Adaptive Nearest Neighbor Classifier Based on Supervised Ellipsoid Clustering

Guo-Jun Zhang[1,2], Ji-Xiang Du[1,2], De-Shuang Huang[1,2] , Tat-Ming Lok [3], and Michael R. Lyu[4]

[1] Intelligent Computing Lab, Hefei Institute of Intelligent Machines, Chinese Academy of Sciences, P.O. Box 1130, HeFei Anhui 230031, China
{zhanggj, dshuang}@iim.ac.cn
[2] Department of Automation, University of Science and Technology of China, Hefei , China
[3] Information Engineering Dept., The Chinese University of Hong Kong, Hong Kong
[4] Computer Science & Engineering Dept., The Chinese University of Hong Kong, Hong Kong

Abstract. Nearest neighbor classifier is a widely-used effective method for multi-class problems. However, it suffers from the problem of the curse of dimensionality in high dimensional space. To solve this problem, many adaptive nearest neighbor classifiers were proposed. In this paper, a locally adaptive nearest neighbor classification method based on supervised learning style which works well for the multi-classification problems is proposed. In this method, the ellipsoid clustering learning is applied to estimate an effective metric. This metric is then used in the *K*-NN classification. Finally, the experimental results show that it is an efficient and robust approach for multi-classification.

1 Introduction

One of most popular classification approaches is the Nearest Neighbor (NN) method. However, the NN rule becomes less appealing in the case of limited training samples and high dimensional feature space due to the curse of dimensionality. Severe bias can be caused in such situations. Recently, several methods [1-4] characterized by query-based local distance functions were proposed to reduce this bias. However, the computation seems not to be efficient and there are too many parameters despite of the relatively high accuracy they achieve. Are there any better alternatives which could estimate the local relevance efficiently?

Our method uses the supervised ellipsoid clustering(SEC) to produce boundary and utilizes its boundary to estimate the local features relevance of the query with a scheme provided in LAMANN[1]. The relevance is then used in the weighted Euclidean distance during the *K*-NN classifications. The algorithm is referred to as Ellipsoids-boundary weighting adaptive Nearest Neighbor Algorithm (EWANN Algorithm) because the class boundary is constructed according to the surface of ellipsoids.

L. Wang et al. (Eds.): FSKD 2006, LNAI 4223, pp. 582 – 585, 2006.

This paper is organized as follows. Section 2 presents the feature relevance theory. Some empirical evaluation of our method is given in Section 3. Finally, a concluding remark is included in Section 4 .

2 Ellipsoids-Boundary Weighting Adaptive Nearest Neighbor Algorithm

It is often noted that different classes may have different discriminant features. This means that distance computation does not always have equal strength on features in the feature space. Considering this, many locally adaptive nearest neighbor classifiers are designed to find these relevant dimensions.

Supervised ellipsoid clustering (SEC) is to cover each class region with ellipsoids. Each ellipsoid will represent a set of points [5]. Ellipsoids can be generated by an incremental learning procedure. The ellipsoids are created, constricted, or enlarged gradually at the present of each training sample. After SEC, each class is stuffed with a set of ellipsoids. In other views, the outer surface of the ellipsoids of a certain class composes the boundary of that class, i.e. the boundary of the class against the other classes is formed by the surface of its ellipsoids.

The gradient vector of the point on the boundary identifies a direction along which nearby data points are well separated. We use this gradient vector to measure local feature relevance and weighting features accordingly.

The feature relevance $R(x)$ can be given as $R_i(x) = \left| N_d \bullet u_i \right| = \left| N_{di} \right|$, where d is the nearest point on the boundary to the query, N_d the gradient vector at point d , u_i the vector unit. After r is transformed from $R(x)$ by a scheme given by Jing Peng [1], it could be applied in the weighted distance computation during the NN classification. The resulting algorithm is summarized in Fig. 1.

INPUT: Ellipsoids of all classes produced by SEC, query q and parameter K for K-NN

Step 1. Find ellipsoid E_n which is nearest to query q

Step 2. Find the point d which is the nearest point to q on the E_n .

Step 3. Compute the gradient vector N_d

Step 4. Compute $R(x)$ and transform it to r by the scheme given by Jing Peng

Step 5. Use r in weighted distance computation and apply K-NN rule

Fig. 1. Ellipsoids-boundary Weighting Adaptive Nearest Neighbor Algorithm

3 Empirical Evaluation

In the following, we compare several competing classification methods using a number of data sets. The classifiers are 5NN which uses the simple five NN rule, DANN using the discriminant adaptive NN rule.[2], Morph which uses he morphing rule [3],ADAMENN using the adaptive Metric Nearest Neighbor Algorithm [4] and LAMANN using the large margin nearest neighbor.[1]. The data sets were taken from the UCI Machine Learning Database Repository. We randomly select 60% of samples of each class as training samples and other 40% for testing.

Table 1. Average classification error rate

	Iris	Heart	Diabetes	Cancer
EWANN	**5.8**	24.4	**24.3**	23.6
5NN	**5.8**	24.8	25.4	24.1
DANN	**5.8**	23.7	24.8	22.7
Morph	**5.8**	**22.7**	25.7	**22.8**
ADAMENN	**5.8**	22.9	25.0	25.0
LAMNN	**5.8**	24.0	24.8	23.1

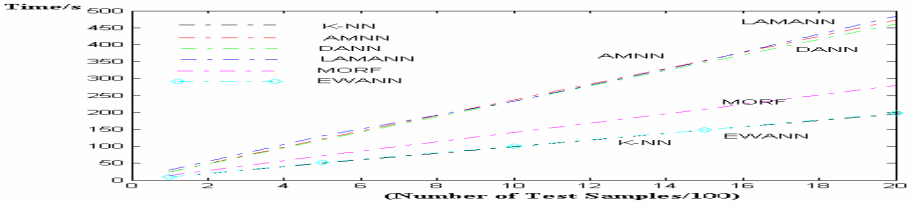

Fig. 2. On-line computing time performance

From Table 1, it can be found that EWANN achieved the best performance of the 2/4, followed closely by LAMANN. The result shows that our method is as competing as other adaptive classifiers. What's more, as it is shown in Fig. 2, the proposed method is superior to other adaptive nearest neighbor methods in terms of online computing.

4 Conclusions

This paper presents a new flexible metric method for effective and efficient pattern classification. It employs SEC to generate class boundary, and use its information for the following nearest neighbor classification. The experimental results show clearly that the proposed algorithm can potentially improve the performance of *K*-NN in several classification problems.

References

1. Domeniconi, C., Gunopulos ,D., and Peng, J.: Large Margin Nearest Neighbor Classifiers, IEEE Transaction on Neural Networks, VOL 16, No. 4 , (2005), 899-909
2. Hastie, T., Tibshirani, R.: "Discriminant Adaptive Nearest Neighbor Classification", IEEE Trans. on PAMI, Vol. 18,No. 6, (1996), 607-615
3. Peng, J., Heisterkamp, D. R., Dai, H.K.: LDA/SVM Driven Nearest Neighbor Classifiers, IEEE Transaction on Neural Networks, VOL 14, No. 4, (2003), 940-942
4. Domeniconi, C., Peng, J., Gunopulos, D.: Locally Adaptive Metric Nearest Neighbor Classification, IEEE Transaction on PAMI, (2002), 1281-1285
5. Kositsky, M., Ullman, S.: Learning Class Regions by the Union of Ellipsoids, Proceedings of the 13th ICPR, IEEE Computer Society Press, (1996), 750-757