CONSISTENT SUBSETS IN SPEECH RECOGNITION SYSTEMS

Stefan Grocholewski
Institute of Computing Science, Poznań University of Technology
Piotrowo 3a, 60-965 Poznań, Poland
Tel: +48 (0)61 782 373; fax: +48 (0)61 771 525
grocholew@pozn1v.tup.edu.pl

ABSTRACT

In the paper the method of the transformation of the learning samples into their representatives is presented. The proposed algorithm combines the features of the neural nets approach, i.e. the representatives lie near the boundaries separating the classes, and cluster seeking approach - each representative corresponds to the group of elements lying close to each other. By using the *consistent subset* the drawbacks of those approaches (cluster can comprise samples from different classes; the sophisticated network is not appropriate in the regions where the classes overlap) can be avoided in some cases. Several applications in the area of speech recognition are presented.

1. INTRODUCTION

Speech recognition systems require learning sets in the learning phase. Most frequently the greater is the set the better is the recognition performance. The NN rule is rarely applied in practical pattern recognition because of great amount of data. The aim of the learning step is to transform the great amount of learning data into the smallest possible set. Let us consider two transformation procedures: cluster analysis and neural nets [1]. By cluster analysis several objects close to each other are replaced by the appropriate centroid. Neural nets have the ability to transform the learning set into the non-linear boundaries between the recognised classes by using the piecewise linear or quadratic discriminant functions [2]. In the first case the objects close to each other can represent different classes; it causes that even the learning set cannot be recognised without errors. In the second case the sophisticated network can correctly classify the learning set of objects in the regions where the classes overlap; however the quality of performance in the recognition stages will be very poor in these regions.

In 1968 Hart [3] introduced the *consistent subset* of a learning set. This is the smallest possible subset, which, when used as a stored reference set for the NN rule, correctly classifies all of the remaining objects in the learning set. Although in his rather heuristic iterative algorithm the number of patterns to be stored as representatives is dependent on the sequence in which the learning samples are used to build up the consistent subset, the results were encouraging for others [4],[5].

In the paper the novel method of the transformation of the learning samples into its representatives is presented. The proposed algorithm combines the features of the neural nets approach (the representatives lie near the boundaries separating the classes) and cluster seeking approach (each representative corresponds to the group of the elements lying close to each other). By using the *consistent subset* the drawbacks of those approaches (cluster can comprise samples from different classes; the sophisticated network is not appropriate in the regions where the classes overlap) can be avoided in some cases. The main idea of the algorithm along with the possible applications of the consistent subsets will be presented in the next section.

2. DESCRIPTION OF THE METHOD

Let us assume the learning set from the Fig.1a. The Figs. 1b) and 1c) present the well known description of the learning set by linear boundaries (in general - non-linear) and centroids (marked by asterisks in the Fig. 1c) respectively. The Fig. 1d) presents our approach that consists in leaving only

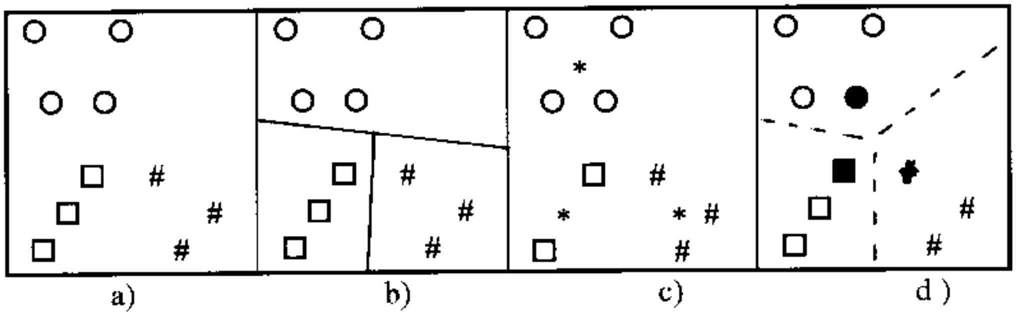

Fig.1. The different approach to the three class learning set transformation

the minimal set (marked by black symbols) necessary to classify correctly all elements of the learning set. Our method combines the neural net approach (three black symbols are equivalent to the decision boundaries marked by dashed lines) and cluster approach (three representatives).

The aim of the paper is to present an appropriate algorithm for finding the set of representatives. The results of the experiments, connected with speaker independent vowels and words recognition, using our method of the transformation of the learning set will be presented.

2.1 Idea of the algorithm

Let us have an object A_1 and its nearest neighbour A_2. The main idea of the algorithm is based on the following remark. If the object A_2 belongs to the same class as A_1 does, object A_1 can be rejected from the learning set only on condition that object A_2 will not. Therefore the object A_2 can be treated as the A_1's representative. If the rejected object A_1 occurs during recognition it will be recognised correctly because A_2 is its nearest neighbour. So the reduction will not deteriorate the recognition performance of the learning set. The only questions are why and when to choose A_1

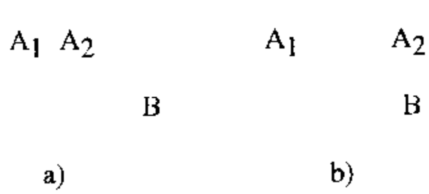

Fig.2. According to our idea in the case a) we can reject A_1 or A_2, whereas in the case b) only A_1 can be rejected.

instead of A_2 and how to find the set of representatives.

Let us remark that in the Fig.2a) we can reject A_1 or A_2, but in the Fig.2b) only A_1 can be rejected. Object B belongs to the another class.

2.2. Description of the algorithm

The following algorithm will solve the above problems.

Step 1.
For each j-th object of the learning set find and store all objects belonging to the same class and situated nearer than the nearest object from another class.

The objects chosen form the potential representatives of the j-th object.

Step 2.
For each object, treated as a representative, form the list of objects which it can represent.

Step 3.
Select the greatest representative in the list as the first representative and reject all these objects that it can represent.

Step 4.
Among the remaining objects repeat step 3 until all objects will be selected or rejected.

As the result of the algorithm we receive the ordered list of the representatives sorted in respect to the decreasing number of objects that they can represent. For these objects, the representative is the nearest neighbour and therefore, in the recognition step, the rejected objects will be correctly classified. The corresponding numbers of represented objects define the power of the representatives and give us the information of their neighbourhood.

2.3. Area of applications

2.3.1. "Seeing" the feature space

The human expert is able to "see" the two- or three dimensional feature space. For this reason the algorithms that transform the n-dimensional feature space to the 2-dimensional one, i.e. Kohonen Feature Map [6], are very appreciable.
Ordered list obtained in our algorithm along with the accompanying numbers is another very attractive description of the feature space. The object which represents many objects indicates the region where the objects belong to the same class. The object which represents only itself indicates the region where the classes overlap (the nearest object belongs to the another class) or it can be treated as the effect of noise.
The ordered list of representatives would lead to better "understanding" of the feature space than with the cluster analysis or neural network approach where it is hard to describe the feature space analysing neural weighting coefficient. We can determine in which regions the particular classes are more frequent and more homogeneous, and in which regions the classes are indistinguishable.

2.3.2. Recognition algorithms

In many cases our algorithm can results in the possibility of using the nonparametric NN rule instead of estimation of the probability distribution for each class. In such a case the NN rule should be modified by considering number of objects represented by the nearest neighbour. The greater the number of represented objects, the greater is the weighting coefficient in the classification rule based on the k-NN rule
The similar approach was proposed in [5] where the representatives are stored with associated radii, i.e. the distance to the nearest object from the another class. The so-called *serial r-nearest neighbour (SER-rNN)* rule consist in considering the patterns serially beginning with the pattern with the greatest radius.
In both cases the aim is not to lose the local information of the neighbourhood of the representatives.

2.3.3. Comparing front-ends

The fact that we can "see" the feature space could be useful for comparing the different front-ends. In the case of the recognition of N classes we tend to construct such a feature extractor that could reduce the learning set to the N representatives.
The great number of representatives, and/or the presence of the representatives which represent the negligible number of objects, should be analysed from the point of view of the feature extractor properties.

3. EXPERIMENTS

We used our algorithm in the relatively difficult task consisting in recognition of Polish vowels pronounced by children in the commercial training aids for deaf children. Because of very high pitch and the difficulties connected with formants we used centers of gravity along with the energies in two frequency bands. Only 26 of the original 397 samples (i.e. less then 7%) uttered by six children were sufficient for correct classification of the learning set. We compared the performance of the consistent subset with the identical number of clusters in the recognition experiment. The use of the consistent subset instead of clusters resulted in about 25 % reduction of errors.
We used also our algorithm in the SPeaker Independent speech REcognition system SPIRE. Instead of using the statistical models of words based on our speech data base, we used the consistent subset of words to obtain the real time system.

4. CONCLUSIONS

In the paper the algorithm to determine a consistent subset of the original sample set is presented. We proposed several applications for consistent subsets. The first experiments in the area of speech recognition show that in some cases the problems with setting the neural network architecture can be avoided, especially when the class overlap and the sophisticated architecture is useless.
The more detailed results of other applications will be presented in the subsequent papers.

References

[1]. D.P.Morgan, C.L.Scofield, Neural Networks and Speech Processing, Kluwer Academic Publishers, 1992.

[2]. G.S.Lim, M.Alder, P.Hadingham, Adaptative quadratic neural nets, Pattern Recognition Letters, vol.13, No 5,1992, pp.325-329.

[3]. P.E.Hart, The Condensed Nearest Neighbour Rule, IEEE Trans. Inf.Theory, vol.14, May 1968, pp.515-516

[4]. C.I.Chang, Finding prototype for nearest neighbour classifiers, IEEE Trans.Comput.vol.C-23,November 1974, pp.1179-1184.

[5]. A.Som, T.K.Das, A.K.Tolukder, A.K.Nath, On a Method of Selection of Representative Patterns Using Binary Distance Matrix, IEEE Trans.Systems, Man, Cybernetics. vol.SMC-10, August 1980, pp.524-529.

[6]. T.Kohonen, Self-organizing feature maps, Proc.IEEE Second Int.Conf. on Neural Networks, San Diego, 1988, p.13.