APPLICATION OF THE HETEROGENEOUS SYSTEM PREDICTION METHOD TO PATTERN RECOGNITION PROBLEM[1]

Tatyana Stupina

Abstract: An application of the heterogeneous system prediction method to solving the problem pattern recognition with respect to the sample size is considered in this paper. The criterion of multivariate heterogeneous variable recognition is used in this approach. The relation of this criterion with probability of error is shown. For the fixed complexities of probability distribution and logical decision function class the examples of pattern recognition problem are presented.

Keywords: the prediction of heterogeneous variables system, the pattern recognition, the complexity of distribution, logical decision function.

ACM Classification Keywords: G.3.

Introduction

The reducing relation problem with respect to sample is one of important problem in data mining. The method quality of constructing sample decision function depends on the size of the sample, the complexity of the distributions, and the complexity of the class of functions used by the algorithm for constructing sample decision functions. When the distribution is known the quality of decision function, for example, is risk function for one prediction variable. For the pattern recognition problem, for example, it is well-known probability of error. We can define a quality of the method so as average of the quality of sample decision function on samples of fixed size. When the distribution is unknown the problem estimating of this function quality with respect to the complexity of the distributions, of the functions class and sample size is appeared. At present time there are approaches solving this problem [Vapnik V.N., Chervonenkis A.Ya, 1970]. In addition the complexity of the functions class is assigned differently [Lbov G.S., Starceva N.G, 1999]. The method quality we can estimate if the class of distribution is known or by mathematical modeling. But that approaches consider the case of one prediction variable and one variable type.

However there are many important applied when we what to predict or recognize several (system) heterogeneous variables. In work [Lbov G.S., Stupina T.A., 2002] was presented this problem statement. It is necessary to construct the sample decision function on the small sample in the multivariate heterogeneous space, so the most proper class is a class of logical decision functions [Lbov G.S., Starceva N.G, 1999]. In this paper for the fixed probability distribution the relation of the criterion with probability of error is shown for pattern recognition problem. The quality of constructing sample decision function method is shown for that problem.

Problem Statement

In the probabilistic statement of the problem, the value *(x,y)* is a realization of a multidimensional random variable (X,Y) on a probability space $<\Omega,\mathrm{B},\mathrm{P}>$, where $\Omega = D_X \times D_Y$ is μ-measurable set (by Lebeg), B is the borel σ-algebra of subsets of Ω, P is the probability measure (we will define such as c, the strategy of nature) on B, D_X is heterogeneous domain of under review variable, $\dim D_x = n$, D_y is heterogeneous domain of objective variable, $\dim D_y = m$. The given variables can be of arbitrary types (quantitative, ordinal, nominal). For the pattern recognition problem the variable Y is nominal. Let us put $\Phi_\circ$ is a given class of decision functions. Class $\Phi_\circ$ is μ-measurable functions that puts some subset of the objective variable $E_y \subseteq D_Y$ to each value of

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

the under review variable $x \in D_X$, i.e. $\Phi_\circ = \{f : D_x \to 2^{D_y}\}$. For example the domain E_y can contains the several patterns $\{\omega_1,...,\omega_k\}$ for pattern recognition problem.

The quality $F(c,f)$ of a decision function $f \in \Phi_\circ$ under a fixed strategy of nature c is determined as $F(c,f) = \int_{D_x} (P(E_y(x)/x) - \mu(E_y(x)))dP(x)$, where $E_y(x) = f(x)$ is a value of decision functions in x, $P(y \in E_y(x)/x)$ is a conditional probability of event $\{y \in E_y\}$ under a fixed x, $\mu(E_y(x))$ is measurable of subset E_y. Note that if $\mu(E_y(x))$ is probability measure, than criterion $F(c,f)$ is distance between distributions. If the specified probability coincides with equal distribution than such prediction does not give no information on predicted variable (entropy is maximum). On the nominal-real space $\Omega = D_н \times D_в$ a measure μ is defined so as any $E \in \mathrm{B}$, $E = \bigcup_{j=1}^{|E_н|} E_в^j \times \{z^j\}$, $\mu(E) = \sum_{j=1}^{|E_н|} \frac{\mu(E_в^i)}{|D_н| \mu(D_в)}$, were $E_н$ is projection of set E on nominal space $D_н$, z^j - item of $E_н$, $E_в^j$ - set in $D_в$ corresponding to z^j, $\mu(E_в^j)$ - lebeg measure of set $E_в^j$. For any subset of domains D_X or D_Y the measure μ is assigned similarly. Clearly, the prediction quality is higher for those E_y whose measure is smaller (accuracy is higher) and the conditional probability $P(y \in E_y(x)/x)$ (certainty) is larger. For a fixed strategy of nature c, we define an optimal decision function $f_\circ(x)$ as such as $F(c,f_\circ) = \sup_{f \in \Phi_\circ} F(c,f)$, where $\Phi_\circ$ is represented above class of decision functions.

When we solve this problem in practice the size of sample is very smaller and type of variables different. In this case is used class of logical decision function Φ_M complexity M [Lbov G.S., Starceva N.G, 1999]. For the prediction problem of the heterogeneous system variables class Φ_M is defined as $\Phi_M = \{f \in \Phi_\circ \mid f \sim <\alpha, r(a)>, \alpha \in \Psi_M, r(\alpha) \in R_M\}$ (the mark '~' denotes the correspondence of pair $<\alpha, r(a)>$ to symbol *f*), were Ψ_M is set of all possible partitioning $\alpha = \{E_X^1,...,E_X^M \mid E_X^t = \prod_{j=1}^n E_{X_j}^t,\ E_{X_j}^t \subseteq D_{X_j}, t = \overline{1,M}, \bigcup E_X^t = D_X\}$ of domain D_X on M noncrossing subsets, R_M is set all possible decisions $r(\alpha) = \{E_y^1,...,E_y^M \mid E_y^t \in \Im_{D_Y}, t = \overline{1,M}\}$, $\Im_{D_Y}$ - set of all possible *m*-measuring intervals. For that class the measure $\mu(E_y(x)) = \frac{\mu(E_y)}{\mu(D_Y)} = \prod_{j=1}^{m} \frac{\mu(E_{y_j})}{\mu(D_{y_j})}$ is the normalized measure of subset E_y and it is introduced with taking into account the type of the variable. The measure $\mu(E_y(x))$ is measure of interval, if we have a variable with ordered set of values and it is quantum of set, if we have a nominal variable (it is variable with finite non-ordering set of values and we have the pattern recognition problem). A complexity of Φ_M class is assigned as M if we have univariant prediction (decision is presented by form: if $x \in E_X^t$, than $y \in E_y^t$), $M_\Phi = M$, and it is assembly $(k_1,...,k_M)$ if we have multivariant, i.e. $E_y^t = \bigcup_{i=1}^{k_t} E_y^i$, $t = 1,...,M$ and $E_y^i \cap E_y^j = \varnothing$ for $i \neq j$ (decision is presented by form: if $x \in E_X^t$, than $y \in E_y^1 \vee E_y^2 \vee ... \vee E_y^{k_t}$). In this work for pattern recognition problem we consider the case $M_\Phi = M$.

Properties of the Criterion

For the fixed strategy of nature *c* the relation of the criterion $F(c,f)$ with probability of error P_f is shown.

Statement 1. For any strategy of nature *c* the quality criterion $F(c,f)$ is represented by risk function such that $1 - R(c,f) = \int_{D_X} \int_{D_Y} (1 - L(y,f(x)))p(x,y)dxdy$, where the loss function $L(y,f)$ such as $L(y,f) = \begin{cases} \mu(E_y), & y \in E_y \\ 1 + \mu(E_y), & y \notin E_y \end{cases}$.

Remark that risk function $R(c,f)$ is probability of error P_f if the loss function $L(y,f)$ is indicator function.

Statement 2. For recognition *k* patterns by decision function *f* the quality criterion is $F(c,f) = \frac{k-1}{k} - P_f$.

Consequence 1. For recognition two patterns we have equation $F(c,f) = 1/2 - P_f$.

Consequence 2. For pattern recognition problem the optimal decision function coincides with bayes function such as $\sup_{x\in(-\infty,+\infty)} F(c,f) = \inf_{x\in(-\infty,+\infty)} P_f$.

Definition 1. Define a nature strategy c_M (generated by logical decision function $f \in \Phi_M$, $f \sim <\alpha, r(\alpha)>$) such as $c_M = \{p^t(x,y) = p_x^t p_{y/x}^t = P(x \in E_X^t)P(y \in E_Y^t / x \in E_X^t), t = 1,...,M\}$, where 1) $\sum_{t=1}^M p_x^t = 1$; 2) $P(E_Y^t / E_X^t) = p_{y/x}^t$, 3) $P(\bar{E}_Y^t / E_X^t) = 1 - p_{y/x}^t$, where $E_X^t \in \alpha$, $E_Y^t \in r(\alpha)$, $<\alpha, r(\alpha)> \in \Phi_M$, 4) $\forall\ A_X \subseteq E_X^t\ P(A_X) = p_x^t \frac{\mu(A_X)}{\mu(E_X^t)}$, $\forall\ A_Y \subseteq E_Y^t\ P(A_Y / E_X^t) = p_{y/x}^t \frac{\mu(A_X)}{\mu(E_Y^t)}$.

In the paper [Lbov G.S., Stupina T.A., 2002] is proved that $F(c,f) = \sum_{t=1}^M p_x^t (p_{y/x}^t - \mu(E_y^t))$ for this nature strategy. Let for *k* pattern recognition the domain D_Y is the set $\{\omega_1,...,\omega_k\}$.

Statement 3. Let the nature strategy c_k for *k* pattern recognition is generated by logical decision function f^* such as $f^*(x) = E_y^i$ for $x \in E_X^i$, then the probability of error P_f for decision function f such as $f(x) = \omega_i$, $\omega_i \in E_y^i$, for $x \in E_X^i$ is $P_f = 1 - \sum_{i=1}^k \frac{1}{k\mu(E_y^i)} p_x^i p_{y/x}^i$.

Consequence 3. From the statement 3 it follows equation $P_f + F(c_k, f^*) = 1 + \sum_{i=1}^k p_x^i \left[p_{y/x}^i \left(1 - \frac{1}{k\mu(E_y^i)}\right) - \mu(E_y^i)\right]$.

Let us illustrate these statements. Let there is *n=1, m=1*, *X*- continuous variable, *Y*-nominal variable, *M*=2. The nature strategy c_2 is generated by f^*, that $\alpha^* = \{E_X^1, E_X^2\}$, $E_X^1 = (0.364, 1.0]$, $E_X^2 = [0.0, 0.364]$, $r(\alpha^*) = \{E_y^1, E_y^2\}$, $E_y^1 = \{1\}$, $E_y^2 = \{1,0\}$, $\omega_1 = '1'$, $\omega_2 = '0'$, $p_x^1 = \frac{19}{50}$, $p_x^2 = \frac{31}{50}$, $p_{y/x}^1 = 0{,}95$, $p_{y/x}^2 = 1$. So we have $F(c_2, f^*) = 0{,}171$. Obviously that c_2 is such that conditional distribution $p(x/\{1\})$ and $p(x/\{0\})$ is intersected for every pattern, if we have f' ($f'(x) = \{1\}$, if $x \in E_X^1$, and $f'(x) = \{0\}$, if $x \in E_X^2$) or f'' ($f''(x) = \{1\}$, if $x \in E_X^1$, and $f''(x) = \{1\}$, if $x \in E_X^2$). Let calculate $P_{f'}$ using definition and compare with criterion $F(c_2, f^*)$: $P_{f'} = P(\{1\})P(E_X^2/\{1\}) + P(\{0\})P(E_X^1/\{0\})$ = $P(E_X^2)P(\{1\}/E_X^2) + P(E_X^1)P(\{0\}/E_X^1)$ = $\frac{31}{50} \cdot \frac{1}{2} \cdot 1 + \frac{19}{50} \cdot (1 - 0{,}95) = 0{,}329$. Similarly we can provide the probability of error $P_{f''}$. For this case we have $F(c_2, f^*) = \frac{1}{2} - P_{f'} = 0{,}5 - 0{,}329 = 0{,}171$ (statement 2 and 3).

The method of Constructing Sample Decision Function

If the strategy of nature is unknown the sampling criterion $F(\bar{f})$ is used by presented method $Q(v_N)$ of constructing sample decision function $\bar{f}$, $\bar{F}(\bar{f}) = \sum_{t=1}^{M'} \bar{p}_x^t (\bar{p}_{y/x}^t - \bar{\mu}_y^t)$, were $\bar{p}_x^t = \frac{N(\tilde{E}_X^t)}{N(D_X)} = \frac{N^t}{N}$, $\bar{p}_{y/x} = \frac{N(\hat{E}_y^t)}{N(\tilde{E}_X^t)} = \frac{\hat{N}^t}{N^t}$, $\bar{\mu}_y = \mu(\hat{E}_y)$, $N('*')$ is number of sample points, generating the set “*”, $\bar{f} \sim <\alpha, r(\alpha)>$, $\alpha = \{\tilde{E}_X^1,...,\tilde{E}_X^{M'}\} \in \Psi_{M'}$, $r(\alpha) = \{\hat{E}_y^1,...,\hat{E}_y^{M'}\} \in R_{M'}$. The optimal sample decision function is $\bar{f}^* = \arg\max_{\alpha\in\Psi_{M'}} \max_{r(\alpha)\in R_{M'}} \bar{F}(\bar{f})$. In order to solver this extreme problem we apply the algorithm *MLRP* of step-by-step increase attachments of decision trees. It do the branching of top point on that value criterion $\bar{F}(\bar{f})$ is maximum and the top point is divisible or $\bar{F}(\bar{f}) \geq F^*$. The top point is indivisible if 1) number of final top point is $M' = M^*$

or 2) $\hat{N}^t \le N^*$. That criterion and parameters F^*, M^*, N^* assign method of constructing sample decision function.

In order to estimate the presented method quality we do statistical modeling. The average of the criterion of sample decision function on samples of fixed size $m_F(c) = \mathrm{E}_{V_N} F(c,\bar{f})$ is estimated for fixed nature strategy.

Conclusion

An approach to solving the problem of heterogeneous multivariate variable recognition with respect to the sample size was considered in this paper. The solution of this problem was assigned by means of presented criterion. The universality of the logical decision function class with respect to presented criterion makes the possible to introduce a measure of distribution complexity and solve this problem for small sample size. For the nature strategy and the class of logical decision function the criterions properties are presented by means of statements and consequences for pattern recognition problem. The relationship of the $\overline{m}_F(c) = \mathrm{E}_{V_N} F(c,\bar{f})$ estimate with respect to decision function class complexity for fixed nature strategy complexity demonstrates the method quality.

Bibliography

[Lbov G.S., Starceva N.G, 1999] Lbov G.S., Starceva N.G. Logical Decision Functions and Questions of Statistical Stability. Inst. Of Mathematics, Novosibirsk.

[Lbov G.S., Stupina T.A., 2002] Lbov G.S., Stupina T.A. Performance criterion of prediction multivariate decision function. Proc. of international conference "Artificial Intelligence", Alushta, pp.172-179.

[Vapnik V.N., Chervonenkis A.Ya, 1970] Vapnik V.N., Chervonenkis A.Ya .Theory of Pattern Recognition, Moscow: Nauka.

Author's Information

Tatyana A. Stupina – Institute of Mathematics SBRAS, Koptuga 4 St, Novosibirsk, 630090, Russia; e-mail: stupina@math.nsc.ru

ON THE QUALITY OF DECISION FUNCTIONS IN PATTERN RECOGNITION

Vladimir Berikov

Abstract: The problem of decision functions quality in pattern recognition is considered. An overview of the approaches to the solution of this problem is given. Within the Bayesian framework, we suggest an approach based on the Bayesian interval estimates of quality on a finite set of events.

Keywords: Bayesian learning theory, decision function quality.

ACM Classification Keywords: I.5.2 Pattern recognition: classifier design and evaluation

Introduction

In the problem of decision functions quality analysis, one needs to find a decision function, not too distinguishing from the optimal decision function in the given family, provided that the probability distribution is unknown, and learning sample has limited size. Under optimal decision function we shall understand such function for which the risk (the expected losses of wrong forecasting for a new object) is minimal. In particular, the following questions should be solved at the analysis of the problem.