Statistical Recognition of a Set of Patterns Using Novel Probability Neural Network

Andrey V. Savchenko

National Research University Higher School of Economics, Nizhniy Novgorod, Russian Federation
avsavchenko@hse.ru

Abstract. Since the works by Specht, the probabilistic neural networks (PNNs) have attracted researchers due to their ability to increase training speed and their equivalence to the optimal Bayesian decision of classification task. However, it is known that the PNN's conventional implementation is not optimal in statistical recognition of a set of patterns. In this article we present the novel modification of the PNN and prove that it is optimal in this task with general assumptions of the Bayes classifier. The modification is based on a reduction of recognition task to homogeneity testing problem. In the experiment we examine a problem of authorship attribution of Russian texts. Our results support the statement that the proposed network provides better accuracy and is much more resistant to change the smoothing parameter of Gaussian kernel function in comparison with the original PNN.

Keywords: Statistical pattern recognition, sets of patterns, probabilistic neural network, hypothesis test for samples homogeneity.

1 Introduction

Pattern recognition [1] is a fundamental aspect of many tasks in artificial intelligence, data mining, computer vision, medical diagnostics, decision-support systems. These tasks may be formulated [2] in terms of statistical recognition [3], [4] of a set of patterns: it is required to estimate the class of an input sample of random variables, with an assumption that all available information about each class is concluded in certain samples of observations [2]. This general formulation could be applied to such acute tasks as image recognition, voice phonemes recognition, authorship attribution, etc.

This problem is usually reduced [5] to a statistical classification of the query sample. The optimal decision is taken with a minimum Bayes risk principle [3]. The unknown probability density, required in this approach, is usually estimated by means of nonparametric techniques [6], [7] like kernel discriminant analysis, e.g. Parzen approach [8]. Such estimations were proved to converge to the real probability density if the training sample size is large [9], [10]. The widely-used parallel implementation of nonparametric approach is a probabilistic neural network (PNN) [11]. This

N. Mana, F. Schwenker, and E. Trentin (Eds.): ANNPR 2012, LNAI 7477, pp. 93–103, 2012.

multilayered feedforward network introduced by Specht [12], [13] is characterized by extremely fast training procedure.

The PNN was proved to be an asymptotically-optimal rule in the classification task [11], [14] if a query object is a single feature vector. Unfortunately, conventional PNN does not provide an optimal solution [2] if the query object is represented by a set of features with the size approximately equal to the training set size. Really, in this case the task should be reduced to a homogeneity testing of query and training samples [15]. In this paper we introduce the modification of the PNN, which saves all advantages of the conventional PNN but yields an optimal decision boundary in statistical recognition of a set of patterns. We experimentally show that the proposed PNN achieves better accuracy and is much more resistant to change the smoothing parameter of Gaussian kernel function [16].

The rest of the paper is organized as follows: Section 2 presents statistical recognition of a set of patterns using the PNN [11]. In Section 3, we introduce our PNN modification. In Section 4, we present the experimental results in the author identification task [17] with well-known texts from Russian literature [18]. Finally, concluding comments are given in Section 5.

2 Statistical Recognition of a Set of Patterns

Let a set $\mathbf{X}=\{\mathbf{x}_j\}, j=\overline{1,n}$ of independent identically distributed (i.i.d.) random variables with unknown $\mathbf{P}$ probability distribution be specified. Here n is a sample size, $\mathbf{x}_j=\{x_{j;1},...,x_{j;M}\}$ - is a vector of features with a fixed dimension $M=const$. The pattern recognition problem is to estimate the class of $\mathbf{X}$. It is assumed that each class $r\in\{1.,,,.R\}$ is defined by a training set of i.i.d. random variables with unknown $\mathbf{P}_r$ probability distribution $\mathbf{X}_r=\left\{\mathbf{x}_j^{(r)}\right\}, j=\overline{1,n_r}$. Here n_r is a training sample size, $\mathbf{x}_j^{(r)}$ is a feature vector with dimension M.

Following the statistical approach [2], [3], we assume that each class is fully determined by the distribution $\mathbf{P}_r$, $r=\overline{1,R}$ of its feature vector. Thus, the problem is referred to a hypothesis testing for distribution of $\mathbf{X}$

$$W_r: \qquad \mathbf{P}=\mathbf{P}_r \quad r=\overline{1,R} \tag{1}$$

To solve the problem (1), the principle of minimum Bayes risk [4] is applied. The query sample $\mathbf{X}$ is assigned to the class ν with maximum a-posterior probability

$$\nu=\underset{r\in\{1,...,R\}}{\arg\max}\, P\left(\mathbf{X}\middle|W_r\right)\cdot P\left(W_r\right) \tag{2}$$

Here $P(W_r)$ is the prior class probability, $P(\mathbf{X}|W_r)$ is a conditional class density (likelihood). In the most practically important pattern recognition tasks [19], [20] it is assumed that each class is equiprobable (prior uncertainty): $P(W_r) = \frac{1}{R}$. Following the nonparametric approach, the likelihood is estimated by the given training set with a kernel trick [21]

$$\hat{P}(\mathbf{X}|W_r) = \frac{1}{(n_r)^n} \prod_{j=1}^{n} \sum_{j_r=1}^{n_r} K_{n_r}\left(\mathbf{x}_j, \mathbf{x}_{j_r}^{(r)}\right) \tag{3}$$

Here $K_{n_r}\left(\mathbf{x}_j, \mathbf{x}_{j_r}^{(r)}\right)$ is a kernel function [14]. For example, the Gaussian Parzen kernel [8] is widely used [11]

$$K_{n_r}\left(\mathbf{x}_j, \mathbf{x}_{j_r}^{(r)}\right) = \frac{1}{\left(2\pi\sigma^2\right)^{M/2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{M} \left(x_{j;i} - x_{j_r;i}\right)^2\right) \tag{4}$$

Here $\sigma = const > 0$ is a fixed smoothing parameter (standard deviation of the Gaussian kernel). Based on the estimate (3), the final decision (2) could be written as

$$\nu = \underset{r \in \{1,\ldots,R\}}{\arg\max} \frac{1}{(n_r)^n} \prod_{j=1}^{n} \sum_{j_r=1}^{n_r} K_{n_r}\left(\mathbf{x}_j, \mathbf{x}_{j_r}^{(r)}\right) \tag{5}$$

The criterion (5) corresponds the PNN for statistical recognition of a set of patterns problem (1). In contrast with the conventional four-layered PNN [11], [12], which is used to classify one object $\mathbf{x}_j$, the network (5) contains additional, production, layer to classify the sample of objects $\mathbf{X}$.

3 An Optimal Algorithm

Pattern recognition problem is characterized by the unknown probability distributions $\mathbf{P}_r$ of each class r. Thus, it is required to estimate the conditional density (3). It is the key difference [2] from the conventional classification task, in which distributions $\mathbf{P}_r$ are given. Hence, we believe it is better to follow the Borovkov's approach [2] rather than refer pattern recognition task to a statistical testing for distribution (1). According to this approach [2], the problem is reduced to a testing for homogeneity of input sample $\mathbf{X}$ and training set $\mathbf{X}_r$:

$W_r:\left\{\mathbf{x}_j\right\}, j=\overline{1,n}$ and $\left\{\mathbf{x}_j^{(r)}\right\}, j=\overline{1,n_r}$ have the same probability density

The decision is made with a minimum Bayes risk principle by using the set from the united sample space $\{\mathbf{X},\mathbf{X}_1,...,\mathbf{X}_R\}$ to a class ν

$$\nu=\underset{r\in\{1,...,R\}}{\arg\max}\ \sup_{\mathbf{P}^*}\ \sup_{\mathbf{P}_j^*,\, j=\overline{1,R}} P\left(\left\{\mathbf{X},\mathbf{X}_1,...,\mathbf{X}_R\right\}\middle| W_r\right)\cdot P\left(W_r\right) \tag{6}$$

Here $\mathbf{P}^*$ is a possible probability distribution of a sample $\mathbf{X}$, $\mathbf{P}_j^*$ is a possible distribution of *j*th training sample. Assuming the independence of random variables in a united sample $\{\mathbf{X},\mathbf{X}_1,...,\mathbf{X}_R\}$, the conditional density in (6) could be written as

$$\sup_{\mathbf{P}^*}\ \sup_{\mathbf{P}_j^*,\, j=\overline{1,R}} P\left(\left\{\mathbf{X},\mathbf{X}_1,...,\mathbf{X}_R\right\}\middle| W_r\right)= \\ =\frac{\sup\limits_{\mathbf{P}^*} P\left(\mathbf{X}\middle|W_r\right)\sup\limits_{\mathbf{P}_r^*} P\left(\mathbf{X}_r\middle|W_r\right)}{\sup\limits_{\mathbf{P}_r^*} P\left(\mathbf{X}_r\right)}\cdot\prod_{j=1}^{R}\sup_{\mathbf{P}_j^*} P\left(\mathbf{X}_j\right)$$

As $\prod\limits_{j=1}^{R}\sup\limits_{\mathbf{P}_j^*} P\left(\mathbf{X}_j\right)=const(\mathbf{X})$ does not depend on a query sample $\mathbf{X}$, and $\sup\limits_{\mathbf{P}^*} P(\mathbf{X})$ is independent on $\mathbf{X}_r$, we convert (6) to the following expression

$$\nu=\underset{r\in\{1,...,R\}}{\arg\max}\ \frac{\sup\limits_{\mathbf{P}^*} P\left(\mathbf{X}\middle|W_r\right)\sup\limits_{\mathbf{P}_r^*} P\left(\mathbf{X}_r\middle|W_r\right)}{\sup\limits_{\mathbf{P}^*} P(\mathbf{X})\cdot\sup\limits_{\mathbf{P}_r^*} P\left(\mathbf{X}_r\right)}\cdot P\left(W_r\right) \tag{7}$$

It is known [2], that the supremum of the likelihood is achieved when the valid probability distributions $\mathbf{P}^*$, $\mathbf{P}_r^*$ are equal to their optimal unbiased estimates. Herewith, to evaluate this optimal estimate the combined sample $\{\mathbf{X},\mathbf{X}_r\}$ is used if the condition W_r is true, i.e. $\mathbf{X}$ and $\mathbf{X}_r$ are the samples of the same random variable. Hence, we may use nonparametric kernel estimation (3), e.g.

$$\sup_{\mathbf{P}^*} P(\mathbf{X}|W_r) =$$

$$= \frac{1}{(n+n_r)^n} \prod_{j=1}^{n} \left(\sum_{j_1=1}^{n} K_n\left(\mathbf{x}_j, \mathbf{x}_{j_1}\right) + \sum_{j_r=1}^{n_r} K_{n_r}\left(\mathbf{x}_j, \mathbf{x}^{(r)}_{j_r}\right) \right),$$

$$\sup_{\mathbf{P}^*} P(\mathbf{X}) = \frac{1}{n^n} \prod_{j=1}^{n} \sum_{j_1=1}^{n} K_n\left(\mathbf{x}_j, \mathbf{x}_{j_1}\right).$$

Thus, (7) is equivalent to the following criterion

$$v = \underset{r \in \{1,\dots,R\}}{\arg\max} \frac{n^n \cdot (n_r)^{n_r}}{(n+n_r)^{n+n_r}} \prod_{j=1}^{n} \left(1 + \frac{\sum_{j_r=1}^{n_r} K_{n_r}\left(\mathbf{x}_j, \mathbf{x}^{(r)}_{j_r}\right)}{\sum_{j_1=1}^{n} K_n\left(\mathbf{x}_j, \mathbf{x}_{j_1}\right)} \right) \times$$

$$\times \prod_{j_r=1}^{n_r} \left(1 + \frac{\sum_{j_1=1}^{n} K_n\left(\mathbf{x}^{(r)}_{j_r}, \mathbf{x}_{j_1}\right)}{\sum_{j_{r;1}=1}^{n_r} K_{n_r}\left(\mathbf{x}^{(r)}_{j_r}, \mathbf{x}^{(r)}_{j_{r;1}}\right)} \right) \qquad (8)$$

Expression (8) corresponds to a proposed PNN for recognition of a set of patterns. Its implementation is shown in Fig. 1. Here the input layer contains not only the query sample $\mathbf{X}$, but the united sample $\{\mathbf{X}, \mathbf{X}_1, \dots, \mathbf{X}_R\}$. The kernel function for a query sample is added to a training set in the second, pattern, layer. The new division layer is added according to (8). In the production layer we multiply not only the features of the query object $\mathbf{X}$, but also features of *r*th sample $\mathbf{X}_r$.

It could be noticed that if $n_r \to \infty$, expression (8) is equivalent to (5). Really, in asymptotics, the training set $\mathbf{X}_r$ fully determines the probability distribution $\mathbf{P}_r$. Hence, the united sample $\{\mathbf{X}, \mathbf{X}_r\}$ does not provide any additional information.

The proposed PNN saves all advantages of the classical PNN [11], but the rate of convergence to the optimal decision should be higher for (8) than for a classical implementation (5). The next section provides an experimental evidence to support this claim.

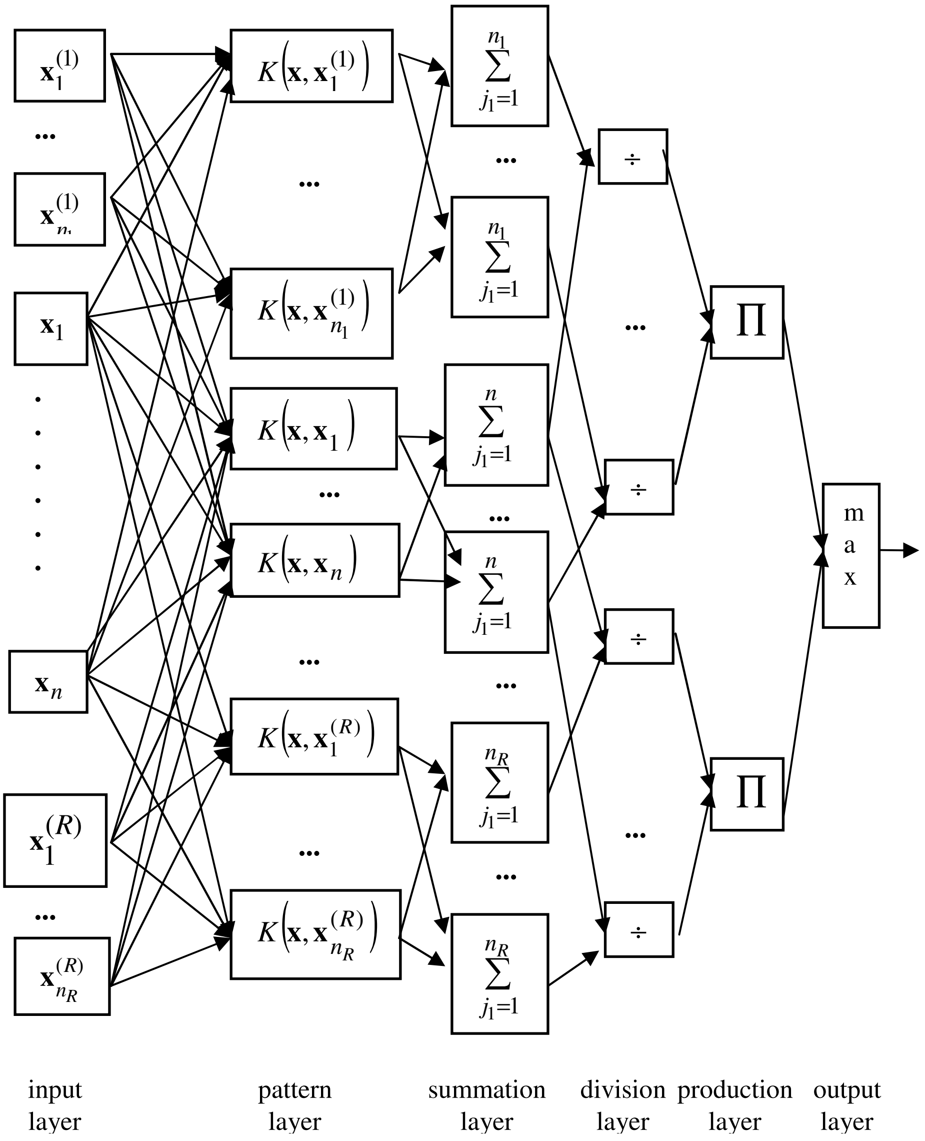

Fig. 1. Proposed modification of the PNN in statistical recognition of a set of patterns

4 Experimental Results

In this section we demonstrate the proposed modification of PNN (8) in the pattern recognition problem from the linguistic analysis. It is required to identify the author of a Russian text fragment [17], [22]. The training sets contain other text extracts. We use the following eight well-known large Russian texts: "Anna Karenina" and "Resurrection" by L. Tolstoy, "Idiot" and "Crime and punishment" by F. Dostoevsky, "Dead souls" and "Taras Bulba" by N. Gogol, "And Quiet Flows the Don" and "Virgin Soil Upturned" by M. Sholokhov. All texts (in original) were taken from the corpus [18].

We compare the accuracy of the proposed PNN (8) with its conventional implementation (5). Additionally, we provide error rate for conventional multilayer perceptron (MLP) with one hidden layer trained by using the backpropagation. To classify the sample of objects $\mathbf{X}$, we use production (5) of MLP outputs for a single pattern $\mathbf{x}_j$. All neural networks were implemented as a parallel application with Java Runtime Environment 1.7 on a modern laptop (Intel Core i7 CPU 2.0 GHz, 6 Gb RAM).

The accuracy was estimated by the following procedure. One randomly chosen fragment of each text with fix size (in characters) was added to the training set. The test set contain 10 randomly chosen fragments of each text (i.e., 8*10=80 samples). The error rate was estimated by these 80 tests. The total recognition quality was estimated as an arithmetical mean of the error rate by 100 such experiments of training and test set selection (i.e., 80*100=8000 classifications).

In the first experiment we used the frequency of punctuation marks in each sentence as a simple but informative feature of Russian language. This feature set is known to show a good quality in Russian authorship attribution [17]. The following punctuation marks were chosen: ".", "?", "!", ",", ".", "-", ":", ";", "("; i.e. the feature vector $\mathbf{x}$ contains $M=9$ features. In the first case both training and test sets were generated by fragments of 25000 characters (the sample size n=900...1150).

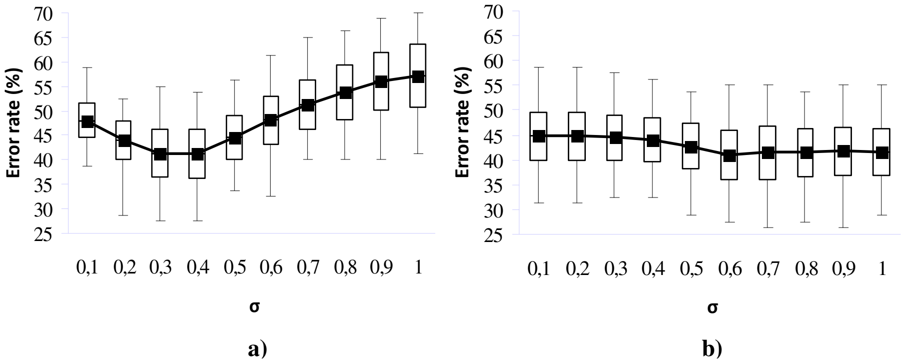

Fig. 2. Dependence of average error rate (in %) on σ for 25000 characters of each fragment both in training and test sets. a. criterion (5); b. criterion (8).

The box-plot diagrams of dependence of classification error rate on the smoothing parameter σ of Gaussian kernel function (4) are shown in Fig. 2. For comparison, the best error rate for MLP was achieved with 100 neurons in a hidden layer and is equal to 58,3%±20%.

The average recognition time per one set t_{rec} here is equal to 0.1s., 0.25s. and 0.4s. for the original PNN (5), proposed criterion (8) and MLP, respectively. The average training time t_{tr} is equal to 0.01s. for both (5), (8) and 34s for MLP.

As we could see in Fig.2, the minimal error rate of criterion (8) is equal to 40.8% , which is a bit less then the minimal error rate 41.3% of the conventional PNN (5). The most significant quality indicator of proposed network (Fig. 1) is the robustness of error rate dependence on the smoothing parameter. Really, the error rate for the proposed criterion (8) is always less than 45%. At the same time, accuracy of the traditional PNN varies enormously. At worst, the average error rate is equal to 59%.

In the second case the experiment was repeated, but the training sets were generated from 100000 characters in a fragment (i.e. the sample size n=3400...4200). The test set was still generated by 25000 character-fragments. The results are illustrated with a box-plot diagram in Fig. 3. The best error rate for MLP was achieved here with 350 neurons in a hidden layer and is equal to 52,2%±23,4%.

The average recognition time per one set t_{rec} is equal to 0.5s., 1.1s. and 0.9 s. for the original PNN (5), proposed criterion (8) and MLP respectively. The average training time t_{tr} is equal to 0.05s. for both (5), (8) and 920s. for MLP.

In Fig. 3 one could see that the recognition accuracy is extremely better in comparison with the previous experiment. The minimal average error rate (23%) of (8) is a bit less than the minimal error rate 23.6% of the PNN (5). Again, the proposed network is more robust to change the smoothing parameter than the traditional PNN.

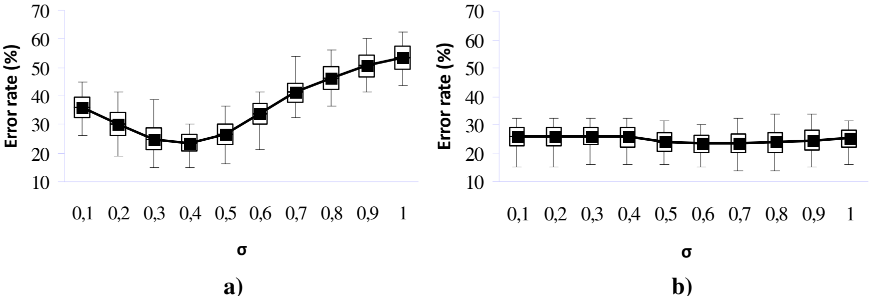

Fig. 3. Dependence of average error rate (in %) on σ for 100000 characters of fragment in training set and 25000 characters of fragment in test sets. a. criterion (5); b. criterion (8).

In the last experiment we chose the frequency of the bigram of words [17]. This feature is widely used in natural language processing and showed good quality of author attribution [22]. We extract 500 most frequent word from the union of all training fragments and calculate the frequency of each word from each fragment. As a preliminary processing we perform stemming, i.e. removing the known Russian endings. Words beginning with capitals were omitted (including the first words of sentences). Finally, we retain only words with length greater than 3 characters. The frequency of the same 500 words were calculated for each test fragment. The kernel function in this experiment is a limit of Gaussian kernel (4) when $\sigma \to 0$, i.e. the discrete delta function. Here the author identification error rate of the proposed PNN (8) is 13.3% if the training set was generated from 25000 character-fragments. It is much less than the error rate of the conventional PNN (27.8%). The best recognition results for different number of characters in a training text fragments is summarized in Table 1.

Table 1. The average error rate of the author identification by text fragment

Features	Frequency of punctuation in a sentence		Frequency of word bigram	
Number of characters	Criterion (5)	Criterion (8)	Criterion (5)	Criterion (8)
25000 - training set, 25000 - test set	41,3%±7,5%	40,8%±7,4%	26,4%±8,3%	13,1%±6,9%
100000 - training set, 25000 - test set	23,6%±4,5%	23,0%±4,4%	19,8%±5,7%	3,6%±3,0%

Based on this results we could draw the following conclusion. First, the classification accuracy of the conventional PNN is not greater than the accuracy of the proposed modification (8). It is particularly noticeable for recognition with discrete features and discrete delta function as a kernel. Second, the task of proper choice of the best value of the smoothing parameter σ of Gaussian kernel (4) for proposed criterion (8) is not as acute as for traditional criterion (5). Third, the author identification quality of synthesized criterion (8) with frequency of words as a feature (the last column of Table 1) is rather good even in comparison with known best results for Russian texts [17]. And, fourth, conventional MLP is not the best choice in this task because model sets of objects may contain equal patterns. As a matter of fact, MLP should be used to compare patterns extracted from the model sets. such as an estimation of distribution (3). Unfortunately, this procedure shows good recognition quality only if the database contains several samples per class. In our experiment we have one class per sample, hence this approach has not been applied.

5 Conclusion

The proposed network (Fig. 1) is a generalization of the conventional PNN in statistical recognition problem of a set of patterns. Our modification has all known advantages of the PNN [11] over classifiers based on other neural networks. First of all, it is an excellent training speed in comparison with back propagation. The new sample can be added even in real time applications. The network begins to generalize each new observed set of patterns causing the decision boundary to become closer to the optimal one. Unlike many networks, the PNN (8) does not contain recursive connections from the neurons back to the inputs. Thus, it could be implemented completely in parallel [12].

The key difference of (8) from other approaches is the usage of the query sample $\mathbf{X}$ to estimate the joint probabilistic quantity of the united sample $\{\mathbf{X}, \mathbf{X}_1, ..., \mathbf{X}_R\}$. Hence, the distribution of each model class r is estimated by the united sample $\{\mathbf{X}, \mathbf{X}_r\}$. The query sample is the part of the second, pattern, layer of the PNN, and each model sample became a member of the first, input, layer.

The most significant advantage of the proposed classifier (8) is that its decision boundary converges to the Bayes optimal solution [2]. The rate of the convergence is essentially higher than the rate for the classical PNN (5) in the most acute case of a pattern recognition when the training sample size is approximately equal to the size of an input sample $n_r \approx n$. At last, the proper choice of the smoothing parameter σ of the Gaussian kernel (4) is not as complex as for the conventional PNN [16], [23]. Our experimental study showed that the criterion (8) is much more resistant to change σ than the PNN (5), though the maximal accuracy of (8) is practically equal to the maximal accuracy of (5). Thus, our network (Fig. 1) could achieve better quality in time-varying environment. However, the accuracy of the proposed network is 2-5 times better than the accuracy of the PNN (5) with a discrete delta function kernel (8), which is known to be a limit ($\sigma \to 0$) of Gaussian kernel (4).

One other advantage of the proposed PNN is that the measure of similarity in (8) is symmetric as the importance of training and input sets is equivalent in the homogeneity testing. On the other hand, the similarity measure in (5) is asymmetric. Really, in statistical classification the model probability distribution (evaluated by the training sample) is much more important than the input sample distribution. In this task it is supposed that $n_r >> n$, so the quality of the model probability distribution estimation is much higher than the quality of the query sample. This fact is an additional argument in behalf of (8) as symmetry is a desired property in many pattern recognition algorithms [1] (e.g., clustering).

Unfortunately, our network (Fig. 1) possesses the same shortcoming as the PNN. First of all, our network requires large memory to store all training samples. Second, the classification speed is low as the network is based on an exhaustive search through all training samples [20]. Moreover, the proposed network classifies the input sample twice slower than the original PNN. However, this fact is not a real obstacle in practical pattern recognition tasks (author identification, image recognition), as the training sample size is not usually large. Third, our network is not as general as the traditional PNN because we require the network input to be a sample with the size which is the same order of magnitude as the training sample size [1], [20].

Thus, in this study we proposed the novel modification of the PNN (8) in recognition of a set of patterns. We experimentally proved that this network is in some terms better than the conventional PNN (5). The PNN modification proposed here can be used in various pattern recognition tasks [1], such as image and speech recognition.

Acknowledgements. This research is supported by Federal Grant-in-Aid Program "Research and development on priority directions of scientific-technological complex of Russia for 2007-2013" (Governmental Contract No. 07.514.11.4137).

References

1. Theodoridis, S., Koutroumbas, C.: Pattern Recognition, 4th edn. Elsevier Inc. (2009)
2. Borovkov, A.A.: Mathematical Statistics. Gordon and Breach Science Publishers (1998)
3. Vapnik, V.N.: Statistical Learning Theory. Wiley, New York (1998)
4. Webb, A.R.: Statistical Pattern Recognition. Wiley, New York (2002)
5. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification. Wiley, New York (2001)
6. Efromovich, S.: Nonparametric Curve Estimation. Methods, Theory and Applications. Springer, New York (1999)
7. Murthy, V.K.: Estimation of probability density. Annals of Mathematical Statistics 36, 1027–1031 (1965)
8. Parzen, E.: On estimation of a probability density function and mode. Annals of Mathematical Statistics 33, 1065–1076 (1962)
9. Greblicki, W.: Asymptotically optimal pattern recognition procedures with density estimates. IEEE Transactions on Information Theory IT-24, 250–251 (1978)
10. Wolverton, C.T., Wagner, T.J.: Asymptotically optimal discriminant functions for pattern classification. IEEE Transactions on Information Theory 15, 258–265 (1969)
11. Specht, D.F.: Probabilistic neural networks. Neural Networks 3, 109–118 (1990)
12. Specht, D.F.: Probabilistic Neural Networks for Classification, Mapping, or Associative Memory. In: IEEE International Conference on Neural Networks, vol. I, pp. 525–532 (1988)
13. Specht, D.F.: A general regression neural network. IEEE Transactions on Neural Networks 2(6), 568–576 (1991)
14. Rutkowski, L.: Adaptive Probabilistic Neural Networks for Pattern Classification in Time-Varying Environment. IEEE Transactions on Neural Networks 15(4), 811–827 (2004)
15. Kullback, S.: Information Theory and Statistics. Dover Pub. (1997)
16. Jones, M.C., Marron, J.S., Sheather, S.J.: A brief survey of bandwidh selection for density estimation. Journal of the American Statistical Association 91, 401–407 (1996)
17. Kukushkina, O.V., Polikarpov, A.A., Khmelev, D.V.: Using Literal and Grammatical Statistics for Authorship Attribution. Problems of Information Transmission 37(2), 172–184 (2001)
18. The e-library of Maxim Moshkov, `http://www.lib.ru`
19. Savchenko, A.V.: Image Recognition with a Large Database Using Method of Directed Enumeration Alternatives Modification. In: Kuznetsov, S.O., Ślęzak, D., Hepting, D.H., Mirkin, B.G. (eds.) RSFDGrC 2011. LNCS (LNAI), vol. 6743, pp. 338–341. Springer, Heidelberg (2011)
20. Savchenko, A.V.: Directed enumeration method in image recognition. Pattern Recognition 45(8), 2952–2961 (2012)
21. Aizerman, M.A., Braverman, E.M., Rozonoer, L.I.: Theoretical foundations of the potential function method in pattern recognition learning. Automation and Remote Control 25, 821–837 (1964)
22. Stamatatos, E.: A survey of modern authorship attribution methods. Journal of the American Society for Information Science and Technology 60(3), 538–556 (2009)
23. Mao, K.Z., Tan, K.-C., Ser, W.: Probabilistic neural-network structure determination for pattern classification. IEEE Transactions on Neural Networks 11, 1009–1016 (2000)