NONPARAMETRIC METHODS OF REGULARITY DETECTION IN SMALL SAMPLES CONDITIONS

Alexandr Lapko, Vasiliy Lapko, and Maxim Sharkov

Institute of Computational Modeling of Siberian Branch of Russian Academy of Sciences, Krasnoyarsk, Russia, *lapko@icm.krasn.ru*

The majority of statistical methods of pattern recognition is focused on representative training samples. However, in the case of applied problems a researcher often have the limited quantity of supervision - short or small sample that is caused with researched object unstability, high cost and complexity of the additional information reception. Solving rules received on their basis not always provide comprehensible results of classification because the small training samples information is insufficient for stochastic characteristics estimation of investigated regularity.

For "detour" of small samples problems the wide popularity was received with principles of systems decomposition and consecutive procedures of decisions formation.

The results of researches on generation of random variables with probability density representing a nuclear estimation of Rozenblatt-Parzen type are shown in work [1]. The offered procedures of continuation of casual sequences prove an opportunity of artificial increase in training sample volume n. The asymptotic properties of a nuclear estimation of probability density received in this conditions are investigated in [2], that is a basis of synthesis of nonparametric algorithms of pattern recognition in conditions of small samples.

Probability density estimation in small samples conditions. We shall consider, that the restored probability density $p(x)$ and its first two derivatives are limited and continuous.

For "detour" of small samples problems at probability density estimation $p(x)\forall x \in R^1$ we shall increase the initial data volume $V = (x^i,\, i = \overline{1,n})$ due to results of statistical modelling. To achieve this purpose we shall do at *β*-vicinity of each *i-th* situation of sample *m* imitations for the random variable $\bar{x}$ with the distribution law $p_i(\bar{x})$ and average of distribution equal to zero.

It is easy to notice, that the received statistical sample $V_2 = (x^i + \bar{x}^j,\, j = \overline{1,m},\, i = \overline{1,n})$ equals to a mix of probability density

$$\hat{p}(x) = \frac{1}{n}\sum_{i=1}^{n} p_i(\bar{x}).$$

Its nonparametric estimation can be written in the in the form of

$$\bar{p}(x) = (nmc)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m}\Phi\left(\frac{x - x^i - \bar{x}^j}{c}\right), \tag{1}$$

where $\Phi(u)$ is nuclear function (positive, symmetric and normed), $c = c(n)$ is blurriness parameter of nuclear function [3].

In a multivariate case $x \in R^k$ the probability density estimation looks like

$$\bar{p}(x) = (nm)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m}\prod_{\nu=1}^{k}\frac{1}{c_\nu}\Phi\left(\frac{x_\nu - x_\nu^{\,i} - \bar{x}_\nu^{\,j}}{c_\nu}\right) \tag{2}$$

In work [2] are certain asymptotic expressions for displacement

$$M\left(\hat{p}(x) - p(x)\right) \sim \frac{p^{(2)}(x)}{2}\left(c^2 + \bar{\mu}^2\right) \tag{3}$$

and root-mean-square deviations

$$\widetilde{W}(x,\bar{x}) = M\left(\bar{p}(x) - p(x)\right)^2 \sim \frac{p(x)\|\Phi(u)\|^2}{2nmc}\bar{\mu}^2 + \frac{1}{n}[p^2(\bar{x}) - 2p(\bar{x})p^{(1)}(\bar{x})\mu^1 +$$
$$+((p^{(1)}(\bar{x}))^2 + p(\bar{x})p^{(2)}(\bar{x}))\mu^2 - p^{(1)}(\bar{x})p^{(2)}(\bar{x})\mu^3 + \frac{(p^{(2)}(\bar{x}))^2}{4}\mu^4] + \quad (4)$$
$$+\frac{1}{m}[p^2(x) - (p^{(1)}(x))^2\bar{\mu}^2 + \frac{(p^{(2)}(x))^2}{4}\bar{\mu}^4 + p(x)p^{(2)}(x)\bar{\mu}^2] + \frac{(p^{(2)}(x))^2}{4}(c^2 + \bar{\mu}^2)^2.$$

Here M - a sign of average of distribution; $p^{(\nu)}(x), p^{(\nu)}(\bar{x}), \nu = 1,2$ - the first and second derivatives of probabilities density of random variables x and $\bar{x}$;

$$\mu^\nu = \int x^\nu p(x)dx, \nu = \overline{1,4} \quad \bar{\mu}^\nu = \int_{-\beta}^{\beta} \bar{x}^\nu p(\bar{x})d\bar{x}, \nu = 1,2 \quad \|\Phi(u)\|^2 = \int \Phi^2(u)\, du.$$

The proof of asymptotic convergence of nonparametric statistics, intended for regularities detection in small samples conditions, allows to prove a technique of their synthesis analytically. On this basis there is an opportunity to define quantitative interrelation between characteristics of training sample, parameters of imitating procedure of additional statistical information formation and investigated algorithms of their processing that is necessary for an explanation of results of computing experiments.

Let's make the analysis of asymptotic expression for the root-mean-square deviations $\widetilde{W}(x,\bar{x})$ on all change range of variables x and $\bar{x}$. Thus we shall believe

$$p(\bar{x}) = (2\beta)^{-1} \forall x \in [-\beta;\beta]$$

and let's neglect during transformations in variables $\beta^3, \beta^2 c^2, c^4\ \beta^3/(nmc)$. Then, integrating expressions $W(x,\bar{x})$ on variables $x, \bar{x}$, we shall receive

$$\widetilde{W} \sim \frac{\Delta}{2n\beta} + \frac{2\|p(x)\|^2\beta}{m}, \quad (5)$$

where Δ - length of an x change interval.

Let's note natural dependence of asymptotic properties of the density estimation $\bar{p}(x)$ from volume of the initial information and results of statistical modelling.

As one would expect, with growth m the root-mean-square deviations estimation (5) aspires to a limit $\Delta/(2n\beta)$. The received conclusions confirm a basic opportunity of use statistics of (1) type at small samples processing. At the same time, conditions of finite *n* also $m \to \infty$ do not provide convergence $\bar{p}(x)$ to $p(x)$.

And dependence W from β at concrete values m also n has extreme character and at optimum

$$\beta^* = \left(\frac{m\Delta}{4n\|p(x)\|^2}\right)^{\frac{1}{2}} \quad (6)$$

reaches the minimum.

From the analysis (6) quite obvious parities between parameters β^*, m and Δ follow. The interval $2\beta^*$ of artificial training sequence generating increases with growth of a definition range of $p(x)$ and quantity of imitations m, decreases in process of increase in volume of initial sample n.

Let's define conditions at which the offered technique of increase in volume of sample provides increase of efficiency of probability density estimation.

It is known, that the minimal value of asymptotic expressions for root-mean-square deviations [3] corresponds to a traditional nonparametric estimation of nuclear type

$$\widetilde{W} = \left\| \widetilde{p}(x) - p(x) \right\|^2 \sim \frac{5}{4} \left(\left(\frac{\|\Phi\|^2}{n} \right)^4 \left\| p^{(2)}_{(x)} \right\|^2 \right)^{\frac{1}{5}}.$$

For the statistics efficiency estimation (1) we shall calculate expression (5) for its root-mean-square deviations at optimum value β^* (6), we shall receive

$$\overline{W} \sim 2 \left(\frac{\Delta \| p(x) \|^2}{nm} \right)^{\frac{1}{2}}.$$

Then, concerning the condition $\widetilde{W} / \overline{W} > 1$, it is possible to determine the requirement for imitations quantity m of procedure of artificial training sample generating

$$m > 2{,}56 \Delta \| p(x) \|^2 \left(n^3 / \left(\left(\|\Phi\|^2 \right)^4 \left\| p^{(2)}_{(x)} \right\|^2 \right) \right)^{\frac{1}{5}}, \tag{7}$$

then the statistics (1) will possess higher approximation properties in comparison with a traditional nonparametric estimation of nuclear type probability density.

Synthesis of the dividing surface equation in small samples conditions. We shall consider a construction technique of the dividing surface equation in small samples conditions on an two-alternative task example of pattern recognition in the space of continuous attributes.

In this case the deciding rule corresponding, for example, to criterion of the maximal credibility looks like

$$\overline{m}(x): \begin{cases} x \in \Omega_1, \text{если } \bar{f}_{12}(x) > 0, \\ x \in \Omega_2, \text{если } \bar{f}_{12}(x) \le 0, \end{cases}$$

where

$$\bar{f}_{12}(x) = \left(nm \prod_{\nu=1}^{k} c_\nu \right)^{-1} \sum_{i=1}^{n} \sigma(i) \sum_{j=1}^{m} \prod_{\nu=1}^{k} \Phi\left(\frac{x_\nu - x_\nu^{\,i} - \overline{x}_\nu^{\,j}}{c_\nu} \right), \quad \sigma(i) = \begin{cases} (n_1/n), \text{если } x \in \Omega_1, \\ -(n_2/n), \text{если } x \in \Omega_2, \end{cases} \tag{8}$$

During the decision function optimization (8) parameters β of training sample generating procedure $\overline{V} = (x^i + \overline{x}^j, \sigma(i), j = \overline{1,m}, i = \overline{1,n})$ are defined first, where $\sigma(i)$ - instructions on a belonging of a situation $x^i + \overline{x}^j$ to one of classes. Their choice is carried out for each class from a condition of a full covering of a definition range of a corresponding part of training sample with the β-vicinities. Thus crossing of all β-vicinities should be minimal.

At the second stage parameters m and c_ν, $\nu = \overline{1, k}$ of nonparametric estimation of the dividing surface equation are defined from a condition of a minimum of an pattern recognition empirical estimation in a «sliding examination » mode on sample $\overline{V}$. Preliminary value of parameter m commensurable with volume of initial sample n is not excluded.

For increase of pattern recognition nonparametric algorithms efficiency in small samples conditions usage of collective estimation principles is possible. Let $\widetilde{m}_{12}^{j}(x)$, $j = \overline{1, M}$ - nonparametric solving rules for a two-alternative pattern recognition task constructed on the samples $(x^i + \overline{x}^j, \sigma(i), j = \overline{1,m}, i = \overline{1,n})$, which are distinguished from each other with the casual sequences "expanding" the initial training sample at same values of imitation parameters m and β.

Let's take advantage of well-known approach of collective estimation, for example, a method of "voting" and we shall construct deciding rule

$$\tilde{\tilde{m}}_{12}(x):\begin{cases} x \in \Omega_1, \textit{если} & \dfrac{M_1}{M} \geq \dfrac{M_2}{M} \\ x \in \Omega_2, \textit{если} & \dfrac{M_1}{M} < \dfrac{M_2}{M}; \end{cases},$$

where M_j, $j = 1, 2$ are numbers of "decisions" which are made with members of a collective about a belonging object with a feature set x to the j-th class.

Results of computing experiment. The analysis of computing experiment results confirms authentic advantage of the investigated qualifier in small samples conditions in comparison with traditional nonparametric pattern recognition algorithm. It is enough of final imitations quantity m in a β-vicinity of each situations of initial training sample for significant reduction in a mistake of classification. Presence of threshold value m is confirmed and agreed with analytical result (7).

It is possible to explain the found out law if it is considered, that the statistics (1) is a nonparametric estimation of a density probability mixture $p_i(\bar{x}), i = \overline{1, n}$. Each component $p_i(\bar{x})$ of the mixture is determined on β-vicinity of an initial situation x^i and it is restored on imitating data $V_i = \left(x^i + \bar{x}^j, j = \overline{1, m}\right)$. It is obvious, that at small values m the information of sample V_i is obviously insufficient for estimation $p_i(\bar{x})$ and deforms the distribution law of x in the classes presented by initial data. Therefore in the field of small values m the advantage of the offered nonparametric qualifier is not observed.

Presence of an extremum-minimum for the dependence of an pattern recognition probability estimation mistake $\bar{\rho}$ on value of β-vicinity of artificial training sample generating procedure is discovered. The given fact is agreeed with the analytical conclusion (6) received at the analysis of asymptotic expression for the root-mean-square deviations of the nonparametric probability density estimations $\bar{p}(x)$ from $p(x)$.

The results of computing experiment confirm, that the parameter β value should provide a full covering of the classes definition range at their minimal mutual crossing with β-vicinities.

Literature

1. Devroy L., Diorthy L. Nonparametric density estimation (The L_1-approach). (In Russian) "Mir", Moscow (1988) 407 p.
2. Lapko A.V., Lapko V.A., Sokolov M.I., Chentsov S.V. Nonparametric classification systems. (In Russian) "Nauka", Novosibirsk (2000) 240 p.
3. Epanenchikov V.A. Nonparametric estimation of multivariate probability density. (In Russian) The probability theory and its application, Moscow (1969), Vol. 1. pp.156-161.