

논문 2016-53-3-7

# 비유사도-기반 분류를 위한 차원 축소방법의 비교 실험

## ( A Comparative Experiment on Dimensional Reduction Methods Applicable for Dissimilarity-Based Classifications )

김 상 운\*

( Sang-Woon Kim<sup>Ⓢ</sup> )

## 요 약

이 논문에서는 비유사도-기반 분류(dissimilarity-based classifications: DBC)를 효율적으로 수행할 수 있는 차원 축소 방법들을 비교 평가한 실험 결과를 보고한다. DBC에선 분류를 위해 대상 물체를 측정된 결과 값들(특징 요소들의 집합)을 이용하는 대신에 각 대상 물체들 사이의 비유사도를 측정하여 분류한다. 현재 DBC와 관련된 이슈들 중의 하나는 대규모 데이터를 취급할 경우에 비유사도 공간의 차원이 고차원으로 되는 문제가 있다. 이 문제를 해결하기 위하여 현재 프로토타입 선택(prototype selection: PS)방법이나 차원 축소(dimension reduction: DR)방법을 이용하고 있다. PS는 전체 학습 데이터에서 프로토타입을 추출하여 비유사도 공간을 구성하는 방법이고, DR은 전체 학습 데이터로 먼저 비유사도 공간을 구성한 다음 이 공간의 차원을 축소하는 방법이다. 이 논문에서는 PS이나 DR 대신에, 학습 데이터에 대한 주성분 분석으로 적절한 차원의 고유 공간(Eigen space: ES)을 구성한 다음, 이 고유 공간으로 매핑된 벡터들 사이의  $l_p$ -놈(norm) 거리를 비유사도 거리로 측정하여 이용하는 DBC를 제안한다. 인터넷에 공개된 인공 및 실세계 데이터를 이용하여 최 근방 이웃 분류규칙으로 ES에서 수행한 DBC의 분류 성능을 측정된 결과, 고유공간의 차원을 적절하게 선정하였을 경우 PS와 DR를 이용한 DBC보다 분류 성능이 더 향상되었음을 확인하였다.

## Abstract

This paper presents an empirical evaluation on dimensionality reduction strategies by which dissimilarity-based classifications (DBC) can be implemented efficiently. In DBC, classification is not based on feature measurements of individual objects (a set of attributes), but rather on a suitable dissimilarity measure among the individual objects (pair-wise object comparisons). One problem of DBC is the high dimensionality of the dissimilarity space when a lots of objects are treated. To address this issue, two kinds of solutions have been proposed in the literature: prototype selection (PS)-based methods and dimension reduction (DR)-based methods. In this paper, instead of utilizing the PS-based or DR-based methods, a way of performing DBC in Eigen spaces (ES) is considered and empirically compared. In ES-based DBC, classifications are performed as follows: first, a set of principal eigenvectors is extracted from the training data set using a principal component analysis; second, an Eigen space is expanded using a subset of the extracted and selected Eigen vectors; third, after measuring distances among the projected objects in the Eigen space using  $l_p$ -norms as the dissimilarity, classification is performed. The experimental results, which are obtained using the nearest neighbor rule with artificial and real-life benchmark data sets, demonstrate that when the dimensionality of the Eigen spaces has been selected appropriately, compared to the PS-based and DR-based methods, the performance of the ES-based DBC can be improved in terms of the classification accuracy.

**Keywords** : dissimilarity-based classification(DBC), prototype selection based methods, dimension reduction based methods, Eigen space based methods

\*정회원-교신저자, 명지대학교 컴퓨터공학과(Department of Computer Engineering, Myongji University)

Ⓢ Corresponding Author(E-mail: kimsw@mju.ac.kr)

※ 이 논문은 2012년 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구의 일부임(2012R1A1A2041661).

※ 이 논문의 초본 [8]은 학술회의(IPCV2015, Las Vegas, Nevada, USA)에서 발표되었음.

Received ; November 23, 2015 Revised ; December 31, 2015 Accepted ; February 12, 2016

## I. 서 론

비유사도기반 분류법(dissimilarity-based classification: DBC)<sup>[1~2]</sup>은 분류대상(학습패턴)을 특징짓는 요소들(특징벡터)을 이용하는 기존의 (통계적 또는 구문적) 분류 방법과는 달리, 학습패턴을 대표하는 프로토타입( $P$ )과 학습패턴( $T$ )들 간의 비유사도(dissimilarity)를 측정하여 테스트 패턴의 클래스를 결정하는 방법이다. 즉,  $m$  개 프로토타입과  $n$  개 학습패턴들 사이의 비유사도를 측정하여  $m \times n$  차원의 비유사도 행렬( $D$ )을 구성한 다음, 이 행렬을 이용하여 학습한 분류기로 테스트 패턴( $z$ )의 클래스를 결정하는 방법이다. 이 때,  $T$ 가  $P$ 를 이용하여  $D$ 로 변환된 것과 같은 방법으로  $z$ 를 새로운 벡터(이하,  $\delta(z)$ 로 표기함)로 변환하여 처리한다. 여기서 DBC로부터 관찰되는 분류특성은 다음과 같다. 먼저 벡터 차원의 유해로움을 회피할 수 있고<sup>[1]</sup>, 비유사도를 측정하고 분류기를 설계할 때 관련 분야 전문가의 지식을 활용할 수 있다. 또한, 예러-프리 분류기를 구현할 수 있고, 설계할 때 테스트할 데이터까지도 학습 데이터로 이용할 수 있는 전달 학습(transductive learning)<sup>[2]</sup>이 가능하다.

한편, DBC의 구현과 관련된 이슈는 다음과 같다. 우선  $D$  행렬을 작성하기 위하여 응용 분야별 최적인 비유사도 측정법을 결정해야 하고, 비유사도 공간이 비정의적(indefinite)일 경우 비유사도 공간을 의사-유클리드 공간으로의 변환-embedding 해야 한다<sup>[3]</sup>. 또한 기존 통계적 처리를 위해 개발된 벡터 공간의 툴<sup>[4]</sup>들을 활용하고 비유사도 공간에 적합한 분류기를 설계해야 한다.

특히, 대규모 데이터를 대상으로 DBC를 구현할 때 데이터 셋의 크기에 따라  $D$ 의 차원이 증가한다. 이를 해결하기 위하여 프로토타입 선택(prototype selection: PS)<sup>[5]</sup> 방법이나 차원 축소(dimension reduction: DR)<sup>[6~7]</sup> 방법을 이용한다. PS 방법은 학습 데이터에서 프로토타입을 추출하여 비유사도 공간을 구성하고, DR 방법은 전체 학습 데이터로 먼저 비유사도 공간을 구성한 다음 이 공간의 차원을 축소하는 방법이다.

이 논문에서는  $D$ 의 차원을 축소하기 위해 PS나 DR 대신에 고유 공간 (Eigen space: ES)에서 DBC를 수행하는 방법<sup>[8]</sup>의 활용을 검토한다. 즉, 학습 데이터에 대한 주성분 분석으로 적절한 차원의 고유 공간을 구성한 다음, 분류할 패턴 벡터를 고유 공간으로 변환한다. 그리고 변환된 벡터들 사이의  $l_p$ -놈(norm) 을 측정하여

DBC를 수행하고 그 결과를 기존의 PS나 DR에 기반한 DBC의 분류 결과와 비교한다.

현재, 저차원 소규모 데이터의 경우 응용분야에 따라 DR을 이용한 DBC가 PS의 DBC보다 양호하다고 알려져 있다<sup>[8]</sup>. 그 이유는 DR의 DBC에서는 PS의 경우와 달리 프로토타입을 선정할 때 발생하는 정보손실을 방지할 수 있기 때문이다. 즉, DR법에서는  $T$ 의 모든 샘플을 활용하여  $D$ 를 계산한 후, PCA(principle component analysis)나 LDA(linear discriminant analysis) 등을 수행하여 차원을 축소한다. 그러나 DR법에서  $D$ 의 차원축소를 위해 PCA를 적용할 경우 계산 복잡도는  $O(8n^3)$ 으로, 데이터 셋의 크기( $n$ )에 민감하게 된다. 따라서 대규모 데이터 셋 ( $n \rightarrow \infty$ )일 경우 시간 복잡도를 줄이기 위한 별도의 방법이 필요하게 된다.

또한  $D$ 의 비유사도를 측정하는 방법은 유클리디언(Euclidean)거리 뿐만 아니라 하우스도르프(Hausdorff) 거리와 스펙트럼과 같은 통계분포의 비교에 유용한 EMD(Earth mover's distance) 등과 같은 다양한 비-유클리디언 거리를 이용할 수 있다. 특히, 유클리드 거리를 적용할 경우 민코프스키(Minkowski)  $l_p$ -놈을 이용할 수 있다. 이상의 고찰로부터 대규모 고차원 데이터의 경우  $D$ 의 차원을 축소하기 위하여 고유공간에서 DBC를 수행하는 방법(이를 ES법이라 한다)을 검토한다. 특히, 고유공간에서  $l_p$ -놈으로 비유사도를 측정할 때  $p$ 에 따른 성능변화를 고찰한다. 즉, 고유공간을 구성할 때 적용하는 메트릭과 비유사도를 측정하는 메트릭의 관계를 고찰한다.

본 논문의 구성은 다음과 같다. 제 II장에서는 비유사도 표현 및 비유사도에 기반한 분류법을 간단히 소개한다. 제 III장에서는 DBC를 대규모 데이터에 적용할 때 비유사도 공간 행렬의 차원을 축소하기 위한 기법인 PS, DR, ES법을 차례로 설명한다. 제 IV장에서는 인공 및 실세계 실험 데이터를 이용하여 데이터 셋 크기에 따른 DBC의 처리시간 변화, 비유사도 벡터 차원에 따른 DBC의 분류오차의 변화, 그리고 비유사도 측정 방법에 따른 차원 축소방법들의 분류오차 및 처리시간을 비교 분석한다. 끝으로 제 V장에서 결론을 맺는다.

1  $d$ -차원 특징벡터에 대한 PCA를 위해 parallel QR 알고리즘을 이용할 경우 계산 복잡도는  $O(8d^3)$ 이다<sup>[9]</sup>.

## II. 관련 연구

비유사도 표현은 식별대상을 학습패턴에서 추출한 표준패턴(프로토타입)들과의 유사도 또는 비유사도를 측정하여 표현하는 방법이다. 패턴 클래스를 미리 알고 있는  $n$ 개의  $d$ -차원 벡터로 이루어진 학습패턴을  $T = \{x_1, \dots, x_n\}$ 이라 할 때,  $T$ 를  $c$ 개의 부분집합  $T_1, T_2, \dots, T_c$ 로 나눈다. 여기서 각 클래스  $c_i$ 의 패턴 집합  $T_i = \{x_1, \dots, x_{n_i}\}$ 에서 추출 또는 선정한 표준패턴의 집합을  $Y_i = \{y_1, \dots, y_{m_i}\}$ 이라 한다. 이 때, 두 벡터  $x_i \in T$ 와  $y_j \in Y$ 간의 비유사도(예, 유클리드 벡터 거리 등)을  $\delta_{ij} \equiv \delta(x_i, y_j)$ 라 하면,  $T$ 와  $Y$ 의 비유사도 행렬  $D_{[T, Y]}$ 는  $(n \times m)$ -차원으로

$$D_{[T, Y]} = \begin{matrix} & \begin{matrix} y_1^{(1)} & y_2^{(1)} & \dots & y_i^{(2)} & y_{i+1}^{(2)} & \dots & y_m^{(c)} \end{matrix} \\ \begin{matrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{matrix} & \begin{pmatrix} \delta_{11} & \delta_{12} & \dots & \delta_{1i} & \delta_{1i+1} & \dots & \delta_{1m} \\ \delta_{21} & \delta_{22} & \dots & \delta_{2i} & \delta_{2i+1} & \dots & \delta_{2m} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \delta_{n1} & \delta_{n2} & \dots & \delta_{ni} & \delta_{ni+1} & \dots & \delta_{nm} \end{pmatrix} \end{matrix} \quad (1)$$

이 되며, 임의의 벡터  $x_i \in T$ 에 대한  $D$ 의 행 벡터(row vector)  $\delta(x_i) (\equiv \delta(x_i, Y))$ 는

$$\delta(x_i) = (\delta(x_i, y_1), \delta(x_i, y_2), \dots, \delta(x_i, y_m))^T \quad (2)$$

이 된다. 즉,  $\delta(x_i)$ 는 벡터  $x_i$ 와 프로토타입  $Y$ 와 비유사도 거리를 측정된 값의 배열로 하나의 새로운 벡터가 된다. 여기서, 식 (2)를 비유사도 벡터라 정의하고, 이 벡터들로 이루어지는 벡터 공간을 비유사도 공간이라 정의한다. 즉,  $\delta(x)$ 는  $d$ -차원 특징 공간의 벡터  $x$ 를  $m$ -차원 비유사도 공간으로 사상시킨 벡터이다. 따라서 DBC란 입력 특징공간의  $d$ -차원 벡터  $x$ 대신에 비유사도 공간의  $m$ -차원 벡터  $\delta(x)$ 를 대상으로 하는 분류방법이다. 이 때, 학습샘플은  $\{\delta(x_i)\}_{i=1}^n$ 이 되고,  $k$ -최근방 이웃( $k$ -nearest neighbor:  $k$ -NN) 분류기를 비롯한 선형 분류기, 베이시언 2-차 분류기 등 기존 특징 기반의 식별규칙을 이용할 수 있다. 그리고 테스트 샘플 벡터  $z$ 도 비유사도 공간의  $m$ -차원 벡터  $\delta(z)$ 로 변환하여 분류한다.

또한, 식 (2)에서  $x_i$ 와  $y_j$ 사이의 비유사도 거리를 나타내는  $\delta(x_i, y_j)$ 는 다양한 방법으로 측정할 수 있다.

예를 들면, 민코프스키 거리도 이용할 수 있으며,

$$\delta(x_i, y_j) = \left( \sum_{k=1}^d |x_{ik} - y_{jk}|^p \right)^{1/p} \quad (3)$$

이 된다. 여기서  $p = 1$ 일 때, 식 (3)의 거리를 맨해튼 거리 또는 시티-블록 거리라 하고,  $p = 2$ 일 경우에는 유클리드 거리가 된다. 한편, 최적 비유사도 공간을 위해 거리 메트릭을 학습하는 방법도 있다<sup>[10]</sup>.

## III. DBC를 위한 차원 축소법

### 1. Prototype Section 방법(PS법)

학습 데이터  $T = \{x_i\}_{i=1}^n$ ,  $x_i \in R^d$ 에서 프로토타입  $Y = \{y_i\}_{i=1}^m$ ,  $y_i \in R^d$ , ( $m \leq n$ )를 선택하는 처리는 DBC의 분류효율(성능)과 계산시간(비용) 사이의 손익 분기점을 선택하는 처리에 해당한다. 즉,  $(n \times n)$ -차원의  $D$  대신에  $(n \times m)$ -차원의  $D$ 에서 DBC를 수행하는 방법은  $m \ll n$ 의 경우 계산시간을 줄일 수 있는 반면에 분류성능의 저하로 결과 될 수 있다. 따라서 성능저하를 최소로 하면서 최소 차원의  $D$ 를 구성할 수 있는 최적의  $Y$ 를 선택하여야 한다.

이 문제를 해결하기 위하여 클러스터링에 기초한 다양한 방법들이 제안되었다<sup>[5]</sup>. 예를 들어, *RandomC*는 일정한 개수의 프로토타입을 각 클래스 별로 랜덤하게 선정하는 방법으로, 몇몇 응용분야에서  $(n \times n)$ -차원의  $D$ 와 비슷한 분류성능이 보고되었다<sup>[5]</sup>.

PS에 기반 한 DBC 알고리즘(이하, PS법)은 다음과 같다. 여기서 입력은 학습 데이터  $T$ ,  $m (= |Y|)$ , 테스트 샘플  $z$ 이고, 출력은  $z$ 의 클래스 *class*이다.

1. 학습 데이터  $T$ 에서 프로토타입  $Y$ 를 추출한다. 이 때, *RandomC* 등을 이용할 수 있다.
2. 식 (1)를 이용하여 주어진 비유사도 측정법으로  $T$ 와  $Y$ 로부터 비유사도 행렬  $D_{[T, Y]}$ 를 계산한다. 또 테스트할 입력 벡터  $z$ 에 대해 같은 방법으로 비유사도 벡터  $\delta(z)$ 를 계산한다.

3. 비유사도 행렬  $D_{[T, Y]}$ 에서 하나의 분류기(예, NN)를 학습한 후, 이 분류기를 이용하여 벡터  $\delta(z)$ 의 클래스를 결정하여 *class*로 출력한다.

위 알고리즘에서 *RandomC*와 NN을 이용할 경우, 각 단계별 시간 복잡도는  $O(1)$ ,  $O(dn^2)$ ,  $O(nd)$ 이다.

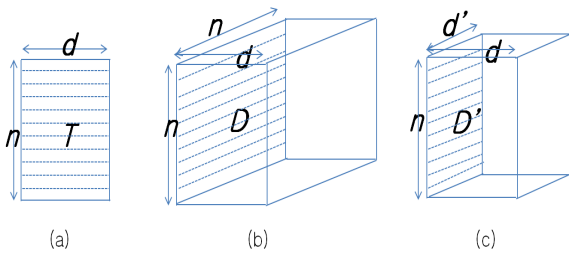


그림 1. DR법의 계산과정(시간 복잡도)을 나타내는 도형: (a) 학습 데이터  $T$ , (b)  $T$ 를 이용하여 작성한  $n$ -차원 비유사도행렬  $D$ , (c)  $D$ 의 차원을 축소한  $d'$ -차원 비유사도 행렬  $D'$ . 여기서  $d$  ( $d'$ ): 점선으로 표시한 벡터의 차원 (축소된 차원),  $n$ : 벡터 개수

Fig. 1. Plots denoting the computational complexity of DR: (a) training data  $T$ , (b)  $n$ -dimensional dissimilarity matrix  $D$  obtained using  $T$  in (a), (c)  $d'$ -dimensional dissimilarity matrix  $D'$  reduced from  $D$  in (b). Here,  $d$  ( $d'$ ): dimensionality of the vectors represented in dotted lines (reduced dimensionality),  $n$ : number of the vectors.

## 2. Dimension Reduction 방법(DR법)

PS법에서  $Y$ 의 선정은 분류성능에 민감하다. 즉,  $Y$ 의 샘플들은 중복을 피하면서 가능한 한  $T$ 에 내재되어 있는 분류정보를 최대한 포함하여야 한다. 그러나  $T$ 에서 부분집합을  $Y$ 로 선택하기 때문에 정보손실이 불가피하게 된다. 이를 해결하기 위하여 먼저  $T$ 의 모든 샘플을  $Y$ 로 선택하여  $D$ 를 작성한 다음, 전통적인 차원 축소법으로  $D$ 의 차원을 축소시키는 방법이 제안되었다<sup>[6~7]</sup>. 즉,  $T$ 의 모든 샘플로  $(n \times n)$ -차원의  $D$ 를 작성한 다음, PCA나 LDA를 이용하여  $n$ -차원 벡터  $\delta(x)$ 를  $m$ -차원 ( $m \ll n$ )으로 축소하는 방법이다.

DR에 기반 하는 DBC 알고리즘(이하, DR법)은 다음과 같다. 여기서 입력은 학습 데이터  $T$ , 축소할 비유사도 벡터의 차원  $d'$ , 테스트 샘플  $z$ 이고, 출력은  $z$ 의 클래스  $class$ 이다.

1. 학습 데이터  $T = \{x_i\}_{i=1}^n$ ,  $x_i \in R^d$  전체를  $Y$ 로 이용하여 주어진 비유사도 측정법으로 비유사도 행렬  $D_{[T,T]}$ 를 계산한다. 또 테스트 벡터  $z$ 에 대해 같은 방법으로  $\delta(z)$ 를 계산한다.

2. 하나의 차원 축소법(예, PCA)을 이용하여  $n$ -차원 벡터  $\{\delta(x_i)\}_{i=1}^n$ 와  $\delta(z)$ 를 각각  $d'$ -차원  $\{\delta'(x_i)\}_{i=1}^n$ 와  $\delta'(z)$ ,  $d' (\ll n)$ 으로 변환한다(그림 1 참조).

3.  $d'$ -차원 비유사도 행렬  $D'$ 에서 하나의 분류기(예,

NN)를 학습한 후, 이 분류기를 이용하여  $d'$ -차원 벡터  $\delta'(z)$ 의 클래스를 결정하여  $class$ 로 출력한다.

위 알고리즘에서 PCA와 NN을 이용할 경우, 단계 별 시간 복잡도는 각각  $O(dn^2)$ ,  $O(8n^3)$ ,  $O(nd')$ 이다.

## 3. Eigen Space 방법(ES법)

DR법은  $T$ 의 모든 정보를 이용하여  $D$ 를 구성할 수 있으나, 생성된 비유사도 공간의 차원이  $n = |T|$ 으로 학습 데이터가 대규모( $n \rightarrow \infty$ )일 경우 시간 복잡도가 크게 증가한다. 여기서  $D$ 의 계산 량은  $n \times n \times d$ 로 도형의 부피에 해당한다(그림 1(b) 참조).

그림 1에서 (a)는  $d$ -차원 벡터  $n$ 개로 구성된 학습 데이터  $T$ 이고, (b)는  $(n \times n)$ -차원 비유사도 행렬  $D_{[T,T]}$ 의 계산 량, (c)는  $d'$ -차원 비유사도 행렬  $D'$ 이다. 여기서 (b)의 계산 량(부피)를 줄이기 위해선  $d$ 를 줄여야 한다는 것을 알 수 있다. 이 때, (b)의 식별정보를 그대로 유지하기 위하여  $T$ 를 고유 공간(ES)으로 변환한 후  $l_p$ -놈으로 비유사도를 측정하는 DBC를 검토한다.

ES에 기반 한 DBC 알고리즘(이하, ES법)은 다음과 같다. 여기서 입력은 학습 데이터  $T$ , 축소할 학습벡터 차원  $q$ , 테스트 샘플  $z$ 이고, 출력은  $z$ 의 클래스  $class$ 이다.

1. 변환 식  $x'_i = A^T(x_i - m)$ 를 이용하여  $T$ 의 모든 샘플  $x_i \in R^d$ 를  $x'_i \in R^q$ 로 변환한다. 여기서 변환행렬  $A \in R^{d \times q}$ 는  $T$ 의 공분산 행렬에 대한  $q$ 개의 고유벡터이고,  $m \in R^d$ 은  $T$ 의 평균이다. 또 테스트 벡터  $z$ 에 대해 같은 방법으로  $z'$ 을 계산한다.

2. 식 (1)를 이용하여 주어진 비유사도 측정법으로 비유사도 행렬  $\{\delta(x'_i)\}_{i=1}^n \in D_{[T,T]}$ 와 비유사도 벡터  $\delta(z')$ 을 계산한다.

3. 비유사도 행렬  $D_{[T,T]}$ 에서 하나의 분류기(예, NN)를 학습한 후, 이 분류기를 이용하여 벡터  $\delta(z')$ 의 클래스를 결정하여  $class$ 로 출력한다.

위 알고리즘에서 PCA와 NN을 이용할 경우, 각 단계 별 시간 복잡도는  $O(8d^3)$ ,  $O(qn^2)$ ,  $O(nq)$ 가 된다.

## IV. 실험

### 1. 실험 #1

먼저, 공개 소프트웨어 PRTools<sup>2</sup>에서 제공하는  $d$ -

<sup>2</sup> <http://prtools.org/> 참조.

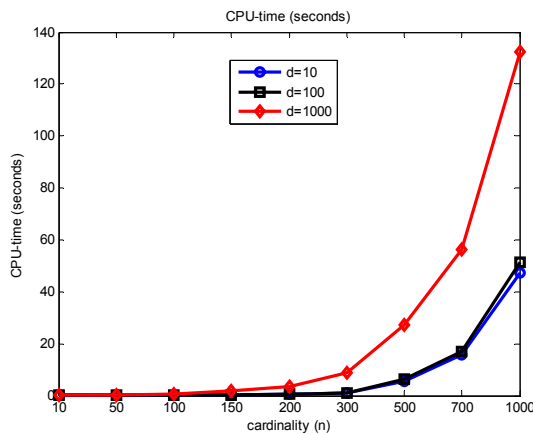


그림 2.  $d$ -차원 2-클래스 인공데이터 Difficult Data에 대한 DR법의 처리시간 비교. 여기서 원, 사각형, 마름모 기호는 각각  $d=10, 100, 1000$ 을 표시함  
Fig. 2. Comparison of CPU-time (seconds) measured using DR-DBC for  $d$ -dimensional 2-class artificial data, Difficult Data datasets. Here, circle, square, diamond markers denote  $d=10, 100, 1000$ , respectively.

차원 2-클래스 인공데이터 Difficult Data에 대한 DR법의 처리시간(학습 및 분류 시간)을 비교하였다. Difficult Data의 차원( $d$ )을 10, 100, 1000으로 정하고 각각의 차원에 대하여 샘플 수( $n$ )를 10부터 1000까지 증가시키면서 DR법의 처리시간을 측정할 결과는 그림 2와 같다.

그림 2에서  $d=10$ 이나  $d=100$ 인 경우 DR법의 수행시간은  $n$ 의 증가에 따라 서서히 증가하면서 두 커브는 매우 유사함을 보인다(그림 2에서 서클기호와 네모기호 참조). 그러나  $d=1000$ 인 경우 샘플 수  $n$ 이 500이나 1000일 때 처리시간이 급격하게 증가함을 보인다(그림 2에서 다이아몬드기호 참조). 즉, DR법에서 학습 데이터가 고차원이면서 대규모일 경우 수행시간은 큰 폭으로 증가하게 됨을 알 수 있다.

## 2. 실험 #2

다음, PRTtools에서 제공하는 숫자 '3'과 '8'에 대한  $32 \times 32$  화소의 필기체 숫자 데이터 Nist38(여기서,  $d=1024$ )을 이용하여 PS, DR, ES법으로 학습한 분류기의 분류오차를 비교하였다. 그림 3은 PS의 프로토타입 개수  $m$ (DR 및 ES법의 경우 비유사도 공간의 차원  $d'$ 과 특징벡터 차원  $q$ )를  $2^1, 2^2, \dots, 2^{10}$ 로 증가시키면서 DBC로 학습한 NN 분류기의 분류오차비율을 측정할 결과이다. 여기서 비유사도는 PS와 DR법의 경우 유클리디언 거리( $l_2$ )로 측정하였고, ES법에서는 유클리디

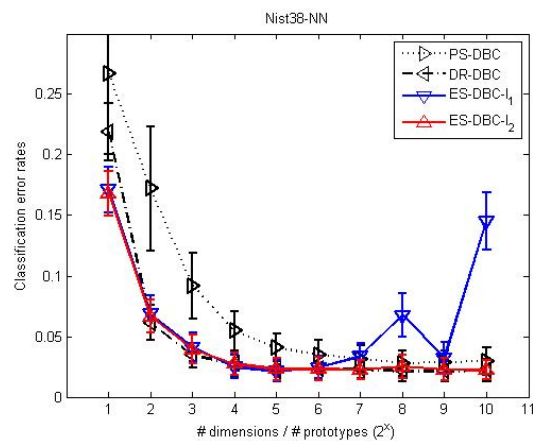


그림 3. Nist38 데이터에 대해 PS, DR 및 ES( $l_2$ 와  $l_1$ )법으로 학습한 NN 분류기의 분류오차비율 비교. 여기서  $x$ -축은 PS법에서는 프로토타입 수( $n$ )를, DR과 ES법의 경우엔 축소된 비유사도 공간의 차원 ( $d', q$ )을 나타냄  
Fig. 3. Comparison of classification error rates of NN classifiers trained using PS, DR, and ES( $l_2$  and  $l_1$ ) for Nist38 data. Here, in PS, the  $x$ -axis denotes the number of prototypes ( $n$ ), while, in DR and ES, the axis denotes the reduced dimensionality ( $d', q$ ).

언 거리( $l_2$ )와 시티-블록 거리( $l_1$ )의 두 방법으로 측정하였다. 또한 학습 및 테스트 과정을 30회 반복하여 평균과 편차를 산출하였다.

그림 3에서 ES-DBC- $l_2$  오차비율은  $x$ -축의 전체 영역에서 DR-DBC와 겹쳐져 있고, ES-DBC- $l_1$ 의 경우엔  $d \leq 2^6$ 의 저차원 공간에서만 DR과 겹치고 있다. 여기서  $d > 2^6$ 인 고차원의 경우 오차율이 불안정하게 큰 폭으로 증가하는 현상을 보인다(이는 통계적 패턴인식에서 관찰되는 피킹 현상(peaking phenomenon)에 해당한다. 이와 달리 PS법의 경우 모든 구간에서 DR법보다 높은 오차를 보이고 있다. 따라서 ES( $l_2$ )법은 고유공간의 모든 차원에서, 그리고 ES( $l_1$ )법은 고유공간의 저차원에서 DR법과 비슷한 성능을 보이고 있다.

그림 3에서 ES-DBC- $l_2$  오차비율은  $x$ -축의 전체 영역에서 DR-DBC와 겹쳐져 있고, ES-DBC- $l_1$ 의 경우엔  $d \leq 2^6$ 의 저차원 공간에서만 DR과 겹치고 있다. 여기서  $d > 2^6$ 인 고차원의 경우 오차율이 불안정하게 큰 폭으로 증가하는 현상을 보인다(이는 통계적 패턴인식에서 관찰되는 피킹 현상(peaking phenomenon)에 해당한다. 이와 달리 PS법의 경우 모든 구간에서 DR법보다 높은 오차를 보이고 있다. 따라서 ES( $l_2$ )법은 고유공간의 모든 차원에서, 그리고 ES( $l_1$ )법은 고유공간의 저차원에서 DR법과 비슷한 성능을 보이고 있다.

## 3. 실험 #3

인식을 위한 표준화 영상 데이터베이스 PASCAL VOC 2007 Dataset(이하 VOC'07로 표기함)<sup>3</sup>에 대해 PS, DR, ES법 NN의 분류오차와 처리시간을 비교하였다. VOC'07 데이터베이스는 다음과 같은 5개 그룹으로 되어 있다: person (1025/4000), horse (139/4000), car

3 <http://host.robots.ox.ac.uk/pascal/VOC/>

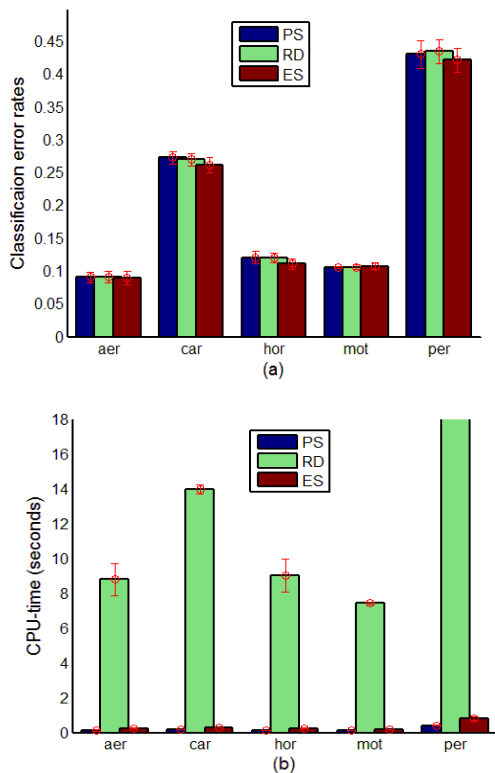


그림 4. VOC'07 데이터에 대해 PS, DR 및 ES( $l_2$ )로 학습한 NN 분류기 성능: (a) 분류오차비율, (b) CPU-처리시간(초)

Fig. 4. Performances of NN classifiers trained using PS, DR, and ES( $l_2$ ) for VOC'07 data: (a) Classification error rates, (b) CPU-times (seconds).

(376/4000), aeroplane (112/4000), motorbike (120/4000). 여기서 괄호안의 숫자는 ( $n/d$ )이다. 각 그룹은 background (1019/4000)와 결합되어 2-클래스 데이터 셋을 구성한다. 즉, person(간단히 per로 표기함)은  $n = [1025, 1019]$ 과  $d = 4000$ 인 2-클래스 데이터 셋이 된다. 그림 4(a)는 VOC'07의 다섯 데이터 셋에 대해 ES, DR, ES법으로 학습한 NN 분류기의 분류오차비율이고, 그림 4(b)는 처리시간(초)를 측정된 결과이다. 여기서, 비유사도는 모두 유클리디언 거리( $l_2$ )로 비유사도를 측정하였으며,  $m = d' = q = 16$ 으로 실험하였다. 또한 학습 및 테스트 과정을 30회 반복하여 평균과 편차를 산출하였다.

그림 4(a)에서 car, horse, person 데이터에 대해서는 ES법의 분류성능이 PS와 DR법 보다 우수함을 보이고 있고, 나머지 두 데이터 셋에 대해서는 세 방법의 성능이 아주 비슷함을 보인다(분류오차율 바의 높이가 같다). 그러나 그림 4(b)에서 세 방법의 처리시간은 큰 차이를 보인다. 예를 들면, person의 경우 PS, DR, ES법의 처리시간은 각각 0.4, 52.5, 0.8(초)이다. 따라

서 VOC'07 데이터의 경우 세 방법의 분류오차는 비슷한 반면, 처리시간은 크게 차이를 알 수 있다.

#### 4. 실험 #4

끝으로, UCI Irvine Machine Learning Repository<sup>4</sup>의 데이터를 이용하여 PS, DR, ES법으로 학습한 NN 분류기의 분류오차와 처리시간을 측정하여 비교하였다. UCI 실험 데이터는 다음과 같다: diabetes (8/ 2/ 768), heart (13/ 2/ 297), liver (16/ 2/ 345), sonar (60/ 2/ 208), Laryngeal1(16/ 2/ 213), wdbc(30/ 2/ 569), waveform(40/ 2/ 3344). 여기서 괄호 안의 세 숫자는 ( $d/c/n$ )이다. 이 데이터에 대해 ES, DR, ES법으로 학습한 NN 분류기의 분류오차비율(%)은 표 1과 같고, 처리시간(초)은 표 2와 같다.

이 실험에서 PS의 프로토타입 개수  $m$  (DR 및 ES의 경우 비유사도 공간 차원  $d'$ 와  $q$ )은 고유 값 누적(cumulative proportion of eigenvalues)<sup>[7]</sup>을 참조하여 위의 실험 데이터에 대해 각각 5, 4, 5, 29, 4, 8, 8으로 정하였다. 또한 비유사도는 PS와 DR법에서는 유클리디언 거리( $l_2$ )로 측정하였고, ES법에서는 네 종류 민코프스키 거리( $l_1, l_{1.5}, l_2, l_3$ )로 측정하였다. 그리고 고유공간을 생성하기 위한 메트릭과 비유사도 거리 메트릭의 관계를 조사하기 위하여 전통적인  $l_2$ -거리 PCA(이하 L2-PCA로 표기)와 함께  $l_1$ -거리 PCA(이하, L1-PCA<sup>[11]</sup>로 표기)를 이용하여 실험하였다.

표 1의 실험결과에 대한 고찰은 다음과 같다. 첫 째, 각 데이터 셋에 대해 최소 분류오차비율을 조사한 결과(표에서 밑줄 친 진한 숫자), ES법의 L2-PCA( $l_1$ )이 4회, ES법의 L2-PCA( $l_1$ ), L2-PCA( $l_{1.5}$ ), L1-PCA( $l_1$ )이 각각 1회 씩 최소 분류오차를 출력하였음을 보인다. 이 결과로부터 ES법이 PS와 DR법 보다 (미세하나마) 성능이 개선되었음을 알 수 있다.

둘 째, ES의 L2-PCA와 L1-PCA는 모든  $p$ (차수)에서 비슷한 분류오차를 보인다. 즉, L2-PCA와 L1-PCA를 쉽게 비교하기 위해 L2-PCA와 L1-PCA의 네 종류 오차(즉,  $p = 1, p = 1.5, p = 2, p = 3$ )를 평균하였다. 이 결과로부터  $l_p$ -놈을 비유사도로 이용할 경우  $p$ 가 작을수록 더 안정적인 측정이 가능함을 알 수 있고, 이 결과는 고차원 응용의 경우  $l_p$ -놈의  $p$ 가 작을수록 견고하다는 기존의 연구결과와 일치한다. 이는 고유 공간으로 투영할

4 <http://archive.ics.uci.edu/ml/>

표 1. UCI 데이터에 대해 PS, DR 및 ES법으로 학습한 NN 분류기의 분류오차 비율(%). 여기서  $p$ 는 민코프스키 거리 (Minkowski distance) 차수이고, 괄호 안에 표시한 숫자는 표준편차임

Table 1. Classification error rates(%) of NN classifiers trained using the PS, DR, and ES methods for UCI datasets. Here,  $p$  denotes the order of Minkowski distance, while the numbers in the brackets are the standard deviations.

data sets	PS	DR	ES							
			L2-PCA				L1-PCA			
			p=1	p=1.5	p=2	p=3	p=1	p=1.5	p=2	p=3
diabetes	30.3 (3.3)	29.8 (2.1)	<b>26.3</b> (2.4)	27.9 (2.4)	29.6 (2.7)	30.4 (2.0)	28.2 (2.7)	29.8 (2.5)	29.7 (2.4)	28.0 (2.4)
heart	39.4 (5.3)	39.3 (4.1)	<b>36.2</b> (5.4)	38.2 (4.7)	39.2 (5.0)	39.5 (5.3)	39.5 (5.1)	38.4 (5.0)	39.3 (5.1)	38.6 (4.3)
liver	40.6 (6.4)	39.5 (4.5)	<b>35.0</b> (4.8)	37.8 (4.6)	38.0 (4.7)	39.4 (4.2)	36.9 (3.8)	37.5 (3.5)	37.7 (4.3)	38.1 (4.7)
sonar	24.4 (6.8)	21.8 (5.1)	18.4 (5.5)	<b>17.6</b> (5.6)	21.3 (5.0)	23.7 (5.4)	19.0 (5.0)	18.6 (4.6)	21.5 (4.8)	23.8 (5.2)
Laryngeal1	33.7 (6.8)	36.5 (5.2)	<b>26.2</b> (3.7)	26.9 (4.8)	30.6 (5.3)	31.7 (4.6)	26.8 (5.6)	27.4 (4.3)	30.4 (4.5)	32.5 (5.3)
wdbc	7.8 (2.6)	7.7 (2.6)	6.5 (2.6)	7.5 (2.4)	7.4 (2.4)	7.4 (2.6)	<b>6.3</b> (2.1)	6.9 (1.8)	7.4 (2.4)	7.8 (2.6)
waveform	10.5 (2.3)	7.1 (0.9)	7.8 (1.1)	7.2 (1.0)	7.0 (1.1)	<b>6.9</b> (1.1)	7.5 (1.0)	7.1 (1.0)	7.1 (0.9)	7.0 (0.9)

표 2. UCI 데이터에 대한 PS, DR 및 ES법으로 학습한 NN 분류기의 처리시간(측정시간×100초)

Table 2. CPU-times (measured values × 100 seconds) of NN classifiers trained using PS, DR, and ES for UCI datasets.

data sets	PS	DR	ES							
			L2-PCA				L1-PCA			
			p=1	p=1.5	p=2	p=3	p=1	p=1.5	p=2	p=3
diabetes	10	170	98	114	113	109	98	114	112	109
heart	5	19	10	11	11	11	10	11	10	10
liver	5	23	13	13	13	13	12	13	13	13
sonar	5	12	7	7	7	7	7	7	7	7
Laryngeal1	5	13	7	7	8	7	7	7	7	7
wdbc	7	77	8	8	8	8	8	8	8	8
waveform	96	23300	234	234	235	235	234	239	237	236

경우 벡터를 구성하는 특이요소가 비유사도 거리에 미치는 영향이 축소되기 때문으로 보인다.

셋 째, PS, DR, ES(L2-PCA- $l_1$ )법을 비교하기 위해 해당 컬럼의 분류오차를 평균 한 결과 ES법이 PS나 DR법보다 분류성능이 개선되었음을 알 수 있다.

끝으로, 표 2로부터 실험데이터에 대한 처리시간은 ES법이 DR법보다 매우 단축되었음을 알 수 있다.

위의 실험결과를 요약하면 다음과 같다. 먼저, DBC를 대규모 데이터에 응용할 경우 시간 복잡도를 줄이기

위해 PS와 DR법과 함께 ES법을 이용할 수 있다. 이때 DR법은 분류오차 성능이 PS법보다 우수하다고 알려져 있으나, 학습 데이터가 고차원 대규모가 되면 수행시간이 큰 폭으로 증가하게 된다. 반면에 ES법은 고유평간의 저차원에서 DR법과 유사한 성능을 고속으로 얻을 수 있다. 또한 고유평간에서 민코프스키 거리로 비유사도를 측정할 경우 차수가 적을수록 효과적이다. 끝으로, 고유 공간의 생성 메트릭과 비유사도 측정 메트릭 사이에 특별한 관계는 관찰되지 않았다.

#### IV. 결 론

본 논문에서는 빅 데이터 처리 등과 같이 고차원 대규모 데이터에 DBC를 효율적으로 적용하기 위해 비유사도 행렬의 차원을 축소하는 방법들을 비교 검토하였다. 특히, 프로토타입 추출에 기반 한 방법(PS법)과 전체 학습 데이터를 이용하여 비유사도 행렬을 작성한 다음 차원 축소법을 적용하는 방법(DR법), 그리고 고유공간에서 비유사도를 측정하여 DBC를 수행하는 방법(ES법)을 실험적으로 비교 고찰하였다. 인공 데이터 및 실세계 데이터를 대상으로 실험한 결과 고차원의 대규모 데이터 셋의 경우 일반적으로 ES법이 PS나 DR법보다 분류성능(분류오차와 처리시간)이 우수하였다. 또한, 고유 공간에서  $l_p$ -놈을 이용하여 비유사도를 측정할 때  $p$ 가 작을수록 더 안정적인 측정이 가능함을 확인하였다. 그러나 고유 공간을 구성하기 위한 메트릭과 비유사도를 측정하는 메트릭 사이의 관계 규명은 앞으로의 과제이다. 또한 고유 공간의 최적 차원을 선정하는 문제, ES법을 효과적으로 적용할 수 있는 응용 분야를 자동 결정하는 문제도 앞으로의 과제이다. 끝으로, 본 논문에서는  $l_p$ -놈으로만 비유사도를 측정하였다. 따라서 ES법에 다양한 거리측정법과 최적 메트릭 학습법을 적용하고, 더 나아가 다양한 비유사도 행렬의 결합법을 활용하는 연구도 앞으로의 과제이다.

#### REFERENCES

- [1] S. -W. Kim and R. P. W. Duin, "On optimizing dissimilarity-based classifier using multi-level fusion strategies (in Korean)," Journal of the Institute of Electronics Engineers of Korea, vol. 45, no. CI-5, pp. 15-24, 2008.
- [2] R. P. W. Duin, "The dissimilarity representation for finding universals from particulars by an anti-essentialist approach," Pattern Recognition Letters, vol. 64, pp. 37 - 43, 2015.
- [3] J. Laub, V. Roth, J. M. Buhmann, and K. -R. Müller, "On the information and representation of non-Euclidean pairwise data," Pattern Recognition, vol. 39, pp. 1815-1826, 2006.
- [4] S. -W. Kim and S. -H. Kim, "On optimizing dissimilarity-based classifications using a DTW and fusion strategies (in Korean)," Journal of the Institute of Electronics Engineers of Korea, vol. 47, no. CI-2, pp. 212-219, 2010.
- [5] E. Pekalska, R. P. W. Duin, and P. Paclik, "Prototype selection for dissimilarity-based classifiers," Pattern Recognition, vol. 39, pp. 189-208, 2006.
- [6] K. Riesen, V. Kilchherr, and H. Bunke, "Reducing the dimensionality of vector space embeddings of graphs," in Proc. of 5th Int'l Conf. on Machine Learning and Data Mining, vol. LNAI-4571 pp. 563-573, 2007.
- [7] S. -W. Kim, "An empirical evaluation on dimensionality reduction schemes for dissimilarity-based classifications," Pattern Recognition Letters, vol. 32, pp. 816 - 823, 2011.
- [8] S. -W. Kim, "Dissimilarity representations using  $l_p$ -norms in Eigen spaces," in Proc. of the 2015 Int'l Conf. on Image Processing, Computer Vision, Pattern Recognition, Las Vegas, Nevada, CSREA Press, pp. 242-248, 2015.
- [9] D. He and J. Wang, "Parallel computing of eigenvalue of doubly stochastic matrix," in Proc. of the 5th Int'l Conf. on Algorithms and Architecture for Parallel Processing, Beijing, China, pp. 355-358, 2002.
- [10] R. P. W. Duin, M. Bicego, M. Orozco-Alzate, S. -W. Kim, and M. Loog, "Metric learning in dissimilarity space for improved nearest neighbor performance," in Proc. of S+SSPR2014, Joensuu, Finland, LNCS8621, pp. 183-192, 2014.
- [11] N. Kwak, "Principal component analysis based on  $l_1$ -norm maximization," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 30, no. 9, pp. 1672-1680, 2008.

#### 저 자 소 개

김 상 운(정회원)-교신저자  
제 51권 제 10호 참조