ON REDUCING DIMENSIONALITY OF DISSIMILARITY MATRICES FOR OPTIMIZING DBC
An Experimental Comparison *

Sang-Woon Kim
Department of Computer Science and Engineering, Myongji University, Yongin, 449-728 South Korea

Keywords: Dissimilarity-based Classifications, Dimensionality reduction schemes, Prototype selection methods.

Abstract: One problem of dissimilarity-based classifications (DBCs) is the high dimensionality of dissimilarity matrices. To address this problem, two kinds of solutions have been proposed in the literature: prototype selection (PS) based methods and dimensionality reduction (DR) based methods. The DR-based method consists of building the dissimilarity matrices using *all* the available training samples and subsequently applying some of the standard DR schemes. On the other hand, the PS-based method works by directly choosing a small set of representatives from the training samples. Although DR-based and PS-based methods have been explored separately by many researchers, not much analysis has been done on the study of comparing the two. Therefore, this paper aims to find a suitable method for optimizing DBCs by a comparative study. In the experiments, four DR and four PS methods are used to reduce the dimensionality of the dissimilarity matrices, and classification accuracies of the resultant DBCs trained with two real-life benchmark databases are analyzed. Our empirical evaluation on the two approaches demonstrates that the DR-based method can improve the classification accuracies more than the PS-based method. Especially, the experimental results show that the DR-based method is clearly more useful for nonparametric classifiers, but not for parametric ones.

1 INTRODUCTION

Dissimilarity-based classifications (DBCs) (Pekalska and Duin, 2005), (Pekalska and Paclik, 2006) are a way of defining classifiers among the classes, and the process is not based on the feature measurements of individual object samples, but rather on a suitable dissimilarity measure among the individual samples. The advantage of this method is that it can avoid the problems associated with feature spaces, such as the curse of dimensionality and the issue of estimating a number of parameters (Kim and Oommen, 2007).

In DBCs, a good selection of prototypes seems to be crucial to succeed with the classification algorithm in the dissimilarity space. The prototypes should avoid redundancies in terms of selection of similar samples, and prototypes should include as much information as possible. However, it is difficult for us to find the optimal number of prototypes. Furthermore, there is a possibility that we lose some useful information for discrimination when selecting the prototypes. To avoid these problems, in (Bicego and Figueiredo, 2004), (Riesen and Bunke, 2007), and (Kim and Gao, 2008), the authors separately proposed an alternative approach where *all* of the available samples were selected as prototypes, and, subsequently, a scheme, such as linear discriminant analysis, was applied to the reduction of dimensionality. This approach is more principled and allows us to completely avoid the problem of finding the optimal number of prototypes (Bunke and Riesen, 2007).

In this paper, we perform an empirical evaluation on the two approaches of reducing the dimensionality of dissimilarity matrices for optimizing DBCs: prototype selection (PS) based methods and dimension reduction (DR) based methods. In PS-based methods, we first select the representative prototype subset from the training data set by resorting to one of the prototype selection methods as described in (Pekalska and Duin, 2005) and (Pekalska and Paclik, 2006). Then, we compute the dissimilarity matrix, in which each individual dissimilarity is computed on the basis of the measures described in (Pekalska and Paclik,

*This work was supported by the National Research Foundation of Korea funded by the Korean Government (NRF-2009-0071283).

Kim S. (2010).
ON REDUCING DIMENSIONALITY OF DISSIMILARITY MATRICES FOR OPTIMIZING DBC - An Experimental Comparison.
In *Proceedings of the 2nd International Conference on Agents and Artificial Intelligence - Artificial Intelligence*, pages 235-240
DOI: 10.5220/0002713002350240

2006). In addition, for a testing sample, z, we compute a dissimilarity column vector, $\delta(z)$, by using the same measure. Finally, we perform the classification by invoking a classifier built in the dissimilarity space and by operating the classifier on $\delta(z)$.

On the other hand, in DR-based methods, we prefer not to directly select the prototypes from the training samples; rather, we employ a way of using a standard DR scheme, after computing the dissimilarity matrix with the *entire* training samples. Then, as in PS-based methods, we compute a dissimilarity column vector for a testing sample and perform the classification of the vector by invoking a classifier built in the dissimilarity space. Here, the point to be mentioned is how to choose the optimal number of prototypes and the subspace dimensions to be reduced. In PS-based methods, we select the same number of (or twice as many) prototypes as the number of classes in heuristic. In DR-based ones, on the other hand, we can use a cumulative proportion technique (Laaksonen and Oja, 1996) to choose the dimensions.

The main contribution of this paper is to present an empirical evaluation on the two methods of reducing the dimensionality of dissimilarity matrices for optimizing DBCs. This evaluation shows that DBCs can be optimized by employing a dimensionality reduction scheme as well as a prototype selection method. Here, the aim of using the dimensionality reduction scheme instead of selecting the prototypes is to accommodate some useful information for discrimination and to avoid the problem of finding the optimal number of prototypes. Our experimental results demonstrate that the DR-based method can generally improve the classification accuracy of DBCs more than the prototype selection based method. Especially, the results indicate that the DR-based method is clearly more useful for nonparametric classifiers, but not for parametric ones.

2 RELATED WORK

Foundations of DBCs. A dissimilarity representation of a set of samples, $T = \{x_i\}_{i=1}^{n} \in \Re^d$, is based on pairwise comparisons and is expressed, for example, as an $n \times m$ dissimilarity matrix $D_{T,Y}[\cdot,\cdot]$, where $Y = \{y_j\}_{j=1}^{m}$, 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, $\langle x_i, y_j \rangle$, where $x_i \in T$ and $y_j \in Y$. Consequently, an object, x_i, is represented as a column vector as follows:

$$[d(x_i,y_1), d(x_i,y_2), \cdots, d(x_i,y_m)]^T, 1 \leq i \leq n. \quad (1)$$

Here, the dissimilarity matrix, $D_{T,Y}[\cdot,\cdot]$, is defined as a *dissimilarity space*, on which the d-dimensional object, x, given in the feature space, is represented as an m-dimensional vector, $\delta(x,Y)$, where if $x = x_i$, $\delta(x_i,Y)$ is the i-th row of $D_{T,Y}[\cdot,\cdot]$. In this paper, the column vector, $\delta(x,Y)$, is simply denoted by $\delta(x)$.

Prototype Selection Methods. The intention of selecting prototypes is to guarantee a good tradeoff between the recognition accuracy and the computational complexity when the DBC is built on $D_{T,Y}[\cdot,\cdot]$ rather than $D_{T,T}[\cdot,\cdot]$. Various prototype selection (PS) methods have been proposed in the literature (Loz,), (Pekalska and Duin, 2005), (Pekalska and Paclik, 2006). The well-known eight selection methods experimented in (Pekalska and Duin, 2005) and (Pekalska and Paclik, 2006) are *Random*, *Random_C*, *KCentres*, *ModeSeek*, *LinProg*, *PeatSeal*, *KCentres-LP*, and *EdiCon*. In the interest of compactness, the details of these methods are omitted here, but can be found in the existing literature (Pekalska and Paclik, 2006).

DBCs summarized previously, in which the representative prototype subset is selected with a PS, are referred to as PS-based DBCs or simply PS-based methods. An algorithm for PS-based DBCs is summarized in the following:

1. Select the representative set, Y, from the training set, T, by resorting to one of the prototype selection methods.

2. Using Eq. (1), compute the dissimilarity matrix, $D_{T,Y}[\cdot,\cdot]$, in which each individual dissimilarity is computed on the basis of the measures described in (Pekalska and Duin, 2005).

3. For a testing sample z, compute $\delta(z)$ by using the same measure used in Step 2.

4. Achieve the classification by invoking a classifier built in the dissimilarity space and by operating the classifier on the dissimilarity vector, $\delta(z)$.

From these four steps, we can see that the performance of the DBCs relies heavily on how well the dissimilarity space, which is determined by the dissimilarity matrix, $D_{T,Y}[\cdot,\cdot]$, is constructed. To improve the performance, we need to ensure that the dissimilarity matrix is well designed.

Dimensionality Reduction Schemes. With regard to reducing the dimensionality of the dissimilarity matrix, we can use a strategy of employing the dimensionality reduction (DR) schemes after computing the dissimilarity matrix with the entire training samples. Numerous DRs have been proposed in the literature, some of which are (Belhumeour and Kriegman, 1997), (Yu and Yang, 2001), (Loog and Duin, 2004), and (Wei and Li, 2009). The most well known DRs

are the class of linear discriminant analysis (LDA) strategies, such as Fisher LDA (Yu and Yang, 2001), Two-stage LDA (Belhumeour and Kriegman, 1997), Chernoff distance based LDA (Loog and Duin, 2004), (Rueda and Herrera, 2008), and so on.

In the interest of brevity, the details of the LDA strategies are again omitted here, but we briefly explain below the Chernoff distance based LDA (in short CLDA) that is pertinent to our present study. It is well-known that LDA is incapable of dealing with the heteroscedastic data in a proper way (Loog and Duin, 2004). To overcome this limitation, in CLDA, the square of Euclidian distance, $S_E = S_B/(p_1 p_2)$, is replaced with the Chernoff distance defined as: $S_C = S^{-\frac{1}{2}}(m_1 - m_2)(m_1 - m_2)^T S^{-\frac{1}{2}} + (\log S - p_1 \log S_1 - p_2 \log S_2)/(p_1 p_2)$, where S_i, m_i, and p_i are the scatter matrix, the mean vector, and a priori probability of class i, respectively; $S = p_1 S_1 + p_2 S_2$. Using S_C, instead of S_E, the Fisher separation criterion, J_H, can be defined (see Eq. (5) of (Loog and Duin, 2004)). To obtain a matrix, A, that maximizes J_H, recently, some researchers (Rueda and Herrera, 2008) have developed a gradient-based algorithm named *Chernoff_LDA_Two*, which consists of three steps: (a) initiate $A^{(0)}$, (b) compute $A^{(k+1)}$ from $A^{(k)}$ by applying the secant method to J_H, and (c) terminate the iteration by checking the convergence (Rueda and Herrera, 2008).

DBCs, in which the dimensionality of dissimilarity matrices is reduced with a DR, are referred to as DR-based DBCs or DR-based methods. An algorithm for DR-based 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 (Pekalska and Duin, 2005). After computing the $D_{T,T}[\cdot,\cdot]$, reduce its dimensionality by invoking a dimensionality reduction scheme.

3. This step is the same as in PS-based DBC.

4. This step is the same as in PS-based DBC.

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

In the attempt to provide a comparison between PS-based DBCs and DR-based DBCs, we are required to analyze their computational complexities. In light of brevity, the details of the analysis are omitted here. From analyzing the algorithms, however, we can observe that the time complexities of PS-based, LDA (and PCA)-based, and CLDA-based DBCs are $O(nmd)$, $O(n^2 d)$, and $O(n^2 d + n^3)$, respectively.

3 EXPERIMENTAL RESULTS

Experimental Data. PS-based and DR-based methods were tested and compared with each other by conducting experiments for a handprinted character data set and a well-known face database, namely Nist38 (Wilson and Garris, 1992) and Yale (Georghiades and Kriegman, 2001). The data set captioned Nist38 consists of two kinds of digits, 3 and 8, for a total of 1000 binary images. The size of each image is 32×32 pixels, for a total dimensionality of 1024 pixels. The Yale database contains 165 gray scale images of 15 individuals. The size of each image is 243×320 pixels, for a total dimensionality of 77760 pixels. To reduce the computational complexity of this experiment, facial images of Yale were down-sampled into 178×236 pixels and then represented by a centered vector of normalized intensity values.

Experimental Method. All of our experiments were performed with a "leave-one-out" strategy; to classify an image, we removed the image from the training set and computed the dissimilarity matrix with the $n-1$ images. This process was repeated n times for every image, and a final result was obtained by averaging the results of each image.

To measure the dissimilarity between two objects, we used Euclidean distance (ED), Hamming distance (HD), regional distance (RD) (Adini and Ullman, 1997), and spatially weighted gray-level Hausdorff distance (WD) (Kim, 2006) measuring systems [2].

To construct the dissimilarity matrix, in PS-based methods, we employed *Random* (in short *Rand*), *Random_C* (in short *RandC*), *KCentres* (in short *KCenter*), and *ModeSeek* (in short *ModeS*) to select the prototype subset. Here, the number of prototypes selected was heuristically determined as c or $2c$.

On the other hand, in DR-based methods, to reduce the dimensionality, we used direct LDA (in short LDA), PCA, two-stage LDA (in short PCALDA), and Chernoff distance-based LDA (in short CLDA). In addition, to select the dimensions for the systems, we used the cumulative proportion, α, which is defined as follows (Laaksonen and Oja, 1996): $\alpha(q) = \sum_{j=1}^{q} \lambda_j \Big/ \sum_{j=1}^{d} \lambda_j$. Here, the subspace dimension, q, (where d and λ_j are the dimensionality and the eigenvalue, respectively) of the data sets is determined by considering the cumulative proportion $\alpha(q)$. The eigenvectors and eigenvalues are computed, and the cumulative sum of the eigenvalues is compared to a fixed number, k. In other words, the subspace

[2] In this experiment, we employed only four measuring systems, namely ED, HD, RD, and WD. However, other numerous solutions could also be considered.

dimensions are selected by considering the relation $\alpha(q) \leq k \leq \alpha(q+1)$. In PCALDA, however, we reduced the dimensions in two steps: we first reduced the dimension $d(=n-1)$ into an intermediate dimension $n-c+1$ using PCA; we then reduced the intermediate dimension $n-c+1$ to q using LDA [3].

To maintain the diversity among the DBCs, we designed different classifiers, such as k-nearest neighbor classifiers (k = 1), nearest mean classifiers, regularized normal density-based linear/quadratic classifiers, and support vector classifiers. All of the DBCs mentioned above are implemented with PRTools [4] and denoted in the next section as *knnc*, *nmc*, *ldc*, *qdc*, and *svc*, respectively. Here, *ldc* and *qdc* are regularized with $(R,S) = (0.01, 0.01)$. Also, *svc* is implemented using a polynomial kernel function of degree 1.

Experimental Results. The run-time characteristics of the empirical evaluation on the two data sets are reported below and shown in figures and tables. In this section, we first investigate the rationality of employing a PS (i.e., *KCenter*) or a DR(i.e., PCA) methods in reducing the dimensionality. Then, we present classification accuracies of the PS and DR-based methods. Consequently, based on the classification results, we grade and rank the methods. Finally, we introduce a numerical comparison of the processing CPU-times.

First, the experimental results of PS and DR-based methods were probed. Fig. 1 shows plots of the classification accuracies obtained with *knnc* for Yale. In Fig. 1(a), the dissimilarity matrix, in which the classifiers were evaluated, was generated with the prototype subset selected with a PS, such as *KCenter*. In Fig. 1(b), on the other hand, after generating the dissimilarity matrix with the entire data set, the dimensionality was reduced by invoking a DR, such as PCA.

In the figure, it is interesting to note that PS and DR-based DBCs (*knnc*) can be optimized by means of choosing the number of prototypes and reducing the dimensions, respectively. For example, both classification accuracies of RD for *knnc* are saturated when having 16 and 8 as the number of prototypes and the subspace dimension, respectively. Here, the problem to be addressed is how to choose the optimal number of prototypes and the dimension to be reduced. In PS-based methods, we selected the number of prototypes as $2c$ (which is an experimental parameter). In DR-based methods, on the other hand, using the cumulative technique, we chose the subspace dimensions for Yale and Nist38 as follows: (1)

[3]Similar to the approaches with prototype selection methods, the number of dimensions is not given beforehand. From this point of view, we could say that the problem of selecting the optimal dimension remains unresolved.

[4]http://prtools.org/

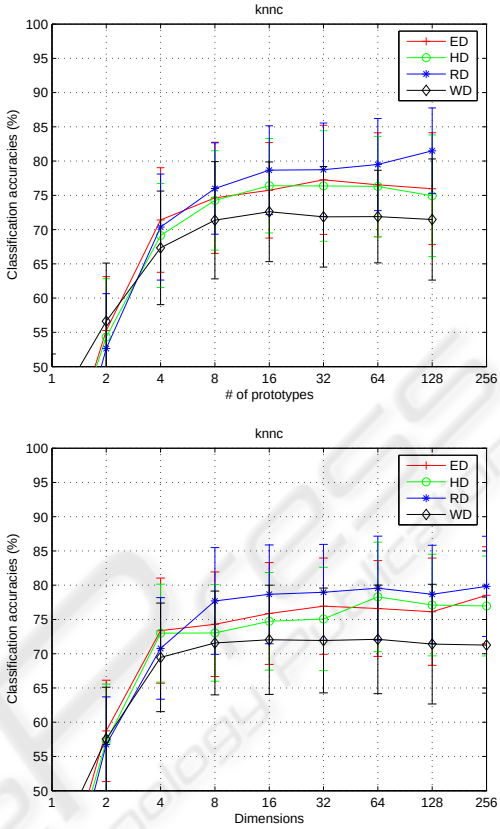

Figure 1: Plots of the classification accuracy rates (%) of PS-based and DR-based DBCs (*knnc*) for Yale database: (a) top (PS-based DBC); (b) bottom (DR-based DBC). Here, the prototype subsets of (a) are selected from the training data set with *KCenter* and the subspace dimensions of (b) are obtained (selected) with a PCA.

Yale: $q_{WD} = 12$; $q_{RD} = 15$; $q_{ED} = 76$; $q_{HD} = 133$, (2)
Nist38: $q_{WD} = 133$; $q_{RD} = 22$; $q_{ED} = 41$; $q_{HD} = 41$.

Although it is hard to quantitatively evaluate the various PS and DR-based methods, we have attempted to do exactly this; we have given a numerical grade to every method tested here according to its classification accuracy to render this comparative study more complete. Tables 1 and 2 show, respectively, the classification accuracies of PS and DR-based DBCs, where the values underlined are the highest ranks in the four accuracies of each classifier.

From the two tables, we can see that *almost* all the highest rates (underlined) achieved with DR-based DBCs are higher than those of PS-based ones. This observation confirms the possibility that the classification performance of DBCs can be improved by effectively reducing the dimensionality after constructing dissimilarity matrices with all of the training samples. To observe how well the methods work, we picked the best three among the eight (four of PSs and

Table 1: Classification accuracies (%) of PS-based DBCs.

data	methods		*knnc*	*nmc*	*ldc*	*qdc*	*svc*
Yale	ED	Rand	80.00	77.58	84.24	75.76	84.85
		RandC	80.61	81.82	96.97	79.39	56.97
		KCenter	77.58	80.00	89.70	75.76	85.45
		ModeS	78.18	78.79	88.48	76.36	89.09
	HD	Rand	75.15	73.94	71.52	76.36	76.36
		RandC	75.76	80.00	89.70	73.33	75.76
		KCenter	73.94	74.55	76.97	75.15	80.61
		ModeS	75.15	78.18	75.76	80.00	87.88
	RD	Rand	77.58	71.52	90.91	76.36	86.06
		RandC	78.79	70.91	98.79	75.15	86.67
		KCenter	78.79	76.97	96.97	76.36	86.67
		ModeS	79.39	74.55	97.58	76.36	89.09
	WD	Rand	72.12	52.12	80.00	71.52	78.79
		RandC	71.52	49.09	79.39	70.91	76.36
		KCenter	74.55	49.70	79.39	70.30	76.36
		ModeS	71.52	49.09	76.97	69.70	74.55
Nist38	ED	Rand	79.10	77.10	80.80	82.00	71.10
		RandC	90.40	84.90	90.00	91.00	89.10
		KCenter	80.00	79.50	83.60	84.80	76.80
		ModeS	97.40	85.40	97.80	99.30	96.50
	HD	Rand	80.70	80.20	82.70	84.00	81.70
		RandC	90.00	82.00	88.50	90.00	86.80
		KCenter	81.30	78.30	83.00	84.00	73.60
		ModeS	97.40	85.40	97.80	99.30	97.80
	RD	Rand	84.50	80.70	86.30	85.70	0
		RandC	93.80	85.10	91.60	93.00	0
		KCenter	90.80	86.40	91.10	91.40	90.90
		ModeS	97.00	86.90	97.80	98.90	97.70
	WD	Rand	78.90	70.50	77.10	78.50	75.10
		RandC	87.40	73.90	84.40	86.60	80.80
		KCenter	79.90	74.70	79.70	81.80	64.00
		ModeS	93.70	77.50	95.50	96.80	95.20

Table 2: Classification accuracies (%) of DR-based DBCs.

data	methods		*knnc*	*nmc*	*ldc*	*qdc*	*svc*
Yale	ED	LDA	89.70	93.94	93.94	83.03	89.70
		PCA	79.39	80.61	93.94	79.39	46.67
		PCALDA	90.30	92.73	89.09	83.64	96.97
		CLDA	89.70	89.70	91.52	82.42	93.33
	HD	LDA	82.42	81.21	81.21	79.39	84.85
		PCA	75.76	77.58	86.67	78.79	89.70
		PCALDA	79.39	84.24	80.00	72.12	87.88
		CLDA	81.21	81.21	81.21	76.36	89.70
	RD	LDA	90.91	96.97	96.97	84.24	94.55
		PCA	79.39	75.15	96.97	77.58	86.06
		PCALDA	98.79	98.79	93.33	89.70	99.39
		CLDA	89.09	89.09	87.88	71.52	81.21
	WD	LDA	80.00	83.03	83.03	73.33	77.58
		PCA	70.91	49.70	72.73	70.91	73.33
		PCALDA	75.76	76.97	70.91	60.61	63.03
		CLDA	64.24	64.24	60.00	48.48	42.42
Nist38	ED	LDA	84.80	87.50	87.50	87.80	87.00
		PCA	98.10	87.50	98.00	99.30	96.60
		PCALDA	71.50	71.60	71.60	70.80	71.40
		CLDA	63.80	63.80	63.80	80.10	63.80
	HD	LDA	84.80	87.50	87.50	87.80	87.00
		PCA	98.10	87.50	98.00	99.30	97.70
		PCALDA	71.50	71.60	71.60	70.80	0
		CLDA	63.80	63.80	63.80	80.10	63.80
	RD	LDA	86.90	88.30	88.20	88.90	85.60
		PCA	97.30	88.20	97.10	98.60	0
		PCALDA	58.10	58.10	58.10	58.00	0
		CLDA	50.40	50.40	49.10	49.50	50.40
	WD	LDA	68.00	76.90	76.90	77.30	76.20
		PCA	92.50	76.90	97.00	97.00	96.30
		PCALDA	55.80	55.80	55.80	55.70	0
		CLDA	52.70	52.70	50.10	59.60	52.70

four of DRs) methods per each classifier and ranked them in the order from the highest to the lowest classification accuracies. Although this comparison is a very simplistic model of comparison, we believe that it is the easiest approach a researcher can employ when dealing with algorithms that have different characteristics.

From the rankings obtained from Tables 1 and 2, we can clearly observe the possibility of improving the performance of DBCs by utilizing the DRs. In most instances, the averaged classification accuracies of DR-based DBCs are increased compared to those of PS-based ones (note that almost all of the highest rankings are those of DR-based methods.) However, some DR-based DBCs failed to improve their classification accuracies [5]. From this consideration, we can see that it is difficult for us to grade the methods as they are. Therefore, for simple comparisons, we first assigned marks of 3, 2, or 1 to all DR and PS methods according to their ranks; 3 marks are given to the 1^{st} rank; 2 marks for the 2^{nd}, and 1 mark for the 3^{rd}. Then, we added up all the marks that each method earned with the *five* classifiers and the *four* measuring methods. For example, the marks that LDA gained in ED, HD, RH, and WD rows are $10(=2+3+2+2+1)$, $8(=3+2+1+2+0)$, $9(=2+2+1+2+2)$, and $14(=3+3+3+3+2)$, respectively. Thus, the total mark that the LDA earned is 41. Using the same system, we graded all the other DR (and PS) methods, and, as a final ranking, we obtained the followings:

(1) For Yale, 1^{st}: LDA (41); 2^{nd}: PCALDA (32); 3^{rd}: CLDA (17).

(2) For Nist38, 1^{st}: PCA (51); 2^{nd}: *ModeS* (45); 3^{rd}: *RandC* (14).

Here, the number $(\cdot)$ of each DR (or PS) method represents the final grade it obtained. From this ranking, we can see that all of the highest ranks are of DR methods. Thus, in general, it should be mentioned that more satisfactory optimization of DBCs can be achieved by applying a DR after building the dissimilarity space with all of the available samples rather than by selecting the representative subset from them.

As analyzed in Section 2, choosing the entire training set as the representative prototypes leads to higher computational complexity as more distances have to be calculated. In comparing PS-based and DR-based methods, we simply measured the processing CPU-times (seconds) of the DBCs designed with the two databases. In the interest of space, the details of the measured times are omitted here. From the measures, however, we can observe that the processing CPU-times increased when DR-based methods were applied. An instance of this change is the processing times of ED for Yale. The processing times of LDA, PCA, PCALDA, and CLDA methods are, respectively, 0.0250, 0.2161, 0.2521, and 27.9266 (seconds), while those of *Rand*, *RandC*, *KCenter*, and *ModeS* are, respectively, 0.0182, 0.0099, 0.0870, and

[5] For this failure, we are currently investigating why it occurs and what the cause is.

0.0141 (seconds). The same characteristic could also be observed in HD, RD, and WD methods. In light of brevity, the results of the others are omitted here again. However, it is interesting to note that the processing time of CLDA increases radically as the number of samples increases.

In review, the experimental results show that when the DR-based method is applied to the dissimilarity representation, the classification accuracy of the resultant DBCs increases, but so does the processing CPU-time. In addition, in terms of the classification accuracies, the DR-based method is more useful for the nonparametric classifiers, such as *knnc* and *nmc*, but not for the parametric ones, such as *ldc* and *qdc*.

4 CONCLUSIONS

In this paper, we performed an empirical comparison of PS-based and DR-based methods for optimizing DBCs. DBCs designed with the two methods were tested on the well-known benchmark databases, and the classification accuracies obtained were compared with each other. Our experimental results demonstrate that DR-based method is generally better than PS-based methods in terms of classification accuracy. Especially, the DR-based method is more useful for the nonparametric classifiers, but not for the parametric ones. Despite this success, problems remain to be addressed. First, in this evaluation, we employed a very simplistic model of comparison. Thus, developing a more scientific model, such as the one in (Sohn, 1999), is an avenue for future work. Next, the classification accuracy of DR-based DBCs increases, but so does the processing CPU-time. To solve this problem, therefore, developing a new dimensionality reduction scheme in the dissimilarity space is required. Future research will address these concerns.

REFERENCES

Adini, Y. Moses, Y. and Ullman, S. (1997). Face recognition: The problem of compensating for changes in illumination direction. *IEEE Trans. Pattern Anal. and Machine Intell.*, 19(7):721–732.

Belhumeour, P. N. Hespanha, J. P. and Kriegman, D. J. (1997). Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Anal. and Machine Intell.*, 19(7):711–720.

Bicego, M. Murino, V. and Figueiredo, M. A. T. (2004). Smilarity-based classification of sequences using hidden markov models. *Pattern Recognition*, 37:2281–2291.

Bunke, H. and Riesen, K. (2007). A family of novel graph kernels for structural pattern recognition. *Lecture Note in Computer Science*, 4756:20–31.

Georghiades, A. S. Belhumeur, P. N. and Kriegman, D. J. (2001). From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. Pattern Anal. and Machine Intell.*, 23(6):643–660.

Kim, S. W. (2006). Optimizing dissimilarity-based classifiers using a newly modified hausdorff distance. *Lecture Note on Artificial Intelligence*, 4303:177–186.

Kim, S. W. and Gao, J. (2008). On using dimensionality reduction schemes to optimize dissimilarity-based classifiers. *Lecture Note in Computer Science*, 5197:309–316.

Kim, S. W. and Oommen, B. J. (2007). On using prototype reduction schemes to optimize dissimilarity-based classification. *Pattern Recognition*, 40:2946–2957.

Laaksonen, J. and Oja, E. (1996). Subspace dimension selection and averaged learning subspace method in handwritten digit classification. In *Proceedings of ICANN*, pages 227–232, Bochum, Germany.

Loog, M. and Duin, R. P. W. (2004). Linear dimensionality reduction via a heteroscedastic extension of lda: The cherno criterion. *IEEE Trans. Pattern Anal. and Machine Intell.*, 26(6):732–739.

Pekalska, E. and Duin, R. P. W. (2005). *The Dissimilarity Representation for Pattern Recognition: Foundations and Applications*. World Scientific Pub., Singapore.

Pekalska, E. Duin, R. P. W. and Paclik, P. (2006). Prototype selection for dissimilarity-based classifiers. *Pattern Recognition*, 39:189–208.

Riesen, K. Kilchherr, V. and Bunke, H. (2007). Reducing the dimensionality of vector space embeddings of graphs. *Lecture Note on Artificial Intelligence*, 4571:563–573.

Rueda, L. and Herrera, M. (2008). Linear dimensionality reduction by maximizing the chernoff distance in the transformed space. *Pattern Recognition*, 41:3138–3152.

Sohn, S. Y. (1999). Meta analysis of classification algorithms for pattern. *IEEE Trans. Pattern Anal. and Machine Intell.*, 21(11):1137–1144.

Wei, C., L. Y. C. W. and Li, C. (2009). Trademark image retrieval using synthetic features for describing global shape and interior structure. *Pattern Recognition*, 42:386–394.

Wilson, C. L. and Garris, M. D. (1992). *Handprinted Character Database 3*. National Institute of Standards and Technology, Gaithersburg, Maryland.

Yu, H. and Yang, J. (2001). A direct lda algorithm for high-dimensional data - with application to face recognition. *Pattern Recognition*, 34:2067–2070.