

On the Problem of Prediction

Evgenii Vityaev^{1,2} and Stanislav Smerdov²

¹ Sobolev Institute of Mathematics, Russian Academy of Sciences,
Koptug prospect 4, Novosibirsk, 630090, Russia

² Novosibirsk State University
evgenii.vityaev@math.nsc.ru
<http://www.math.nsc.ru/AP/ScientificDiscovery>

Abstract. We consider predictions provided by Inductive-Statistical (I-S) inference. It was noted by Hempel that I-S inference is statistically ambiguous. To avoid this problem Hempel introduced the Requirement of Maximal Specificity (RMS). We define the formal notion of RMS in terms of probabilistic logic, and maximally specific rules (MS-rules), i. e. rules satisfying RMS. Then we prove that any set of MS-rules draws no contradictions in I-S inference, therefore predictions based on MS-rules avoid statistical ambiguity. I-S inference may be used for predictions in knowledge bases or expert systems. In the last we need to calculate the probabilistic estimations for predictions. Though one may use existing probabilistic logics or “quantitative deductions” to obtain these estimations, instead we define a semantic probabilistic inference and prove that it approximates logical inference in some sense. We also developed a program system ‘Discovery’ which realizes this inference and was successfully applied to the solution of many practical tasks.

Keywords: scientific discovery, probability and logic synthesis, probabilistic logic programming, machine learning.

1 Introduction

1.1 The Statistical Ambiguity Problem

One of the major results of the Philosophy of Science is the so-called *Covering Law Model*, which was introduced by Hempel in the early sixties in his famous article ‘Aspects of Scientific Explanation’ (see Hempel [1], [2], and Salmon [3] for a historical overview). The basic idea of this covering law model is that a fact is explained by subsumption under the so-called *covering law*, i.e. the task of an explanation is to show that a fact can be considered as an instantiation of a law. In the covering law model two types of explanation are distinguished: *Deductive-Nomological* explanations (D-N explanations) and *Inductive-Statistical* explanations (I-S explanations). In D-N explanations the laws are *deterministic*, whereas in I-S explanations the laws are *statistical*. Right from the beginning it was clear to Hempel that two I-S explanations can yield contradictory conclusions. He called this phenomenon the *statistical ambiguity* of I-S explanations [1], [2]. Let us consider the following example of the statistical ambiguity.

Suppose that we have the following statements about Jane Jones. ‘Almost all cases of streptococcus infection clear up quickly after the administration of penicillin’(L1). ‘Almost no cases of penicillin resistant streptococcus infection clear up quickly after the administration of penicillin’(L2). ‘Jane Jones had streptococcus infection’(C1). ‘Jane Jones received treatment with penicillin’(C2). ‘Jane Jones had a penicillin resistant streptococcus infection’(C3). From these statements it is possible to construct two contradictory arguments, one explaining why Jane Jones recovered quickly (E), and the other one, explaining its negation why Jane Jones did not recover quickly ($\neg E$).

$$\begin{array}{cc} \text{Argument1} & \text{Argument2} \\ \frac{L1}{\frac{C1, C2}{E}[r]} & \frac{L2}{\frac{C2, C3}{\neg E}[r]} \end{array}$$

The premises of both arguments are consistent with each other, they could all be true. However, their conclusions contradict each other, making these arguments rival ones.

Hempel hoped to solve this problem by forcing all statistical laws in an argument to be maximally specific. That is, they should contain all relevant information with respect to the domain in question. In our example, then, premise C3 of the second argument invalidates the first argument, since the law L1 is not maximally specific with respect to all information about Jane Jones (presented treatment is intuitively clear, but not formal, because we don’t have a precise definition of specificity yet – it will appear in the following sections). So, we can only explain $\neg E$, but not E.

1.2 Inductive-Statistical Inference

Hempel proposed the formalization of the statistical inference as Inductive-Statistical Inference (I-S inference) and the property of maximally specific statistical laws as the Requirement of Maximal Specificity (RMS). The Inductive-Statistical Inference has the form:

$$\frac{\frac{L_1, \dots, L_m}{C_1, \dots, C_n}[r]}{G}$$

It satisfies the following conditions:

- $L_1, \dots, L_m, C_1, \dots, C_n \vdash G$;
- $L_1, \dots, L_m, C_1, \dots, C_n$ are consistent;
- $L_1, \dots, L_m \not\vdash G$; $C_1, \dots, C_n \not\vdash G$;
- L_1, \dots, L_m are composed of statistical quantified formulas.
- C_1, \dots, C_n are quantifier-free;
- RMS: All laws L_1, \dots, L_m are maximally specific.

In Hempel's [1], [2] the RMS is defined as follows. An I-S argument of the form:

$$\frac{p(G; F)}{\frac{F(\mathbf{a})}{G(\mathbf{a})}}[r]$$

is an acceptable I-S explanation with respect to a "knowledge state" K, if the following Requirement of Maximal Specificity is satisfied. For any class H for which the corresponding two sentences are contained in K

$$\begin{aligned} \forall x(H(x) \Rightarrow F(x)), \\ H(\mathbf{a}), \end{aligned} \tag{1}$$

there exists a statistical law $p(G; H) = r'$ in K such that $r = r'$. The basic idea of RMS is that if F and H both contain the object \mathbf{a} , and H is a subset of F, then H provides more specific information about the object \mathbf{a} than F, and therefore the law $p(G; H)$ should be preferred over the law $p(G; F)$.

1.3 The Requirement of Maximal Specificity in Default Logic

Nowadays the same problems arise in non-monotonic logic and especially in default logic. Hempel's RMS produces also non-monotonic effects in inductive statistical reasoning. The streptococcus infection example is non-monotonic in the following sense. It was observed that the conflict between argument 1 and the argument 2 depends on the knowledge state K. If K contains only the information that John is infected, then RMS determines that argument 1 is the best (or the most specific) explanation: since no additional information (such as C3) is given, so L1 is maximally specific according to K. In that case, K implies the conclusion that John will recover quickly. However, if K is expanded with the premise C3, i.e. the information that John had a penicillin resistant streptococcus infection, then RMS determines that argument 2 explains that John will not recover quickly. Hence, the conclusion that John will recover quickly is not preserved under expansion of K.

Yao-Hua Tan [4] showed that there is a remarkable resemblance between two research traditions: default logic and inductive-statistical explanations. Both research traditions have the same research objective; to develop formalisms for reasoning with incomplete information. In both research traditions the crucial problem that had to be dealt with is the problem of *Specificity*, i.e. when two arguments conflict with each other the most specific argument has to be preferred to the less specific argument. This criterion of specificity, that was proposed in AI research, is very similar to the criterion of maximal specificity suggested by Hempel in the early sixties.

Let us formulate the Requirement of Maximal Specificity (RMS*) in default logic. Essentially, default logic is an ordinary first-order predicate logic extended with extra inference rules that are called default rules. The logical form of a *default rule* is:

$$(\alpha(x) : \beta_1(x), \dots, \beta_n(x) / \omega(x))$$

The subformulas $\alpha(x)$, $\beta_i(x)$, and $\omega(x)$ are predicate logical formulas with free variable x . The subformula $\alpha(x)$ is called the *prerequisite*, $\beta_i(x)$ are the *justifications* and $\omega(x)$ is the *consequent* of the default rule. The intuitive interpretation of a default rule follows: if the prerequisite $\alpha(x)$ is valid, and all justifications $\beta_i(x)$ are consistent with the available information (i.e. $\neg\beta_i(x)$ is not derivable from the available information), then one can assume that the consequent $\omega(x)$ is valid.

A set of formulas E is an *extension* of the default theory $\Delta = \langle W; D \rangle$, D – the set of default rules, W – a set of predicate logical formulas, if E is the smallest set such as: $W \subseteq E$; $E = \text{Th}(E)$; for each default rule $(\alpha(x); \beta_1(x), \dots, \beta_n(x) / \omega(x)) \in D$, and each term t : if $\alpha(t) \in E$, and $\neg\beta_1(t), \dots, \neg\beta_n(t) \notin E$, then $\omega(t) \in E$.

RMS*: If a default theory has multiple conflicting extensions, then the extension is preferred which is generated by the most specific defaults [4].

The default rule with the ‘most specific’ prerequisite is preferred in case of conflicts. Let $A(x)$ and $B(x)$ be the prerequisites of the default rules $D1$ and $D2$. The prerequisite $A(x)$ is *more specific* than $B(x)$ if the set that the predicate A refers to is a subset of the set that B refers to, i.e. if the sentence $\forall x(A(x) \Rightarrow B(x))$ is valid. It is obvious that this criterion can be considered as the analogue of RMS in default logic.

1.4 The Solution of the Statistical Ambiguity Problem

From the previous consideration we see that the statistical ambiguity problem raises in AI in different forms, but it isn’t solved hitherto. We will once again state the problem:

- is it possible to define the RMS in such a way that it solves the statistical ambiguity problem?
- can we define the RMS in such a way that the set of sentences satisfying the RMS will be consistent?

This problem is very important, because it means the consistency of predictions. The predictions nowadays are produced by different AI systems. In this paper we present our solution of this problem. We define the set of Maximally Specific Rules (MSR) and the Requirement of Maximal Specificity (RMS) and prove that sentences from MSR satisfy RMS and that the set MSR is consistent.

1.5 Probabilistic Approximation of Empirical Theories

Let us consider the task of the empirical theory discovery in the presence of noise provided for example the propensity interpretation of probability by Karl Popper.

Let \mathfrak{L} be the first-order logic with signature $\mathfrak{S} = \langle P_1, \dots, P_m \rangle$, $m > 0$, where P_1, \dots, P_m are predicate symbols of arity n_1, \dots, n_m , and fixed tuples of variables attached to each predicate symbol (so every predicate appears only with its own variables — this situation is quite similar to propositional classical logic). An empirical system [5] is taken to mean a finite model $\mathfrak{M} = \langle B, W \rangle$ of the signature \mathfrak{S} , where B is the basic set of the empirical system, and $W = \langle P_1, \dots, P_m \rangle$ is the

tuple of predicates of the signature \mathfrak{S} defined on B . Let $\text{Th}(\mathfrak{M})$ be the set of all rules, which are true on the empirical system \mathfrak{M} and having the form:

$$C = (A_1 \& \dots \& A_k \Rightarrow A_0), \quad k \geq 0 \quad (2)$$

where A_0, A_1, \dots, A_k are literals.

In the next section we define the notion of law and the set of all laws \mathcal{L} and prove that $\mathcal{L} \vdash \text{Th}(\mathfrak{M})$. Hence, we can solve the task of empirical theory discovery by discovering all laws from the set \mathcal{L} . In section 5 we prove that $\mathcal{L} \subset \text{MSR}$ and in section 7 we prove that MSR is consistent. Therefore $\text{MSR} \vdash \text{Th}(\mathfrak{M})$ and MSR provides a probabilistic approximation of the empirical theory $\text{Th}(\mathfrak{M})$. See the review of Jon Williamson [6] for other approaches.

1.6 Approximation of Logical Inference by Semantic Probabilistic Inference

So far we considered I-S-inferences using one rule for the inference. In general, I-S-inference uses many rules for the inference in knowledge bases and expert systems. This inference is based on the logical inference rules. The probability estimations of the inference results are obtained by the probabilistic logics or so called “quantitative deductions” [10], [11]. These estimations not always produce satisfactory results.

We replace the logical inference by the special semantic probabilistic inference, which produces all rules of the sets L, MSR, and also approximates the logical inference. We prove (Theorem 7) that estimations produced by the semantic probabilistic inference are no less (or even greater) than estimations produced by the probabilistic logics based on the logical inference.

2 Laws

Proposition 1. *The rule $C = (A_1 \& \dots \& A_k \Rightarrow A_0)$ logically follows from any rule of the form:*

$$(B_1 \& \dots \& B_h \Rightarrow A_0) \vdash (A_1 \& \dots \& A_k \Rightarrow A_0), \quad (3)$$

$$\{B_1, \dots, B_h\} \subset \{A_1, \dots, A_k\}, \quad 0 \leq h < k$$

Definition 1. *By a subrule of the rule $C = (A_1 \& \dots \& A_k \Rightarrow A_0)$ we mean any logically stronger rule of the form (3).*

Corollary 1. *If a subrule of the rule C is true on \mathfrak{M} , then the rule C is also true on \mathfrak{M} .*

Definition 2. *By a law on \mathfrak{M} , we mean any rule C of the form (2) that satisfies the following conditions [7], [8]:*

- (1) C is true on \mathfrak{M} ;
- (2) the premise of the rule is not always false on \mathfrak{M} ;
- (3) none of its subrules is true on \mathfrak{M} .

Let \mathcal{L} be the set of all laws on \mathfrak{M} .

Theorem 1. [7]. $\mathcal{L} \vdash \text{Th}(\mathfrak{M})$.

3 The Probability of Events and Sentences

Let us generalize the notion of law into the probabilistic case. For this purpose we introduce the probability on the model \mathfrak{M} . For the sake of simplicity we define the probability in the most simple case (follow the paper [9]). More general definitions of probability function μ are considered in [9]. Further considerations don't depend on the selected probability definition and, for example, true for definition 10 below.

We introduce the probability μ as a discrete function on B (B should be countable), $\mu: B \rightarrow [0,1]$ such that

$$\sum_{a \in B} \mu(a) = 1, \text{ and } \mu(a) \neq 0, a \in B; \mu(D) = \sum_{b \in D} \mu(b), D \subseteq B \quad (4)$$

We define the probability μ on the product B^n as a probability function $\mu^n(a_1, \dots, a_n) = \mu(a_1) \times \dots \times \mu(a_n)$.

Let us define the interpretation of the language \mathcal{L} on the empirical system $\mathfrak{M} = \langle B, W \rangle$ as mapping $I: \mathfrak{S} \rightarrow W$, which associates with every signature symbol $P_j \in \mathfrak{S}$, $j = 1, \dots, m$, the predicate P_j from W of the same arity. Let $X = \{x_1, x_2, x_3, \dots\}$ be the set of all variables of the language \mathcal{L} . By the validation ν is meant the function $\nu: X \rightarrow B$, mapping variables into the set of objects B .

Let us define the probability for the sentences of the language \mathcal{L} . Let $U(\mathfrak{S})$ be the set of all atomic formulas of the language \mathcal{L} ; $\mathfrak{R}(\mathfrak{S})$ is the set of all sentences of the language \mathcal{L} , obtained by closure of the set $U(\mathfrak{S})$ with respect to standard Boolean operations $\&, \vee, \neg$. By the $\hat{\varphi}$, $\varphi \in \mathfrak{R}(\mathfrak{S})$ we define the formula, where the predicate symbols of \mathfrak{S} are substituted by the predicates of W via interpretation I and by the $\nu\hat{\varphi}$ we define the formula, where variables of the formula $\hat{\varphi}$ are substituted by the objects of A via the validation ν . In particular, $\nu\hat{P}_j(x_1^j, \dots, x_{n_j}^j)^{\varepsilon_j} = P_j(a_1, \dots, a_j)^{\varepsilon_j}$, $\nu(x_1^j) = a_1, \dots, \nu(x_{n_j}^j) = a_j$. Let us define the probability η of the sentences of $\mathfrak{R}(\mathfrak{S})$. If x_1, \dots, x_n are all variables of the sentence $\varphi \in \mathfrak{R}(\mathfrak{S})$, then

$$\eta(\varphi) = \mu^n(\{(a_1, \dots, a_n) \mid \nu\hat{\varphi} \text{ is true on } \mathfrak{M}, \nu(x_1) = a_1, \dots, \nu(x_n) = a_n\}) \quad (5)$$

4 The Probabilistic Laws on \mathfrak{M}

Let us revise the concept of the law on \mathfrak{M} in terms of probability. We do it in such a way that the concept of the law on \mathfrak{M} would be a particular case of this definition. A law on \mathfrak{M} is a true rule such that all its subrules are false on \mathfrak{M} or, in other words, a law is such a true rule that cannot be made simpler or logically stronger without losing its truth. This property of the law "not to be simplified" allows stating the law not only in terms of truth but also in terms of probability.

For the rule $C = (A_1 \& \dots \& A_k \Rightarrow A_0)$ we define the conditional probability as $\eta(C) = \eta(A_0/A_1 \& \dots \& A_k) = \eta(A_0 \& A_1 \& \dots \& A_k) / \eta(A_1 \& \dots \& A_k)$.

Theorem 2. [7]. For any rule $C = (A_1 \& \dots \& A_k \Rightarrow A_0)$, the following two conditions are equivalent:

1. a rule C is a law on \mathfrak{M} , that is, it satisfies properties (1–3) of Definition 2;
2. (a) $\eta(C) = 1$;
 (b) $\eta(A_1 \& \dots \& A_k) > 0$;
 (c) the conditional probability $\eta(C)$ of the rule C is strictly greater than the conditional probability of each of its subrules.

This theorem gives us the equivalent definition of the law on \mathfrak{M} .

Definition 3. *By a probabilistic law on \mathfrak{M} with conditional probability 1 is meant the rule $C = (A_1 \& \dots \& A_k \Rightarrow A_0)$ of the form (2) satisfying the following conditions:*

1. $\eta(C) = 1$, $\eta(A_1 \& \dots \& A_k) > 0$;
2. conditional probability of the rule $\eta(C)$ is strictly greater than the conditional probability of each of its subrules.

The next corollary follows from Theorem 2.

Corollary 2. *A rule is a probabilistic law on \mathfrak{M} with conditional probability 1 iff it is a law on \mathfrak{M} .*

Let us consider items 1 and 2 of Theorem 2 from the standpoint of the ‘not to be simplified’ law:

- A law is such a true on \mathfrak{M} rule that cannot be simplified or made logically stronger without loss of truth.
- Any logically stronger subrule of a rule **has has** a conditional probability smaller than 1, so the rule cannot be simplified without losing the value 1 of the conditional probability.

A more general definition of a law follows:

Definition 4. *A law is such a rule of the form (2) based on truth values, conditional probability or other evaluations of the sentences, which cannot be made logically stronger without reducing their values.*

Therefore, we can define the probabilistic law for the more general case by omitting the condition $\eta(C) = 1$ from the point (1) of Definition 3.

Definition 5. *By a probabilistic law on \mathfrak{M} we mean such a rule $C = (A_1 \& \dots \& A_k \Rightarrow A_0)$ of the form (2), the conditional probability of which is defined and strictly greater than the conditional probability of each of its subrules. In particular, the conditional probability $\eta(C)$ of the rule C is strictly greater than the probability $\eta(A_0)$, which is the probability of the subrule $(\Rightarrow A_0)$.*

Let us denote by \mathcal{LP} the set of all probabilistic laws. It follows from Theorem 2 and Definition 5 that the set \mathcal{LP} includes the set \mathcal{L} .

Corollary 3. $\mathcal{L} \subseteq \mathcal{LP}$.

Definition 6. *By a Strongest Probabilistic Law (SPL-rule) on \mathfrak{M} , we designate such a probabilistic law $C = (A_1 \& \dots \& A_k \Rightarrow A_0)$, which is not a subrule of any other probabilistic law.*

We define as SPL the set of all SPL-rules.

Proposition 2. $\mathcal{L} \subseteq \text{SPL} \subseteq \mathcal{LP}$.

5 Semantic Probabilistic Inference

Let us define the Semantic Probabilistic inference (SP-inference) of the sets of laws \mathcal{L} and probabilistic laws \mathcal{LP} .

Definition 7. [7], [8], [16]. By a *Semantic Probabilistic inference* of some SPL-rule C we mean a sequence $C_1, C_2, \dots, C_n = C \in \mathcal{LP}$, denoted by $C_1 \sqsubset C_2 \sqsubset \dots \sqsubset C_n$, such that:

$$\begin{aligned} C_i &= (A_1^i \& \dots \& A_{k_i}^i \Rightarrow G), \quad i = 1, 2, \dots, n, \quad n > 0, \\ &\text{the rules } C_i \text{ are subrules of the rules } C_{i+1}, \\ &\eta(C_{i+1}) > \eta(C_i), \quad i = 1, 2, \dots, n-1, \end{aligned} \quad (6)$$

and this sequence is maximal, i.e., there is no $C' \in \mathcal{LP}$ such that $\eta(C') > \eta(C)$ and C is a subrule of C' .

Unlike probabilistic logics [10], [11] the probability of the sentences is strictly increase in the process of SP-inference.

Proposition 3. *Any probabilistic law from \mathcal{LP} belongs to some SP-inference. For any SPL-rule there is some SP-inference of that rule.*

Corollary 4. *For any law from \mathcal{L} there is some SP-inference of that law.*

Let us consider the set of all SP-inferences of the sentence G . This set constitutes the Semantic Probabilistic Inference lattice (SPI-lattice) of this sentence.

Definition 8. *By a maximally specific rule $MS(G)$ of a sentence G we mean a SPL-rule of the SPI-lattice of the sentence G , which has the maximum value of conditional probability among all other SPL-rules of the SPI-lattice.*

We define as MSR the set of all maximally specific rules.

Proposition 4. $\mathcal{L} \subseteq MSR \subseteq SPL \subseteq \mathcal{LP}$

6 Probabilistic Maximally Specific Laws

Now we define the Requirement of Maximal Specificity (RMS). We will suppose that the class H of objects in (1) is defined by some sentence $H \in \mathcal{R}(\mathfrak{S})$ of the language \mathfrak{L} . In this case the RMS says that $p(G;H) = p(G;F) = r$ for this sentence. In terms of probability η it means that $\eta(G/H) = \eta(G/F) = r$ for any $H \in \mathcal{R}(\mathfrak{S})$ satisfying (1).

Definition 9. *The Requirement of Maximal Specificity (RMS): if we add any sentence $H \in \mathcal{R}(\mathfrak{S})$ to the premise of the rule $(F \Rightarrow G)$, $\eta(G/F) = r$, such that $F(\mathbf{a}) \& H(\mathbf{a})$ for some object \mathbf{a} , then for the new rule $(F \& H \Rightarrow G)$ we have $\eta(G/F \& H) = \eta(G/F) = r$.*

In other words, the requirement RMS means that there is no other sentence H in $\mathfrak{R}(\mathfrak{S})$ that increases (or decreases, see lemma 1 below) the conditional probability $\eta(G/F) = r$ by adding it to the premise.

Lemma 1. [8]. If the sentence $H \in \mathfrak{R}(\mathfrak{S})$ decreases the probability $\eta(G/F \& H) < \eta(G/F)$ then the sentence $\neg H$ increases it: $\eta(G/F \& \neg H) > \eta(G/F)$.

Lemma 2. [8]. For any rule $C = (B_1 \& \dots \& B_t \Rightarrow A_0)$, $\eta(B_1 \& \dots \& B_t) > 0$, of the form (2) there is a probabilistic law $C' = (A_1 \& \dots \& A_k \Rightarrow A_0)$ on \mathfrak{M} which is a subrule of the rule C and $\eta(C') \geq \eta(C)$.

Theorem 3. [8]. Any MS(G) rule satisfies the RMS requirement.

Corollary 5. [8]. Any law on \mathfrak{M} satisfies the RMS requirement.

7 The Solution of the Statistical Ambiguity Problem

Theorem 4. [8]. The I-S inference is consistent for any theory $\text{Th} \subseteq \text{MSR}$ in the following sense: it is impossible to obtain a contradiction (ambiguity) in I-S inference using only rules from Th , i.e. there are no $(A \Rightarrow G) \in \text{Th}$ and $(B \Rightarrow \neg G) \in \text{Th}$ such that $\eta(A \& B) > 0$.

Let us illustrate this theorem by the example of Jane Jones. We can define the maximally specific rules $\text{MS}(E)$, $\text{MS}(\neg E)$ for the sentences E , $\neg E$ as follows:

- $\widehat{L1}$: ‘Almost all cases of streptococcus infection, that are not resistant to streptococcus infection, clear up quickly after the administration of penicillin’;
- $L2$: ‘Almost no cases of penicillin resistant streptococcus infection clear up quickly after the administration of penicillin’.

The rule $\widehat{L1}$ has a greater value of conditional probability than the rule $L1$; hence, it is a $\text{MS}(E)$ rule for the sentence E . These two rules can’t be fulfilled on the same data.

So, we can predict without contradictions, if we use the set MSR as statistical laws in I-S inference.

8 Probabilistic Herbrand Models

In the following, we perform all considerations in the frame of logical programming with functional symbols, substitutions, and the countable set of variables. For that purpose we extend our language and redefine some definitions. Next sections 8-12 are updated and translated version of the paper [16].

Consider a first order language L with equality of the finite signature $\Omega = \langle P_1, P_2, \dots, P_{n_1}, f_1, f_2, \dots, f_{n_2}, c_1, c_2, \dots, c_{n_3} \rangle$. Let U denotes a set of all *ground terms* (without free variables), X – a countable set of *variables*, T – a set of *terms*, F_L – a set of *formulas*, F – a set of *formulas without quantifiers*, S – a set

of *sentences* (formulas without free variables), $\mathfrak{R} = F \cap S$ – the set of all *ground sentences* of the signature, B_L – the set of all ground atoms of the signature Ω .

A mapping $\theta: X \rightarrow T$ is called a *substitution*. Denote by Θ the set of all substitutions. The substitution $\theta(x) = x$ is called identical. Substitutions are naturally extended to arbitrarily expressions. Thus, substitutions for the term $t = f(t_1, \dots, t_n)$ and atom $A = P(t_1, \dots, t_n)$ are equal to $\theta t = f(\theta t_1, \dots, \theta t_n)$, $\theta A = P(\theta t_1, \dots, \theta t_n)$ respectively. A rule $\theta A \leftarrow \theta A_1, \dots, \theta A_n$ is a variant of the rule $A \leftarrow A_1, \dots, A_n$ if θ is a permutation of the set X .

Following [9], let us define a probability μ on a subset $F' \subseteq F$, $F' \neq \emptyset$ of sentences closed with respect to logical operations $\&, \vee, \neg$ and term substitutions.

Definition 10. A mapping $\mu: F' \rightarrow [0, 1]$ is called a *probability* provided that the following conditions are satisfied:

- 1) if $\vdash \phi$, then $\mu(\phi) = 1$;
- 2) if $\vdash \neg(\phi \& \psi)$, then $\mu(\phi \vee \psi) = \mu(\phi) + \mu(\psi)$;

Definition 11. A pair $M = \langle U, \mu \rangle$, where μ is a probability on \mathfrak{R} , is called a *probabilistic Herbrand model of signature Ω* .

Definition 12. A pair $M = \langle U, \mu \rangle$, where $I: B_L \rightarrow \{0, 1\}$, is called a *Herbrand model of signature Ω* .

Let there be given a certain class $G \subseteq 2^{B_L}$ of Herbrand models (a set of possible worlds) and a probability μ on F . For every $\varphi \in F$ let $G(\phi) = \{M \mid M \in G, M \models \phi\}$, where \models denotes the satisfaction, and let $D = \{G(\phi) \mid \phi \in F\}$.

Definition 13. A class G of Herbrand models is said to be *coordinated with the probability μ on the set of formulas F'* , and a probabilistic Herbrand model $M = \langle U, \mu \rangle$ is said to be a *probabilistic model of the class G* if $\mu(\phi) = 0$ follows from $G(\phi) = \emptyset$, $\phi \in F$.

We will consider two cases: $F' = F$ and $F' = \mathfrak{R}$. In the first case, the probability expands on sentences with free variables as $\mu(\varphi) = \inf_{\theta \in \Theta G} \{\mu(\varphi\theta)\}$, where ΘG is the set of all substitutions of variables by ground terms.

9 Logical Programs

Let PR denotes a set of all *rules* $A \leftarrow A_1, \dots, A_k$, $k \geq 0$ of the signature Ω , where A, A_1, \dots, A_k are atoms of the signature Ω . If atom A is absent, then the rule $\leftarrow A_1, \dots, A_k$ ($k \geq 0$) is called a *goal (request)*. In requests we will write ' $\&$ ' between atoms instead of ' $,$ '. If $k = 0$, then the rule $A \leftarrow$ is called a *fact*. A *logic program* Pr is a finite collection of rules.

Let fix a *selection rule* R , which selects one of the atoms from a request. Let $N = \leftarrow A_1 \& \dots \& A_i \& \dots \& A_k$, $k \geq 1$ be a request, where the rule R selects the atom A_i , and a rule $C = A \leftarrow B_1, \dots, B_m$ be a variant of some rule of the program

Pr, where all the variables are different from those of the request. Let θ be the most general unification of the atoms A_i , A ($A_i\theta = A\theta$). Then the requests

$$\begin{aligned} \leftarrow (A_1 \& \dots \& A_{i-1} \& B_1 \& \dots \& B_m \& A_{i+1} \& \dots \& A_k)\theta, \text{ if } m \geq 1 \\ \leftarrow (A_1 \& \dots \& A_i \& \dots \& A_k)\theta, \text{ if } m = 0 \end{aligned} \quad (7)$$

are called *inferred* from the request N by the rule $C = A \leftarrow B_1, \dots, B_m$ with the help of the substitution θ and the selection rule R. It is seen from the definition, that the atom A_i is not removed from the request after the unification with a certain program fact. Such atoms will be underlined. Suppose, that the rule R does not select the underlined atoms for the next inference steps.

The set of all possible requests of the signature Ω with the given relation of inference is called a *calculation space of the program Pr and the selection rule R*. A maximal sequence of requests $N = N_0, N_1, N_2, \dots$ together with the sequence of rules C_0, C_1, C_2, \dots and unifications $\theta_0, \theta_1, \theta_2, \dots$ such, that requests N_{i+1} are inferred from the requests N_i by means of the rules C_i , substitutions θ_i and the selection rule R, $i = 1, 2, \dots$ is called a *SLDF-inference* (Linear resolution with Selection rule for Definite clauses and underlined Facts) of the goal N in the calculation space. A SLDF-inference is a maximal path in the calculation space, starting with N. A SLDF-inference ending by a request with all atoms underlined is called *successful*. A finite inference, which is not successful, is called *dead ended*. A set of all SLDF-inferences starting with the goal N can be presented in the form of a tree (a prefix tree of SLDF-inferences). This tree is called the *SLDF-tree of the request N calculations*. The SLDF-tree containing a successful SLDF-inference is called a *successful SLDF-tree*.

10 Estimations of the Probability and Conditional Probability of Requests

Let $M = \langle U, \mu \rangle$ be a probabilistic Herbrand model. Consider a successful SLDF-inference $N = N_0, N_1, \dots, N_k$ of the request N in a calculation space of the program Pr obtained by means of the sequence of rules C_0, C_1, \dots, C_{k-1} , unifications $\theta_0, \theta_1, \dots, \theta_{k-1}$, $\theta \Leftarrow \theta_0\theta_1\dots\theta_{k-1}$ and the selection rule R (here we suppose that any N_i and N_j have no common variables – this modification can be easily performed by choosing the appropriate variants of rules C_0, \dots, C_{k-1}).

It is not difficult to show that the sequence of requests $N\theta, N_1\theta, \dots, N_k\theta = N_k$ is also a successful SLDF-inference of the request $N\theta$ by means of the same sequence of rules $C_0\theta, C_1\theta, \dots, C_{k-1}\theta$, identical unifications and the selection rule R.

The probability of a rule $C = A \leftarrow B_1, \dots, B_m$ ($m \geq 1$) is defined and equal to

$$\mu(C) = \mu(A|B_1 \& \dots \& B_m) = \mu(A \& B_1 \& \dots \& B_m) / \mu(B_1 \& \dots \& B_m)$$

iff $\mu(B_1 \& \dots \& B_m) \neq 0$ and it is undefined otherwise. Represent facts $A \leftarrow$ by the rules $A \leftarrow \text{true}$. Then $\mu(C) = \mu(A|\text{true}) = \mu(A)$. Suppose that $\mu(C)$ means that the probability is defined. Denote by $PR_0 \subseteq PR$ the set of all rules, for which conditional probability μ is defined; $Pr_0 \Leftarrow PR_0 \cap Pr$.

Definition 14. A rule C is true on a Herbrand model $N \in 2^{B_L}$ ($N \models C$) iff it is true on N under any state (for any mapping $\rho: X \rightarrow U$).

Definition 15. A program Pr is true on a Herbrand model N ($N \models Pr$) iff each rule of the program is true on N .

Definition 16. A program Pr is true on a class G of models iff $\forall N \in G, N \models Pr$.

We will write $C \in F'$ for the rule $C \Leftarrow A \leftarrow B_1 \& \dots \& B_m$, as soon as $A, B_1, \dots, B_m \in F'$.

Proposition 5. If $C \in Pr \cap F'$, $C = A \leftarrow B_1, \dots, B_m$, $\mu(B_1 \& \dots \& B_m) > 0$, then

$$\mu(\neg(B_1 \& \dots \& B_m) \vee A) = 1 \Leftrightarrow \mu(C) = 1.$$

Corollary 6. If the program Pr is true on the class of Herbrand models G , which is coordinated with the probability μ on the set of formulas F' , then $\mu(C) = 1$, $C \in Pr \cap F'$, if it is defined.

Denote the conjunction of all non-underlined atoms of the request N_i by N_i^\wedge . If all atoms are underlined (as in the request N_k), then $N_k^\wedge = true$. Denote the conjunction of all underlined atoms of the request N_i by $N_i F^\wedge$. Then, $N_i F^\wedge$ is a conjunction of all facts used in the SLDF-inference of the request $N\theta$.

Consider the inference of the requests (7) from the request

$$N\theta = \leftarrow (A_1 \& \dots \& A_i \& \dots \& A_k)\theta, \quad k \geq 1$$

by means of the rule $C = A \leftarrow B_1, \dots, B_m$. Let us estimate the probabilities $\mu(N\theta^\wedge)$, $\mu(N\theta^\wedge | N_k F^\wedge)$ assuming, that only probabilities $\mu(N_1 \theta^\wedge)$, $\mu(A_i \theta)$, $\mu(B\theta)$ and $p = \mu(A\theta | B\theta^\wedge)$ are known, where B stands for the conjunction $B_1 \wedge \dots \wedge B_m$.

Lemma 3. [16]. If $\mu(N_1 \theta^\wedge) > 0$ and $\mu(B\theta) > 0$, then:

- 1) $\mu(N\theta^\wedge) \leq \mu(\neg B\theta^\wedge) + \min\{\mu(N_1 \theta^\wedge), \mu(A\theta \& B\theta^\wedge)\}$;
- 2) $\mu(N\theta^\wedge) \geq \mu(N_1 \theta^\wedge) - (1 - p)\mu(B\theta^\wedge)$;
- 3) $\mu(N\theta^\wedge | N_1 \theta^\wedge) \leq p / \mu(N\theta^\wedge | B\theta^\wedge)$;
- 4) $\mu(N\theta^\wedge | N_1 \theta^\wedge) \geq 1 - (1 - p) / \mu(N\theta^\wedge | B\theta^\wedge)$.

Corollary 7. [16]. If $\mu(N_1 \theta^\wedge) > 0$, $\mu(B\theta) > 0$ and $p = 1$, then:

- 1) $\mu(N_1 \theta^\wedge) \leq \mu(N\theta^\wedge) \leq \min\{1, \mu(\neg B\theta^\wedge) + \mu(N_1 \theta^\wedge)\}$;
- 2) $\mu(N\theta^\wedge | N_1 \theta^\wedge) = 1$.

Corollary 8. [16]. If $\mu(N_1 \theta^\wedge) > 0$ and the rule is the fact $(A \leftarrow true)\theta$, $\mu(B\theta) = 1$, then:

- 1) $\mu(N\theta^\wedge) \leq \min\{\mu(N_1 \theta^\wedge), \mu(A\theta)\}$;
- 2) $\mu(N\theta^\wedge) \geq \mu(N_1 \theta^\wedge) + \mu(A\theta) - 1$;
- 3) $\mu(N\theta^\wedge | N_1 \theta^\wedge) \leq \mu(A\theta) / \mu(N_1 \theta^\wedge)$;
- 4) $\mu(N\theta^\wedge | N_1 \theta^\wedge) \geq 1 - (1 - \mu(A\theta)) / \mu(N_1 \theta^\wedge)$.

Corollary 9. [16]. If $\mu(B\theta) > 0$, then:

- 1) $\mu(N\theta^\wedge \& B\theta^\wedge) \leq \min\{\mu(N_1\theta^\wedge), \mu(A\theta \& B\theta^\wedge)\};$
- 2) $\mu(N\theta^\wedge \& B\theta^\wedge) \geq \mu(N_1\theta^\wedge) - (1 - p)\mu(B\theta^\wedge).$

Consider the SLDF-inference $N\theta, N_1\theta, \dots, N_k$ of the request $N\theta$ by means of sequence of rules $C_i\theta = (A^i \leftarrow B_1^i, \dots, B_{k_i}^i)\theta, i = 0, \dots, k-1$ and empty unifications. Denote $B^i\theta = (B_1^i \& \dots \& B_{k_i}^i)\theta, p_i = \mu(C_i\theta).$

Theorem 5. [16]. If $\mu(B^i\theta) > 0, i = 0, \dots, k-1$, then under the previous conditions

$$\mu(N\theta^\wedge \& A^0\theta \& \dots \& A^{k-1}\theta) \geq 1 - \sum_{i=0}^{k-1} (1 - p_i)\mu(B^i\theta)$$

Corollary 10. [16]. If $\mu(B^i\theta) > 0, i = 0, \dots, k-1$, then under the previous conditions

$$\mu(N\theta^\wedge) \geq 1 - \sum_{i=0}^{k-1} (1 - p_i)\mu(B^i\theta).$$

For every successful SLDF-inference $N\theta = N_0\theta, N_1\theta, \dots, N_{k-1}\theta, N_k$, there exists a SLDF'-inference $N\theta = N'_0\theta, N'_1\theta, \dots, N'_i\theta, \dots, N'_{k-1}\theta, N'_k = N_k$, where facts are used in the last turn and the rules $C_j\theta$ with $k_j \geq 1; j = 1, \dots, i-1$ are applied before facts. Then the request $N'_i\theta$ has the form $\leftarrow A'_1, \dots, A'_s$, and the request N_k - the form $\leftarrow A'_1, \dots, A'_s$. Such a SLDF'-inference is called *normalized*.

Theorem 6. [16]. If $\mu(B^j\theta) > 0, j = 0, 1, \dots, i-1$, and $\mu(N_k F^\wedge) > 0$, then for a successful SLDF-inference as defined earlier

$$\mu(N\theta^\wedge | N_k F^\wedge) \geq 1 - \sum_{j=0}^{i-1} (1 - p_j)\mu(B^j\theta) / \mu(N_k F^\wedge),$$

where p_j - conditional probabilities, $B^j\theta$ - conditions of the rules $C_j, j = 1, \dots, i-1$.

Let us define the probability estimations $\nu(N), \eta(N)$ of the calculation space requests for the program Pr and selection rule R. Consider the SLDF-tree of some request N in the calculation space. If the SLDF-tree is not successful, then estimations ν and η are not defined. For the successful SLDF-tree consider a set $\{SLDF'_i\}_{i \in I}, I \neq \emptyset$ of all successful normalized SLDF-inferences of the requests $\{N\theta^i\}_{i \in I}$.

Determine the estimations $\{\nu^i\}_{i \in I}$, which are equal to the right-hand side of the inequality of Corollary 10, for the probabilities $\mu(N\theta^{i\wedge}) \geq \nu^i (\forall i \in I)$ of the requests $\{N\theta^i\}_{i \in I}$ obtained by corresponding inferences. Determine also the estimates $\{\eta^i\}_{i \in I}$, which are equal to the right-hand side of the inequality of the theorem for the conditional probabilities $\mu(N\theta^{i\wedge} | N_{k_i}^i F^\wedge) \geq \eta^i (\forall i \in I)$ of the requests $\{N\theta^i\}_{i \in I}$. Define $\nu(N) = \sup_{i \in I} \{\nu^i\}, \eta(N) = \sup_{i \in I} \{\eta^i\}.$

The SLDF-inference of the request N , where the estimation $\eta(N)$ is reached, is called a *prediction of the request N* . The value $\eta(N)$ is called the *estimation of the request N prediction*. If the prediction is not defined, then the estimation of the prediction $\eta(N)$ is not defined.

Define the relation \triangleright - “to be more common” on the set PR. Denote the set of all substitutions, which are not rearrangements, by Θt (the identical substitution belongs to Θt).

Definition 17. The relation $C' \triangleright C$, $C = A \leftarrow B_1, \dots, B_m, C' = A' \leftarrow B'_1, \dots, B'_{m'}$, $m, m' \geq 0$ takes place, iff there exists a substitution $\theta \in \Theta$ such, that $A'\theta = A$, $\{B'_1\theta, \dots, B'_{m'}\theta\} \subseteq \{B_1, \dots, B_m\}$ and either $\theta \in \Theta t$ is not an identical substitution or $m' < m$.

11 Inductive Synthesis of Probabilistic Logic Programs

A full set of facts for the class of models G represents a collection of sets $F(N) = \{A \leftarrow \mid N \models A \text{ for any state of the atom } A\}$, $N \in G$. Any finite collection D of finite subsets $D(N) \subset F(N)$ is called *data*. A probabilistic Herbrand model $M = \langle U, \mu \rangle$, which is in accordance with the class of models G , is called a *probabilistic Herbrand model of data D* .

In what way should the rules $C = A \leftarrow B_1, \dots, B_m$, $m \geq 1$ be used for predictions? If, at a certain substitution, $\theta \in \Theta$ the conjunction $(B_1 \& \dots \& B_m)\theta$ is true on a certain model N , which was chosen randomly from G in accordance with the measure μ , i.e. $\{B_1\theta, \dots, B_m\theta\} \subseteq F(N)$, then the conclusion $A\theta$ is true on N with the probability $\mu(A\theta \mid (B_1 \& \dots \& B_m)\theta) \geq \mu(A \mid B_1 \& \dots \& B_m) = \mu(C)$. Thus, the probability $\mu(C)$ for rules with variables gives a lower bound for prediction probabilities of the atom $A\theta$. Note that only one model N , chosen arbitrarily from G , and the corresponding data $D(N)$ should be used for predictions.

Definition 18. The relation $C' \sqsubset C \Leftrightarrow (C' \triangleright C) \& (\mu(C') < \mu(C))$ is called the *probabilistic inference relation*.

Definition 19. A rule $C \in PR_0$, such that $\forall C' \in PR_0 (C' \triangleright C \Rightarrow C' \sqsubset C)$, is called a *probabilistic regularity (P-rule)*.

Let $PR(M)$ denote the set of all P-rules, and $P(M) \subset PR(M)$ – the set of all P-rules, where the premise contains at least one atom.

Definition 20. A set of rules $PR(M, N) = P(M) \cup D(N)$, where $D(N) \in D$, and N is a certain model chosen arbitrarily from G in accordance with the measure μ , is called a *probabilistic logic program synthesized inductively by data $D(N)$ and the probabilistic model of data M* .

12 Predictions Based on Semantic Probabilistic Inference

Definition 21. [16]. A maximal sequence of rules $C_1 \sqsubset C_2 \sqsubset \dots$, where $C_1, C_2, \dots \in P(M)$, $C_i = A_i \leftarrow B_1^i, \dots, B_{k_i}^i$, $i = 1, 2, \dots$ such that an atom A is unified

with all atoms A_1, A_2, \dots is called a *semantic probabilistic inference* (*P-inference*) of the atom A of the signature Ω . If such a sequence for the atom A does not exist, then the P-inference is empty. Each P-inference produces a sequence of substitutions $\theta_1, \theta_2, \dots$ from the definition of the relation \triangleright . A substitution $\theta = \theta_1\theta_2\dots$ is called a *semantic probabilistic inference result* (calculation). The final rule of the finite P-inference is called a *resulting rule*.

Definition 22. [16]. By a *P-prediction* of some atom A of the signature Ω by the program $PR(M, N) = P(M) \cup D(N)$ we mean such a P-inference $C_1 \sqsubset C_2 \sqsubset \dots \sqsubset C_i \sqsubset \dots$, $C_1, C_2, \dots, C_i, \dots \in P(M)$ of the goal A , where:

1. There exists a rule $C_i = A_i \leftarrow B_1^i, \dots, B_{l_i}^i$ and a substitution θ such that $\{B_1^i\theta, \dots, B_{l_i}^i\theta\} \subseteq D(N)$; $A\theta = A_i\theta$; $\mu(A_i\theta) < \mu(C_i)$;
2. A maximum of conditional probability is reached on the rule C_i among all rules, satisfying condition 1, of all P-inferences of the goal A ;
3. If there is no P-inference of the goal A or there is no required substitution, then a P-prediction is not defined;
4. A substitution $\theta_p = \theta_1\theta_2\dots\theta_{i-1}\theta$, where $\theta_1, \theta_2, \dots, \theta_{i-1}$ are the substitutions of the P-inference $C_1 \sqsubset C_2 \sqsubset \dots \sqsubset C_i$, is called the *P-prediction result*. The value $\eta_p(A) = \mu(C_i)$ is called the *P-prediction value*. If a P-prediction is not defined, then the value $\eta_p(A)$ is not defined.

Proposition 6. *Rules satisfying point 2 are not comparable with respect to the relation \triangleright .*

Though for the sake of simplicity the following statements (Lemmas 4–5, Corollary 11 and Theorem 7) are given in the frame of finite signature with no functional symbols (predicate and constant symbols are allowed) which is standard in probabilistic logic programming [12], a broader case of signature with functional symbols is also investigated (see [18]).

Lemma 4. *A P-prediction is defined iff there is at least one rule $C \in P(M)$ satisfying point 1 of Definition 22.*

Lemma 5. *Let rule $C \in PR_0$ satisfy point 1 of Definition 22 and has at least one atom in the premise, then either C is a P-rule ($C \in P(M)$) or there exists a rule $C' \in P(M)$ such that $C' \triangleright C$, $\mu(C') \geq \mu(C)$ and C' satisfying point 1 of Definition 22.*

Corollary 11. *A P-prediction is defined iff there is a rule $C \in PR_0$ with at least one atom in the premise that satisfies point 1 of Definition 22.*

Let Pr be the logical program with facts belonging to the facts $D(N)$ of the program $PR(M, N) = P(M) \cup D(N)$.

Theorem 7. [16]. If atom A is predicted by the program Pr with estimation $\eta(A) > \mu(A\theta)$, for any $\theta \in \Theta G$, then it is P-predicted by the program $PR(M, N)$ with P-prediction value $\eta_p(A) \geq \eta(A)$.

13 The Relational Data Mining and Program System ‘Discovery’

Based on the semantic probabilistic inference the Relational Data Mining (RDM) approach to the intensive area of applications – Knowledge Discovery in Data Bases and Data Mining (KDD&DM) – was developed in [7], [8], [13], [14], [15], [17]. The program system ‘Discovery’, which utilizes this approach, has been implemented. This system realizes the SP-inference and can discover the sets of laws \mathcal{L} , \mathcal{LP} and the sets SPL, MSR. So, we may discover the full (in the sense of Theorem 1 and Propositions 2,4) and consistent (in the sense of Theorem 4) sets of rules. In [7], [17] we argue that using RDM we may cognise the object domain. The system ‘Discovery’ has been successfully applied to solve many practical tasks: as a cancer diagnostic systems, time series forecasting, psychophysics, bioinformatics, and many others (see www-site Scientific Discovery [19]).

Acknowledgments

The work is partially supported by the Council for Grants (under RF President) and State Aid of Leading Scientific Schools (grant NSh-3606.2010.1); Russian Science Foundation grant 08-07-00272a and Integration projects of the Siberian Division of the Russian Academy of science 47,111,119.

References

1. Hempel, C.G.: Aspects of Scientific Explanation. In: Hempel, C.G. (ed.) Aspects of Scientific Explanation and other Essays in the Philosophy of Science. The Free Press, New York (1965)
2. Hempel, C.G.: Maximal Specificity and Lawlikeness in Probabilistic Explanation. *Philosophy of Science* 35, 116–133 (1968)
3. Salmon, W.C.: Four Decades of Scientific Explanation. University of Minnesota Press, Minneapolis (1990)
4. Tan, Y.H.: Is default logic a reinvention of inductive-statistical reasoning? *Synthese* 110, 357–379 (1997)
5. Krantz, D.H., Luce, R.D., Suppes, P., Tversky, A.: Foundations of measurement, vol. 1, 2, 3, p. 577 (1971) p. 493 (1986) p. 356 (1990); Acad. press, New York
6. Williamson, J.: Probability logic. In: Gabbay, D., Johnson, R., Ohlbach, H.J., Woods, J. (eds.) Handbook of the Logic of Inference and Argument: The Turn Toward the Practical. Studies in Logic and Practical Reasoning, vol. 1, pp. 397–424. Elsevier, Amsterdam
7. Vityaev, E.E., Kovalerchuk, B.Y.: Empirical Theories Discovery based on the Measurement Theory. *Mind and Machine* 14(4), 551–573 (2004)
8. Vityaev, E.E.: The logic of prediction. In: Proceedings of the 9th Asian Logic Conference Mathematical Logic in Asia, Novosibirsk, Russia, August 16–19, 2005, pp. 263–276. World Scientific, Singapore (2006)
9. Halpern, J.Y.: An analysis of first-order logic of probability. In: Artificial Intelligence, vol. 46, pp. 311–350 (1990)

10. Nilsson, N.J.: Probability logic. *Artif. Intell.* 28(1), 71–87 (1986)
11. Ng, R.T., Subrahmanian, V.S.: Probabilistic reasoning in Logic Programming. In: *Proc. 5th Symposium on Methodologies for Intelligent Systems*, pp. 9–16. North-Holland, Knoxville (1990)
12. Ng, R.T., Subrahmanian, V.S.: Probabilistic Logic Programming. *Information and Computation* 101(2), 150–201 (1993)
13. Kovalerchuk, B.Y., Vityaev, E.E.: *Data Mining in finance: Advances in Relational and Hybrid Methods*, p. 308. Kluwer Academic Publishers, Dordrecht (2000)
14. Kovalerchuk, B.Y., Vityaev, E.E., Ruiz, J.F.: Consistent and Complete Data and "Expert" Mining in Medicine. In: *Medical Data Mining and Knowledge Discovery*, pp. 238–280. Springer, Heidelberg (2001)
15. Vityaev, E.E., Kovalerchuk, B.Y.: Data Mining For Financial Applications. In: Maimon, O., Rokach, L. (eds.) *Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers*, pp. 1203–1224. Springer, Heidelberg (2005)
16. Vityaev, E.E.: Semantic approach to knowledge base creating. Semantic probabilistic inference of the best for prediction PROLOG-programs by a probability model of data Logic and Semantic Programming, Novosibirsk. *Computational Systems* 146, 19–49 (1992) (in Russian)
17. Vityaev, E.E.: Knowledge inductive inference. Computational cognition. Cognitive process modelling, p. 293. Novosibirsk State University Press, Novosibirsk (2006) (in Russian)
18. Smerdov, S.O., Vityaev, E.E.: Probability, logic & learning synthesis: formalizing prediction concept. *Siberian Electronic Mathematical Reports* 9, 340–365 (2009)
19. Scientific Discovery, <http://www.math.nsc.ru/AP/ScientificDiscovery>