[10] Hughes G.F. On the mean accuracy of statistical pattern recognizers // IEEE Trans. Inform. Theory. 1968. V. IT-14, N 1. P. 55-63.

[11] W. Buntine. Learning classification trees // Statistics and Computing. 1992. V. 2. P. 63--73.

[12] Lbov, G.S., Startseva, N.G., *About statistical robustness of decision functions in pattern recognition problems.* Pattern Recognition and Image Analysis, 1994. Vol 4. No.3. pp.97-106.

[13] Berikov V.B., Litvinenko A.G. *The influence of prior knowledge on the expected performance of a classifier.* Pattern Recognition Letters, Vol. 24/15, 2003, pp. 2537-2548.

[14] Berikov, V.B. A Priori Estimates of Recognition Accuracy for a Small Training Sample Size // Computational Mathematics and Mathematical Physics, Vol. 43, No. 9, 2003. pp. 1377- 1386

[15] Lbov G.S., Berikov V.B. Stability of decision functions in problems of pattern recognition and heterogeneous information analysis. Inst. of mathematics Press, Novosibirsk. 2005. (in Russian)

[16] Berikov V.B. Bayes estimates for recognition quality on a finite set of events // Pattern Recognition and Image Analysis. 2006. V. 16, N 3. P. 329-343.

Author's Information

Vladimir Berikov – Sobolev Institute of Mathematics SD RAS, Koptyug pr.4, Novosibirsk, Russia, 630090; e-mail: berikov@math.nsc.ru

MEASURE REFUTATIONS AND METRICS ON STATEMENTS OF EXPERTS (LOGICAL FORMULAS) IN THE MODELS FOR SOME THEORY[1]

Alexander Vikent'ev

Abstract. The paper discusses a logical expert statements represented as the formulas with probabilities of the first order language consistent with some theory T. Theoretical-models methods for setting metrics on such statements are offered. Properties of metrics are investigated. The research allows solve problems of the best reconciliation of expert statements, constructions of decision functions in pattern recognition, creations the bases of knowledge and development of expert systems.

Keywords: pattern recognition, distance between experts' statements.

ACM Classification Keywords: I.2.6. Artificial Intelligence - Knowledge Acquisition.

Introduction

As the increasing interest to the analysis of the expert information given as probabilities logic statements of several experts is now shown, questions on knowledge of the experts submitted by formulas of the first order language with probabilities are interesting also. With the help of suitable procedure the statement of experts it is possible to write down as formulas of Sentence Logic or formulas of the first order language. Clearly, the various statements of experts (and the formulas appropriate to them) carry in themselves different quantity of the information. To estimate and analyses this information it is necessary to define the degree of affinity of statements that allows to estimate a measure of refutation statements of experts (a measure of refutation above at formula of the first order language with smaller number of elements satisfying it) and to specify their probabilities (an average share probabilities realizations for formulas with variables). It allows solve problems of the best reconciliation of expert statements, constructions of decision functions in pattern recognition, creation of bases of knowledge and expert systems [1].

[1] This work was financially supported by the Russian Foundation for Basic Research, project no. **04-01-00858a**.

A number of natural metrics on probabilities knowledge of experts is offered with use of suitable class of models (with metrics) some theory and modifications symmetric difference, by analogue, par example [4] for logical Lbov's the predicat for unique model. Properties of these metrics, connected to them measures of refutation of formulas (distance from the formula up to class of equivalence of identically true formula) and probability are established. From the point of view of importance of the information presented by an expert, it is natural to assume that a measure of refutation of the formula (nonempty predicate) the above, than it is les measure of elements satisfying it (i.e. a measure determined on subsets, set by predicate formulas).

We introduce the measure of refutation similarly to a case of formulas without probabilities. We call

$$R\left(P\left(\overline{x}\right)\right)=\rho_{вер}\left(P\left(\overline{x}\right),\ 1\right)$$

the measure of refutation of formula $P\left(\overline{x}\right)$, where 1 is an identical true predicate, that is, $\overline{x}=\overline{x}$. All stated for predicates (and Lbov's predicate aussi without probability) fairly and for formulas of the first order language with probabilities.

The distance between the formulas of Sentence Logic is entered in [1], properties of the entered distance are given and proved in the same place. Ways of introduction of distance between the formulas of the first order language are offered in [2]. Measures of refutation and probabilities of formulas are entered and their properties are formulated in [3]. The distance between the formulas of Sentence Logic with probabilities is entered in [3,5]. In the given work the way of introduction of distances between probabilities statements of experts represented as the formulas of the first order language theory T with probabilities is offered.

Distance between statements of experts represented as the formulas of the first order language with probabilities in theory

Let experts speak about probabilities of predicates on the product $\prod_{j=1}^{p} D_{x_j}$.

Then the given by expert probability is interpreted as follows: "the knowledge" $B_l^i=<P_l^i(x_1,...,x_p),p_l^i>$ means, that the predicate $P_l^i(x_1,...,x_p)$ is true on $n_{P_l^i}=\lfloor n\cdot p_l^i\rfloor$ trains of length p in model M_i , where $n=\prod_{j=1}^{k}\left|D_{x_j}\right|$ - measure of model.

Let's find distance between predicates P_l and P_j . For this purpose all over again we shall calculate distance $\rho^i(B_l^i,B_j^i)$ between probabilities interpretations $B_l^i=<P_l^i(\overline{x}),p_l^i>$ and $B_j^i=<P_j^i(\overline{x}),p_j^i>$ of predicates P_l and P_j in each model M_i . Distances are calculated between predicates of identical district and from the same variables plus measure protjaga (разброса в пространстве модели) p.e. [4], and without stable theory.

Then interpreting the probabilities given by experts the described above way we receive that the predicate $P_l^i(\overline{x})$ is true on $n_{P_l^i}$ trains in model M_i and the predicate $P_j^i(\overline{x})$ is true on $n_{P_j^i}$ trains in model M_i *theory* T. We shall note that is not known on what trains each predicate true and number (or mera) of trains on which these predicates are simultaneously true.

We shall consider the following task. Let the predicate $P_l^i(\overline{x})$ is true on $n_{P_l^i}$ trains in model M_i and the predicate $P_j^i(\overline{x})$ is true on $n_{P_j^i}$ trains in model M_i and k^i - number of trains on which these predicates are simultaneously true. It is required to calculate distance between $B_l^i=<P_l^i(\overline{x}),p_l^i>$ and $B_j^i=<P_j^i(\overline{x}),p_j^i>$.

Distances arising in further we shall designate through $\rho_{k^i}(B_l^i,B_j^i)$, where, $k^i=t,t+1,\dots,\min(n_{P_l^i},n_{P_j^i})$, $t=\max(0,n_{P_l^i}+n_{P_j^i}-n)$ hereinafter.

Distance $\rho_{k^i}(B_l^i, B_j^i)$ we shall define as a modifies (as ask above) symmetric difference, i.e.

$$\rho_{k^i}(B_l^i, B_j^i) = \frac{1}{n}(n_{P_l^i} + n_{P_j^i} - 2k^i), \tag{1}$$

for every one $k^i = t, t+1, \ldots, \min(n_{P_l^i}, n_{P_j^i})$. All properties of distances formulated in [1] are fair for $\rho_{k^i}(B_l^i, B_j^i)$. Let's offer some ways of calculation distance $\rho^i(B_l^i, B_j^i)$ between probabilities interpretations $B_l^i =< P_l^i(\overline{x}), p_l^i >$ and $B_j^i =< P_j^i(\overline{x}), p_j^i >$ of predicates P_l and P_j in each model M_i *theory T*. If the number k^i is not known (the number of trains on which these predicates are simultaneously true in model M_i) and if there are no preferences for value k^i (preference can be stated by experts) it is possible to act as follows. We shall assume, that all values for number k^i are equally probability. Then distance between probabilities interpretations $B_l^i =< P_l^i(\overline{x}), p_l^i >$ and $B_j^i =< P_j^i(\overline{x}), p_j^i >$ of predicates P_l and P_j in model M_i we shall define as average of distances $\rho_{k^i}(B_l^i, B_j^i)$ on all values k^i, i.e.

$$\rho(B_l^i, B_j^i) = \frac{\sum_{k^i=t}^{\min(n_{P_l^i}, n_{P_j^i})} \rho_{k^i}(B_l^i, B_j^i)}{\min(n_{P_l^i}, n_{P_j^i}) + 1 - t}. \tag{2}$$

For this distance all properties of distances formulated in [1] also are executed.
If by experts it is stated what value for k^i is more preferable in quality $\rho^i(B_l^i, B_j^i)$ it undertakes $\rho_{k^i}(B_l^i, B_j^i)$, i.e.

$$\rho(B_l^i, B_j^i) = \rho_{k^i}(B_l^i, B_j^i). \tag{3}$$

In the offered formulas (1) – (3) of distances the kind of formulas between which the distance is calculated is not taken into account. Therefore it is natural to offer distance by which takes into account a kind of formulas. Applying the model approach [1-3] to elements of set $\{M_i\}_{i=1}^s$ we shall find probabilities $P_{M_i}(P_j^i)$, $P_{M_i}(P_j^i)$ ([3]) and distance $\rho_{M_i}(P_l^i, P_j^i)$ in model M_i ([2]), then we shall calculate probability $P_{M_i}(P_l^i \wedge P_j^i) = \frac{1}{2}(P_{M_i}(P_l^i) + P_{M_i}(P_j^i) - \rho(P_l^i, P_j^i))$ ([3]) and we shall find $k_0^i = [P_{M_i}(P_l^i \wedge P_j^i) \cdot n]$ - the number of trains on which predicates are simultaneously true. Having k_0^i (calculated on models), it is possible to reduce number of possible values for k^i. Three cases here are possible: 1) if $t < k_0^i < \min(n_{P_l^i}, n_{P_j^i})$, then $k^i = k_0^i - 1, k_0^i, k_0^i + 1;$ 2) if $k_0^i = t$ or $k_0^i = \min(n_{P_l^i}, n_{P_j^i})$, then, for example, $k^i = k_0^i, k_0^i + 1;$ or $k^i = k_0^i - 1, k_0^i;$ 3) if $k_0^i < t$ or $k_0^i > \min(n_{P_l^i}, n_{P_j^i})$, then $k^i = t$ or $k^i = \min(n_{P_l^i}, n_{P_j^i})$.

And already to these values for k^i apples offered above formulas (1) - (3) of distances. As required probably some expansion of number of values for k^i.
The offered ways it is possible to calculate distance between the following statements: $B_l^i =< P_l^i(\overline{x}), p_l^i >$ - the information received from one expert, and $B_l^j =< P_l^j(\overline{x}), p_l^j >$ - the information received from other expert. Thus we have calculated distance $\rho^i(B_l^i, B_j^i)$ between probabilities interpretations $B_l^i =< P_l^i(\overline{x}), p_l^i >$ and $B_j^i =< P_j^i(\overline{x}), p_j^i >$ of predicates P_l and P_j and degree elongate in each model M_i *theory* T.

Then as distance $\rho_{вер}(P_l, P_j)$ between predicates P_l and P_j we shall take size

$$\rho_{вер}(P_l, P_j) = \frac{1}{s}\sum_{i=1}^{s} \rho^i(B_l^i, B_j^i).$$

For all properties of distance formulated in ([5]) are carried out for $\rho_{вер}(P_l, P_j)$.

Acknowledgements

This work was financially supported by the Russian Foundation for Basic Research, project no. 04-01-00858a.

Bibliography

[1] Lbov G.S., Startseva N.G. Decision Logical Functions and Statistical Robustness. Novosibirsk: Izd. Inst. Math., 1999.

[2] Vikent'ev A.A., Koreneva L.N. “Setting the metric and measures of informativity in predicate formulas corresponding to the statements of experts about hierarchical objects”, *Pattern Recognition and Image Analysis*, V. 10, N. 3, (2000), 303--308.

[3] Vikent’ev A.A., Koreneva L.N. “Model approach to probabilities expert statements”, “*Mathematical Methods for Pattern Recognition – 10”, Moscow,* (2001), 25-28.

[4] G.S.Lbov, M.K.Gerasimov. Determining the distance between logical statements in forecasting problems. In: Artificial Intelligence, 2’2004 [in Russian]. Institute of Artificial Intelligence, Ukraine.

[5] Викентьев А.А., Лбов Г.С., Коренева Л.Н. “Расстояние между вероятностными высказываниями экспертов”, *Искусственный интеллект, 2’2002, НАН Украины,* 58-64.

[6] Keisler G., Chang C. Model theory. M.: Mir, 1977.

Author’s Information

Alexander Vikent’ev – Institute of Mathematics, SB RAS, Acad. Koptyuga St., bl.4, Novosibirsk, Russia; e-mail: vikent@math.nsc.ru

ANALYSIS AND COORDINATION OF EXPERT STATEMENTS IN THE PROBLEMS OF INTELLECTUAL INFORMATION SEARCH[1]

Gennadiy Lbov, Nikolai Dolozov, Pavel Maslov

Abstract: The paper is devoted to the matter of information presented in a natural language search. The method using the statements agreement process is added to the known existing system. It allows the formation of an ordered list of answers to the inquiry in the form of quotations from the documents.

Keywords: Search engine, natural language, coordination of statements, semantic graph

ACM Classification Keywords: I.2.7 Computing Methodologies – Text analysis

Introduction

Efficiency of the search engine is determined by the use of various methods of relevant documents revealing and insignificant ones eliminating, as well as methods peculiar to the specific search engine or their certain kind (for example, specialized search engines). Existing search engines are based on the oversight of index databases of

[1] This work was financially supported by **RFBF-04-01-00858**