H-016 On Reducing Dimensionality of Dissimilarity Representation

Sang-Woon Kim and Jian Gao

Dept. of Computer Engineering, Myongji University, Korea

1 Introduction

One of the most recent and novel developments in pattern classification is the concept of dissimilarity-based classifiers (DBCs) proposed by Duin and his co-authors [1]. DBCs are a way of defining classifiers between the classes, which are not based on the feature measurements of the individual patterns, but rather on a suitable *dissimilarity measure* between them [2]. The problem with this strategy is that we need to select a representative set of data that is both compact and capable of representing the entire data set. However, it is difficult to find the optimal number of prototypes and, furthermore, selecting prototype stage may potentially lose some useful information for discrimination. Recently, to avoid these problems, we proposed an alternative approach, where we used *all* available samples from the training set as prototypes, and subsequently apply dimensionality reduction schemes [3]. For the purpose of finding the most appropriate reduction method, in this paper, we consider a way of applying a newly developed dimensionality reduction scheme [4] after computing the dissimilarity matrix with the *entire* training samples.

2 Dissimilarity Representation

A dissimilarity representation of a set of samples, $T = \{\boldsymbol{x}_i\}_{i=1}^{n} \in \Re^d$, is based on pairwise comparisons and is expressed, for example, as an $n \times m$ dissimilarity matrix $D_{T,Y}[\cdot,\cdot]$, where $Y = \{\boldsymbol{y}_1, \cdots, \boldsymbol{y}_m\}$, a prototype set, is extracted from T, and the subscripts of D represent the set of elements, on which the dissimilarities are evaluated. Thus, each entry $D_{T,Y}[i,j]$ corresponds to the dissimilarity between the pairs of objects $\langle \boldsymbol{x}_i, \boldsymbol{y}_j \rangle$, where $\boldsymbol{x}_i \in T$ and $\boldsymbol{y}_j \in Y$. Consequently, an object $\boldsymbol{x}_i$ is represented as a column vector as follows:

$$[d(\boldsymbol{x}_i, \boldsymbol{y}_1), d(\boldsymbol{x}_i, \boldsymbol{y}_2), \cdots, d(\boldsymbol{x}_i, \boldsymbol{y}_m)]^T, 1 \le i \le n. \quad (1)$$

Here, the dissimilarity matrix $D_{T,Y}[\cdot,\cdot]$ is defined as a *dissimilarity space*, on which the d-dimensional object, $\boldsymbol{x}$, given in the feature space, is represented as an m-dimensional vector $\delta_Y(\boldsymbol{x})$.

To select the representative set that is compact and capable of simultaneously representing the entire data set, Duin and colleagues [1] discussed the following methods : *Random*, *RandomC*, *KCentres*, *ModeSeek*, *LinProg*, *FeatSeal*, *KCentres-LP*, and *EdiCon*. The details of these methods are omitted here in the interest of compactness, but can be found in the existing literature, including [1] and [2].

3 Dimensionality Reduction Schemes

In the proposed method, we use a strategy of reducing the dimensionality after computing the dissimilarity matrix with the entire training samples. With regard to reducing the dimensionality of the dissimilarity matrix, we make use of the well-known dimensionality reduction schemes (DRSs) proposed in the literature, such as direct LDA (DLDA) [5] and heteroscedastic LDA (linear dimensionality reduction via a heteroscedastic extension of LDA) (HLDA) [4].

In LDA, we use the concept of a within-class scatter matrix, $\boldsymbol{S}_W$, and a between-class scatter matrix, $\boldsymbol{S}_B$, to maximize a separation criterion, such as $J = tr(\boldsymbol{S}_W^{-1}\boldsymbol{S}_B)$. The reasoning behind making J as large as possible is to look for a direction that maximizes the difference between the two projected means in $\boldsymbol{S}_B$, while minimizing the variance of the individual classes $\boldsymbol{S}_W$.

The separation criterion can also be extended as $J = tr\{(\boldsymbol{A}\boldsymbol{S}_W\boldsymbol{A}^T)^{-1}(\boldsymbol{A}\boldsymbol{S}_E\boldsymbol{A}^T)\}$, where $\boldsymbol{A}$ is a transformation matrix and $\boldsymbol{S}_E$ is the square of Euclidian distance: $(\boldsymbol{m}_1 - \boldsymbol{m}_2)(\boldsymbol{m}_1 - \boldsymbol{m}_2)^T$, where $\boldsymbol{m}_1$ and $\boldsymbol{m}_2$ are the means of class 1 and class 2, respectively. In the heteroscedastic extension of LDA, on the other hand, the square of Euclidian distance $\boldsymbol{S}_E$ is replaced with the Chernoff distance $\boldsymbol{S}_C$ [4], which is defined as:

$$\begin{aligned}\boldsymbol{S}_C &= \boldsymbol{S}^{-\frac{1}{2}}(\boldsymbol{m}_1 - \boldsymbol{m}_2)(\boldsymbol{m}_1 - \boldsymbol{m}_2)^T\boldsymbol{S}^{-\frac{1}{2}} \\ &\quad + \frac{1}{\alpha(1-\alpha)}(\log \boldsymbol{S} - \alpha \log \boldsymbol{S}_1 - (1-\alpha)\log \boldsymbol{S}_2).\end{aligned}$$

Here, $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ are the scatter matrices of class 1 and class 2; $\boldsymbol{S} = \alpha\boldsymbol{S}_1 + (1-\alpha)\boldsymbol{S}_2$; α is the probability of class 1.

4 Schema for the Proposed Solution

To solve the classification problem, we first compute the dissimilarity matrix with the entire training samples, and then reduce its dimensionality of the dissimilarity matrix with a DRS; finally, DBCs are designed on the dissimilarity space to reduce the classification error rates. The proposed algorithm for DBCs is summarized in the following:

1. Select the entire training samples T as the representative set Y.

2. Using Eq. (1), compute the dissimilarity matrix, $D_{T,T}[\cdot,\cdot]$, in which each individual dissimilarity is computed on the basis of the measures described in [1], [2]. After computing the $D_{T,T}[\cdot,\cdot]$, reduce its dimensionality by invoking a DRS.

3. For a testing sample $\boldsymbol{z}$, compute a dissimilarity column vector, $\delta_Y(\boldsymbol{z})$, by using the same measure used in Step 3.

4. Achieve the classification by invoking a classifier built in the dissimilarity space, and operating it on the dissimilarity vector $\delta_Y(\boldsymbol{z})$.

The rationale of this strategy is presented in a later section together with the experimental results.

5 Experimental Results

The proposed method has been tested and compared with conventional methods. This was done by performing experiments on the well-known benchmark database, namely, "UMIST" face database [1].

After computing the dissimilarity matrix, we reduced the dimensionality of the matrix with a DRS, such as DLDA [5] or HLDA [4]. In the two LDA-based methods, we reduced the dimension $n-1$ to $c-1$, where n is the total number of training samples and c is the number of classes. In the conventional methods of *Random*, *RandomC*, *KCentres*, and *ModeSeek*, on the other hand, we selected $c-1$, c, c, and c samples from the training data set as the prototypes of DBCs.

To maintain the diversity between the dissimilarity-based classifications, we experimented different classifiers, such as the k-nearest neighbor classifiers (1-NN, 3-NN, 5-NN, 7-NN), the nearest mean classifiers (NMC), the support vector classifier (SVC), and the regularized normal density-based linear/quadratic classifiers (RLDC and RQDC). These classifiers were implemented with PRTools [2]. In *Whole* method, the entire training data set T is selected as a representative subset Y. The result of this method is included as a reference.

Table 1 shows a comparison between the classification accuracy rates (%) (for the process of prototype selection or dimensionality reduction) of DBCs for the UMIST database.

From Table 1, it is clear that there is a considerable improvement in the achieved performances. An example of this is the classification accuracies of the *Random* and HLDA methods. For the *eight* classifiers, namely, 1NN, 3NN, 5NN, 7NN, NMC, RLDC, QLDC, and SVC, the classification accuracy rates of the *Random* method are, respectively, 98.00, 98.67, 98.67, 97.33, 95.33, 95.33, 99.33, 98.67(%), while the classification accuracies of the HLDA method are, respectively again, 99.33, 100, 99.33, 99.33, 99.33, 99.33, 100, 100(%). From this consideration, the reader can observe that the classification performances of the classifiers are improved with the HLDA method. It is also interesting to point out that the classification accuracies of the parametric classifiers, such as RLDC and RQDC, are considerably improved. The same characteristic could also be observed in the other methods.

[1] http://images.ee.umist.ac.uk/danny/database.html

[2] PRTools is a MATLAB toolbox for pattern recognition (refer to http://www.prtools.org/).

Table 1: A comparison between the classification accuracy rates (%) of DBCs for the UMIST database in the conventional and LDA based methods.

Experimental Methods	1NN / NMC	3NN / RLDC	5NN / QLDC	7NN / SVC
Whole	99.33	99.33	99.33	99.33
	99.33	99.33	99.33	99.33
Random	98.00	98.67	98.67	97.33
	95.33	95.33	99.33	98.67
RandomC	99.33	99.33	98.00	97.33
	97.33	98.00	99.33	100
KCentres	98.67	98.67	98.67	98.00
	96.00	88.00	93.00	98.67
ModeSeek	99.33	99.33	99.33	99.33
	99.33	99.33	99.33	98.67
DLDA	99.33	99.33	99.33	99.33
	99.33	99.33	99.33	99.33
HLDA	99.33	100	99.33	99.33
	99.33	99.33	100	100

6 Conclusion

In this paper, we proposed a method of reducing the dimensionality of the dissimilarity representation. The proposed scheme was experimented and compared with the conventional methods for the well-known benchmark facial images. Our experimental results demonstrated the possibility that the proposed method could be used efficiently for dissimilarity-based classifiers (DBCs). The research concerning the reduction of the processing CPU-times is a future aim of the authors.

Acknowledgments This work was supported by the Korea Research Foundation Grant funded by the Korea Government(MOEHRD- KRF-2007-313-D00714).

References

[1] E. Pekalska and R. P. W. Duin, *The Dissimilarity Representation for Pattern Recognition: Foundations and Applications*, World Scientific Publishing, Singapore, 2005.

[2] S. -W. Kim and B. J. Oommen, "On using prototype reduction schemes to optimize dissimilarity-based classification", *Patt. Recogn.* 40, 2946–2957, 2007.

[3] S. -W. Kim and J. Gao, "On Using Dimensionality Reduction Schemes to Optimize Dissimilarity-Based Classifiers", *Proc. CIARP 2008*, Cuba, LNCS-5197, 302–309, 2008.

[4] M. Loog and R. P. W. Duin, "Linear dimensionality reduction via a heteroscedastic extension of LDA: The ChernoCriterion", *IEEE Trans. Pattern Anal. and Machine Intell.* PAMI-26(6), 732–739, 2004.

[5] H. Yu and J. Yang, "A direct LDA algorithm for high-dimensional data - with application to face recognition", *Patt. Recogn.* 34, 2067–2070, 2001.