy with the definition of a formal concept in FCA, we define a probabilistic formal concept in terms of the fixed points of probabilistic implications.

As a result, we obtain an inductive probabilistic definition of a formal concept that

• is related to the objects of some general population;

• coincides, under standard restrictions, with the original definition of a formal concept in FCA;

• has predictive force: new objects can be classified under earlier found probabilistic formal concepts;

• minimizes the description of a formal concept by eliminating random attributes from implications;

• is stable with respect to some types of noise in data.

In Section 2, we give all definitions and results of FCA that are necessary for reading the present article. To demonstrate the definitions, at the end of Section 4 we present examples of numerical experiments on the detection of probabilistic formal concepts on data.

2. PRELIMINARY DEFINITIONS

We begin with the main definitions and results of FCA.

Definition 1. *A formal context is a triple* (G, M, I), *where G and M are some sets and* $I \subseteq G \times M$ *is a relation between the elements of G and M. The elements of G are called objects of the context, and the elements of M are called attributes of the context. A formal context is said to be finite if G and M are finite sets.*

Henceforth, we omit, for brevity, the word "formal" and call the triples (G, M, I) mentioned in the definition contexts. Any context can be represented as a table, as we pointed out in the Introduction. If (G, M, I) is a context, then we define the operation ' on subsets $A \subseteq G$ and $B \subseteq M$ as follows:

$$A' = \{m \in M | \forall g \in A (g, m) \in I\},$$

$$B' = \{g \in G | \forall m \in B (g, m) \in I\}.$$

If $g \in G$, then the symbol g' serves as an abbreviation for the set $\{g\}'$.

Definition 2. *By a concept in the context* (G, M, I) *is meant a pair* (A, B), *where* $A \subseteq G$, $B \subseteq M$, $A' = B$, *and* $B' = A$. *Here the set A is called the extent and B, the intent of the concept* (A, B).

In fact, a concept is a classification unit that groups the objects and the attributes of the context.

In the proofs of the main propositions of this paper, we will repeatedly use the following simple fact:

Lemma 1. *The following relations are valid for any context* (G, M, I) *and sets* $B_1, B_2 \subseteq M$:

1. $B_1 \subseteq B_2 \Rightarrow B_2' \subseteq B_1'$,

2. $B_1 \subseteq B_1''$.

Definition 3. *Define a (partial) order* $\le$ *of the concepts of a context as follows: if* (A_1, B_1) *and* (A_2, B_2) *are concepts in a certain context, then we assume that* $(A_1, B_1) \le (A_2, B_2)$ *if* $A_1 \subseteq A_2$ *(or, which is equivalent by Lemma* 1, *if* $B_2 \subseteq B_1$).

Theorem. *The relation* $\le$ *induces a complete lattice on the set of concepts of a context, in which the infimum and supremum of the subsets are respectively defined as follows*:

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

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

Example 1. Consider a finite context $K = (\{g_1, g_2, g_3, g_4\}, \{m_1, m_2, m_3, m_4\}, I)$ represented in a tabular form in Fig. 1. The lattice of all concepts in the context K is also represented in this figure; each element of the lattice is marked by a set of objects and attributes, which are the extent and the intent of a concept, respectively.

The computation of the complete lattice of concepts by a given finite context [9, 10] is one of the key procedures in the FCA method. In fact, this procedure classifies the objects of the context according to appropriate attributes and enables one to find all the existing classes.

Given a context $K = (G, M, I)$, we can speak of the truth of assertions of the following type on K: "all objects possessing attributes $B_1 \subseteq M$ also possess a set of attributes $B_2 \subseteq M$." Since all properties of the context are in a sense symmetric with respect to the sets G and M, we can formulate a similar assertion for the subsets of G: "all attributes whose objects are $A_1 \subseteq G$ also have $A_2 \subseteq G$ as their objects." Without loss of generality, we will consider only assertions of the first type. In fact, these assertions define a monotonic operator, an implication, on the Boolean algebra of the subsets of M. It is clear that if a context K is finite, then the set of all such assertions that are true on K is also finite. Let us formalize the concept of an implication that is true on a context with the use of definitions of Section 2.3 in [1].

Definition 4. *An implication on a set M is an ordered pair of subsets A, $B \subseteq M$ denoted as $A \longrightarrow B$. The set A is called the premise, and B is called the conclusion of the implication $A \longrightarrow B$. A set $T \subseteq M$ is said to respect the implication $A \longrightarrow B$ if either $A \not\subseteq T$ or $B \subseteq T$. A family of subsets of M respects an implication if each set from this family respects this implication.*

If $K = (G, M, I)$ is a context, then the implication $A \longrightarrow B$ is *true* on K (which is denoted as $K \models A \longrightarrow B$) if A, $B \subseteq M$ and the family of sets $\{g' | g \in G\}$ respects $A \longrightarrow B$.

The premise of an implication $A \longrightarrow B$ is said to be *false* on K if there does not exist an $g \in G$ such that $A \subseteq g'$. An implication $A \longrightarrow B$ is called a *tautology* if $B \subseteq A$.

For a context $K = (G, M, I)$, we will denote the set of all implications on M that are true on the context K by $Imp(K)$. One can easily verify that the set of tautologies and the set of implications whose premise is false on K are subsets of $Imp(K)$. When this does not lead to confusion, we will use the same symbol $\models$ to denote the fact that a set or a family of sets respects a certain implication.

Any family L of implications on the set M generates a monotone operator $f_L: 2^M \longrightarrow 2^M$ given by

$$f_L(X) = X \cup \{B | A \longrightarrow B \in L, A \subseteq X\}.$$

It is clear that $f_L(X) = X \Leftrightarrow X \models L$ for any $X \subseteq M$.

Remark 1. *Let L be a family of implications on a set M. Then, for any $X \subseteq M$, there exists a minimal set $Y \subseteq M$ such that $X \subseteq Y$ and $f_L(Y) = Y$.*

Proof. Consider the standard inductive construction of the extensions of a given set $X \subseteq M$. Let $X_0 = X$ be the initial set. If a set X_i has been constructed, then we set $X_{i+1} = f_L(X_i)$. Then the sought set is $Y = \bigcup_{i \in \omega} X_i$.

Thus, any family of implications L on the set M defines an operator $\bar{f}_L: 2^M \longrightarrow 2^M$, which, for every $X \subseteq M$, gives a minimal set $Y \subseteq M$ that satisfies the conditions of the remark. It is obvious that the following relation holds for any set $X \subseteq M$: $f_L(X) = X \Leftrightarrow \bar{f}_L(X) = X$.

Remark 2. *If $K = (G, M, I)$ is a context and $A \longrightarrow B$ is an implication on M, then $K \models A \longrightarrow B \Leftrightarrow \forall m \in B$ $(K \models A \longrightarrow \{m\})$.*

In what follows, we will consider only implications of the form $A \longrightarrow \{m\}$ and denote them by $A \longrightarrow m$.

If K is a context in which a set of objects is compact (below we will assume that all the contexts considered are of this kind), then, for any implication $A \longrightarrow m \in Imp(K)$, there exists a set $\{A_0 \longrightarrow m \in Imp(K) | A_0 \subseteq A$ and, for any set $A_1 \subseteq A$, $A_1 \subset A_0$ implies $A_1 \longrightarrow m \notin Imp(K)\}$. For the context K, denote by $MinImp(K)$ the set of all implications of the form $A_0 \longrightarrow m$ that are true on K, in which the set A_0 is minimal in the above-indicated sense. Note that the definition of implications of this kind is a variant of the definition of implications as a law, given in [13–15].

Next, we give a proof of a slightly modified Proposition 20 from [1], which is central in this paper.

Proposition 1. *Let $K = (G, M, I)$ be a context, $T \subseteq Imp(K)$ be a set of tautologies on M, and $F \subseteq Imp(K)$ be a set of implications whose premises are false on K. Then the following is valid for any set $B \subseteq M$:*

1. $f_{MinImp(K)\backslash T}(B) = B \Leftrightarrow B'' = B$;
2. *if $B' \neq \varnothing$, then $f_{MinImp(K)\backslash\{F \cup T\}}(B) = B \Leftrightarrow B'' = B$.*

Proof. First, we show that, for any subset $B \subseteq M$, $f_{Imp(K)}(B) = B$ if and only if $f_{MinImp(K)}(B) = B$. Indeed, if $f_{Imp(K)}(B) \supset B$ for some B, then (with regard to Remark 2), there exists an implication $A \longrightarrow m \in Imp(K)$ such that $A \subseteq B$ but $m \notin B$. Then, there exists an implication $A \longrightarrow m \in MinImp(K)$, where $A_0 \subseteq A$ and therefore $A_0 \subseteq B$, $m \notin B$, and $f_{MinImp(K)}(B) \supset B$ a contradiction. The converse assertion is obvious because $MinImp(K) \subseteq Imp(K)$.

In a similar way, we can verify that $f_{MinImp(K)\backslash L}(B) = B \Leftrightarrow f_{Imp(K)\backslash L}(B) = B$, where either $L = T$ or $L = F \cup T$. This follows from the fact that, for any implication $A \longrightarrow m$ on M and any subset $A_0 \subseteq A$, the condition $A \longrightarrow m \notin T$ obviously implies $A_0 \longrightarrow m \notin T$, and the condition $A' \neq \varnothing$ implies $A_0' \neq \varnothing$ by Lemma 1. Therefore, we will prove assertions 1 and 2 of this proposition with respect to the set $Imp(K)$ rather than with respect to the set $MinImp(K)$.

1. $\Leftarrow$: Let $B'' = B$, $A_1 \longrightarrow A_2 \in Imp(K)\backslash T$, and $A_1 \subseteq B$. Let us show that then $A_2 \subseteq B$. Indeed, for any $g \in B'$, we have $g' \supseteq A_2$ because, by Lemma 1, $g' \supseteq B'' = B$ and the implication $A_1 \longrightarrow A_2$ is true on K. Hence, $\bigcap\{g'|(g \in B')\} \supseteq A_2$. On the other hand, $\bigcap\{g'|g \in B'\} = B''$; however, since $B'' = B$, we obtain $B \supseteq A_2$.

1. $\Rightarrow$: By Lemma 1, we have $B'' \supseteq B$ in any case; therefore, suppose that $f_{Imp(K)\backslash T}(B) = B$ but $B'' \not\subseteq B$. Then $B \not\models B \longrightarrow B'' \notin T$, and to arrive at a contradiction, it suffices to show that $B \longrightarrow B'' \in Imp(K)$.

(a) If $B' = \varnothing$, this is obviously true because in this case there does not exist a $g \in G$ such that $B \subseteq g'$; i.e., the premise of the implication is false on K.

(b) Let $B' \neq \varnothing$. We have to show that $\forall g \in G(B \subseteq g' \Rightarrow B'' \subseteq g')$. It is clear that $\forall g \in G\ (B \subseteq g' \Leftrightarrow g \in B')$ and, by Lemma 1, $\forall g \in B'\ (B'' \subseteq g')$. Thus, if $B \subseteq g'$ for some $g \in G$, then $B'' \subseteq g'$; i.e., $B \longrightarrow B'' \in Imp(K)$. Moreover, $B \longrightarrow B'' \in Imp(K)\backslash F$ because $B' \neq \varnothing$.

2. The sufficiency follows from the proof of item 1, because it is clear that if $f_{MinImp(K)\backslash T}(B) = B$, then $f_{MinImp(K)\backslash\{F \cup T\}}(B) = B$. The necessity follows from item (b) above.

For any context $K = (G, M, I)$ from Definition 2, it obviously follows that the subset $B \subseteq M$ is the intent of some concept in the context K if and only if $B'' = B$. Thus, given a context $K = (G, M, I)$, we have the set $Imp(K)$ of all implications that are true on K, and the fixed points of the operator $f_{MinImp(K)\backslash T} : 2^M \longrightarrow 2^M$ coincide with the intents of the concepts of the context K. If we exclude the set of those implications K whose premise is false on K from the implications of $MinImp(K)\backslash T$, then the fixed points of the operator $f_{MinImp(K)\backslash\{F \cup T\}}$: $2^M \longrightarrow 2^M$ coincide with the intents of the concepts of the context K except for the single concept $(\varnothing, M)$. This follows from the fact that, for any $B \subseteq M$, the condition $B'' \neq M$ obviously implies $B' \neq \varnothing$.

3. PROBABILISTIC CONCEPTS ON A CLASS OF CONTEXTS

Above we have defined the concept of the truth of an implication on a certain individual context. Let us describe how to generalize this concept by evaluating the truth of an implication on a class of contexts. Let us describe the ideas of the method of semantic probabilistic inference as applied to FCA. Within this method, which is presented in [13–15], regularities on data (in particular, implications) are formalized as universal formulas of the language of first-order logic of a countable signature consisting of predicates and constants. Thus, the standard notion of implication defined in [1] turns out to be far more specific than the concept of regularity on data considered in semantic probabilistic inference (note that implications that go far beyond the definitions given in [1] were also considered in papers on FCA). However, to demonstrate that this method can be applied to FCA, it is convenient to stay within the standard algebraic definitions. Therefore, below we present a restriction of the method of semantic probabilistic inference in terms close to those used in FCA.

Definition 5. *A class of contexts over sets G and M is a family* $\mathcal{K} = \{(G, M, I_j)\}_{j \in J \neq \varnothing}$, *where, for every* $j \in J$, *the triple* (G, M, I_j) *is a context. We use the notation* $\mathcal{K}(G, M)$ *for a class* $\mathcal{K}$ *of contexts over the sets G and M. A probability model of type I is a pair* $\mathcal{M} = (\mathcal{K}(G, M), \rho)$, *where* $G \neq \varnothing$ *and* ρ *is a probability measure on the set* $\mathcal{K}$, *that satisfies the condition*

$$\forall S_1, S_2 \subseteq M\ \forall (G, M, I) \in \mathcal{K}(S_1 \not\subseteq I \text{ or } S_2 \subseteq I)$$
$$\Leftrightarrow$$
$$\rho(\{G, M, I_j\} | S_1 \cup S_2 \subseteq I_j)$$
$$= \rho(\{(G, M, I_j) | S_1 \subseteq I_j\}).$$

If $S \subseteq G \times M$, *then we call the value of the function* $\nu_{\mathcal{M}}(S) = \rho(\{(G, M, I) \in \mathcal{K} | S \subseteq I\})$ *the probability of the set S on* $\mathcal{M}$.

In this section, for brevity, we call the pairs $(\mathcal{K}(G, M), \rho)$ from Definition 5 above *probability models* or, simply, *models*.

Let $\mathcal{M} = (\mathcal{K}(G, M), \rho)$ be a probability model and $A \longrightarrow m$ be an implication on the set M. A pair $\langle g, A \longrightarrow m\rangle$, where $g \in G$, is called an instantiation of the implication $A \longrightarrow m$ on the model $\mathcal{M}$. The value of the function $\mu_{\mathcal{M}}(\langle g, A \longrightarrow m\rangle) =$

$$\begin{cases} \dfrac{\nu_{\mathcal{M}}(S \cup \{\langle g, m\rangle\})}{\nu_{\mathcal{M}}(S)}, & \text{if } \nu_{\mathcal{M}}(S) \neq 0, \\ & \text{where } S = \{\langle g, a\rangle | a \in A\} \\ & \text{undefined otherwise.} \end{cases}$$

The value of the function $\eta_{\mathcal{M}}(A \longrightarrow m) =$

$$\begin{cases} \text{undefined,} & \text{if } \forall g \in G \\ & \mu_{\mathcal{M}}(\langle (g, A) \longrightarrow m\rangle) \\ & \text{undefined} \\ \inf\limits_{g \in G} \mu_{\mathcal{M}}(\langle g, A \longrightarrow m\rangle) & \text{otherwise.} \end{cases}$$

Thus, we adopt a cautious strategy to estimate the probabilities of implications—a strategy involving the least value of probability from among the instantiations.

Remark 3. *Let* $\mathcal{M} = (\mathcal{K}(G, M), \rho)$ *be a probability model and* $A \longrightarrow m$ *be an implication on the set M whose probability on the model* $\mathcal{M}$ *is defined. Then* $\eta_{\mathcal{M}}(A \longrightarrow m) = 1$ *if and only if* $\forall K \in \mathcal{K}(A \longrightarrow m \in Imp(K))$.

Proof. $\Rightarrow$: The condition $\eta_{\mathcal{M}}(A \longrightarrow m) = 1$ implies that, for every $g \in G$, the value of $\mu_{\mathcal{M}}(\langle g, A \longrightarrow m\rangle)$ is either undefined or equal to one. Therefore, for every $g \in G$ and every context $K \in \mathcal{K}$, either $A \not\subseteq g'$ or $m \in g'$ by the definition of the functions $\mu_{\mathcal{M}}$ and ρ. But this means that $\forall K \in \mathcal{K}$ $(A \longrightarrow m \in Imp(K))$.

$\Leftarrow$: Suppose that $\eta_{\mathcal{M}}(A \longrightarrow m) < 1$. Then there exists a $g \in G$ such that the value of $\mu_{\mathcal{M}}(\langle g, A \longrightarrow m\rangle)$ is defined and is also strictly less than one. In turn, there exists a $K \in \mathcal{K}$ in which $A \subseteq g'$ but $m \notin g'$; but this means that $A \longrightarrow m \notin Imp(K)$.

Definition 6. *Let $\mathcal{M} = (\mathcal{K}(G, M)\rho)$ be a probability model and $imp(M)$ be a set of those implications on M whose probability is defined on $\mathcal{M}$. Implications $A \longrightarrow m \in imp(M)$ such that*

- $\eta_{\mathcal{M}}(A \longrightarrow m) \neq 0$ *and*
- *if $A_0 \longrightarrow m \in imp(M)$ and $A_0 \subset A$, then $\eta_{\mathcal{M}}(A_0 \longrightarrow m) < \eta_{\mathcal{M}}(A \longrightarrow m)$ are called probability regularities (probability laws* [13–15]) *on $\mathcal{M}$.*

An implication $A \longrightarrow m \in imp(M)$ is called a maximally specific probability regularity (maximally specific law [13, 14]) *on $\mathcal{M}$ if it is a probability regularity on $\mathcal{M}$, $A \neq \{m\}$ and there does not exist a probability law $A_0 \longrightarrow m$ on $\mathcal{M}$ such that $A \subset A_0$ and $A_0 \longrightarrow m$ is not a tautology.*

Note that the fundamental difference of the probability regularities with probability 1 from the well-known Duequenne–Guigues basis [11] consists in the necessity to ensure the predictive force by the second condition in Definition 6, which is missing in the implications of the Duequenne–Guigues basis. This condition guarantees a strict increase in the conditional probability under the extension of the premise of an implication, which increases its predictive force and precludes the possibility that random attributes that do not increase the probability of prediction fall into the predictive force.

Remark 4. *If an implication is a maximally specific probabilistic law on the model $\mathcal{M}$, then it is not a tautology.*

Definition 7. *A semantic probabilistic inference is a sequence of probabilistic laws $(A_0 \longrightarrow m)$, ..., $(A_n \longrightarrow m)$ such that $A_0 \subset \ldots \subset A_n$, $\eta_{\mathcal{M}}(A_0 \longrightarrow m) < \ldots < \eta_{\mathcal{M}}(A_n \longrightarrow m)$, and $A_n \longrightarrow m$ is a maximally specific probabilistic law.*

The concept of semantic probabilistic inference maximally expresses the inductive essence of the definitions. In addition to the inductive properties that are inherent in probabilistic laws, semantic probabilistic inference implies a directed choice of attributes that substantially increase the prediction probability in the conclusion. Therefore, only specific information significant for predicting is included in the premise of an implication. Note that association rules do not have such properties. In addition, we can argue that logic, probability, and learning are combined in this way [12].

Definition 8. *Let $\mathcal{M} = (\mathcal{K}(G, M), \rho)$ be a probability model and $S(\mathcal{M})$ be the set of all maximally specific probability laws on $\mathcal{M}$. An implication $A \longrightarrow m \in S(\mathcal{M})$ is called the strongest probability law on $\mathcal{M}$ if the value of its probability on $\mathcal{M}$ is maximal among all the implications $B \longrightarrow m \in S(\mathcal{M})$.*

We will use the notation $D(\mathcal{M})$ for the set of all strongest probability laws on the model $\mathcal{M}$.

Note that, in view of the rather arbitrary form of the function ρ in the definition of a probability model, there is nothing to guarantee the existence of a maximum in the sense of Definition 8 and, hence, the existence of the strongest probability laws themselves. However, below we give a method for defining a probability model (on the basis of a finite class of finite contexts) that gives rise to a wide class of models that guarantee the presence of such implications. Note that, in the general case, for an arbitrary m, there may exist several implications of the form $A \longrightarrow m$ that are strongest probabilistic laws.

Informally, every implication on a probability model should be considered as a prediction with some assessment of the truth of the fact that each object possessing a set of attributes from the premise will also possess an attribute from the conclusion of the implication. Just as in FCA (recall Proposition 1), implications in the method of semantic probabilistic inference are directly related to the grouping of objects and attributes into classification units. If data are represented as a class of contexts $\mathcal{K}$, then the type of implications selected among all possible implications on the probability model $(\mathcal{K}, \rho)$ determines the formation of the classes themselves on the basis of the data presented. Minimal implications in the sense of the set $MinImp(K)$, as well as probability laws (maximally specific and the strongest probability laws) are an adaptation of appropriate probability definitions from [13–16] as applied to the FCA method. Such implications possess a number of theoretical and practically useful properties that justify their application:

- the set of all minimal implications that are true on every context from a class $\mathcal{K}$ yields, in a sense, an axiomatization of this class of contexts: from this class, one can semantically derive an implication theory of $\mathcal{K}$ that is restricted to implications with nonfalse premises [13, 15] (an analog of the Duequenne–Guigues theorem on implication base [11]);
- a probabilistic law precludes the prediction of an attribute in its conclusion by a certain proper subset of attributes of the premise with probability greater than (or equal to) that of the law itself; together with the requirement of maximal specificity, this leads in practice to grouping attributes into smaller classes, with maximum probability [12];
- it is proved in [13, 14] that if implications admit negative information, then maximally specific probabilistic laws form a consistent set of propositions (i.e., there is no situation where the presence or absence of some attribute is predicted simultaneously);
- the strongest probabilistic laws lead to assigning an attribute to the class that predicts it with maximum probability; at the same time, a situation is possible

when the same attribute may belong to different classes [17];

• a Discovery software tool is implemented that allows one to find the above-mentioned types of implications on tabular data and construct the corresponding object–attribute classes. This software has been successfully applied to solving a large number of applied problems [18, 13, 16].

Definition 9. *Let* $\mathcal{M} = (\mathcal{K}(G, M), \rho)$ *be a probability model of type I. A pair of sets* (A, B) *is called a probabilistic concept* of *a context* $(G, M, I) \in \mathcal{K}$ *in the model* $\mathcal{M}$ *if it satisfies the following conditions*:

• $A \subseteq G$, $B \subseteq M$,

• $f_{D(\mathcal{M})}(B) = B$,

• $\exists E \subseteq B\ (\bar{f}_{D(\mathcal{M})}(E) = B \text{ and } E \neq \varnothing \neq E')$,

• $A = \bigcup\{E' \mid \varnothing \neq E \subseteq B,\ \bar{f}_{D(\mathcal{M})}(E) = B\}$, *where* ' *denotes an operation within the context* (G, M, I). *The set* A *is called the extent and* B, *the intent of the probabilistic concept* (A, B).

Thus, given a probability model $\mathcal{M} = (\mathcal{K}(G, M), \rho)$, the set of fixed points of the operator $f_{D(\mathcal{M})}$ restricts the set of all possible probabilistic concepts of contexts from the class $\mathcal{K}$ in the model $\mathcal{M}$.

Theorem 1. *Consider a context* $K = (\varnothing \neq G, M, I)$ *and a probability model* $\mathcal{M} = (\{K\}, \rho)$. *Then, for any nonempty subsets* $A \subseteq G$ *and* $B \subseteq M$, *the pair* (A, B) *is a concept in the context* K *if and only if* (A, B) *is a probabilistic concept of the context* K *in the model* $\mathcal{M}$.

Proof. Let $S \subseteq Imp(K)$ be the set consisting of all tautologies on M and all implications whose premises are false on the context K. Let us show that $MinImp(K)\backslash S = D(\mathcal{M})$.

$\subseteq$: Consider an arbitrary implication $A \longrightarrow m \in MinImp(K)\backslash S$. By the definition of the model $\mathcal{M}$, for any subset $S \subseteq G \times M$, we have $\rho(S) = 0 \Leftrightarrow S \not\subseteq I$. Since the premise A is not false on K, we find that the probability of the implication $A \longrightarrow m$ on the model $\mathcal{M}$ is defined; then, according to Remark 3, we have $\eta_{\mathcal{M}}(A \longrightarrow m) = 1$. In view of the minimality of the premise A, any implication $A_0 \longrightarrow m$, where $A_0 \subset A$, is no longer true on K. Moreover, since the premise A is not false on K, A_0 is not false on K either. According to Remark 3, then $\eta_{\mathcal{M}}(A \longrightarrow m) = 0$, and thus the implication $A \longrightarrow m$ is a probabilistic law on $\mathcal{M}$. Taking into account that $\eta_{\mathcal{M}}(A \longrightarrow m) = 1$ and $m \notin A$, hence we find that $A \longrightarrow m \in D(\mathcal{M})$.

$\supseteq$: By the definition of the model $\mathcal{M}$, we have $\forall S \subseteq G \times M\ \rho(S) \in \{0, 1\}$; hence, for any implication $A \longrightarrow m \in D(\mathcal{M})$, the definition of a probabilistic law implies $\eta_{\mathcal{M}}(A \longrightarrow m) = 1$. By the definition of the function $\mu_{\mathcal{M}}$, then we find that the premise A is not false on K; therefore, with regard to Remarks 3 and 4, we obtain $A \longrightarrow m \in Imp(K)\backslash S$. Suppose that there exist an implication $A_0 \longrightarrow m \in Imp(K)$ such that $A_0 \subset A$. Then $A_0 \longrightarrow m \in Imp(K)\backslash S$, $\eta_{\mathcal{M}}(A_0 \longrightarrow m) = 1$, and we arrive at a contradiction to the fact that $A \longrightarrow m$ is a probabilistic law on $\mathcal{M}$. Hence, $A \longrightarrow m \in MinImp(K)\backslash S$.

Suppose that (A, B) is a probabilistic concept of the context K in the model $\mathcal{M}$. Let us show that (A, B) is a concept in the context K. To this end, it suffices to verify that $A' = B$ and $B' = A$. Consider the set $\mathcal{C} = \{E \subseteq B \mid \bar{f}_{D(\mathcal{M})}(E) = B,\ E \neq \varnothing \neq E'\}$. By the definition of a probabilistic concept, this set is nonempty. By virtue of $\bar{f}_{D(\mathcal{M})}(E) = B$ and what has been proved above, for every $E \in \mathcal{C}$, there exists an implication $E \longrightarrow B \in Imp(K)$; therefore, $B' \neq \varnothing$, and, taking into account that $f_{D\{\mathcal{M}\}}(B) = B$, from item 2 of Proposition 1 we find that $B'' = B$. In addition, it follows from $E \longrightarrow B \in Imp(K)$ that $g' \supseteq B$ for every $g \in E'$. This means that $g' \supseteq B$ for every $g \in \bigcup\{E' \mid E \in \mathcal{C}\} = A$; therefore, $A \subseteq B'$. On the other hand, for every $E \in \mathcal{C}$, from the condition $E \subseteq B$ we obtain $B' \subseteq E'$; therefore, $B' \subseteq \bigcup\{E' \mid E \in \mathcal{C}\} = A$. Thus, we have $A = B'$, which, combined with $B'' = B$, yields $A' = B$.

Let (A, B) be a concept in the context K, and suppose that the sets A and B are nonempty. Let us verify that (A, B) is a probabilistic concept of the context K in the model $\mathcal{M}$. Indeed, since $A \neq \varnothing$ and $B' = A$, we have $B' \neq \varnothing$, and, since $B'' = B$, by item 2 of Proposition 1, in view of what has been proved above, we obtain $f_{D(\mathcal{M})}(B) = B$. It remains to verify that $A = \bigcup\{E' \mid E \in \mathcal{C}\}$, where $\mathcal{C} = \{E \subseteq B \mid E \neq \varnothing,\ \bar{f}_{D(\mathcal{M})}(E) = B\}$, because it is obvious that $B \in \mathcal{C}$. We have $\bigcup\{E' \mid E \in \mathcal{C}\} \supseteq B' = A$. Conversely, if $g \in \bigcup\{E' \mid E \in \mathcal{C}\}$, then there exists an $E \in \mathcal{C}$ such that $g \in E'$ and thus $g' \supseteq E$. In view of $\bar{f}_{D(\mathcal{M})}(E) = B$, we have $E \longrightarrow B \in Imp(K)$; therefore, $g' \supseteq B$, and hence $g \in B' = A$. Thus, all the conditions in the definition of a probabilistic concept are satisfied.

Let $\mathcal{K} = \{(\varnothing \neq G, M, I_j)\}_{j \in J \neq \varnothing}$ be a finite class consisting of finite contexts. Let us show a natural method for defining a probability model $(\mathcal{K}, \rho)$ on the class $\mathcal{K}$. For each context $K \in \mathcal{K}$, we set $\rho(\{K\}) = 1/|J|$ and, for a subset $\mathcal{C} \subseteq \mathcal{K}$, define $\rho(\mathcal{C}) = \sum_{K \in \mathcal{C}} \rho(\{K\})$.

Then ρ is a discrete probability measure on $\mathcal{K}$, and, for every $S \subseteq G \times M$, we have $\nu_{\mathcal{M}}(S) = |\tilde{J}|/|J|$, where $\tilde{J}$ is a maximal subset of J satisfying the condition $\forall j \in \tilde{J}$ $(S \subseteq I_j)$. It is easy to verify that then $(\mathcal{K}, \rho)$ is indeed a probability model. We call the model $\mathcal{M}$ defined in this way a *frequency probability model*.

Let us illustrate the definitions given above.

Example 2. *Suppose given sets* $G = \{g_1, g_2\}$ *and* $M = \{m_1, m_2, m_3\}$. *Consider a class* $\mathcal{K} = \{(G, M, I_j)\}_{j \in \{1, 2, 3\}}$ *consisting of three contexts given in the tabular form below*:

I_1	m_1	m_2	m_3
g_1	×		×
g_2		×	

I_2	m_1	m_2	m_3
g_1		×	×
g_2	×	×	

I_3	m_1	m_2	m_3
g_1	×		×
g_2	×	×	

Then the pairs ($\{g_1\}$, $\{m_1, m_2, m_3\}$) *and* ($\{g_1, g_2\}$, $\{m_1, m_2\}$) *are the only probabilistic concepts of the context* (G, M, I_1) *in the frequency probability model* $\mathcal{M} = (\mathcal{K}, \rho)$.

Proof. The probability measure ρ uniquely defines the value $\eta_{\mathcal{M}}(A \longrightarrow m)$ for every implication $A \longrightarrow m$ on the set M. In the tables below, we give the probabilities of all possible implications of the form $A \longrightarrow m$ on $\mathcal{M}$ that are not tautologies.

$A \to m$	$\eta_{\mathcal{M}}(A \to m)$
$\{\varnothing\} \to m_1$	2/3
$m_2 \to m_1$	0
$m_3 \to m_1$	2/3
$m_2, m_3 \to m_1$	0
$\{\varnothing\} \to m_2$	1/3
$m_1 \to m_2$	0
$m_3 \to m_2$	1/3
$m_1, m_3 \to m_2$	0
$\{\varnothing\} \to m_3$	0
$m_1 \to m_3$	0
$m_2 \to m_3$	0
$m_1, m_2 \to m_3$	0

The premises of implications that form the set $D(\mathcal{M})$ of all the strongest probabilistic laws on $\mathcal{M}$ are given in curly brackets. Let us give an example of computing the probability of one of the implications from the table above:

$$\eta_{\mathcal{M}}(m_3 \longrightarrow m_1) = \inf_{g \in G} \mu_{\mathcal{M}}(\langle g, m_3 \longrightarrow m_1 \rangle)$$
$$= \inf_{g \in G} \frac{\nu_{\mathcal{M}}(\{\langle g, m_3\rangle, \langle g, m_1\rangle\})}{\nu_{\mathcal{M}}(\{\langle g, m_3\rangle\})}$$
$$= \frac{\nu_{\mathcal{M}}(\{\langle g_1, m_3\rangle, \langle g_1, m_1\rangle\})}{\nu_{\mathcal{M}}(\{\langle g_1, m_3\rangle\})} = 2/3,$$

because the value of $\mu_{\mathcal{M}}(\langle g_2, m_3 \longrightarrow m_1\rangle)$ is undefined due to $\nu_{\mathcal{M}}(\langle g_2, m_3\rangle) = 0$. Note, however, that the implication $m_3 \longrightarrow m_1$ is not a probabilistic law, because there exists an implication $\varnothing \longrightarrow m_1$ with the same probability on $\mathcal{M}$.

Let us give the values of the operator $f_{D(\mathcal{M})}$ on the subsets $B \subseteq M$:

$B \subseteq M$	$f_{D(\mathcal{M})}(B)$
m_1	m_1, m_2
m_2	m_1, m_2
m_3	m_1, m_2, m_3
m_1, m_2	m_1, m_2
m_1, m_3	m_1, m_2, m_3
m_2, m_3	m_1, m_2, m_3
m_1, m_2, m_3	m_1, m_2, m_3
$\{\varnothing\}$	m_1, m_2

Obviously, there are exactly two subsets $B \subseteq M$ that satisfy the condition $f_{D(\mathcal{M})}(B) = B$, namely, $\{m_1, m_2\}$ and $\{m_1, m_2, m_3\}$. Finally, we have

$$\bigcup \{E' \mid \varnothing \neq E \subseteq \{m_1, m_2\}, \bar{f}_{D(\mathcal{M})}(E) = \{m_1, m_2\}\}$$
$$= \{g_1, g_2\},$$
$$\bigcup \{E' \mid \varnothing \neq E \subseteq \{m_1, m_2, m_3\}, \bar{f}_{D(\mathcal{M})}(E)$$
$$= \{m_1, m_2, m_3\}\} = \{g_1\}.$$

The only subset $E \subseteq \{m_1, m_2, m_3\}$ that satisfies the conditions in the definition of a probabilistic concept is the set $\{m_3\}$, for which we have $\{m_3\}' = g_1$.

Thus, ($\{g_1\}$, $\{m_1, m_2, m_3\}$) and ($\{g_1, g_2\}$, $\{m_1, m_2\}$) are the only probabilistic concepts of the context (G, M, I_1) in the model $\mathcal{M}$. □

4. PROBABILISTIC CONCEPTS ON ONE CONTEXT

In Section 3, we considered the notion of a probability model of type I defined on a class of contexts. In fact, any class $\mathcal{K}$ of contexts that admits the definition of a probability measure generates a set of probability models and thus defines possible families of implications that are strongest probabilistic laws. Using such families of implications, we predicted the existence of attributes in objects in an arbitrarily chosen context from the class $\mathcal{K}$. Similar to this approach, we can define the strongest probabilistic laws on the basis of a single given formal context. To this end, we merely need to slightly modify Definition 5 of a probability model.

Definition 10. A *probability model of type II* (*a probabilistic context*) *is a pair* $\mathcal{M} = (K, \rho)$, *where* $\mathcal{K} = (G, M, I)$ *is a context and* ρ *is a probability measure on the set* G, *that satisfies the condition*

$$\forall B, C \subseteq M \ (B' \subseteq C' \Leftrightarrow \rho((B \cup C)') = \rho(B')).$$

If $B \longrightarrow m$ is an implication on the set M, then its probability on the model $\mathcal{M}$ is the value of the function

$$\eta_{\mathcal{M}}(B \longrightarrow m) = \begin{cases} \dfrac{\rho((B \cup \{m\})')}{\rho(B')}, & \text{if } \rho(B') \neq 0 \\ \text{undefined} & \text{otherwise.} \end{cases}$$

For brevity, in this section we will call the pairs (K, ρ) from the definition above *probability models*, or, simply, *models*.

If $K = (\varnothing \neq G, M, I)$ is a finite context, then we call a model a *frequency probability model* if, for every $g \in G$, we have $\rho(\{g\}) = 1/|G|$ and, for every subset $A \subseteq G$, we have $\rho(A) = \sum_{g \in A} \rho(\{g\})$. Thus, $\forall B \subseteq M$ ($\rho(B') = |B'|/|G|$). We stress that $\mathcal{M}$ is indeed a model because the equality $B' \subseteq C' \Leftrightarrow (B \cup C)' = B' \Leftrightarrow |(B \cup C)'| = |B'|$ holds for any subsets $B, C \subseteq M$.

Remark 5. *For any probability model $\mathcal{M} = (K, \rho)$, where $K = (G, M, I)$, and any implication $B \longrightarrow m$ on the set M, we have $\eta_{\mathcal{M}}(B \longrightarrow m) = 1$ if and only if $B \longrightarrow m \in Imp(K)$ and $B' \neq \varnothing$ (where ' is an operation within the context K).*

Proof. If $\eta_{\mathcal{M}}(B \longrightarrow m) = 1$, then $\rho(B') \neq \varnothing$, and hence $B' \neq \varnothing$; i.e., the premise B is not false on K. On the other hand, this condition implies that $\rho((B \cup \{m\})') = \rho(B')$; therefore, we have $B' \subseteq \{m\}'$, which is equivalent to $B \longrightarrow m \in Imp(K)$. This argument also proves the converse proposition.

Let us define the notions of a probabilistic law, maximally specific probabilistic law, and the strongest probabilistic law on a model of type II in complete analogy with Definitions 6 and 8. We use the same notation $D(\mathcal{M})$ for the set of all strongest probabilistic laws on a model $\mathcal{M}$ of type II as that used in Section 3.

Proposition 2. *Let $\mathcal{M} = (K, \rho)$ be a probability model, where $K = (G, M, I)$, and $S \subseteq Imp(K)$ be the set consisting of all tautologies on M and all implications whose premise is false on K. Then we have $MinImp(K) \backslash S \subseteq D(\mathcal{M})$.*

Proof. Indeed, $B' \neq \varnothing$ for every implication $B \longrightarrow m \in MinImp(K)\backslash S$; therefore, in view of Remark 5, we have $\eta_{\mathcal{M}}(B \longrightarrow m) = 1$. The maximality condition of the value of probability for the implication $B \longrightarrow m$ on the model $\mathcal{M}$ is satisfied; hence, there cannot exist a probabilistic law $B_1 \longrightarrow m$ on $\mathcal{M}$ such that $B \subset B_1$. Moreover, the implication $B \longrightarrow m$ itself is a probabilistic law, because the condition $B \longrightarrow m \in MinImp(K)\backslash S$ and Remark 5 imply that $\eta_{\mathcal{M}}(B_0 \longrightarrow m) < 1$ for any set $B_0 \subset B$. Thus, all the conditions in the definition of the strongest probabilistic law are satisfied, and $B \longrightarrow m \in D(\mathcal{M})$.

Definition 11. *Let $\mathcal{M} = (K, \rho)$ be a probability model of type* II, *where $K = (G, M, I)$. A pair of sets (A, B) is called a probabilistic concept in the model $\mathcal{M}$* (a *concept in the probabilistic context $\mathcal{M}$*) *if it satisfies the conditions of Definition* 9.

Let $\mathcal{M} = (K, \rho)$ be a probability model, where $K = (G, M, I)$. Consider a context $\bar{K} = (G, M, \bar{I})$, where $\bar{I} = \{\langle g, m\rangle \mid g \in G, m \in \bar{f}_{D(\mathcal{M})}(g')\}$ and ' is an operation within the context K. In other words, $I \subseteq \bar{I}$, and the ratio $\bar{I}$ is obtained from the original I by adding pairs $\langle g, m\rangle$ predicted by the family of implications $D(\mathcal{M})$. To clarify the relationship between the concepts in the context K and probabilistic concepts in the model $\mathcal{M}$, it is important to note that *the following assertion is false in both directions*:

for any nonempty subsets $A \subseteq G$ and $B \subseteq M$, the pair (A, B) is a probabilistic concept in the model $\mathcal{M}$ if and only if (A, B) is a concept in the context $\bar{K}$.

To prove this, it suffices to consider any of the contexts $K_1 = (\{g_1, g_2\}, \{m_1\}, I_1)$ and $K_2 = (\{g_1, g_2\}, \{m_1, m_2, m_3\}, I_2)$ given below, together with the corresponding frequency probability models $\mathcal{M}_1 = (K_1, \rho_1)$ and $\mathcal{M}_2 = (K_2, \rho_2)$.

I_1	m_1
g_1	×
g_2	

I_2	m_1	m_2	m_3
g_1	×		
g_2		×	×

For these models, we have $D(\mathcal{M}_1) = \{\varnothing \longrightarrow m_1\}$ and $D(\mathcal{M}_2) = \{f \longrightarrow m_1, \{m_2\} \longrightarrow m_3, \{m_3\} \longrightarrow m_2\}$. Therefore, the set of all probabilistic concepts in the model $\mathcal{M}_1$ consists of a single concept $(\{g_1\}, \{m_1\})$, and the set $\{(\{g_1\}, \{m_1\}), (\{g_2\}, \{m_1, m_2, m_3\})\}$ represents all the probabilistic concepts in the model $\mathcal{M}_2$.

We can easily verify that, for every $j = 1, 2$, the context $\bar{K}_j$ is obtained from K_j by setting $\bar{I}_j = I_j \cup \{\langle g_2, m_1\rangle\}$. It remains to note that the set of all concepts in the context $\bar{K}_1$ consists of a single pair $(\{g_1, g_2\}, \{m_1\})$, while the set $\{(\{g_1, g_2\}, \{m_1\})\}$ represents all the concepts in the context $\bar{K}_2$.

Nevertheless, we can guarantee the following property, which characterizes the relationship between the concepts in the context K and the probabilistic concepts in the model $\mathcal{M} = (K, \rho)$:

Theorem 2. *The following properties are valid for any probability model $\mathcal{M} = (K, \rho)$, where $K = (G, M, I)$:*

1. if (A, B) is a concept in the context K with $A \neq \varnothing \neq B$, then there exists a probabilistic concept (A_1, B_1) in the model $\mathcal{M}$ such that $A \subseteq A_1$ and $B \subseteq B_1$;

2. if (A_1, B_1) is a probabilistic concept in the model $\mathcal{M}$, then there exists a concept (A, B) in the context K such that $\varnothing \neq A \subseteq A_1$ and $\varnothing \neq B \subseteq B_1$. Moreover, the set A_1 is a union of the extents of some of these concepts.

Proof. 1. Let $S \subseteq Imp(K)$ be the set consisting of all tautologies on M and all implications whose premise is false on K. Since (A, B) is a concept in the context K, we have $B'' = B$ and $B' = A \neq \varnothing$, and, by Proposition 1, we find that $f_{MinImp(K)\backslash S}(B) = B$.

By Proposition 2, the inclusion $MinImp(K)\backslash S \subseteq D(\mathcal{M})$ is valid. Moreover, for any families of implications L_1 and L_2 on the set M and any subset $B \subseteq M$, $L_1 \subseteq L_2$ implies $\bar{f}_{L_1}(B) \subseteq \bar{f}_{L_2}(B)$; therefore, $B \subseteq \bar{f}_{D(\mathcal{M})}(B)$. Denote $B_1 = \bar{f}_{D(\mathcal{M})}(B)$, $\mathcal{C} = \{E \subseteq B_1 | \bar{f}_{D(\mathcal{M})}(E) = B_1, E \neq \varnothing \neq E'\}$, and $A_1 = \cup\{E' | E \in \mathcal{C}\}$. It is clear that $\bar{f}_{D(\mathcal{M})}(B_1) = B_1$. In addition, notice that $B \in \mathcal{C}$, $A = B'$, and $B' \subseteq A_1$. Therefore, we have $A \subseteq A_1$; thus, (A_1, B_1) is the sought probabilistic concept in the model $\mathcal{M}$.

2. Consider the set $\mathcal{C} = \{E \subseteq B_1 | \bar{f}_{D(\mathcal{M})}(E) = B_1, E \neq \varnothing \neq E'\}$ and an arbitrary $E \in \mathcal{C}$. We have $MinImp(K)\backslash S \subseteq D(\mathcal{M})$; therefore, $\bar{f}_{MinImp(K)\backslash S}(E) \subseteq B_1$. Denote $B = \bar{f}_{MinImp(K)\backslash S}(E)$; it is clear that $D(\mathcal{M})$; therefore, $\bar{f}_{MinImp(K)\backslash S}(E) \subseteq B_1$. Denote $B = \bar{f}_{MinImp(K)\backslash S}(B) = B$. Moreover, it follows from $E \neq \varnothing \neq E'$ that $B \neq \varnothing \neq B'$; therefore, by Proposition 1, we obtain $B'' = B$. On the other hand, $E \subseteq B$; therefore, $E' \supseteq B'$ and $A_1 = \cup\{E' | E \in \mathcal{C}\} \supseteq B'$. We find that (B', B) is the sought concept in the context K.

Now, notice that the condition $B = \bar{f}_{MinImp(K)\backslash S}(E)$ implies $E \longrightarrow B \in Imp(K)$, which is equivalent to $E' \subseteq B'$; therefore, we obtain $E' = B'$. In view of arbitrary choice of the set $E \in \mathcal{C}$ and the fact that $A_1 = \cup\{E' | E \in \mathcal{C}\}$, we conclude that the set A_1 is a union of the extents of some concepts (A, B) in the context K such that $\varnothing \neq B \subseteq B_1$.

Below, we give computational schemes for finding probabilistic laws and probabilistic concepts for a given frequency probability model $\mathcal{M} = (K, \rho)$, where $K = (G, M, I)$.

Let $S \subseteq Imp(K)$ be the set consisting of all tautologies on M and all the implications whose premise is false on K. Note that, for a given context K, the cardinality of the set $MinImp(K)\backslash S$ may exponentially depend on the value of $|G| \times |M|$. This follows from Theorem 1 in [19], where an example of constructing such a context is given. By Proposition 2, we have $MinImp(K)\backslash S \subseteq D(\mathcal{M})$, and the set of all probabilistic laws on $\mathcal{M}$ contains $D(\mathcal{M})$ by definition. Therefore, the procedure of finding probabilistic laws is based on heuristics.

Let us introduce a few auxiliary definitions. The length of an implication $A \longrightarrow m$ on a set M is the cardinality of its premise, i.e., the cardinality of the set A; we will denote it by $len(A \longrightarrow m)$. We say that an implication $A_2 \longrightarrow m$ is a specification of implication $A_1 \longrightarrow m$ if $A_2 = A_1 \cup \{n\}$, where $n \in M \backslash A_1$. If L is a family of implications, we denote by $Spec(L)$ the set of all possible specifications of implications from L.

The computational procedure for finding probabilistic laws is based on the concepts of semantic probabilistic inference. The main idea consists in the successive specification of implications and checking if the conditions for a probabilistic law can be fulfilled. In fact, this implements a directed search for implications that allows one to considerably reduce the search space. The reduction is achieved due to the application of the following heuristics: starting from the time when the length of the generated implications reaches a certain prescribed value (called the base enumeration depth), the specification is applied only to those implications that are probabilistic laws.

For simplicity, we describe the computational procedure for finding probabilistic laws of the form $A \longrightarrow m$ on the model $\mathcal{M}$ for a chosen attribute $m \in M$. In addition to the probability model $\mathcal{M}$ and the element $m \in M$, the base enumeration depth d, $1 \leq d \leq |M|$, is also an input parameter of this procedure. The output of the procedure is the set of the probabilistic laws on the model $\mathcal{M}$ with the element m in the conclusion.

At step $k = 0$, a set $imp(\mathcal{M})_{(k)}$ of implications is generated that consists of a single implication of zero length of the form $R = \varnothing \longrightarrow m$. The implication R is checked as to if the conditions on a probabilistic law that are formulated in Definition 6 are satisfied. Denote the set of all probabilistic laws found at step k of the computational procedure by $REG_{\mathcal{M}}^{(k)}(m)$. If R is a probabilistic law, then $REG_{\mathcal{M}}^{(0)}(m) = \{R\}$; otherwise, $REG_{\mathcal{M}}^{(0)}(m) = \varnothing$ and the procedure outputs the empty set. Indeed, in this case we have $\eta_{\mathcal{M}}(\varnothing \longrightarrow m) = 0$; therefore, by the definition of the model $\mathcal{M}$, the probability of any implication of the form $B \longrightarrow m$ is either undefined or vanishes on $\mathcal{M}$. This means that none of such implications can be a probabilistic law on the model $\mathcal{M}$.

At step $1 \leq k \leq d$, the set $imp(\mathcal{M})_{(k)}$ of all specifications is computed for all implications obtained at the previous step, whose probability is defined but not equal to zero or one: $imp(\mathcal{M})_{(k)} = Spec(\{R | R \in imp(\mathcal{M})_{(k-1)}, 0 < \eta_{\mathcal{M}}(R) < 1\})$. Each implication in this set has length k. Each implication from $imp(\mathcal{M})_{(k)}$ is checked as to if it satisfies the conditions in the definition of a probabilistic law. Then, the set $REG_{\mathcal{M}}^{(k)}(m)$ is formed.

At step $d < k \leq |M|$, the set $imp(\mathcal{M})_{(k)}$ of all specifications is generated for all probabilistic laws, found at the previous step, that have a probability strictly less than one: $imp(\mathcal{M})_{(k)} = Spec(\{R | R \in REG_{\mathcal{M}}^{(k-1)}(m), \eta_{\mathcal{M}}(R) < 1\})$. All the implications obtained are checked as to if they satisfy the conditions for the probabilistic laws. Then, the set $REG_{\mathcal{M}}^{(k)}(m)$ is formed. The computational procedure ends either at the step $k = |M|$ or when no probabilistic law is obtained at some step $d < k < |M|$, i.e., when $REG_{\mathcal{M}}^{(k)}(m) = \varnothing$. The sought set of probabilistic

	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}
g_1	×	×	×	×	×					
g_2	×	×	×	×	×					
g_3	×	×	×	×	×					
g_4	×	×	×	×	×					
g_5	×	×	×	×	×					
g_6	×	×	×	×	×					
g_7	×	×	×	×	×					
g_8	×	×	×	×	×					
g_9	×	×	×	×	×					
g_{10}	×	×	×	×	×					
g_{11}	×	×	×	×	×					
g_{12}	×	×	×	×	×					
g_{13}	×	×	×	×	×					
g_{14}	×	×	×	×	×					
g_{15}	×	×	×	×	×					
g_{16}	×	×	×	×	×					
g_{17}	×	×	×	×	×					
g_{18}	×	×	×	×	×					
g_{19}	×	×	×	×	×					
g_{20}	×	×	×	×	×					
g_{21}						×	×	×	×	×
g_{22}						×	×	×	×	×
g_{23}						×	×	×	×	×
g_{24}						×	×	×	×	×
g_{25}						×	×	×	×	×
g_{26}						×	×	×	×	×
g_{27}						×	×	×	×	×
g_{28}						×	×	×	×	×
g_{29}						×	×	×	×	×
g_{30}						×	×	×	×	×
g_{31}						×	×	×	×	×
g_{32}						×	×	×	×	×
g_{33}						×	×	×	×	×
g_{34}						×	×	×	×	×
g_{35}						×	×	×	×	×
g_{36}						×	×	×	×	×
g_{37}						×	×	×	×	×
g_{38}						×	×	×	×	×
g_{39}						×	×	×	×	×
g_{40}						×	×	×	×	×

Context K_1

	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}
g_1	×	×	×	×	×	×				
g_2	×	×	×	×	×			×		
g_3	×	×	×	×	×					×
g_4	×	×	×	×	×	×				
g_5	×	×	×	×						
g_6	×	×	×	×	×	×				
g_7	×		×	×	×					
g_8	×	×	×	×	×			×		
g_9	×	×	×	×						
g_{10}	×	×		×	×					
g_{11}	×	×		×	×					
g_{12}	×		×	×	×					
g_{13}	×	×	×	×	×				×	
g_{14}	×	×	×	×	×				×	
g_{15}	×	×	×	×	×	×				
g_{16}	×	×	×	×	×					×
g_{17}	×	×	×	×						
g_{18}	×	×	×	×	×			×		
g_{19}	×	×	×	×	×				×	
g_{20}	×	×	×	×	×		×			
g_{21}							×	×	×	×
g_{22}						×	×	×	×	
g_{23}						×	×		×	×
g_{24}		×				×	×	×	×	×
g_{25}	×					×	×	×	×	×
g_{26}						×		×	×	×
g_{27}						×	×	×		×
g_{28}						×		×	×	×
g_{29}							×	×	×	×
g_{30}							×	×	×	×
g_{31}					×	×	×	×	×	×
g_{32}				×		×	×	×	×	×
g_{33}	×					×	×	×	×	×
g_{34}	×					×	×	×	×	×
g_{35}						×		×	×	×
g_{36}					×	×	×	×	×	×
g_{37}	×					×	×	×	×	×
g_{38}						×		×	×	×
g_{39}					×	×	×	×	×	×
g_{40}	×					×	×	×	×	×

Context K_2

Fig. 2. Recovery of a concept in a noised context.

laws for the attribute m is given by the union $\bigcup_k REG_{\mathcal{M}}^{(k)}(m)$; this is the output of the procedure.

In order to choose a family of the strongest (with respect to the input parameters) probabilistic laws from the set of implications obtained, it suffices to directly verify the conditions of Definition 8.

The steps $k \le d$ of the procedure are called base enumeration steps, and the steps $k > d$ are called additional enumeration steps. According to experiments, for a large number of practical problems, it suffices to use a base enumeration depth of $d \le 3$. Note that, in practice, the inequalities in Definition 6 are verified with regard to Fisher's statistical criterion (Fisher's exact test for contingency tables), which is applied with some (user-defined) confidence level α.

Let L be a nonempty set of probabilistic laws on the model $\mathcal{M}$. Note that if L is the output of the above procedure for the base enumeration depth $d = |M|$, then we have $L = D(\mathcal{M})$.

Let us describe the iterative procedure for finding probabilistic concepts in the model $\mathcal{M}$ with respect to the family L of implications.

At step $k = 1$, the following set is generated: $C^{(1)} = \{\bar{f}_L(A \cup \{m\}) | A \longrightarrow m \in L\}$.

At step $k > 1$, in case of $C^{(k-1)} = \varnothing$, the procedure outputs the list of probabilistic concepts found. Otherwise, for each $B \in C^{(k-1)}$, we consider the family of implications $L_B = \{A \longrightarrow m \in L | A \subseteq B\}$ and compute the set $A = \{g \in G | g' \cap B \neq \varnothing, f_{L_B}(g' \cap B) = B\}$. If $A \neq \varnothing$, then the pair (A, B) is added to the list of probabilistic concepts obtained. Then, the set $C^{(k)} = \{\bar{f}_L(B \cup C) | B, C \in C^{(k-1)}, \bar{f}_L(B \cup C) \notin C^{(k-1)}\}$ is generated, and the procedure goes to the next iteration step. The description of the procedure is complete.

Example 3. Consider the contexts K_1 and K_2 shown in Fig. 2. Concepts with nonempty extent and intent in the context K_1 are given by the pairs $(\{g_1, ..., g_{20}\}, \{m_1, ..., m_5\})$ and $(\{g_{21}, ..., g_{40}\}, \{m_6, ..., m_{10}\})$. The context K_2 was obtained from K_1 by adding random noise. The task is to recover the initial concepts in a noisy context K_2. In accordance with the algorithms described, the set of the strongest probabilistic laws is computed on the frequency model $\mathcal{M} = (K_2, \rho)$; it consists of 22 implications. The set of probabilistic concepts in the model $\mathcal{M}$ turns out to be equal to the set of concepts in the initial context K_1 with nonempty extents and intents.

Now we present an example in which probabilistic concepts are applied to the problem of classification of symbols displayed in different fonts.

Example 4. Consider the symbols of the letters A and B presented in Fig. 3 in three standard fonts where each symbol is depicted in a pixel representation as a matrix of dimension $m = 5 \times 6$. We can naturally enumerate the cells of a matrix and the matrices themselves and find a one-to-one correspondence between

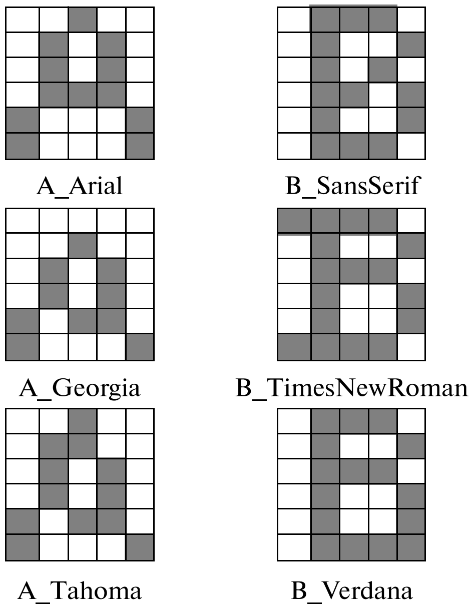

Fig. 3. Symbols of letters in a pixel representation.

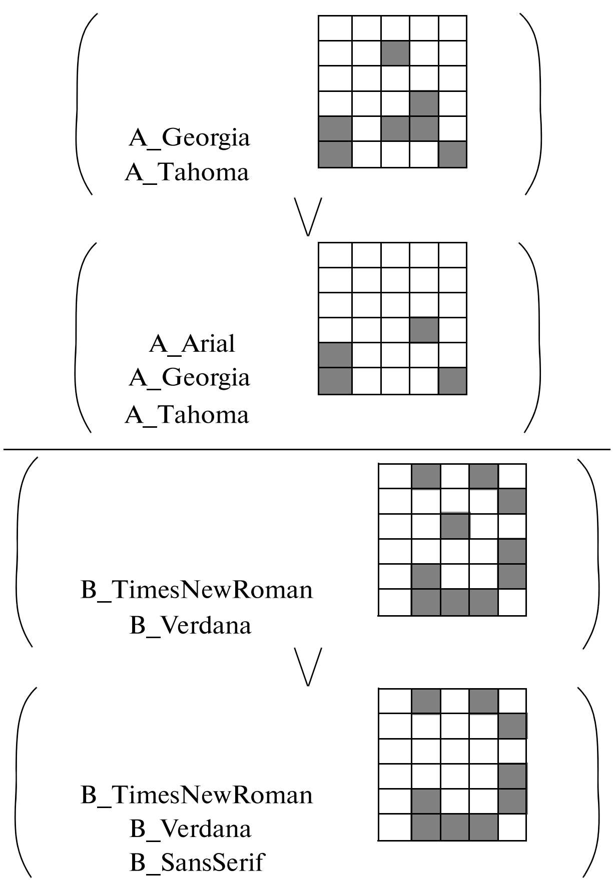

Fig. 4. Probabilistic concepts in the problem of classification of symbols.

the families of cardinality n of this kind of matrices and a family of formal contexts of the form $(\{1...n\}, \{1...m\}, I)$, where $\langle i, j\rangle \in I \Leftrightarrow$ matrix with number i in the cell with number j has a value of 1. Then it becomes possible to apply probability models to the symbols in the pixel representation and consider the contexts obtained as frequency probability models.

For the probabilistic context constructed by the family of six matrices shown in Fig. 3, we computed the set of the strongest probabilistic laws and the set of probabilistic concepts in accordance with the algorithms described. The set of probabilistic concepts together with their ordering by the inclusion of the intents of concepts is shown in Fig. 4. For convenience, the intents of the concepts are given in the matrix representation according to the above-mentioned one-to-one correspondence.

The family of concepts obtained demonstrates that the symbols of the letters A and B are assigned to different classification units and that different ways of depicting the same letter exhibit common parts.

CONCLUSIONS

It easy to notice from Definitions 5 and 10 that, from the theoretical point of view, the distinction between probability models of type I and II is quite relative. In particular, for any model $\mathcal{M}_2 = (K, \rho_2)$ of type II with $K = (\varnothing \neq G, M, I)$, one can define a model $\mathcal{M}_1 = (\mathcal{K}, \rho_1)$ of type I so that $D(\mathcal{M}_1) = D(\mathcal{M}_2)$. Indeed, it suffices to set $\mathcal{K} = \{\mathcal{K}_g | g \in G, K_g = (\{h\}, M, I_g), I_g = \{\langle h, m\rangle | \langle g, m\rangle \in I\}\}$ and define $\forall \mathcal{C} \subseteq \mathcal{K}\ \rho_1(\mathcal{C}) = \rho_2(\{g | K_g \in \mathcal{C}\})$. Then, for every implication $B \longrightarrow m$ on the set M, we have $\eta_{\mathcal{M}_1}(B \longrightarrow m) = \mu_{\mathcal{M}_1}(\langle h, B \longrightarrow m\rangle) = \dfrac{\rho_1(\{K_g | \{\langle h, n\rangle | n \in B \cup \{m\}\} \subseteq I_g\})}{\rho_1(\{Kg | \{\langle h, n\rangle | n \in B\} \subseteq I_g\})} = \dfrac{\rho_2((B \cup \{m\})')}{\rho_2(B')}$; thus, $\eta_{\mathcal{M}_1}(B \longrightarrow m) = \eta_{\mathcal{M}_2}(B \longrightarrow m)$, which obviously implies $D(\mathcal{M}_1) = D(\mathcal{M}_2)$. Nevertheless, in practice it is important to distinguish between the analysis of data represented by a class of contexts and the analysis of data on the basis of a single given context. In the first case, we have a problem of classification of objects that are observed in a number of experiments each of which establishes whether an object has a certain attribute. In the second case, the classification of objects is based on a single context, which represents the whole body of experimental data on these objects. A context uniquely determines whether an object has a particular attribute, and the FCA method provides tools for constructing a precise classification of objects on the basis of a given context. In turn, revealing probabilistic laws on a model defined on a given context allows one to obtain noiseimmune classification units.

Example 3 demonstrates that noise of a certain level does not change the set of concepts in a context; i.e., the set of concepts with nonempty extent and intent in a given initial context is equal to the set of probabilistic concepts in a new context obtained by adding noise to the initial one. There exist types of noise (a formal definition is given in [13]) such that any level of noise of this kind does not change the set of concepts in a context; the sets of concepts and probabilistic concepts coincide for this kind of noise; such noise is called concept preserving noise [13]. This raises the problem of characterization of these types of noise.

The definitions of implications and probabilistic laws considered in this paper do not involve the notion of negation. Therefore, the formulation of Theorem 2 seems to be weaker than expected. This is because the very fundamentals of FCA lack negation; in this paper, we aimed at obtaining the simplest generalization of the basic concepts of this method. The generalization of FCA within the class of given ideas will allow one to formalize the notions of "natural classification" and "idealization" as they are defined in [13, 20].

The semantic probabilistic inference, which is central in the definitions of probabilistic concepts, has been first introduced for first-order logic and provides a method for revealing rather complicated regularities on data compared with those considered in this paper. Moreover, in the relational approach described in monographs [13, 18], it is argued that the formalization of regularities in the language of first-order logic is essential for analyzing the whole body of information contained in data. Some examples of such regularities are given on the website [16] at http://math.nsc.ru/AP/Scientific Discovery/pages/Examples_of_rules.html.

ACKNOWLEDGMENTS

This work was supported by a grant of the President of the Russian Federation (project no. MK-2037.2011.9), by the Russian Foundation for Basic Research (project nos. 11-07-00560-a and 11-07-00388-a), and by Integration Projects of the Siberian Branch, Russian Academy of Sciences (project nos. 3, 87, and 136).

REFERENCES

1. Ganter, B. and Wille, R., *Formal Concept Analysis*: *Mathematical Foundations*, Berlin: Springer, 1999.
2. *Formal Concept Analysis*: *Foundations and Applications*, Ganter, B., Stumme, G., and Wille, R., Eds., Heidelberg: Springer, 2005.
3. Pasquier, N., Bastide, Y., Taouil, R., and Lakhal, L., Efficient Mining of Association Rules Using Closed Item Set Lattices, *Inf. Syst.*, 1999, no. 24 (1), pp. 25–46.
4. Tan, P.-N., Steinbach, M., and, Kumar, V., *Introduction to Data Mining*, Boston: Addison-Wesley, 2006, Chapter 6.
5. Webb, G.I., Discovering Significant Patterns, *Mach. Learn.*, 2007, no. 68 (1), pp. 1–33.
6. Webb, G.I., Association Rules (Chapter 2) in *The Handbook of Data Mining*, Nong Ye, Ed., Mahwah, N.J.: Lawrence Earlbaum Associates, 2003.
7. Deogun, J., Jiang, L., Xie, Y., and Raghavan, V., *Probability Logic Modeling of Knowledge Discovery in Databases*, *Proc. ISMIS'2003*, Springer, 2003, pp. 402–407.
8. Halpern, J.Y., An Analysis of First-Order Logic of Probability, *Artif. Intell.*, 1990, no. 46, pp. 311–350.
9. Kuznetsov, S. and Obiedkov, S., Comparing Performance of Algorithms for Generating Concept Lattices, *J. Exp. Theor. Artif. Intell.*, 2002, no. 14 (2–3), pp. 189–216.
10. Merwe, D., Obiedkov, S., and Kourie, D., AddIntent: A New Incremental Algorithm for Constructing Concept Lattices, *Proc. ICFCA'2004*, Springer, 2004, pp. 372–385.
11. Guigues, J.-L. and Duquenne, V., Families minimales d'implications informatives resultant d'un tableau de données binaries, *Math. Sci. Hum.*, 1986, no. 95, pp. 5–18.
12. Smerdov, S.O. and Vityaev, E.E., Synthesis of Logic, Probability, and Training: Formalization of Prediction, *Sib. Elektron. Mat. Izv.*, 2009, no. 6, pp. 340–365.
13. Vityaev, E.E., *Kompyuternoe poznanie* (Computer Cognition), Novosibirsk: Novosibirsk. Gos. Univ., 2006.
14. Vityaev, E., The Logic of Prediction, *Proc. of the 9th Asian Logic Conference*, World Sci., 2006, pp. 263–276.
15. Vityaev, E. and Kovalerchuk, B., Empirical Theories Discovery Based on the Measurement Theory, *Mind Mach.*, 2005, no. 14 (4), pp. 551–573.
16. Vityaev, E.E., Scientific Discovery. http://math.nsc.ru/AP/Scientific Discovery. Cited December 2011.
17. Vityaev, E. and Smerdov, S., New Definition of Prediction without Logical Inference, *Proc. IASTED Int. Conf. on Computational Intelligence (CI 2009)*, 2009, pp. 48–54.
18. Kovalerchuk, B. and Vityaev, E., *Data Mining in Finance*: *Advances in Relational and Hybrid Methods*, Kluwer, 2000.
19. Kuznetsov, S., On the Intractability of Computing the Duquenne–Guigues Base, *J. Univ. Comput. Sci.*, 2004, no. 10 (8), pp. 927–933.
20. Vityaev, E., Lapardin, K., Khomicheva, I., and Proskura, A., Transcription Factor Binding Site Recognition by Regularity Matrices Based on the Natural Classification Method, *Intell. Data Anal.* Special Issue *New Methods in Bioinformatics*, Vityaev, E. and Kolchanov, N., Eds., 2008, no. 12 (5), pp. 495–512.