Proceedings of the IASTED International Conference
Computational Intelligence (CI 2009)
August 17 - 19, 2009 Honolulu, Hawaii, USA

NEW DEFINITION OF PREDICTION WITHOUT LOGICAL INFERENCE

Evgenii E. Vityaev
Department or mathematical logic
Sobolev Institute of Mathematics of the Russian Academy of Sciences
630090, Novosibirsk, Russia
email: vityaev@math.nsc.ru

Stanislav O. Smerdov
Novosibirsk State University
630090, Novosibirsk, Russia

ABSTRACT

Predictions are very important for many Artificial Intelligence tasks and systems, such as expert systems, decision support systems, control systems and robotics. But prediction notion encounters with some deep problems are to be solved yet. We will consider Deductive-Nomological (D-N) and Inductive-Statistical explanations/predictions. D-N explanations/predictions are treated as predictions in accordance with 'The Logic of Scientific Discovery' by K. Popper [1]. According to this work we cannot apply D-N explanations/predictions to inductively obtained knowledge. We argue that logical inference of predictions from inductively obtained knowledge induces some problems relating to probability and logic synthesis. To avoid this complications we propose an inductive inference of predictions without logical inference. We will define an inductive inference of predictions (Semantic Probabilistic Inference (SPI)) and a p-prediction. For any literal A p-prediction inductively infers a rule, that predicts this literal with estimation no less than the corresponding estimations obtained by probabilistic logic or probabilistic logic programming. Moreover, we prove that inductively inferred rules possess many important properties: for example, predictions based on these rules are free from the problem of statistical ambiguity. Finally, we will mention the program system 'Discovery', implementing SPI, which was successfully applied for solution of many practical tasks (see www.math.nsc.ru/AP/ScientificDiscovery)

KEY WORDS

Prediction, Probabilistic Inference, Probabilistic Logic Programming

1. Introduction

1.1 Problems of inductive-statistical inference

The definition of explanations/predictions was introduced by Hempel in the early sixties in his article 'Aspects of Scientific Explanation' (see Hempel [2, 3], and Salmon [4] for a historical overview). He distinguished two types of explanations: *Deductive-Nomological* (D-N explanations) and *Inductive-Statistical* (I-S explanations). The first type explanations use *deterministic* laws, while the second type laws employs *statistical*.

Note that D-N explanations of future events are considered as predictions according to K. Popper [1], but I-S explanations are not considered as predictions in [1]. Inductively inferred knowledge is of statistical character. If we consider this knowledge as hypotheses and, hence, as deterministic knowledge, then we can use a logical inference to obtain explanations/predictions and verification of this knowledge, but calculation of its (probabilistic) estimations have no sense (in accordance with [1]).

Another problem of I-S explanations/predictions concerning the probability and logic synthesis, which is discussed at the series of workshops on "Combining Probability and Logic" [5, 6]. The problem appears in calculation of inferred explanations/predictions estimations (probabilistic, evidence support, confirmation). The calculation either needs a rather strong assumptions like independence or a priori distributions [7] or leads to unregulated decries of estimations during the inference, like in probabilistic logic [8], 'quantitative deductions' [8], [9] or probabilistic logic programming [10].

We agree with [1] that a logical inference may be applied only to infer explanations/predictions from hypothetical knowledge. We can't apply logical inference to inductively inferred knowledge that leads to problems discussed at "Combining Probability and Logic" and paradoxes of confirmation.

This consideration brings up a problem: how should prediction based on inductively inferred knowledge be performed without logical inference?

In this paper we introduce an inductive inference of predictions (Semantic Probabilistic Inference (SPI), see definition 12, [11], [12], [13]), that infer predictions without logical inference. For any literal A it inductively infer such a maximum specific rule (see explanation in the next section) that predicts this literal with estimation no less than the corresponding estimations obtained by probabilistic logic or probabilistic logic programming (see theorem 5).

Predictions in probabilistic logic or probabilistic logic programming for inductively inferred knowledge is usually obtained in tree steps: (1) an inductive inference of knowledge; (2) logical inference of predictions from this knowledge; (3) estimations of the inference, using probabilistic logic.

Introduced inductive inference allows us to predict in one step: for every literal A a special rule that predicts this literal in the most specific way is inferred inductively. We will prove that this special rule possess the following properties:

- it is maximally specific (uses all available information that is relevant to prediction, see definition 13);
- it is the best for prediction of this literal by means of available data in probabilistic sense (see definition 15);
- its estimation (conditional probability) is no less than the corresponding estimation obtained by probabilistic logic, when prediction is performed in three steps (see theorem 5);
- this rule avoid the statistical ambiguity problem (see theorem 4).

The introduced Semantic Probabilistic Inference combines probability and logic for predictions as follows [14]:

- truth values (for closed formulas, rules, facts, etc) are generalized to probabilistic ones as in ”quantitative deductions” [9] or probabilistic logic programming [10];
- logical inference is generalized to semantic inference based on probabilities as truth values;
- the actual purpose of the inference is not to keep truth values, but to increase the conditional probability estimations of this rules (which are valid for accessible data);
- obtained as a result of the SPI the maximal specific rule solves the problem of statistical ambiguity and allows us to make consistent (in logical sense) predictions (see theorem 4).

1.2 Statistical ambiguity problem

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 the I-S explanations [2, 3]. 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. He introduced the property of the maximal specific statistical laws as the Requirement of Maximal Specificity (RMS). Hempel proposed the following formalization of I-S inference:

$$\cfrac{\cfrac{L_1, \ldots, L_{\mathrm{m}}}{C_1, \ldots, C_{\mathrm{n}}}}{G}[r]$$

It satisfies the following conditions:

- $\mathrm{L}_1,\ldots,\mathrm{L}_\mathrm{m}$, $\mathrm{C}_1,\ldots,\mathrm{C}_\mathrm{n} \vdash \mathrm{G}$;
- $\mathrm{L}_1,\ldots,\mathrm{L}_\mathrm{m}$, $\mathrm{C}_1,\ldots,\mathrm{C}_\mathrm{n}$ are consistent;
- $\mathrm{L}_1,\ldots,\mathrm{L}_\mathrm{m} \nvdash \mathrm{G}$; $\mathrm{C}_1,\ldots,\mathrm{C}_\mathrm{n} \nvdash \mathrm{G}$;
- $\mathrm{L}_1,\ldots,\mathrm{L}_\mathrm{m}$ are statistical quantified formulas.
- $\mathrm{C}_1,\ldots,\mathrm{C}_\mathrm{n}$ are quantifier-free;
- RMS: All laws $\mathrm{L}_1,\ldots,\mathrm{L}_\mathrm{m}$ are maximal specific.

In Hempel's works [2, 3] the RMS is defined as follows. An I-S argument of the form:

$$\cfrac{\cfrac{p(G;F)}{F(\boldsymbol{a})}}{G(\boldsymbol{a})}[r]$$

is an acceptable I-S explanation with respect to a ”knowledge state” K and statistical law $F(x) \Rightarrow G(x)$ with probability $p(G;F) = r$, if the following Requirement of Maximal Specificity is satisfied. For any class H if the following two sentences are contained in K

$$\forall x(H(x) \Rightarrow F(x)),\ H(\boldsymbol{a}), \tag{1}$$

there exists a statistical law $H(x) \Rightarrow G(x)$ with probability $p(G;H) = r'$ in K such that $r = r'$. The basic idea of RMS is that if both F and H contain the object ***a*** and H is a subset of F, then H provides more specific information about the object ***a*** than F, and therefore the law $p(G;H)$ should be preferred over the law $p(G;F)$. But, if any more preferred law has the same probability r as the law $F(x) \Rightarrow G(x)$ then it is maximaly specific - it can't be made more precise.

The Semantic Probabilistic Inference may be considered as the maximal specific rules inference because if the condition (1) is satisfied for some sentence H&F, but equality p(G; H&F) = r is not, it choose the rule

$$H(x)\&F(x) \Rightarrow G(x)$$

as the next step of the SPI (see details in definition 12). It means that with the SPI we add to the rules any additional conditions H that increases the (conditional) probability r of these rules.

The statistical ambiguity problem occurs in AI in different forms, but it hasn't been solved hitherto. So it leads us to the following questions:

- how to define the RMS that solves the statistical ambiguity problem?
- is it possible to define RMS that theory of sentences satisfying RMS be logically consistent?

The stated problems are very important, because their solution means consistency of predictions produced by different AI systems: expert systems, knowledge bases, robotics, intelligent data analysis and etc.

In this paper solution is presented: we define the Requirement of Maximal Specificity (RMS) and the set of

Maximum Specific Rules (MSR), inferred by SPI, so that we can prove, that sentences from the MSR satisfy RMS (see theorem 3) and that the MSR-set is logically consistent (see theorem 4).

2. Probabilistic Herbrand models

We perform all considerations in the frame of logical programming with functional symbols. Consider a first order language L with the equality of the finite signature

$\Omega = \langle P_1, P_2, ..., P_{n_1}, f_1, f_2, ..., f_{n_2}, c_1, c_2, ..., c_{n_3} \rangle$.

Let X denote a countable set of *variables*, T_L — be set of *terms*, U_L — be a set of all *ground terms* (without free variables), A_L — be a set of *literals* (atoms and negated atoms), B_L — be a set of all *ground literals*, F_L — be a set of *formulas*, G_L — be a set of *formulas without quantifiers*, S_L — be a set of *sentences* (formulas without free variables), $\Re = G_L \cap S_L$ — a set of all *ground sentences* of the signature Ω.

An arbitrary mapping $\theta : X \to T_L$ is called a *substitution*. The set of all substitutions is denoted by Θ. The substitution θ(x) = x is called *identical*. Substitutions are naturally extended to arbitrary expressions. Thus, substitutions for term $t = f(t_1, ..., t_n)$ and literal $A = P(t_1, ..., t_n)$ are equal to $t\theta = f(t_1\theta, ..., t_n\theta)$ and $A\theta = P(t_1\theta, ..., t_n\theta)$ respectively. This remark can be also applied to constructions of *rule*, *fact* and *query* (arriving from logic programming). A rule $A\theta \leftarrow A_1\theta, ..., A_n\theta$, where θ is rearrangement of the set X, is called a *variant* of the rule $A \leftarrow A_1, ..., A_n$.

Following [15], let us define the probability μ on a non-empty subset $F' \subset \Re$ closed with respect to logical operations $\&, \vee, \neg$.

Defiition 1 . A mapping $\mu : F' \to [0, 1]$ is called a *probability* on F', 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 2 A pair $M = \langle U_L, \mu \rangle$, where μ is a probability on $\Re$, called a *probabilistic Herbrand model* of the signature Ω.

Definition 3. A pair $M = \langle U_L, I \rangle$, where $I : B_L \to \{0, 1\}$, is called a *Herbrand model* of the 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 $\Re$. Then we define the set D of subsets $G(\phi) = \{M | M \in G, M \models \phi\}$, $\varphi \in \Re$. Here $\models$ states for satisfaction of considered sentence on the model.

Definition 4. A class of Herbrand models G is said to be *coordinated with the probability μ on the set of formulas F'* (a probabilistic Herbrand model $M = \langle U_L, \mu \rangle$ is said to be a *probabilistic model of the class* G), if $\mu(\phi) = 0$ immediately follows from $G(\phi) = \emptyset$, where $\phi \in F'$.

We will also consider extension of μ for free variables: assume $\mu(\varphi) = \inf\limits_{\theta \in \Theta G} \{\mu(\varphi\theta)\}$, where $\varphi \in G_L$, ΘG — a set of all ground substitutions for variables.

3. Logical programs.

Let PR denote a set of all *rules* $A \leftarrow A_1, ..., A_k$, $k \geq 0$, $A_i \neq A$, $i = 1, ..., k$ of the signature Ω, where $A, A_1, ..., A_k$ are literals of the signature Ω. If the literal A is absent, then the rule $\leftarrow A_1, ..., A_k$ $(k > 0)$ is called a *goal* (or *query*). In queries we will write '& ' between literals 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 us fix a *selection rule R*, which selects one of the literals from a query. Let $N = \leftarrow A_1\&...\&A_i\&...\&A_k$, $k \geq 1$ be a query, where the rule R selects the literal A_i, and a rule $C = A \leftarrow B_1, ..., B_m$ be a variant of some rule of the program Pr, where all the variables are different from those of the query. Let θ be the most general unification of the literals A_i and A $(A_i\theta = A\theta)$. Then new query

$$\leftarrow (A_1\&...\&A_{i-1}\&B_1\&...\&B_m\&A_{i+1}\&...\&A_k)\theta \text{ , if } m \geq 1$$
$$\text{and} \leftarrow (A_1\&...\&A_i\&...\&A_k)\theta, \text{ if } m = 0$$

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

The set of all possible queries of the signature Ω with the given relation of inference is called a *calculation space for the program Pr and the selection rule R*. The maximal sequence of queries $N = N_0, N_1, N_2, ...$ together with the sequence of rules $C_0, C_1, C_2, ...$ and unifications $\theta_0, \theta_1, \theta_2, ...$ such, that every query N_{i+1} is inferred from the query N_i by means of the rule C_i, substitution θ_i and the selection rule R $(i = 1, 2, ...)$ is called an *SLDF-inference* (Linear resolution with Selection rule for Definite clauses and underlined Facts) of the goal N in a given calculation space. Note that each SLDF-inference is a maximal way in the calculation space starting with N. A finite SLDF-inference, for which the final query only consists of underlined literals, 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 viewed as a tree (a prefix tree of SLDF-inferences). This tree is called an *SLDF-tree of the request N calculations*. A SLDF-tree containing at least one successful SLDF-inference is called a *successful SLDF-tree*.

4. 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 $N, N_1, ..., N_k$ of the request N in a calculation space of the program Pr obtained by means of the sequence of rules $C_0, C_1, ..., C_{k-1}$, unifi-

cation's $\theta_0, \theta_1, ..., \theta_{k-1}, \theta \leftrightharpoons \theta_0\theta_1...\theta_{k-1}$ and the selection rule R.

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

The probability of the rule $C = A \leftarrow B_1, ..., B_m,\ m \geq 1$ is defined if $\mu(B_1\&...\&B_m) \neq 0$ and equal to $\mu(C) = \mu(A|B_1\&...\&B_m) = \mu(A\&B_1\&...\&B_m)/\mu(B_1\&...\&B_m)$ and undefined otherwise. Represent facts $A \leftarrow$ by the rules $A \leftarrow true$. Then, $\mu(C) = \mu(A|true) = \mu(A)$. Suppose that $\mu(C)$ means that the probability is defined. Denote as $PR_0 \subseteq PR$ a set of all rules, for which the probability μ is defined $Pr_0 \leftrightharpoons PR_0 \cap \mathrm{Pr}$.

Definition 5 The *rule C is true on the Herbrand model* $N \in 2^{B_L}$ $(N \models C)$ iff it is true on N under any state (for any mapping ρ: X $\rightarrow$ U).

Definition 6. The *program Pr is true on the Herbrand model* $(N \models \mathrm{Pr})$ iff each rule of the program is true on N.

Definition 7. The *program Pr is true on the class of models G* iff $\forall N \in G, N \models \mathrm{Pr}$.

We will write down $C \in F'$ for the rule $C \leftrightharpoons A \leftarrow B_1\&...\&B_m, A, B_1, ..., B_m \in F'$.

Proposition 1. If $C \in \mathrm{Pr} \cap F'$, $C = A \leftarrow B_1, ..., B_m, \mu(B_1\&...\&B_m) > 0$, then

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

Corollary 1. 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 $\mu(C) = 1$, $C \in \mathrm{Pr} \cap F'$, if it is defined.

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

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

Theorem 1. If $\mu(B^i\theta) > 0$, $i = 0, ..., k-1$, then under the above conditions

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

Corollary 5. If $\mu(B^i\theta) > 0$, $i = 0, ..., k-1$, then under the above conditions

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

For every successful SLDF-inference $N\theta = N_0\theta, N_1\theta, ..., N_{k-1}\theta, N_k$, there exists a SLDF$'$-inference $N\theta = N_0'\theta, N_1'\theta, ..., N_i'\theta, , ..., N_{k-1}', N_k' = N_k$, where facts are used in the last turn and the rules $C_j\theta$ with $k_j \geq 1$; $j = 1, ..., i-1$ are applied before facts. Then the request $N_i'\theta$ has the form $\leftarrow A_1', ..., A_s'$, and the request N_k - the form $\leftarrow A_1', ..., A_s'$. Such a SLDF$'$-inference we call *normalized*.

Theorem 2. If $\mu(B^j\theta) > 0$, j = 0,1, ..., i-1, and $\mu(N_kF^\wedge) > 0$, then for the successful SLDF-inference, defined earlier

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

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

Let us define the probability estimations $\nu(N)$, $\eta(N)$ of the calculation space requests for the program Pr and selection rule R. We will 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 $\{N\theta^i\}_{i\in I}$.

Let us determine the estimations $\{\nu^i\}_{i\in I}$, which are equal to the right-hand side of inequalities of the 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 the corresponding inferences. Determine also the estimates $\{\eta^i\}_{i\in I}$, which are equal to the right-hand side of inequalities of theorem for the conditional probabilities $\mu(N\theta^{i\wedge}|N_{ki}^iF^\wedge) \geq \eta^i(\forall i \in I)$ of the requests $\{N\theta^i\}_{i\in I}$. Define $\nu(N) = \sup\limits_{i\in I}\{\nu^i\}$, $\eta(N) = \sup\limits_{i\in I}\{\eta^i\}$.

The SLDF-inference of the request N, where the estimation η(N) is reached, we call a *prediction of the request N*. The value η(N) we call the *estimation of the request N prediction*. If the prediction is not defined, then the estimation of the prediction η(N) is not defined.

5. Inductive synthesis of probabilistic logic programs

A full set of facts for the class of models G is represented by collection of sets $F(N) = \{A|\ N \models A$ for any ground literal $A\}$, where $N \in G$. Any finite collection D of finite subsets $D(N) \subseteq F(N)$ is called *data*. A probabilistic Herbrand model $M = \langle U_L, \mu\rangle$, which is coordinated with the class G, we call a *probabilistic Herbrand model of data* $D = \{D(N)|N \in G\}$.

How should the rules $C = A \leftarrow B_1, ..., B_m$ $(m \geq 1)$ be used for predictions? If in the case of ground substitution $\theta \in \Theta$ the conjunction $(B_1\&...\&B_m)\theta$ is true on a certain model N, which is chosen randomly from G in accordance with the measure μ, i.e. $\{B_1\theta, ..., B_m\theta\} \subseteq F(N)$, then the conclusion $A\theta$ is true on N with the probability $\mu(A\theta|(B_1\&...\&B_m)\theta) \geq \mu(A|B_1\&...\&B_m) = \mu(C)$.

Thus, probability $\mu(C)$ for the rules with free variables gives a lower (or equal) boundary for prediction of the literal $A\theta$. Note that only one model N, chosen arbitrary from G, and the corresponding data $D(N)$ should be used for predictions in a particular model.

Definition 8. For two rules $C = A \leftarrow B_1, ..., B_m$ and $C' = A' \leftarrow B'_1, ..., B'_{m'}$ $(m, m' \geq 0)$ relation $C' \rhd C$ (*to be more common*) takes place iff there exists a substitution $\theta \in \Theta$ such that $A'\theta = A$, $\{B'_1\theta, ..., B'_{m'}\theta\} \subseteq \{B_1, ..., B_m\}$ and C is not a variant of C'.

Definition 9. A *probabilistic inference relation* $C' \sqsubset C$ is holds iff $C' \rhd C$ and $\mu(C') < \mu(C)$.

Definition 10. A *probabilistic regularity* (or *P-rule*) is such a rule $C \in PR_0$ that for any other C' from PR_0: if $C' \rhd C$, then $C' \sqsubset C$.

Denote $PR(M)$ is the set of all P-rules; $P(M) \subseteq PR(M)$ is the set of all P-rules with premises containing at least one literal.

Definition 11. We will refer to the set $PR(M,N) = P(M) \cup D(N)$, where $D(N) \in D$ and N is a certain model chosen arbitrary from G in accordance with the measure μ, as a *probabilistic logic program synthesized inductively by the data* $D(N)$ *and the probabilistic model* M.

6. Semantic probabilistic inference

Definition 12. A maximal sequence of rules $C_1 \sqsubset C_2 \sqsubset ...$, where $C_i = (A_i \leftarrow B^i_1, ..., B^i_k) \in P(M)$ $(i = 1, 2, ...)$, such that the literal A is unified with all literals $A_1, A_2, ...$, we call a *semantic probabilistic inference* (or *P-inference*) of this A of the signature Ω. If there is no such a sequence for some literal A, then its P-inference is empty. Each P-inference produces a sequence of substitutions $\theta_1, \theta_2, ...$ from the definition of relation $\rhd$. We will refer to the substitution $\theta = \theta_1\theta_2...$ as the *result of semantic probabilistic inference* (calculation). The final rule of the finite P-inference is called the *resulting rule*.

Unlike probabilistic logics (see [8] and [9]) the probability of rules strictly increases during the process of P-inference. Let us consider the set of all P-inferences of an literal A. This set constitutes the *semantic probabilistic inference tree (SPI-tree)* of this literal.

7. The solution of the statistical ambiguity problem

Now we will define the Requirement of Maximal Specificity (RMS). In this section we will consider the language L without functional symbols. Suppose that the class H of objects in (1) is defined by some sentence $H \in \Re$ in L.

Defiition 13 . By the maximum specific rule MS(A) for the P-inference of the literal A we mean such resulting rule (if it exists) of the SPI-tree for A, that has the maximum value of conditional probability. We define the set of all maximum specific rules as MSR.

Definition 14. The *Requirement of Maximal Specifiity* : if we add any sentence $H \in \Re$ to the premise of the rule $(F \Rightarrow G)$, $\mu(G|F) = r$, such that $F(\mathbf{a})\&H(\mathbf{a})$ for an object $\mathbf{a}$, then the new rule $(F\&H \Rightarrow G)$ will have the same probability $\mu(G|F\&H) = \mu(G|F) = r$.

In other words the requirement RMS means that there is no other sentence H in $\Re$ that increases (or decreases, see lemma 2 below) the conditional probability $\mu(G|F) = r$ (by adding it to premise).

Lemma 2. If the sentence $H \in \Re$ decreases the probability $\mu(G|F\&H) < \mu(G|F)$ then its negation $\neg H$ increases it: $\mu(G|F\&\neg H) > \mu(G|F)$.

Lemma 3. For any rule $C = (A_0 \leftarrow B_1, ..., B_t$, $\mu(B_1\&...\&B_t) > 0$ there is a P-rule $C' = (A_0 \leftarrow A_1, ..., A_k$, which is subrule of C and $\mu(C') \geq \mu(C)$.

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

Assume that rules in MSR are all standardized apart: any two $C_1, C_2 \in MSR$ have no common variables.

Theorem 4. For any theory $Th \subset MSR$ there exist a substitution $\theta \in \Theta$ such that I-S inference is consistent with $Th\theta$.

So, we can predict without contradictions if only elements of the MSR-set are used as statistical laws in I-S inference.

8. Predictions based on the semantic probabilistic inference

Let $D^*(N) = \{\neg A | A \in D(N)\} \cup D(N)$ be *a complete collection of alternatives*.

Definition 15. By a *P-prediction* of some literal A of the signature Ω by the program $PR(M,N) = P(M) \cup D(N)$ we mean such a P-inference $C_1 \sqsubset ... \sqsubset C_i \sqsubset ... \in P(M)$ of the goal A, where:

1. There exists a rule $C_i = A_i \leftarrow B^1_i, ..., B^l_i$ and a substitution θ such that $\{B^1_i\theta, ..., B^l_i\theta\} \subseteq D^*(N)$, $A\theta = A_i\theta$; $\mu\{B^1_i\theta\&...\&B^l_i\} > 0$ and $\mu(A\theta) < \mu(C_i)$;

2. A maximum of conditional probability among all rules, satisfying the condition 1, of all P-inferences of the goal A is attained by the rule C_i , ;

3. If there is no a P-inference of the goal A or there is no required substitution, then a P-prediction is not defined;

4. The substitution θ for C_i (from the first condition) is called the *result of P-prediction*. The value $\eta_p(A) = \mu(C_i)$ we call the *estimation of P-prediction*. If P-prediction is not defined, then estimation value $\eta_p(A)$ is not defined.

Proposition 2. Rules satisfying the point 2 are not comparable with respect to relation $\rhd$.

Lemma 4. A P-prediction (of some literal A) is defined iff there exists at least one rule$C \in P(M)$ satisfying the fist condition of definition.

Lemma 5. Let the rule $C \in PR_0$ satisfies the point 1 of definition and has at least one literal in the premise, then either C is P-rule $C \in P(M)$ or there exists a rule

$C' \in P(M)$, $C' \rhd C$, $\mu(C') \geq \mu(C)$ satisfying the point 1 of definition.

Corollary 6. P-prediction is defined iff there exists at least one rule $C \in PR_0$ (with at least one literal in the premise) that satisfies the point 1 of definition of P-prediction.

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

Theorem 5. If literal A is predicted by the program Pr with estimation $\eta(A) > \mu(A\theta)$ (where $\theta \in \Theta G$ is corresponding substituition for SLDF-prediction), then it is P-predicted by the program $PR(M, N)$ with P-prediction estimation value $\eta_p(A) \geq \eta(A)$.

It follows from the definition that there is $C_i\theta$ (where theta is a result) – a maximal specific regularity for A, given available data, – such that $\eta_p(A) = \mu(C_i\theta)$. Note, we predict a property A according to $C_i\theta$ only if premise of $C_i\theta$ is contained in $D(N)$ (it also might include negated literals). Thus, we have an opportunity to eliminate contradictory predictions appearing in the case of SLD-constructions.

9. The Relational Data Mining and program system 'Discovery'

The Relational Data Mining (RDM) approach to the intensive area of applications – Knowledge Discovery in Data Bases and Data Mining (KDD&DM) – was developed on the basis of the semantic probabilistic inference [12], [13], [16], [17]. The program system 'Discovery', which utilizes this approach, was implemented. This system realizes the SP-inference and is able to discover the sets of rules $PR(M)$, MSR. In [12], [13] we prove that using RDM we may cognise the object domain. The system 'Discovery' has been successfully applied for solving many practical tasks in such fields as the cancer diagnostic systems, time series forecasting, psychophysics, bioinformatics, and many others (see www-site Scientific Discovery [18]).

Acknowledgments

The work is partially supported by the Scientific Schools grant of the President of the Russian Federation 335.2008.1; Russian Science Foundation grant 08-07-00272 and Integration projects of the Siberian Division of the Russian Academy of science 47, 111, 119.

References

[1] Karl Popper. The Logic of Scientific Discovery, London, 1959.

[2] Hempel, C. G. Aspects of Scientific Explanation, in C. G. Hempel, *Aspects of Scientific Explanation and other Essays in the Philosophy of Science*, The Free Press, New York, 1965.

[3] Hempel, C.G. Maximal Specificity and Lawlikeness in Probabilistic Explanation, *Philosophy of Science*, 35, 1968, 116-133.

[4] Salmon, W. C. *Four Decades of Scientific Explanation* (University of Minnesota Press, Minneapolis, 1990).

[5] Special issue on Combining Probability and Logic. *Journal of Applied Logic*, 1(3-4), 2003, 135-308.

[6] Combining Probability and Logic. *Journal of Logic, Language and Information*. 15(1-2), Jule 2006.

[7] Kristian Kersting and Luc De Raedt, Bayesian Logic Programs, *Proceedings of the Work-in-Progress Track at the 10th International Conference on Inductive Logic Programming*,2000, 138-155.

[8] Nils J. Nillson, Probability logic, *Artif. Intell.*, 28(1), 1986, 71-87.

[9] R.T.Ng, V.S.Subrahmanian, Probabilistic reasoning in Logic Programming, in *Proc. 5th Symposium on Methodologies for Intelligent Systems*, Knox-ville, North-Holland, 1990, 9-16.

[10] Raymond Ng and V.S. Subrahmanian. Probabilistic Logic Programming, *Information and Computation*, 101(2),1993, 150-201.

[11] 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, *Comp. Syst.*, 146, Novosibirsk, 1992, 19-49. (Russian);

[12] Evgenii Vityaev, Boris Kovalerchuk, Empirical Theories Discovery based on the Measurement Theory. *Mind and Machine*, 14(4),2004, 551-573.

[13] Evgenii Vityaev. The logic of prediction, in *Mathematical Logic in Asia*. Proceedings of the 9th Asian Logic Conference (August 16-19, 2005, Novosibirsk, Russia), World Scientific, Singapore, 2006, 263-276.

[14] E. Vityaev Logic, probability and learning synthesis in the semantic probabilistic inference, in *Proceedings of the 4-th international conference "Integrated models and soft computing in artificial intelligence"* (28-30 may, Colomna, 2007), 1, 133-140. (Russian)

[15] Halpern, J.Y. An analysis of first-order logic of probability, *Artifiial Intelligence* 46, 1990, 311-350.

[16] Kovalerchuk, B., Vityaev, E., *Data Mining in finance: Advances in Relational and Hybrid Methods* (Kluwer Academic Publishers, 2000, 308p).

[17] Kovalerchuk, B., Vityaev, E., Ruiz, J.F., Consistent and Complete Data and ”Expert” Mining in Medicine, in *Medical Data Mining and Knowledge Discovery*, Springer, 2001, 238-280.

[18] Website devoted to Scientific Discovery http://www.math.nsc.ru/AP/ScientificDiscovery