Barley Seeds Classification with a Genetically Optimized Kernel Density Estimator

VASILE GUI, FLORIN ALEXA, CATALIN CALEANU
Department of Electronics and Telecommunications
„Politehnica" University Timisoara
Bd. V. Parvan No. 2, 30223 Timisoara
ROMANIA

Abstract: - This paper summarizes our new research on a unified design of the feature extraction and classification processes. We have developed a multiple class classifier, based on a genetically optimized kernel density estimator. The genetic algorithm provides the tool to solve the bandwidth matrix optimization problem. The bandwidth matrix of the kernel function plays a similar role with the matrix used in a linear feature extraction, since it weights differently the data vector components. More, the bandwidth matrix controls the smoothness of decision surfaces. Tests are made with both, the standard nonparametric k-Nearest Neighbour classifier and the genetically optimized kernel based density estimator. Results on a barley seed image feature data show the utility of the proposed approach.

Key-Words - Pattern recognition, density estimation, kernel function, bandwidth matrix, genetic algorithm, feature extraction

1 Introduction

Designing good classifiers for applications with complex patterns, such as textures, text categorization, video indexing, or data mining is a challenging task. Lacking a deep understanding of class specificities and facing pattern complexity, designers usually end up by measuring a broad range of features that hopefully would capture enough information for correct classification. Unfortunately, this approach raises several problems. Perhaps one of the most obvious is the increase of computing time needed to generate each feature vector. More importantly, as the dimensionality of the feature space is increased and a significant part of the features might contain more noise than useful information, the classifier design become much more difficult. Theoretically, any new and statistically independent feature contributes to reducing the probability of classification error, as long as the number of feature samples available for designing the classifier is unlimited [1]. The higher the dimension of the feature space, the more feature samples are needed for proper classifier design. In practice, the number of training samples available is always limited and having to compute longer feature vectors means actually increasing the cost of generating feature samples. Most often than not, this contributes to reducing the number of samples generated for training and testing the classifier. With limited training sample size, the inclusion of new features beyond some limit can result in actually reducing the probability of correct classification, a paradox referred to as *curse of dimensionality* or *peaking phenomenon*.

Kernel probability density estimation techniques [2] have been used longtime ago in pattern recognition and are most often referred to as Parzen window techniques. Their popularity was longtime shadowed by the high memory requirements and computational complexity. However, recent research gave rise to very efficient computing algorithms, like the Fast Gauss Transform and it's improved version proposed by Yang [3][4]. Recent interest in nonparametric density estimation techniques in the field of pattern recognition is sustained by the papers [5][6].

The main advantage of nonparametric over parametric methods consists of the relative insensitivity of the nonparametric estimators to the exact shape of the multivariate distribution of the data vectors being classified. Parametric methods may lead to significant classification errors when the chosen distribution model does not fit well the underlying distribution. Finding a good model for a high dimensional complex distribution is not a trivial task and the chances to pick the wrong model cannot be ignored. Therefore, nonparametric methods are a good candidate to solve high dimensional problems.

In this work we study the effect of scale parameters used by nonparametric classifiers in the context of high dimensional noisy data. We find an optimized bandwidth matrix for the multivariate kernel function through evolutionary search, which is equivalent with a feature extraction transform and evaluate the improvements in accuracy, obtained over the traditional k-NN classifier. Our experimental results show that the proposed approach is effective.

2 Nonparametric classifier using kernel density estimation

Nonparametric density estimators can be designed without having to define a specific distribution law. Instead, a kernel function is used to estimate the probability density at the desired location in the feature space. The most popular nonparametric density estimator used in pattern recognition is the k-NN classifier. It classifies a data sample from an unknown class based on its proximity to available samples with known class. The class with most samples within the set of the nearest k samples to the unknown sample is assigned to the unknown sample. The number of neighbors used, k, is the only one parameter of this classifier. A high value of k results in smoother decision surfaces. In spite of its simplicity, the k-NN classifier proved to work very well in many applications. However, the underlying assumption of uniform distribution allowing the k-NN classifier to be interpreted as a density estimation based classifier is restrictive. In the present work, we abandon this restriction in favor of a more general nonparametric density model.

Given a sample of N d-dimensional data points, $\mathbf{x}_i$, drawn from a distribution with multivariate probability density function $p(\mathbf{x})$, the kernel estimate of this density at $\mathbf{x}$ can be written as [7]:

$$\hat{p}_{\mathbf{H}}(\mathbf{x}) = \frac{1}{N}\sum_{i=1}^{N} K_{\mathbf{H}}(\mathbf{x}-\mathbf{x}_i) \quad (1)$$

where

$$K_{\mathbf{H}}(\mathbf{x}) = |\,\mathbf{H}\,|^{-1/2}\; K\left(\mathbf{H}^{-1/2}\mathbf{x}\right) \quad (2)$$

is the kernel function depending on a symmetric positive definite $d\times d$ matrix, called bandwidth matrix. To do pattern classification, equation (1) is used in the form:

$$\hat{p}_{\mathbf{H}}(\mathbf{x}\mid\omega_i) = \frac{1}{N}\sum_{i=1,x_i\in\omega_i}^{N} K_{\mathbf{H}}(\mathbf{x}-\mathbf{x}_i) \quad (3)$$

and the feature vector $\mathbf{x}$ is classified into class c, such that

$$c = \arg\max_i \hat{p}_{\mathbf{H}}(\mathbf{x}\mid\omega_i) \quad (4)$$

Finding the right bandwidth is a main topic in nonparametric density estimation literature. In the present work, we used a kernel function with a diagonal bandwidth matrix:

$$H = \begin{bmatrix} h_0 & 0 & 0 & \cdots & 0 \\ 0 & h_1 & 0 & 0 & 0 \\ 0 & 0 & h_2 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & h_{d-1} \end{bmatrix} \quad (5)$$

It can be seen that the bandwidth matrix acts like a feature extraction transform matrix, multiplying the input data prior to density estimation through the kernel function. Parameters h_i scale the contributions of the data vector components and control the amount of smoothing introduced by the kernel function. In this work we used the Gaussian kernel function. As a result, our nonparametric density estimator can be thought of also as a Gaussian mixture model, with every training data giving rise to one Gaussian.

3 Genetic optimization of the bandwidth matrix

Theoretically, nonparametric classifiers like the k-NN, do not need to learn. However, in spite of being nonparametric, kernel probability density estimators use their own parameters, namely, the coefficients of the bandwidth matrix. In this work, we use a learning procedure to find a set of optimal parameters of the bandwidth matrix, h_i, considering the classification task. This is in contrast with the main stream of the work in data driven bandwidth selection [8], where the mean squared error or mean integrated squared error. The resulting, genetically optimized, kernel density estimator will be called GOKDE.

Many optimization techniques have been proposed to solve hard optimization problems, with multiple extremes, as proved to be our search space. Our choice went to genetic algorithms [9], considering their ability to avoid being trapped in local extreme points. Genetic algorithms have already been used in pattern classifier design for feature selection in several works, such as [10], [11].

Genetic algorithms are search procedures inspired by the mechanisms of natural selection and genetics. The search starts from a set of randomly generated candidate solutions. Throughout iterations, each solution is *evaluated* by a *fitness function*. Then, the set of solutions (genes) undergoes an evolution process through *selection*, *crossover* and *mutation* steps. The fitness function used in the present work is the classification accuracy.

4 Experiments with barley seed data

Barley seeds are the main ingredient in beer production. In order to obtain a high quality beer, the technological parameters of the brewing process have to be adjusted to make the best of the available species. When more barley species are mixed together, the quality is compromised. Ideally, there should be only one species used in one process. In practice, farmers have difficulties in separating different species and therefore manufacturers receive seeds from one predominant species mixed with some other contaminating species. The purity of the seed set is an important measure of the seed set quality and therefore has to be evaluated. In this work, we classify four barley species, grown in Finland:

- Artturi maha,
- Innari maha,
- Saana maha,
- Prisma maha.

Finding an appropriate set of image descriptors for barley species recognition is a challenging task, due to high similarity between species and high variety within species. In this paper, we use shape and texture features extracted from barley seed images. Prior to texture feature extraction, we segment the seeds from the background, then filter the seed images with a Laplacian of Gaussian filter, followed by a nonlinear processing method in order to retain the important textural information. The details of the processing steps are not within the scope of this paper, but an illustration of the processing results is given in figure 1. Texture features extracted from seed images include statistical moments and edge direction histogram. Shape parameters are extracted from a hexagonal approximation of the seed, as illustrated in figure 2. A set of 25 dimensional feature vectors was generated for 400 images used in the experiments. Each class was represented by 100 labeled samples.

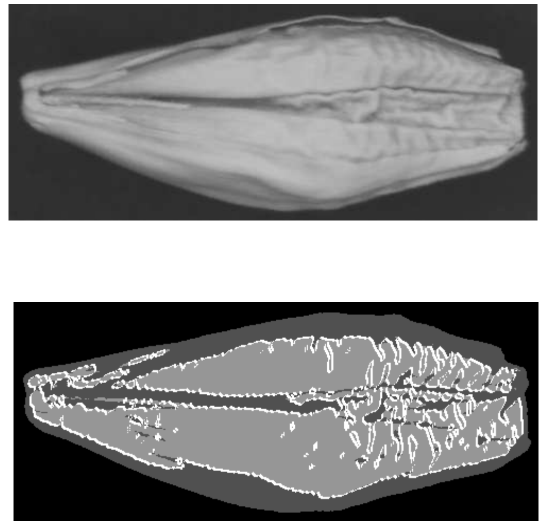

Fig.1. a) Artturi maha species barley seed; b) Processed Artturi maha seed image

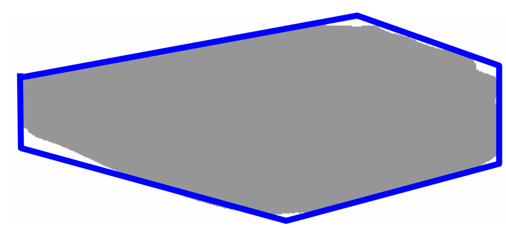

Fig.2. Hexagonal approximation of the barley seed shape

To test the performances of the proposed GOKDE classifier, we made two sets of experiments. In the first set of experiments, we compared our classifier with the k-NN classifier. To this end, we split our data set in a training set and a test set. The training set contained 80 seeds, while the test set contained 20 seeds for each class. We formed 10 such pairs of sets with a random choice of the training set. In the training phase, the nonparametric classifier accuracy was used as a fitness function, and the solution searched was the set of bandwidth matrix parameters, grouped in a 25 dimensional vector. With the optimal bandwidth matrix parameters, we evaluated the error rate on the test set. The k-NN classifier was run on the same data. The best results for the k-NN classifier were obtained for *k*=5. In figure 3, compare these results with the results of the proposed nonparametric.

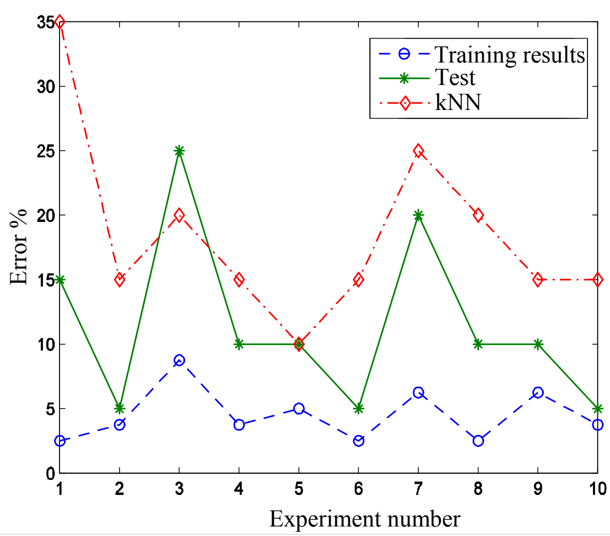

Fig. 3. Error rates for the GOKDE classifier and the k-NN classifier on barley seed data

Average errors and standard deviations of the results are given in Table 1.

Table 1.

	Training	Test	kNN
Mean error	4.375	11.5	18.5
Standard deviation of error	1.7922	6.6875	5.7975

From the experiments, it can be seen that the proposed GOKDE outperforms the k-NN classifier on the barley seed data. The generalization capability of the proposed classifier is however rather modest. This can be explained by the over-fitting phenomenon, generated by the relatively low number of training samples, considering the high dimensional feature space. The average error rate obtained is marginally acceptable from the point of view of practical needs.

4 Conclusions

In this work, we presented the results of barley seed classification based on a genetically optimized kernel density estimator. The proposed genetically optimized kernel density estimator based classifier outperformed the traditional nonparametric k-NN classifier, as a result of its capability to weight vector components according to their usefulness in the classification task and to reshape the decision space after additional optimization. We believe that the performances of the proposed classifier can be further improved by using a larger seed data base, thus the avoiding the over-fitting problem.

References:

[1] R. O. Duda, P. E. Hart, D. Stork, “Pattern Classification” Second edition, Wiley New York, 2001.

[2] D.W. Scott, Multivariate Density Estimation, New York: Wiley, 1992.

[3] A.Elgamal, R.Duraiswami, L.S.Davis, “Efficient kernel density estimation using the Fast Gauss Transform with applications to color modeling and tracking”, *IEEE Trans. Pattern Anal. Machine Intell*. Vol. 25, No. 11, 2003, pp. 1499-1504.

[4] J. Yang, R. Duraiswami, N. Gumerov, L. Davis, “Improved Fast Gauss Transform for efficient kernel density estimation”, *IEEE Intl. Conference on Computer Vision, ICCV*, 2003, pp. 464-471.

[5] M. Browne, “a geometric approach to non-parametric density estimation”. *Pattern Recognition*, Vol. 40, Issue 1, January 2007, pp. 134-140.

[6] C. Fang, X. Ding, “Improvement of ICA based Probability density estimation for pattern recognition”. *17th International Conference on Pattern recognition (ICPR’04)*, Vol. 3, pp. 466-469.

[7] D.Comaniciu, P.Meer, “Mean shift: A robust approach toward feature space analysis”, *IEEE Trans. Pattern Anal. Machine Intell,* Vol. 24, No.5, 2002, pp. 603-619.

[8] S.J. Sheather, M.C. Jones, A Reliable Data-based Bandwidth Selection Method for Kernel Density Estimation," J. R. Statist. Soc. B, 53(3):683{690, 1991.

[9] D. E. Goldberg, “Genetic algorithms in search, optimization and machine learning”. Addison-Wesley, Reading, Mass., 1989.

[10] W. Siedlecki and J. Sklansky, “A note on genetic algorithms for large-scale feature selection”, Pattern recognition Letters, Vol. 10, 1989, pp 335-347.

[11] M.L. Raymer et al., “Dimensionality reduction using genetic algorithms”. IEEE Trans. Evolutionary Computing, Vol. 4, No. 2, 2000, pp 164-171.