CLASS OVERLAP RATE AS A DESIGN CRITERION FOR A PARALLEL NEAREST NEIGHBOUR CLASSIFIER

Adam Jóźwik[1], Leszek Chmielewski[2,3], Marek Skłodowski[2,3], Waldemar Cudny[2,3]

[1] **Institute of Biocybernetics and Biomedical Engineering, PAS**
Trojdena 4 02-109 Warsaw www.ibib.waw.pl
[2] **Institute of Fundamental Technological Research, PAS**
Świętokrzyska 21 00-049 Warsaw www.ippt.gov.pl
[3] **Association for Image Processing**
www.tpo.org.pl

ABSTRACT

It is well known that the classifiers based on the 1-NN and the k-NN rules offer good performance and that their implementation is simple. However, classification of a single object requires calculation of a large number of distances, equal to the power of the reference set. The classification speed depends also on the number of features of the objects. This paper deals with the problems of feature selection and reference set reduction as the methods of improving the classification speed. As a classifier the parallel net of two-decision k-NN classifiers is considered. Each of the component k-NN classifiers is approximated by a 1-NN classifier with the reduced reference set. The proposed classifier is experimentally compared with the standard k-NN one. a very requiring data set obtained in an industrial application to quality inspection of the surfaces of ferrite products has been used in the experiments.

INTRODUCTION

The proposed method of desinging a nearest neighbour classifier has been developed within the context of visual quality inspection of the surfaces of ferrite products. In this papper we focus our attention on the latest results concerning the applied pattern recognition method. Other aspects of the general quality inspection approach have been discussed elsewhere [4, 5, 6, 7, 10, 11, 12]. The previous versions of the classifier have been described in [4, 10].

The applied concept of surface inspection consists in classification of the pixels of a numerical image of the product. Features of each pixel are calculated from its neighbourhood [5, 10, 12]. The classification algorithm is not applied to the pixels of the whole image, but only to the potentially defective regions indicated by a series of detectors: the morphological pyramid [5, 7, 8, 9, 12], the morphological detector of elongated defects (cracks) [12], and adaptive thresholding [12]. After detection and classification, the map of classes is postprocessed and segmented, and the dimensions of defects are compared to respective standards. Requirements concerning the classification speed are

Proc. *1st Polish Conference on Computer Pattern Recognition Systems KOSYR'99*, Trzebieszowice, Poland, May 24-27, 1999, 109-115. http://www.zssk.pwr.wroc.pl/zssk/kosyr/kosyr.htm

very high since a large number of pixels must be analysed to assign a quality class to a single product.

The statistical object recognition theory offers some methods in which a *training set* or a *reference set* is used. The training set contains objects with known class membership. Information contained in this set serves for construction of the classifier which is used to classify the new objects. The whole computational process needed for the construction of the classifier is called *training*. Usually it consumes a lot of time. The fastest training is promised by methods based on 1-NN or k-NN rules. This is the main reason why such kind of classifiers appears as the one most interesting to us.

To evaluate the classification quality a *testing set* can be used, or the well known *leave one out* method can be applied to the training set. It consists in classification of each object from the reference set by the classifier derived from the reference set reduced by exclusion of the classified object (the classified object do not take part in the training). The process is repeated for all the objects from the reference set. The number of misclassified objects divided by the power of the reference set gives a ratio that estimates the misclassification probability, *i.e.* the error rate. We shall consider the error rate estimation with the *leave one out* method since it explores the training set more extensively than the *resubstitution* method – dividing the reference set into subsets larger than one object, and taking each of them as a testing set, in sequence.

The standard 1-NN or k-NN rule can operate for an arbitrary number of classes. No training is required for the fixed feature set if we decide to use the 1-NN classifier. It assigns the classified object to the same class as its *nearest neighbour* (in the feature space) from the reference set. The time of the error rate calculation depends then only on the reference set size. The k-NN rule assigns the classified object to the class most heavily represented from among its k nearest *neighbours* in the reference set. Usually the fast to calculate *city* distance measure is used. If the k-NN rule is used we need to determine the value of k. Thus, the error rate must be calculated for all possible values of k, and the k that offers the smallest error rate is taken. If the reference set contain m objects then m-1 different k-NN rules must be evaluated. Although all m-1 error rates can be calculated simultaneously it is obvious that evaluation of the fixed feature set by the k-NN rule takes much more time when compared with 1-NN one.

To reject the redundant features which may decrease the classification quality the feature selection is strongly recommended. It requires a review a large number of feature combinations. For each combination a separate error rate calculation session is required. The review of all possible feature combination is the only strategy which ensures finding the optimum feature set. For the large number n of features, realisation of this feature selection strategy can be impossible. In such cases forward or backward feature selection strategy may be used [1]. The former consists in sequential adding one, currently best feature, starting with a single feature which offers the smallest error rate. Finally, the full set of n features is evaluated. The later strategy starts with the full set of n features. Then n combinations of n-1 features are evaluated. Next, n-1 feature combinations by n-2 features are analysed. This process is stopped with the single feature. In both strategies the best one of all the reviewed feature sets is selected.

THE PROPOSED FEATURE SELECTION CRITERION

It is well known that the objects which are close to the class boundaries are more "difficult" to be recognised than the ones lying far from it. As a feature selection criterion we can use the number of the difficult objects.

Proc. *1st Polish Conference on Computer Pattern Recognition Systems KOSYR'99*, Trzebieszowice, Poland, May 24-27, 1999, 109-115. http://www.zssk.pwr.wroc.pl/zssk/kosyr/kosyr.htm

Let us define the *difficult* objects. If *nc* is the number of classes then the reference set consists of *nc* subsets: X_1, X_2, ..., X_{nc} and each X_i contains objects from the class *i* only. We associate these sets with the positive real numbers e_1, e_2, ..., e_{nc} defined as follows:

$$e_i = \max_{x_j \in X_i} d(X_i - x_j, x_j) \tag{1}$$

where *d* denotes a distance function between an object x_j and a set $X_i - x_j$, and x_j is an element of X_i, *i.e.*, *d* is the distance between x_j and the nearest object in $X_i - x_j$. We also define the areas A_1, A_2, ..., A_{nc}:

$$A_i = \{x: d(X_i, x) \le e_i\} \tag{2}$$

Let us denote by *A* the set of all points in the feature space each of which belongs to only one of the areas A_i, $i = 1, 2, ..., nc$. Now, we can define the meaning of the *difficult* objects. The object is *difficult* for recognition if and only if it falls simultaneously in the two or more areas A_i. A ratio obtained by division of the number of difficult objects by the power of the training set will be called the *class overlap rate*. This rate is easily calculated. For two classes with equal numbers of objects in the training set the class overlap rate requires twice less number of distances than those which must be calculated when compared with the classical error rate. Hence, it is worthwhile to verify its usefulness to the feature selection problem.

THE PARALLEL NET OF TWO-DECISION K-NN CLASSIFIERS

The k-NN classifier obtained by performing feature selection and finding the optimum *k* determines *nc* decision regions in the feature space. Each region corresponds to a different class. The boundary between any two classes *i* and *j* depends on the selected features and on the value of *k*. The selected features as well as the value of *k* depend on objects from the third classes, *i.e.* from the classes other than *i* and *j*. Thus, the objects from the classes different than *i* and *j* influence the shape of the boundary separating the classes *i* and *j*. This influence acts as noise.

To reduce the influence of the third classes on the boundary between the two classes *i* and *j* a parallel net of two-decision k-NN classifiers has been recommended [3], each of them corresponding to a different pair of classes. Hence, there are *nc*(*nc*-1)/2 different component classifiers. Feature selection as well as the determination of *k* is performed for each of the component classifiers separately. The object to be recognised is presented simultaneously to all the component classifiers. Final decision of the whole net is created by voting of the component classifiers and the class that gathers the largest number of voices is finally assigned.

NECESSITY OF REDUCTION OF THE REFERENCE SET

The use of the k-NN instead of the 1-NN classifier increases the classification quality but makes it slower. To keep the classification speed of the 1-NN classifier and the classification quality close to that offered by the k-NN one we can approximate the k-NN classifier by the 1-NN one. The only thing we need to do is to reclassify the whole reference set by the (k+1)-NN rule, where *k* is the value found by the *leave one out* method as the one which minimises the error rate. The use of (k+1)-NN instead of k-NN results from the fact that this time the classified object belongs simultaneously to the

Proc. *1st Polish Conference on Computer Pattern Recognition Systems KOSYR'99*, Trzebieszowice, Poland, May 24-27, 1999, 109-115. http://www.zssk.pwr.wroc.pl/zssk/kosyr/kosyr.htm

reference set. So, now for each object of the reference set the (k+1) *nearest neighbours* occupy exactly the same hypersphere as *k nearest neighbours* during the realisation of the *leave one out* method.

The speed of the 1-NN classifier depends linearly on the reference set size. Hence, it is reasonable to use the reference sets of small sizes, instead of the whole training set. Some objects, mainly those lying far from the class boundaries, can be rejected from the training set without significantly changing the classifier performance. Numerous procedures for reducing the reference set in the 1-NN can be found in the literature. The fastest one is the Hart's algorithm.

HART'S ALGORITHM AND ITS MODIFICATION

The Hart's algorithm [2] operates as follows. The reduced set is initialised with the first object from the original reference set. Then, all the remaining objects are classified by the 1-NN rule with the present reduced set as a reference set. Each misclassified object is included into the reduced set. After presenting all the objects, again all the objects from the original set are classified by a 1-NN classifier, with the current reduced set. These steps are continued until all the objects from the original reference set are correctly classified, *i.e.*, till the reduced reference set is *consistent* with the original one.

It can be noticed that mainly those objects which are close to the class boundary are selected to the reduced set, except those chosen at the beginning. To remove this undesired phenomenon, in each step of reduction the reference set is renumbered in such a way that objects previously chosen as the last ones are now presented as the first ones.

PREPARATION OF THE TRAINING SET

The training set has been prepared with the use of a series of images of the mating surfaces of ferrite products (in the described experiments these were magnets; in other experiments also ferrite cores were used). Pixels belonging to the classes of defects of the product (referred to as *irregularities* by the ferrite specialists), and to the classes of good surface, have been marked with colours corresponding to classes. Looking at the surface of the original magnets through a looking glass is necessary in this tedious work. The training set consists of 4297 pixels (objects), 64 features each. There are 14 classes, like *good object*, *background*, *chip*, *pull-out* and *crack*. Some of them were divided, as *e.g. bright pull-out – edge with good object*, or *dark pull-out*. One of the training images is shown in Figure 1b. The most detailed description of classes and features can be found in [12]. For features, see also [5, 10].

COMPUTATIONS

The first experiment has been made with the standard k-NN classifier with the optimum value of k and all 64 features. In the second attempt the same kind of the classifier has been applied with the feature set selected by the use of error rate as a criterion. The overlap rate has also been used in the standard k-NN classifier. Results of these three experiments can be seen in Table 1, rows 1-3. It can be noticed that for this kind of classifier the feature selection with the error rate as a criterion remarkably improves the classification quality. The error rate decreased from 23.81% to 16.62%. However, the use of the overlap rate as a criterion offered the error rate (25.51%) even higher than the one obtained for all features (23.81%).

The fourth experiment has been made with the parallel net of two-decision k-NN classifiers. As the feature selection criterion the error rate offered by k-NN rules with

Proc. *1st Polish Conference on Computer Pattern Recognition Systems KOSYR'99*, Trzebieszowice, Poland, May 24-27, 1999, 109-115. http://www.zssk.pwr.wroc.pl/zssk/kosyr/kosyr.htm

the optimum values of k has been used. So, the feature selection criterion was based on the same type of classifiers as those used as the components of the considered net.

In the last, fifth experiment we have applied as a criterion the class overlap rate. Furthermore, all the component classifiers were replaced by the 1-NN ones. As we can see from the Table 1, rows 4 and 5, for this kind of classifier the results of feature selection based on the class overlap rate was significantly better. The error rate could be reduced from 4.84% to as little as 1.4%.

In the experiments 4 and 5, for each component classifier a different subset of features has been selected, but all the available features were used in the whole net.

Table 1. Results of experiments with the standard k-NN classifier (experiments 1-3) and with the parallel net of two-decision classifiers (experiments 4-5).

Experiment no.	Type of classifier	Feature selection criterion	Number of features	Error rate
1	k-NN with optimum k	no selection	64	23.81%
2	k-NN with optimum k	error rate	16	16.62%
3	k-NN with optimum k	overlap rate	24	25.51%
4	parallel net of k-NN classifiers	error rate	64	4.84%
5	parallel net of 1-NN classifiers	overlap rate	64	1.40%

An example of final results of classification for one image can be seen in Figure 1c. The reader interested in more examples and details is kindly invited to visit the web site of the project SQUASH at `www.tpo.org.pl/squash/`.

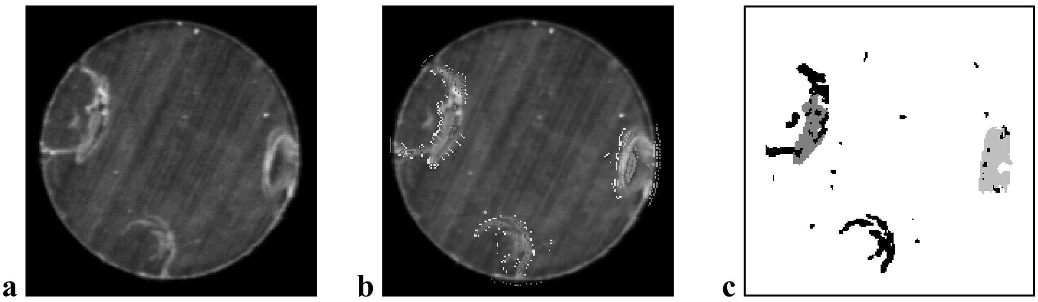

Figure 1. Training image and classification results. a: input grey image; b: training pixels – classes marked with colours; c: result of classification (postprocessed [12]); black: *crack*, dark grey: *pull-out*, light grey: *chip*; colours different from those in b. Note that in c much more pixels are classified than marked in b.

CONCLUSIONS

Our aim was to construct a fast classifier which would also offer a very good performance. The obtained results imply that to perform this task we need the following: 1^o feature selection with the overlap rate as a criterion, 2^o reclassification of the objects in the reference sets for each pair of classes with the 1-NN rule, 3^o reduction of these sets with the above suggested modification of the Hart's algorithm.

As it was mentioned previously the speed of classification depend on the number of features and on the powers of the reference sets for the class pairs. This dependence is nearly linear in the number of features as well as in the set powers. It is then necessary to compare the total sizes of the reference sets for all class pairs before and after feature selection, and before and after reference sets reduction. We have done this for the best classifier number five – the last row of the Table 1.

The training set occupies about 3.5 MB. The full reference set (after forming all the class pairs) occupies 45.2 MB (row 1 in Table 1). Feature selection reduces this size to 6.7 MB, *i.e.*, about 7 times, and Hart's algorithm further reduces it to 0.6 MB (row 5). In the latter case, classification speed is $\approx$170 obj/s (GNU C, Pentium 200 MHz).

No theoretical explanations have been attempted to find out why the feature selection with the use of the overlap rate gives remarkably better results. Although the results for the component classifiers are the worst, the final misclassification rate is a few times lower. The reason should be in better voting of the component classifiers. For the parallel classifier the gap between classes in a pair is higher if the class overlap criterion is used. In the case of the error rate treated as a criterion, the training stops at the feature set which offers error rate equal to zero. However, another feature set could exist which would also offer zero error rate, plus a greater gap between the classes. Such feature set could have been found with the class overlap criterion.

Acknowledgement This research was supported by the European Commission within the INCO-COPERNICUS Programme, project *Standard compliant QUAlity control System for High-level ceramic material manufacturing (SQUASH)*, contract no. ERBIC 15CT 96 0742.

REFERENCES

[1] Devijver P.A., Kittler J., *Pattern Recognition: a Statistical Approach*, Prentice Hall, London 1982.

[2] Hart P.E., The condensed Nearest Neighbor rule, IEEE Trans. *Information Theory*, IT-14, 3, May 1968, pp. 515-516.

[3] Jóźwik A., Object recognition method based on k Nearest Neighbor rule, *J. of Communications*, XLV, Jul-Aug 1994, pp. 27-29.

[4] Jóźwik A., Chmielewski L., Cudny W., Skłodowski M., a 1-NN preclassifier for fuzzy k-NN rule, *Proc. 13th Int. Conf. Pattern Recogn.*, Wien, Austria, Aug 25-29, 1996, vol. 4, pp. D-234 - D-238.

[5] Chmielewski L., Skłodowski M., Cudny W., Nieniewski M., Jóźwik A., Optical system for detection and classification of surface defects in ferrites, *Proc. 3rd Symp. Image Processing Techniques (3TPO)*, Serock, Poland, Oct 29-31, 1997, pp. 1-13.

[6] *Light systems for testing the surface defects of ferrite cores*, EPSILON ARG Ltd, Jan 1997. Annex to [7].

[7] Fontana F., Mari M., Chetverikov D., Lugg M., Postupolski T., Jóźwik A., Skłodowski M., Nieniewski M., Cudny W., Chmielewski L., *CRAck and SHape defect detection in ferrite cores – Final Technical Report of the project CRASH*, CRASH Consortium, Jan 1997.

[8] Nguyen T.C., Nieniewski M., Using genetic algorithm for optimisation of the thresholds in the morphological pyramid for detecting defects on the surface of ferrite cores, *Proc. 3rd Symp. Image Processing Techniques (3 TPO)*, Serock, Poland, Oct 29-31, 1997, pp. 229-242.

[9] Nieniewski M., Morphological method of detection of defects on the surface of ferrite cores, *Proc. 10th Scandinavian Conf. Image Analysis*, Lappeenranta, Finland, Jun 9-11, 1997, pp. 323-330.

[10] Jóźwik A., Chmielewski L., Skłodowski M., Cudny W., a parallel net of (1-NN, k-NN) classifiers for optical inspection of surface defects in ferrites, *Machine Graphics & Vision*, 7, 1-2, 1998, pp. 99-112.

[11] Chmielewski L., Skłodowski M., Cudny W., Nieniewski M., Detection and classification of surface irregularities in black ceramics. *Proc. Workshop on Standardisation in the Field of Non-Destructive Testing and Quality Control of Sintered Powder Materials*, Cagliari, Italy, Nov 6, 1998, pp. 1-13. Università di Cagliari. Extended summary.

[12] Mari M., Fontana F., Vernazza G., Chetverikov D., Verestoy J., Lugg M, Postupolski T., Wiœniewska A., Okoniewska E., Jóźwik A., Chmielewski L., Skłodowski M., Nieniewski M., Cudny W., Thang N.C., Standard compliant QUAlity control System for High-level ceramic material manufacturing – Final Technical Report of the project SQUASH, SQUASH Consortium, Jan 1999.

Proc. *1st Polish Conference on Computer Pattern Recognition Systems KOSYR'99*, Trzebieszowice, Poland, May 24-27, 1999, 109-115. `http://www.zssk.pwr.wroc.pl/zssk/kosyr/kosyr.htm`