KYBERNETIKA — VOLUME 34 (1998), NUMBER 4, PAGES 473-478

PROBABILITY DISTRIBUTION OF TRANSFORMED RANDOM VARIABLES WITH APPLICATION TO NONLINEAR FEATURE EXTRACTION

LUBOMÍR SOUKUP[1]

A method for estimation of probability distribution of transformed random variables is presented. The proposed approach admits an approximation of the transformation of the random variables. The approximate probability density function (pdf) is corrected to obtain a resulting pdf which incorporates a prior knowledge of approximation errors. The corrected pdf is not contaminated by any uncontrollable approximation. The method is applied to pattern recognition. It is shown that class conditional pdf of features can be easily computed even when the feature extraction was performed with nonlinear mapping of an input pattern.

1. INTRODUCTION

Estimation of probability distributions is general problem that is extensively studied in many scientific branches where mathematical statistics is applied. Generally, joint probability density function (pdf) of random variables Y_j, $j \in \{1, 2, \ldots, n\}$ that depend on other random variables X_i, $i \in \{1, 2, \ldots, m\}$ should be estimated. The relationship between the two sets of random variables is usually described by a system of equations

$$Y_j = b_j(X_1, X_2, \ldots, X_m), \quad j \in \{1, 2, \ldots, n\}, \tag{1.1}$$

where b_j is real function for $\forall j \in \{1, 2, \ldots, n\}$. The problem is how to determine pdf of random variables Y_j when a joint pdf of random variables X_i is given. If $n \leq m$, no more information is needed.

Many simplified versions of this problem have been studied. For example, if uncertainty of variables X_i is given by their maximal errors, then interval aritmetics can be used for estimation of maximal errors of Y_j. Such an approach was applied in computer vision and was described in [2]. Functions b_j were implemented as a computer program and a software solution was designed there.

In this contribution, functions b_j have to be in form of mathematical expressions. Then pdf of random variables Y_j is easy to derive and express in a simple formula. However, the formula is not so easy to evaluate even when functions b_j are only slightly complicated. Therefore approximate functions a_j are used instead of the original functions b_j. Then an approximate pdf is obtained which differs from the exact pdf. It is expected that the difference between the approximate and exact pdf's

[1]Supported by the Grant of the Academy of Sciences No. A 2075603

is not significant if functions a_j are close to b_j. This expectation is not used to prove in practice and the difference between the both pdf's is usually neglected without any assessment of its magnitude. Hence, the decisions made on the approximate pdf cannot be reliable.

This difficulty is resolved in this contribution. The approximate pdf is corrected with respect to the approximation error to obtain reliable pdf of random variables Y_j. The core of this idea in one-dimensional case ($n = m = 1$) was published in [4] and [3].

2. FORMULATION OF THE PROBLEM

The problem introduced above will be exactly formulated in this section. Vector notation should be used to make subsequent formulae well-arranged.

$$\boldsymbol{X} \triangleq [X_1, X_2, \ldots, X_m], \qquad \boldsymbol{Y} \triangleq [Y_1, Y_2, \ldots, Y_n],$$
$$\boldsymbol{b} \triangleq [b_1, b_2, \ldots, b_n]; \qquad m, n \in \mathbb{N},\ m \geq n.$$

The vector function $\boldsymbol{b}$ is mapping

$$\boldsymbol{b} : \mathcal{X} \to \mathcal{Y} : \boldsymbol{x} \mapsto \boldsymbol{b}(\boldsymbol{x}), \quad \mathcal{X} \subseteq \mathbb{R}^m, \ \mathcal{Y} \subseteq \mathbb{R}^n. \tag{2.1}$$

Then equations (1.1) can be written in vector form:

$$\boldsymbol{Y} = \boldsymbol{b}(\boldsymbol{X}). \tag{2.2}$$

Probability distribution of random vector $\boldsymbol{X}$ is given by means of its pdf

$$f : \mathcal{X} \to \mathbb{R} : \boldsymbol{x} \mapsto f(\boldsymbol{x}).$$

Such correspondence will be denoted as $\boldsymbol{X} \sim f$. For some $\boldsymbol{x} \in \mathcal{X}$ which is realization of random vector $\boldsymbol{X}$, notation $\boldsymbol{x} \rightsquigarrow \mathcal{X}$ will be used too.

Let a subset $\mathcal{U}_f \subseteq \mathcal{X}$ be support of f, i.e.: $\mathcal{U}_f \triangleq \{\boldsymbol{x} \in \mathcal{X} \mid f(\boldsymbol{x}) > 0\}$.

Pdf of random vector $\boldsymbol{Y}$ can be easily derived if mapping $\boldsymbol{b}$ is regular and injective on $\mathcal{X}$.

$$\boldsymbol{Y} \sim g_{\boldsymbol{b}},$$

$$g_{\boldsymbol{b}}(\boldsymbol{y}) = \int_{\mathcal{Z}} f(\widetilde{\boldsymbol{b}}^{-1}(\boldsymbol{y}, \boldsymbol{z})) \frac{1}{\left|\det\left(\dfrac{\partial \widetilde{\boldsymbol{b}}}{\partial \boldsymbol{x}}(\widetilde{\boldsymbol{b}}^{-1}(\boldsymbol{y}, \boldsymbol{z}))\right)\right|} \, \mathrm{d}\boldsymbol{z}. \tag{2.3}$$

Mapping $\widetilde{\boldsymbol{b}}$ is vector function $\widetilde{\boldsymbol{b}} \triangleq [\boldsymbol{b}, \boldsymbol{b}'']$, where $\boldsymbol{b}''$ is an auxillary mapping

$$\boldsymbol{b}'' : \mathcal{X} \to \mathcal{Z} : \boldsymbol{x} \mapsto \boldsymbol{b}''(\boldsymbol{x}), \quad \mathcal{Z} \subseteq \mathbb{R}^{m-n}.$$

Mapping $\boldsymbol{b}''$ can be chosen in an arbitrary way that makes $\widetilde{\boldsymbol{b}}$ regular and injective on $\mathcal{X}$.

To evaluate formula (2.3) easily, the original mapping b is approximated by a simpler mapping a.

$$a : \mathcal{X} \rightarrow \mathcal{Y} : x \mapsto a(x) , \quad \forall x \in \mathcal{X} : a(x) \approx b(x) . \tag{2.4}$$

Substituting b with a in (2.3) pdf g_a of random vector $a(X)$ outcomes. Approximation error $v(x) \triangleq b(x) - a(x)$ is not known precisely, so that a random vector V_x can be assigned to $v(x)$ to describe its uncertainty. Since V_x depends on x, its pdf is conditional pdf.

$$v(x) \rightsquigarrow V_x \sim p(\,.\,|\,x) , \quad x \in \mathcal{U}_p \supseteq \mathcal{U}_f . \tag{2.5}$$

Pdf $p(\,.\,|\,x)$ represents a prior knowledge of approximation error $v(x)$.

In summary, pdf of a random vector Y' which incorporates uncertainties of random vectors X, V_x is requested. Random vector X is given by its pdf f. An acceptable pdf $p(\,.\,|\,x)$ and a set $\mathcal{U}_p \supseteq \mathcal{U}_f$; $x \in \mathcal{U}_p$ have to be designed to represent the pdf of V_x.

3. SOLUTION OF THE PROBLEM

Random vectors X, V_x should be summed since the approximation error $v(x)$ was defined as a subtraction. Therefore

$$Y' = V_x + a(X) . \tag{3.1}$$

Let pdf's of random vectors Y', $[\,V_x\,,\,X\,]$ be introduced:

$$Y' \sim g\,, \quad [\,V_x\,,\,X\,] \sim \tilde{f}\,.$$

To express pdf g, an auxillary mapping $\tilde{a} \triangleq [\,a'\,,\,a''\,]$ have to be introduced similarly to $\tilde{b}$ in (2.3).

$$\tilde{a} : \mathcal{V} \times \mathcal{X} \rightarrow \mathcal{Y}' \times \mathcal{Y}'' : [\,v\,,\,x\,] \mapsto \tilde{a}(v\,,\,x\,) ,$$

$$a'(v\,,\,x\,) \triangleq v + a(x) , \quad \mathcal{V} \supseteq v(\mathcal{U}_p) , \; \mathcal{Y}' \triangleq a(\mathcal{X}) , \; \mathcal{Y}'' \triangleq a''(\mathcal{V}, \mathcal{X}) .$$

Then pdf g can be carried out with the aid of formula (2.3).

$$g(y) = \int_{\mathcal{Y}''} f(\tilde{a}^{-1}(y\,,\,z)) \frac{1}{\left|\det\left(\frac{\partial \tilde{a}}{\partial \tilde{x}}(\tilde{a}^{-1}(y\,,\,z))\right)\right|} \,\mathrm{d}z\,, \tag{3.2}$$

where $\tilde{x} \triangleq [\,v\,,\,x\,]$.

The simplest way of choosing a'' is: $a''(v\,,\,x\,) \triangleq x$. Then

$$\tilde{a}^{-1}(y\,,\,z\,) = [\,y - a(z)\,,\,z\,] . \tag{3.3}$$

Pdf $\tilde{f}$ can be factorized with the aid of conditonal pdf $p(\,.\,|\,x)$ of random vector V_x.

$$\tilde{f}(v\,,\,x) = p(v\,|\,x)\,f(x) . \tag{3.4}$$

Furthermore,

$$\det\left(\frac{\partial\tilde{a}}{\partial\tilde{x}}(\tilde{b}^{-1}(y\,,z))\right)=\begin{vmatrix}\dfrac{\partial a'}{\partial v}, & \dfrac{\partial a'}{\partial x}\\ \dfrac{\partial a''}{\partial v}, & \dfrac{\partial a''}{\partial x}\end{vmatrix}=\begin{vmatrix}E_n & , \dfrac{\partial a}{\partial x}\\ O_{m,n}, & E_m\end{vmatrix}=|E_n|\,|E_m|=1\,. \quad (3.5)$$

E_n is identity matrix of order n, and $O_{m,n}$ is $m \times n$ zero matrix.

After substitutions (3.3), (3.4), (3.5), and with regard to the inclusion in (2.5), formula (3.2) changes to

$$g(y)=\int_{\mathcal{U}_p} p(y-a(x)\,|\,x)\,f(x)\,\mathrm{d}x\,. \quad (3.6)$$

To evaluate the integral in (3.6) easily, a suitable pdf $p(\,.\,|\,x)$ and set $\mathcal{U}_p$ has to be chosen. A solution of this subproblem will now be shown in a subsequent pattern-recognition example.

4. APPLICATION TO NONLINEAR FEATURE EXTRACTION

In pattern recognition, Bayesian approach is frequently used to classify a pattern to one of classes ω_k, $k \in \mathcal{K} \subset \mathbb{N}$. The pattern is represented by a vector of measured parameters $x \in \mathcal{X}$. Parameters that admits easy discrimination among the classes ω_k are rarely identical with components of the pattern vector x. Therefore feature extraction is usually necessary before classification. Feature extraction is a mapping b (see (2.1)) that can be nonlinear in principle. Choice of a mapping b is based on the knowledge of the physical process that generates the pattern vector x. Instead of the vector x a feature vector

$$y=b(x)$$

has to be dealt with in classification. A decision rule for classification utilizes a posterior probabilities $P(\omega_k\,|\,y)$ that depend on class conditional pdf's $g(y\,|\,\omega_k)$ besides a prior probabilities $P(\omega_k)$ according to Bayes formula

$$P(\omega_k\,|\,y)=\frac{g(y\,|\,\omega_k)\,P(\omega_k)}{\sum_{l\in\mathcal{K}} g(y\,|\,\omega_l)\,P(\omega_l)}\,, \quad k\in\mathcal{K}\,.$$

(See e.g. [1])

Now, for some given class conditional pdf $f(x\,|\,\omega_k)$, pdf $g(y\,|\,\omega_k)$ can be computed by means of formula (3.6). For simplicity, substitutions $f(x\,|\,\omega_k)\to f(x)$, $g(y\,|\,\omega_k)\to g(y)$ can be adopted. Approximate mapping a is assumed to be linear. Pdf of the pattern vector x is truncated Gaussian, i.e.:

$$f(x)=\begin{cases} q\exp\left(-\frac{1}{2}(x-\mu)^T M(x-\mu)\right) & \Leftarrow x\in\mathcal{U}_f\subset\mathcal{X}\,,\\ 0 & \Leftarrow x\notin\mathcal{U}_f\,,\end{cases}$$

where q is normalizing constant, μ is mean vector, and M inverse covariance matrix.

Support $\mathcal{U}_f$ has to be given in form of a m-dimensional cuboid parallel to eigenvectors of matrix M.

Typically, the only information about the approximation error is its range, namely its lower and upper bound. It means that the pertinent pdf $p(\,.\,|\,\boldsymbol{x})$ should be uniform. In the least informative case, support $\mathcal{V}$ of pdf $p(\,.\,|\,\boldsymbol{x})$ is the same for all $x \in \mathcal{U}_p$, i.e.:

$$p(\boldsymbol{v}\,|\,\boldsymbol{x}) \triangleq \begin{cases} c_{\mathcal{V}} \triangleq \dfrac{1}{|\mathcal{V}|} & \Leftarrow \boldsymbol{v} \in \mathcal{V}\,, \\ 0 & \Leftarrow \boldsymbol{v} \notin \mathcal{V}\,. \end{cases} \tag{4.1}$$

Set $\mathcal{V}$ can be estimated on the basis of knowledge of the approximation error $\boldsymbol{v}(\boldsymbol{x})$, so that $\boldsymbol{v}(\mathcal{X}) \subseteq \mathcal{V}$. If mapping $\boldsymbol{a}$ is part of Taylor series expansion of mapping $\boldsymbol{b}$ up to the linear terms, the well-known formula for the rest (Taylor's theorem) can be used for the estimation of $\mathcal{V}$.

Under these assumptions, formula (3.6) can be further modified:

$$\begin{aligned} g(\boldsymbol{y}) &= q\, c_{\mathcal{V}} \int_{\mathcal{U}_p \cap\, \mathcal{U}_{\mathcal{W}}(\boldsymbol{y})} \exp\left(-\frac{1}{2}\,(\boldsymbol{x}-\boldsymbol{\mu})^T\, \boldsymbol{M}(\boldsymbol{x}-\boldsymbol{\mu})\right) \\ &= q\, c_{\mathcal{V}} \int_{\boldsymbol{K}^{-1}(\mathcal{U}_p \cap\, \mathcal{U}_{\mathcal{W}}(\boldsymbol{y}))} \exp\left(-\frac{1}{2}\,\boldsymbol{z}^T\, \boldsymbol{K}^T\, \boldsymbol{M}\, \boldsymbol{K}\, \boldsymbol{z}\right) \mathrm{d}\boldsymbol{z}\,, \end{aligned}$$

where $\mathcal{U}_{\mathcal{W}}(\boldsymbol{y}) \triangleq \{\boldsymbol{x} \in \mathcal{X} \mid \boldsymbol{y} - \boldsymbol{a}(\boldsymbol{x}) \in \mathcal{W} \subseteq \mathcal{V}\}$. The last equality was derived after substitution $\boldsymbol{z} = \boldsymbol{K}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$ that diagonalizes matrix $\boldsymbol{M}$.

$$\boldsymbol{K}^T\, \boldsymbol{M}\, \boldsymbol{K} = \mathrm{diag}(\kappa_1\,, \dots\,, \kappa_m)\,.$$

Matrix $\boldsymbol{K}$ consists of eigenvectors of matrix $\boldsymbol{M}$. Sets $\mathcal{U}_p$, $\mathcal{W}$ have to be chosen in form of appropriate parallelepipeds so that $\boldsymbol{K}^{-1}(\mathcal{U}_p \cap \mathcal{U}_{\mathcal{W}}(\boldsymbol{y}))$ is parallel to the m-dimensional cuboid $\mathcal{U}_f$. Pdf g can be finally expressed:

$$.\; g(\boldsymbol{y}) = q\, c_{\mathcal{V}} \prod_{i=1}^{m} \int_{r_i(\boldsymbol{y})}^{s_i(\boldsymbol{y})} \exp\left(-\frac{1}{2}\,\kappa_i\, t_i^2\right) \mathrm{d}t_i\,,$$

where r_i, s_i are linear functions.

Hence, the corrected pdf can be easily computed with the aid of error function.

5. EXAMPLE

As an illustrative example let us consider a simple case of feature extraction in two-dimensional feature space. Figure 1 shows three classes with normally distributed features x_1, x_2. It is apparent, that a feature selection (or another linear feature extraction) cannot provide a clasifier with a reasonable single feature to distinguish among the three classes, since their projections onto the axes overlap themselves. The Euclidean distance to the origin, i.e. the nonlinear function $y = \sqrt{x_1^2 + x_2^2}$, seems to be more convenient. The proposed method allows to reliably estimate the pdf of the distance y for all the three classes. Consequently, the distinguish-ability of feature y can be tested and the risk of misclassification can be evaluated in advance.

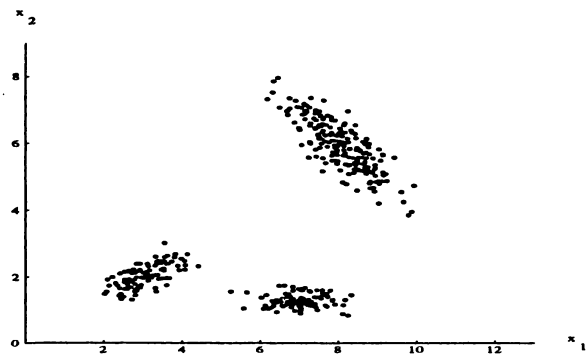

Fig. 1. Feature extraction by means of Euclidean distance to the origin.

6. CONCLUSION

The proposed method enables safe and reliable use of approximation in transformation of random variables. Uncertainty of the approximate functions is integrated with uncertainty of the approximation errors by means of superposition of two random vectors, so that the approximate pdf is corrected with regard to the approximation errors. Computation of the corrected pdf is very effective because it does not include any iterative procedures.

Beside the presented application to pattern recognition, the suggested approach can also be applied to a wide area of scientific and technical problems dealing with non-linear parameter estimation.

(Received December 18, 1997.)

REFERENCES

[1] J. Kittler: Statistics and Images: 2. (K. V. Mardia, ed.), volume 2 of Advances in Applied Statistics, chapter Statistical pattern recognition in image analysis. Carfax Publishing Company, 1994.

[2] R. Mařík, J. Kittler and M. Petrou: Error sensitivity assesment of vision algorithms based on direct error–propagation. In: Workshop on Performance Characteristics of Vision Algorithms (H. I. Christensen, W. Forstner, and C. B. Madsen, eds.), Cambridge 1996, pp. 45–58.

[3] L. Soukup: Probability Distribution of an Approximated Function with Uncertain Argument. Research Report No. 1877, Institute of Information Theory and Automation, Prague 1996.

[4] L. Soukup: Probability distribution of approximation. In: The 2nd European Workshop on Computer–Intensive Methods in Control and Signal Processing (L. Berec, J. Rojíček and M. Kárný, eds.), Institute of Information Theory and Automation, Prague 1996, pp. 141–144.

Ing. Lubomír Soukup, Institute of Information Theory and Automation – Academy of Sciences of the Czech Republic, Pod vodárenskou věží 4, 182 08 Praha 8. Czech Republic. e-mail: soukup@utia.cas.cz