Adaptive Analysis of Experts' Statements in Pattern Recognition[*)]

Lbov G.S. [1)], Gerasimov M.K. [2)]

1) Institute of Mathematics, SB RAS; Koptyug St., bl.4, Novosibirsk, Russia; lbov@math.nsc.ru

2) Institute of Mathematics, SB RAS; Koptyug St., bl.4, Novosibirsk, Russia; max_post@ngs.ru

Abstract: *In this paper, we consider some problems related to forming a consensus of experts' statements for the case of probabilistic forecasting of qualitative feature. We assume that decision rule is constructed on the base of analysis of empirical information represented in the form of probabilistic statements from several experts. The criterion of a quality of experts' statements is suggested. The method of forming of united expert decision rule using distances / similarities between multidimensional sets in heterogeneous feature space is proposed.*

Keywords**:** pattern recognition, expert statements, coordination.

1. INTRODUCTION

In this work we assume that objects under investigation are described by some set of qualitative and quantitative features, and some independent experts give probabilistic predictions of estimated qualitative feature. Their statements may be partially or completely identical, supplementary, and/or contradictory. Also, experts' statements may vary from time to time as well as new "knowledge" from new experts may be obtained. Hence, decision rule is constructed on the base of analysis of empirical information, represented in the form of several experts' statements. Obtained decision rule must be free from anomalies as conflict and redundancy.

2. SETTING OF A PROBLEM

Let Γ be a population of elements or objects concerned by the problem of recognition. By assumption, some experts give probabilistic predictions of unknown belonging classes k of objects $a \in \Gamma$, being already aware of their description $X(a)$. It can be assumed that $X(a) = (X_1(a),...,X_j(a),...,X_n(a))$, where the set X may simultaneously contain qualitative and quantitative features X_j, $j=\overline{1,n}$. Let D_j be the domain of the feature X_j, $j=\overline{1,n}$. It can be assumed that the feature space D is a subset of the product set $\prod_{j=1}^{n} D_j$.

Note that D may be not equal to $\prod_{j=1}^{n} D_j$.

<u>Example.</u> $D_1 = \{a,b,c,d\}$, $D_2 = [10,20]$, $D = [a,c]\times[10,15] \bigcup [b,d]\times[12,20]$.

We shall say that a set E is a *rectangular set* in D when $E = \prod_{j=1}^{n} E_j$, $E_j \subseteq D_j$, $E_j = [\alpha_j, \beta_j]$ if X_j is a quantitative feature, E_j is a finite subset of feature values if X_j is a nominal feature.

In this paper, we consider statements S^i, $i=\overline{1,M}$; represented as sentences of type "if $X(a) \in E^i$, then the object a belongs to the k^i-th pattern with probability p^i", where $k^i \in \{1,...,K\}$, E^i is a rectangular set in D. By assumption, each statement S^i has its own weight w^i. Such a value is like a measure of "confidence". Each statement S^i corresponds to $\left\langle l^i, E^i, k^i, p^i, w^i \right\rangle$, where l^i is a code of expert from which statement is obtained.

Denote the sets of statements concerned to the k-th pattern by Ω^k, the set of initial statements by Ω, $\Omega = \bigcup_{k=1}^{K} \Omega^k$.

The problem consists in constructing decision rule that reflects information synthesized from an organized group of expert opinions.

3. ON CRITERION OF A QUALITY OF EXPERTS' STATEMENTS

Without loss of generality, we can limit our discussion to the case of two patterns, $K=2$.

Let $p_0(1|x)$ be the probability of the first pattern at the point $x \in D$, $p_0(1|x) = P(k=1 | X(a) = x)$. Let $p_l(1|x)$ be the estimation of the $p_0(1|x)$ made by l-th expert.

Since $K=2$, it follows that the probability of the second pattern may be simply obtained from $p_0(1|x)$.

We shall say that the set of the values $p_0(1|x)$ on D is a *strategy of nature* (denote it by c), and the set of the values $p_l(1|x)$ on D is a *strategy of l-th expert* (denote it by g_l.

In this paper we assume for simplicity that there exists rectangular sets $E^1,\ldots,E^T \subseteq D$ such that $D = \bigcup_{t=1}^{T} E^t$, $E^{t_i} \bigcap E^{t_j} = \varnothing$ if $i \neq j$, $p_0(1|x) \equiv \alpha^t \;\; \forall x \in E^t$, where α^t is a constant.

Similarly, assume that there exists rectangular sets $V^1,\ldots,V^{T_l} \subseteq D$ such that $D = \bigcup_{t=1}^{T_l} V^t$, $V^{t_i} \bigcap V^{t_j} = \varnothing$ if $i \neq j$, $p_l(1|x) \equiv \beta^t \;\; \forall x \in V^t$, where β^t is a constant.

Thus, we assume that the strategies c and g_l are piecewise constant in D.

We shall say that l-th expert (a strategy g_l) has a *competence* h if $|p_0(1|x) - p_l(1|x)| \leq 1-h \;\; \forall x \in D$.

Define the criterion of a quality of strategy g_l as the integral

[*)] The work was supported by the RFBR under Grant N07-01-00331a.

$$\eta(g_l)=\frac{\int_D (p_0(1|x)-p_l(1|x))^2 dx}{\mu(D)}, \qquad (1)$$

where $\mu(D)$ is a measure of the set D.

Note that strategies c and g_l are piecewise constant in D, therefore the value $\eta(g_l)$ is a sum of items of the type $\alpha\mu(E)$, where α is a constant, E is a rectangular set in D.

Consider strategies $g_1,\ldots,g_m$. Let A be some algorithm of constructing decision rule on the base of these strategies. Denote the resulted strategy by g^A, $g^A = A(g_1,\ldots,g_m)$.

We shall say that an algorithm A is a *linear combination* of strategies $g_1,\ldots,g_m$ if $\exists\alpha_1,\ldots,\alpha_m \geq 0$ such that $\sum_{l=1}^m \alpha_l = 1$, $p^A(1|x)=\sum_{l=1}^m \alpha_l p_l(1|x)$ $\forall x\in D$.

Proposition 1. If strategies $g_1,\ldots,g_m$ have a competence h, then their linear combination has a competence at least equal to h.

The proof is trivial.

Proposition 2. There exists an algorithm A such that for any strategies g_1 and g_2 we have

$$\eta(A(g_1,g_2)) \leq \frac{\eta(g_1)+\eta(g_2)}{2}. \qquad (2)$$

Proof. Consider algorithm A such that

$$p^A(1|x)=\frac{p_1(1|x)+p_2(1|x)}{2} \quad \forall x\in D.$$

Since strategies g_l are piecewise constant in D, strategy g^A is piecewise constant in D.

Take any point $x\in D$. Then

$$\left(p_0(1|x)-\frac{p_1(1|x)+p_2(1|x)}{2}\right)^2 =$$

$$=\frac{1}{4}\left(p_0(1|x)-p_1(1|x)+ p_0(1|x)-p_2(1|x)\right)^2 =$$

$$=\frac{1}{2}\left(p_0(1|x)-p_1(1|x)\right)\left(p_0(1|x)-p_2(1|x)\right)+$$

$$+\frac{1}{4}\left(p_0(1|x)-p_1(1|x)\right)^2+\frac{1}{4}\left(p_0(1|x)-p_2(1|x)\right)^2 \leq$$

$$\leq\frac{\left(p_0(1|x)-p_1(1|x)\right)^2+\left(p_0(1|x)-p_2(1|x)\right)^2}{2}.$$

■

Proposition 3. There exists an algorithm A such that for any strategies $g_1,\ldots,g_m$ we have

$$\eta(A(g_1,\ldots,g_m)) \leq \frac{\eta(g_1)+\ldots+\eta(g_m)}{m}. \qquad (3)$$

Proof. Consider algorithm A such that

$$p^A(1|x)=\frac{p_1(1|x)+\ldots+p_m(1|x)}{m} \quad \forall x\in D.$$

Further proof is similar to the proof of Proposition 1.

■

Note that equality in (3) is obtained if and only if $p_1(1|x)\equiv\ldots\equiv p_m(1|x)$ $\forall x\in D$.

Suppose that strategy of nature c is unknown and there are independent experts with the same competence. From propositions 1 and 3 it follows that the decision rule obtained by the considered algorithm A has at least the same competence and the quality better than average experts' quality.

Proposition 4. Let A be the linear combination of strategies g_1, g_2; then the minimum of the value $E\eta(g^A)$ is obtained if $\alpha_1=\alpha_2=\frac{1}{2}$.

The proof is omitted.

4. A "DEFAULT" ALGORITHM OF FORMING OF UNITED EXPERTS' DECISION RULE

Further on, we assume that strategy of nature c is unknown.

Let for some point $x\in D$ we have probabilistic statements of several experts. Consider some "reasonable" algorithm of forming a consensus of experts' statements (denote it by A). For simplicity, weights of these statements are omitted.

Firstly, the algorithm A coordinates each l-th expert's statements concerned to certain k-th pattern separately. Suppose that $S^1,\ldots,S^m \in \Omega^k$. Then put

$$p(k,l)=\frac{\sum_{i=1}^{m} p^i}{m}. \qquad (4)$$

Secondly, the algorithm A coordinates all expert's statements concerned to certain k-th pattern. Suppose that we have statements from l^k experts. Then put

$$p(k)=\frac{\sum_{l=1}^{l^k} p(k,l)}{l^k}. \qquad (5)$$

Thirdly, the algorithm A coordinates probabilities of all patterns. Consider the case of two classes, $K=2$. Suppose that $p(1)+p(2)\neq 1$. Then it can be assumed that

$$p^A(1|x)=p(1)-\frac{p(1)+p(2)-1}{2}=\frac{p(1)+(1-p(2))}{2}, \qquad (6)$$

where $p^A(1|x)$ is the probability of the first pattern prescribed to the point $x\in D$ by the algorithm A.

Let us notice that resulted decision rule may suffer from redundancy. Since there are M initial statements, we have up to 2^M sets in D with different predictions. These sets are in the form of $\tilde{E}^1$ or $\tilde{E}^1\setminus(\tilde{E}^2\cup\tilde{E}^3\cup\ldots)$, where $\tilde{E}^i$ are rectangular sets in D.

Consider algorithms B of forming a consensus of experts' statements under restrictions on amount of resulted statements. The value

$$F(B)=\frac{\int\limits_{D}\sum\limits_{k=1}^{K}\left(p^{A}(k\mid x)-p^{B}(k\mid x)\right)^{2}dx}{\mu(D)} \qquad (7)$$

estimates a quality of the algorithm B. Here $p^{A}(k\mid x)$ and $p^{B}(k\mid x)$ are the probabilities of the k-th pattern prescribed to the point $x\in D$ by the algorithms A and B, respectively. In the general case, the best algorithm $B^{*}=\arg\min_{B}F(B)$ is unknown. Further on, the heuristic algorithm of forming a consensus of experts' statements is considered (see, for example, [1, 2]).

5. DISTANCE BETWEEN MULTIDIMENSIONAL SETS IN HETEROGENEOUS FEATURE SPACE

First let us introduce some notation.

Let E^{i_1} and E^{i_2} be the rectangular sets. Denote by $E^{i_1 i_2}:=E^{i_1}\oplus E^{i_2}=\prod_{j=1}^{n}(E_j^{i_1}\oplus E_j^{i_2})$, where $E_j^{i_1}\oplus E_j^{i_2}$ is the *Cartesian join* of feature values $E_j^{i_1}$ and $E_j^{i_2}$ for feature X_j and is defined as follows [3]. When X_j is a nominal feature, $E_j^{i_1}\oplus E_j^{i_2}$ is the union: $E_j^{i_1}\oplus E_j^{i_2}=E_j^{i_1}\bigcup E_j^{i_2}$. When X_j is a quantitative feature, $E_j^{i_1}\oplus E_j^{i_2}$ is a minimal closed interval such that $E_j^{i_1}\bigcup E_j^{i_2}\subseteq E_j^{i_1}\oplus E_j^{i_2}$.

Denote by $r^{i_1 i_2}:=d(E^{i_1 i_2},E^{i_1}\bigcup E^{i_2})$. The value $d(E,F)$ is defined as follows

$$d(E,F):=\max_{E'\subseteq E\setminus F}\frac{diam(E')}{diam(E)}, \qquad (8)$$

where E' is a rectangular set. Note that this value is like a measure of “insignificance” of the set $E\setminus F$.

In the works [4, 5] a method to measure the distances between sets in heterogeneous feature space was proposed. Consider some modification of this method.

Let E^{1} and E^{2} be the rectangular sets, $E^{s}=\prod_{j=1}^{n}E_j^{s}$, $E_j^{s}\subseteq D_j$, $s=1,2$. Denote by $\mu(E_j)$ the measure of the set E_j.

By definition, put

$$\rho(E^{1},E^{2})=\sum_{j=1}^{n}\lambda_j\rho_j(E_j^{1},E_j^{2}), \qquad (9)$$

or, by analogy with Euclidean metrics,

$$\rho(E^{1},E^{2})=\sqrt{\sum_{j=1}^{n}\lambda_j(\rho_j(E_j^{1},E_j^{2}))^{2}}, \qquad (10)$$

where $0\le\lambda_j\le 1$, $\sum_{j=1}^{n}\lambda_j=1$.

Values $\rho_j(E_j^{1},E_j^{2})$ are given by

$$\rho_j(E_j^{1},E_j^{2})=\frac{\mu(E_j^{1}\Delta E_j^{2})}{\mu(D_j)} \qquad (11)$$

if X_j is a nominal feature;

$$\rho_j(E_j^{1},E_j^{2})=\frac{\Psi_j^{12}+\theta\mu(E_j^{1}\Delta E_j^{2})}{\mu(D_j)} \qquad (12)$$

if X_j is a quantitative feature, where

$\Psi_j^{12}=\left|\frac{\alpha_j^{1}+\beta_j^{1}}{2}-\frac{\alpha_j^{2}+\beta_j^{2}}{2}\right|$, if $E_j^{s}=[\alpha_j^{s},\beta_j^{s}]$, $s=1,2$.

It can be proved that the triangle inequality is fulfilled if and only if $0\le\theta\le 1/2$.

The proposed measure ρ satisfies the requirements of distance there may be.

Note that someone can use another distance in multidimensional space (see, for example, [6]).

6. CONSTRUCTING OF UNITED EXPERTS' DECISION RULE

By definition, put

$$\lambda_j=\frac{\gamma_j}{\sum\limits_{i=1}^{n}\gamma_i}, \text{ where} \qquad (13)$$

$$\gamma_j=\sum_{u|S^{u}\in\Omega^{1}}\sum_{v|S^{v}\in\Omega^{2}}\rho_j(E_j^{u},E_j^{v}),\ j=\overline{1,n}. \qquad (14)$$

Let us remark that values λ_j are like a measure of difference between two patterns by features X_j.

We first treat single expert's statements concerned to a certain pattern class.

Consider the sets $E^{i_1},\dots,E^{i_q}$ such that $r^{i_u i_v}\le\varepsilon$ $\forall u,v=\overline{1,q}$, where ε is a threshold decided by the user, $0\le\varepsilon\le 1$, $q=\overline{2,Q}$, Q is an amount of this expert's statements concerned to this pattern class k. Suppose that there is not exist another set $E^{i_{q+1}}\mid\forall u=\overline{1,q}$ $r^{i_u i_{q+1}}\le\varepsilon$.

Denote by $J_q=\{i_1,\dots,i_q\}$, $E^{J_q}=E^{i_1}\oplus\dots\oplus E^{i_q}$, $c^{iJ_q}=1-\rho(E^{i},E^{J_q})$, where $\rho(E,F)$ is a distance between rectangular sets E and F. Now, we can aggregate the statements S^{i_1}, …, S^{i_q} into the statement $S^{J_q}=\left\langle -,E^{J_q},k,p^{J_q},w^{J_q}\right\rangle$, where

$$p^{J_q}=\frac{\sum_{i\in J_q}c^{iJ_q}w^{i}p^{i}}{\sum_{i\in J_q}c^{iJ_q}w^{i}}, \qquad (15)$$

$$w^{J_q}=\left(1-d(E^{J_q},\bigcup_{i\in J_q}E^{i})\right)\frac{\sum_{i\in J_q}c^{iJ_q}w^{i}}{\sum_{i\in J_q}c^{iJ_q}}. \qquad (16)$$

The value p^{J_q} evaluates contributions of probabilities p^i from statements S^i in accordance with their weights w^i and measures of difference c^{iJ_q} between corresponding sets.

The procedure of forming a consensus of single expert's statements consists in aggregating into statements S^{J_q} for all J_q under previous conditions, $q = \overline{2,Q}$.

After coordinating each expert's statements separately, we can construct an agreement of several independent experts for each pattern class. The procedure is as above, except the weights:

$$w^{J_q} = \sum_{i \in J_q} c^{iJ_q} w^i \tag{17}$$

(the more experts give similar statements, the more we trust in resulted statement).

Denote the sets of coordinated statements concerned to the k-th pattern by Ω_1^k, the set of coordinated statements by Ω_1, $\Omega_1 = \bigcup_{k=1}^{K} \Omega_1^k$.

After constructing of a consensus for each pattern, we must make decision rule in the case of contradictory statements.

Take any statements $S^u \in \Omega_1^1$ and $S^v \in \Omega_1^2$ such that $E^{uv} := E^u \bigcap E^v \neq \varnothing$, $p^u + p^v \neq 1$. Consider the sets $I(k) = \left\{ i \mid S^i \in \Omega^k \text{ and } \rho(E^i, E^{uv}) \leq \varepsilon^* \right\}$, where ε^* is a threshold.

Denote by $c^i = 1 - \rho(E^i, E^{uv})$;

$$p(k) = \frac{\sum_{i \in I(k)} c^i w^i p^i}{\sum_{i \in I(k)} c^i w^i}; \tag{18}$$

$$w(k) = \frac{\sum_{i \in I(k)} c^i w^i}{\sum_{i \in I(k)} c^i}; \tag{19}$$

$$p^{uv} := \frac{p(1)w(1) + (1 - p(2))w(2)}{w(1) + w(2)}; \tag{20}$$

$$w^{uv} = \frac{w(1) + w(2)}{2(1 + | p(1) + p(2) - 1 |)}. \tag{21}$$

Prescribe to the set E^{uv} evaluated probability of the first pattern p^{uv} and weight w^{uv}.

Now, we can make resulted decision statement $\left\langle -, E^{uv}, 1, p^{uv}, w^{uv} \right\rangle$.

This iterative algorithm does not guarantee an optimal solution, but experience has shown that a reasonable approximation is obtained.

7. CONCLUSION

Suggested method of forming of united decision rule can be used for coordination of several experts statements and different decision rules obtained from learning samples and/or time series. Applications of this method are relevant to many areas, such as medicine, economics and management.

8. REFERENCES

[1] Lbov G., Gerasimov M. Constructing of a Consensus of Several Experts Statements. *Proceedings of the XII International Conference "Knowledge-Dialogue-Solution"*, Varna, Bulgaria, 2006, pp. 193-195.

[2] Gerasimov M.K. Forming of United Experts' Decision Rule in Pattern Recognition. *Proceedings of the IX International Conference "Pattern Recognition and Information Processing"*. Minsk, 2007, Vol. 1, pp. 136-139.

[3] Diday E. *Symbolic Data Analysis*. INRIA Roquencourt. Paris, 1995. pp. 1-136.

[4] Lbov G., Gerasimov M. Determining of Distance Between Logical Statements in Forecasting Problems [in Russian]. *Artificial Intelligence*, 2 (2004), Ukraine, pp. 105-108.

[5] Lbov G., Berikov V. *Decision Functions Stability in Pattern Recognition and Heterogeneous Data Analysis* [in Russian]. Institute of Mathematics, Novosibirsk, 2005. p. 218.

[6] Vikentiev A.A., Lbov G.S. Metrication of Boolean Algebra of Statements and Informativeness of Experts' Statements [in Russian]. *J. Doklady RAS*, Vol. 361, No.2 (1998), Russia, pp. 174-176.