A PROPOSITION OF THE NEW FEATURE SPACE AND ITS USE TO CONSTRUCTION OF A FAST MINIMUM DISTANCE CLASSIFIER

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

[1]Technical University of Łódź, Computer Engineering Department
Al. Politechniki 11, 90-924 Łódź,
[2]Institute Biocybernetics and Biomedical Engineering, PAS
ul. Trojdena 4, 02-109 Warszawa
[3] Institute of Fundamental Technological Research, PAS
ul. Świętokrzyska 21, 00-49 Warsaw

ABSTRACT

The paper presents a new approach to the reduction of the reference set size for minimum distance classifiers. An idea of the proposed method consists in mapping of each point from the reference set into a new feature space. Feature selection in the new feature space is equivalent to reduction of a number of hyperplane pieces which form the separating hypersurface. An effectiveness of the considered approach is shown on a simple artificial example. Furthermore, it is verified on a real classification problem.

INTRODUCTION

The k-NN classifiers offer very good performance and their training is very fast [1]. For a fixed feature set, the value of k is the only parameter that ought to be established on the basis of the training set. The value of k is chosen experimentally in such a way that it minimizes the probability of misclassification, which can be estimated with the *leave-one-out* method.

There are some applications with very high requirements concerning the speed of classification. If the parameter k is greater than one then certain acceleration can be reached by an approximation of a k-NN classifier by a 1-NN one. To make this approximation, it is sufficient to reclassify the primary reference set (the training set) by applying the k-NN rule and then to use the 1-NN rule with this reclassified reference set. Further, more effective acceleration can be reached by the reduction of the reference set [2, 3, 4]. Most of the methods of reduction are based on the consistency idea, *i.e.*, the proposition that all points from the primary reference set ought to be correctly classified by the 1-NN rule with the reduced reference set. None of the algorithms promises the reduction of the reference set to a minimum possible size.

We propose an approach that consists in transforming the set to be reduced into a new feature space. Feature selection in this space leads to the size reduction of the reference set. The reduction algorithms existing in the literature can be applied to obtain

Proc. 2nd Polish. Conf. Computer Pattern Recognition Systems KOSYR 2001, May 28-31, 2001, Miłków, Poland, pp. 381-386. http://www.zssk.pwr.wroc.pl/zssk/kosyr/kosyr.html

a better start point to the algorithm proposed here. Hence, our method will be used for further reduction of the sets produced by the existing algorithms.

Our considerations will be constrained to the two classes. In the case of more than two classes, we can recommend a parallel net of two-decision 1-NN classifiers [5].

REFERENCE SET REDUCTION ALGORITHMS

Now, the three most popular algorithms for the reference set reduction will be described. They will be applied in our approach as the initial steps.

Hart's algorithm

The first point from the reference set is qualified to the (initially empty) reduced reference set. Next, the remaining points of the primary reference set are classified by the 1-NN rule with the current reduced reference set. Each misclassified point is added to the reduced reference set. Such classification of all points from the primary reference set is repeated as long as *m* subsequent classifications do not increase the size of the reduced reference set. The first points selected to the reduced reference set can lie far away from the class boundary. This disadvantage of the Hart's algorithm has been removed by Gowda-Krishna modification.

Gowda-Krishna algorithm

A mutual distance measure *mdm*(**x**) is associated with each point **x** of the primary reference set. The *mdm*(**x**) is calculated in the following way. For the point **x** a nearest point **y** from the opposite class is found. A number of points from the same class as **x** that lie closer to **y** than **x** is the value of *mdm*(**x**). Next, the all points of the primary reference set are arranged according to growing values of *mdm*(**x**). Finally, the Hart's algorithm is applied to the ordered in this way reference set.

Tomek's algorithm

Each point **x** for which exists a point **y**, from the opposite class than **x**, such that the internal part of the ball spanned by the points **x** and **y** does not contain any points from the reference set, is qualified to the reduced reference set.

The authors of the three above described algorithms tried to construct the so called *consistent reduced reference set*, *i.e.* the set which, when used as the reference set with the 1-NN rule leads to correct classification of all points from the primary reference set. However, in the case of Tomek's algorithm the consistency is not guaranteed. Gowda-Krishna algorithm produces the smallest size of the reduced set.

Description of the proposed approach

We carry out our considerations on a simple numerical example presented in Table 1 and in Fig 1. Hart's algorithm selects 6 points: 1, 2, 3, 5, 6 and 7. The remaining two algorithms choose 5 points: 2, 3, 4, 6 and 7. The piecewise linear line (1), shown in Fig. 1, represents the obtained classifier based on these 5 points.

Each piece of the separating line (1) is determined by the pairs of points (2,3), (2,5), (5,6) and (6,7). It is worth to examine if it is possible to find smaller number of such pairs sufficient for separation of all the points from the reference set. There are 12 possible pairs of points from the different classes. Let us consider all such pairs ($\mathbf{x_i}$,$\mathbf{y_i}$) of points, where $\mathbf{x_i}$ comes from the class with a lower number and $\mathbf{y_i}$ is from the class with a higher number. For each pair ($\mathbf{x_i}$,$\mathbf{y_i}$) we create one feature and for any point **p** in the

Proc. 2nd Polish. Conf. Computer Pattern Recognition Systems KOSYR 2001, May 28-31, 2001, Miłków, Poland, pp. 381-386. http://www.zssk.pwr.wroc.pl/zssk/kosyr/kosyr.html

feature space, the value of which should be determined. This feature assumes the value of 1 if **p** is closer to $\mathbf{x_i}$ than to $\mathbf{y_i}$, the value of 0 if its distance to $\mathbf{x_i}$ is the same as its distance to $\mathbf{y_i}$ and the value of –1 if **p** is closer to $\mathbf{y_i}$ than to $\mathbf{x_i}$.

Point	Class	Feature 1	Feature 2
1	1	1.0	5.0
2	2	4.0	2.0
3	1	3.0	2.0
4	2	6.0	4.0
5	1	3.0	4.0
6	2	4.0	5.0
7	1	4.0	7.0

Table 1. The simple numerical example

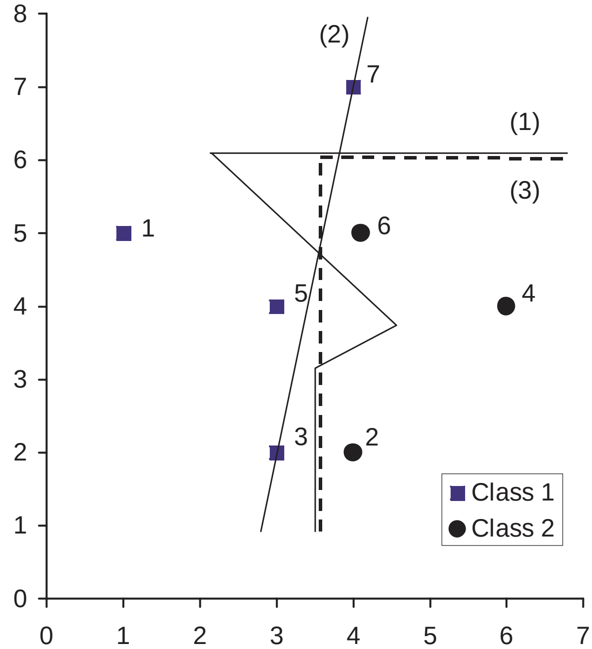

Figure 1. (1) 1-NN classifier based on the points 2, 3, 5, 6 and 7, (2) 1-NN classifier based on the points 1 and 4, (3) 1-NN classifier (dashed line) obtained by Tomek's algorithm improved by the proposed approach

In this way, each point of our simple reference set can be transformed into a new 12-dimensional feature space. The result of such transformation is presented in Table 2. The points that have the same values of all 12 features of the new feature space ought to be classified to the same class. We assume that there are no points in the primary feature space, coming from different classes and having the same feature values.

Our example satisfies this requirement. This property defines the transformation into the new feature space. For this reason, the whole reference set in the new feature space is consistent. Our problem now consists in such selection of features, in the new feature space, that the consistency is maintained. It is obvious that we are interested in preserving a possibly small number of features.

Pair of points		1/2	1/4	1/6	2/3	2/5	2/7	3/4	3/6	4/5	4/7	5/6	6/7
Point	Feature → Class ↓	*f1*	*f2*	*f3*	*f4*	*f5*	*f6*	*f7*	*f8*	*f9*	*f10*	*f11*	*f12*
1	1	1	1	1	1	1	1	1	-1	1	1	1	-1
2	2	-1	-1	-1	-1	-1	-1	1	1	1	-1	1	-1
3	1	-1	0	-1	1	-1	-1	1	1	1	-1	1	-1
4	2	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
5	1	0	1	-1	1	1	-1	1	-1	1	-1	1	-1
6	2	0	-1	-1	-1	1	1	-1	-1	1	1	-1	-1
7	1	1	0	-1	-1	1	1	-1	-1	1	1	-1	1

Table 2. The form of the reference set in the new feature space

Each of the selected features corresponds to a pair of points, *i.e.* two points. All points, which correspond to the selected features, gathered in one set, form the reduced reference set being the result of the proposed approach.

Let us come back to our illustrating example. Since the data set and the number of features is very small, we can review all the possible feature combinations. As the result of feature selection we have obtained only one feature: feature number 2 which corresponds to the pair containing the points 1 and 4. These two points form the final reduced reference set for our illustrating training set. Looking at the Tab. 1, we notice that if the entry in that table is equal 0 or 1 then the point is classified to the class 1, and if it is equal to –1 then the point is assigned to the class 2. If the feature, of a given point **p,** equals 0 then the hyperplane passes through that point. See Tab. 1 for the points 3 and 7 and the feature 2. The classifier determined by the points 1 and 4 is presented in Fig. 1 as the straight-line (2).

In the real task we can have an extremely great number of features in the new space, so the proposed approach would be hardly applicable. However, we can start with the reduced reference sets found by the existing algorithms.

If we start with the solution found by Tomek's algorithm then our task is constrained to feature selection out of 3 features only, *i.e.* to the features 4, 11 and 12, see Tab. 1. They correspond to the following pairs of points: (2,3), (5,6) and (6,7), respectively. After feature selection only two features remained - the feature 4 and the feature 12. So, the reduced reference set will finally contain 4 points: 2, 3, 6 and 7. The classifier based on the reduced reference set composed of these points has been shown in Fig. 1 as the dashed line (3).

<u>The solution of a real problem</u>

To test the proposed approach we have applied it to the real data set, which concerns quality control of ferrite cores [5]. The image of the inspected ferrite core was analyzed pixel by pixel, so the classified objects were pixels of a ferrite core surface. Each pixel could belong to the good part of the core, to the background or to one of six types of defects. Thus eight classes of the pixels were considered. The features of the classified pixels were found as functions defined on its square neighborhood. To have direction-invariant features also a rotated neighborhood was used. Finally, 30 features described each pixel.

Proc. 2nd Polish. Conf. Computer Pattern Recognition Systems KOSYR 2001, May 28-31, 2001, Miłków, Poland, pp. 381-386. http://www.zssk.pwr.wroc.pl/zssk/kosyr/kosyr.html

The reference sets, of the sizes shown in the column 2, Tab. 3, corresponding to the class pairs, were reclassified with by the (k+1)-NN rule, where *k* was established by the *leave one out* method [1]. In the *leave one out* method the classified point does not appear in the reference set currently used. However, during the reclassification the classified point ought to be taken into account too. This is the reason why we should use the (k+1)-NN rule instead of the k-NN one. In this way the 1-NN rule with the reclassified reference set approximates the k-NN rule.

Pair	**Size**	**k-NN**	**Error %**	**G.K.**	**Dim1**	**Dim2**	**F.S.**	**P.A.**
1	2	3	4	5	6	7	8	9
1/2	1706	3	0.00	**19**	90	58	18	**13**
1/3	1861	13	0.05	**12**	35	25	6	**9**
1/4	2577	5	0.47	**62**	952	-	13	**19**
1/5	2631	1	0.04	**7**	12	3	2	**4**
1/6	2270	1	3.13	**233**	13230	171	30	**48**
1/7	1563	9	0.00	**11**	30	-	2	**4**
1/8	1587	5	5.10	**108**	2907	-	18	**32**
2/3	803	3	1.99	**44**	475	33	15	**20**
2/4	1519	6	2.24	**63**	972	-	15	**21**
2/5	1573	4	0.19	**12**	36	35	4	**6**
2/6	1212	1	0.17	**21**	104	41	6	**10**
2/7	505	4	0.99	**22**	117	-	7	**10**
2/8	529	4	0.19	**20**	91	-	16	**12**
3/4	1674	1	1.08	**50**	621	49	21	**29**
3/5	1728	4	0.64	**38**	357	56	13	**19**
3/6	1367	3	0.00	**10**	25	17	5	**7**
3/7	660	1	0.97	**109**	2970	-	29	**35**
3/8	684	7	0.00	**13**	42	6	4	**7**
4/5	2444	1	1.88	**135**	4466	367	37	**54**
4/6	2083	8	0.10	**18**	80	-	8	**11**
4/7	1376	1	0.15	**20**	96	-	7	**9**
4/8	1400	1	0.21	**20**	99	-	11	**14**
5/6	2137	12	0.09	**11**	28	-	6	**9**
5/7	1430	1	1.26	**55**	744	-	24	**30**
5/8	1454	12	0.07	**10**	25	6	1	**2**
6/7	1069	7	0.00	**16**	64	55	1	**2**
6/8	1093	5	12.17	**172**	6612	95	25	**35**
7/8	386	4	0.52	**18**	80	9	6	**8**

Table 3. The results for the real data set

The optimum values of *k* and the error rates are presented in the columns 3 and 4. Since Gowda-Krishna's (G.K.) algorithm produced the smallest reduced reference set we decided to use it to create the preliminary reduced reference sets. The sizes of the reduced sets obtained by G.K. algorithm are shown in the column 5.

From the points of each such set we could create *Dim1* different pairs ($\mathbf{x}_i$,$\mathbf{y}_i$), *i*=1,2,...,*Dim1*, of points, defined previously. Thus, we would need to choose the feature subset out of *Dim1* features, where values of *Dim1* are presented in the column 6.

To constrain the preliminary set of features we applied Tomek's algorithm to the reduced sets obtained by the G.K. algorithm. If the obtained twice-reduced sets were consistent in relation to the primary reference set then we could decrease the number of features to *Dim2*, given in the column 7. Unsuccessful reductions are marked by blanks. Further reduction of the preliminary reduced sets was performed by feature selection out of *Dim1* or *Dim2* features. The numbers of features that remained after selection (F.S.) are given in the column 8. The sizes of such obtained reduced reference sets, *i.e.* by use of the proposed approach (P.A.), are presented in the column 9. The applied feature selection strategy was very similar to the *backward strategy* [1]. We removed sequentially all the features, the rejection of which did not destroy the consistency of the whole reference set transformed to the new feature space, *i.e.* any two points with the same components remained in the same class. Finally, for all the pairs of the classes the size of the reference set was significantly reduced with the proposed approach.

CONCLUSIONS

Comparing the results of Gowda-Krishna's algorithm and the proposed approach we can see a significant advantage of transforming the reference set reduction problem into selection of artificially created features.

BIBLIOGRAPHY

[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 (Corresp.), May 1968, 515-516.

[3] Gowda K. C. and Krishna G., The condensed nearest neighbor rule using the concept of nutual nearest neighborhood, *IEEE Trans. Information Theory*, IT-25, 4, July 1979, 488-490.

[4] Tomek I., Two modifications of CNN, *IEEE Trans. Systems, Man, and Cybernetics*, SMC-7, 2, February 1977, 92-94.

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

Proc. 2nd Polish. Conf. Computer Pattern Recognition Systems KOSYR 2001, May 28-31, 2001, Miłków, Poland, pp. 381-386. http://www.zssk.pwr.wroc.pl/zssk/kosyr/kosyr.html