

Hybrid Pattern Recognition Using a Combination of Different Features

Sang-Il Choi *

Abstract

We propose a hybrid pattern recognition method that effectively combines two different features for improving data classification. We first extract the PCA (Principal Component Analysis) and LDA (Linear Discriminant Analysis) features, both of which are widely used in pattern recognition, to construct a set of basic features, and then evaluate the separability of each basic feature. According to the results of evaluation, we select only the basic features that contain a large amount of discriminative information for construction of the combined features. The experimental results for the various data sets in the UCI machine learning repository show that using the proposed combined features give better recognition rates than when solely using the PCA or LDA features.

▶ Keyword : Pattern classification; Feature extraction; Feature selection; Hybrid method; Discriminant analysis; Combined features

I. Introduction

센서 및 인터넷 통신망의 발달로 인해 수집되고 전송되는 데이터의 양이 급속히 증가함에 따라, 방대한 양의 데이터들로부터 사용하고자 하는 용도에 적합한 정보를 추출, 가공하는 기술들이 개발되고 있다. 이러한 기술들은 홍채인식, 얼굴인식, 지문인식과 같은 생체 인식, 데이터 마이닝, 진단 시스템 등 다양한 패턴 인식 분야에서 활용되고 있다[1-3].

패턴 인식을 위해, 각각의 데이터 샘플은 입력특징(input feature) 또는 원시변수(primitive variable)로 불리는 성분들로 구성된 한 개의 벡터로 표현된다[1,3,4]. 입력 특징은 데이터 수집에 사용된 센서의 종류나 측정 내용에 따라 다양한 종류로 구성될 수 있다. 그런데, 해결하고자 하는 문제에 따라 문제 해결에 적합한 입력특징들과 그렇지 못한 특징들이 혼재하거나, 데이터를 수집하는 과정에서 발생한 노이즈 성분으로 인해 입력특징 값이 훼손되는 경우, 데이터 샘플들을 그대로 사용하면 인식 성능이 저하된다[1,5]. 따라서 많은 경우, 데이터의 분류를 보다 효과적으로 수행하기 위해 여러 종류의 특징 선택 방법 [5-10] 또는 특징 추출 방법[11-17]들을 사용한다.

특징 선택 방법은 입력특징들 중에서 문제 해결의 목적에 부합하는 특징들로만 구성된 입력특징의 부분 집합을 구성하는 것이다. 유지할 입력특징과 제거할 입력특징을 결정하기 위해 먼저 각 특징들이 문제해결에 적합한지 여부를 측정하는데, 측정 기준으로 엔트로피(entropy), 확률적 거리 측정자, 상호 정보(mutual information)와 같은 정보 이론적 측정자[17,18]를 이용하거나 판별 분석에 근거한 특징의 분별 능력을[7,8] 사용한다. 특징을 선별적으로 사용하는 것은 데이터의 시각화(visualization)와 해석, 고차원 데이터 측정과 저장 양의 축소, 처리시간 감소 면에서의 효율성과 함께 분류 성능을 향상시킬 수 있는 이점이 있다.

특징 선택 방법은 불필요한 입력특징을 제거하지만 선택된 입력특징은 그 값을 그대로 사용한다. 반면, 특징 추출 방법은 입력특징들로부터 새로운 특징들을 찾아낸다[14,19,20]. 특징 추출 과정은 보통 입력특징들의 선형결합(linear combination)과 같은 변환(transformation)에 의해 이루어지며, 데이터 샘플을 투영벡터(projection vectors)들을 이용하여 일종의 특징공간(feature space)에서 특징벡터(feature vector)로 표현한다.

• First Author: Sang-Il Choi, Corresponding Author: Sang-Il Choi

*Sang-Il Choi(choisi@dankook.ac.kr), Dept. of Computer Science and Engineering, Dankook University

• Received: 2015. 08. 21, Revised: 2015. 09. 14, Accepted: 2015. 10. 13.

• This research was supported by the MSIP(Ministry of Science, ICT and Future Planning), Korea, under the ITRC(Information Technology Research Center) support program (IITP-2015-R0992-15-1012) supervised by the IITP(Institute for Information & communications Technology Promotion) and also supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) the Ministry of Science, ICT & Future Planning (2015R1A1A1A05001065).

그동안, 주어진 데이터 샘플로부터 유용한 특징들을 찾아내기 위한 많은 방법들이 개발되어 왔다. 그 가운데 PCA (Principal Component Analysis)와 LDA(Linear Discriminant Analysis)는 대표적인 방법들로서, 이들을 기초로 하여 2DPCA (2-Dimensional PCA) [14], NLDA (Null space LDA) [15], DLDA (Direct LDA) [16]와 같이 다양한 형태로 변형된 방법들도 많이 제안 되어 왔다.

이와 같이, 패턴 인식 문제를 해결하기 위해 다양한 방법들이 제안되어 왔으나 이들 방법들은 구성하고자 하는 특징공간을 위한 목적함수 개념 및 수학적 특성으로 인해, 주어진 문제의 종류에 따라 나름대로의 장점 및 한계점을 가지고 있다. 예를 들면, PCA는 데이터 샘플들의 분포를 최대화 할 수 있도록 변환함으로써 데이터 샘플들 간의 구분이 쉽게 할 수 있는 장점이 있지만, 주성분(principal component)를 찾는 과정에서 클래스 정보를 사용하지 않는 기본 속성 상, 지도학습(supervised learning)에 의한 분류(classification) 문제보다는 최소 MSE(Mean Square Error) 관점에서의 데이터 표현(representation) 또는 시각화(visualization)에 보다 적합하다 [19]. 반면, LDA는 최적의 선형 분류 함수(linear discriminant function)를 찾는 데 초점을 맞추는 방법으로서, 학습(training) 데이터의 클래스 정보를 활용하여 클래스 내의 분포를 좁히고 클래스 간의 거리를 넓히는 특징 공간을 구성한다.

PCA와 LDA의 기본적인 속성으로만 비추어 볼 때, 데이터의 분류를 다루는 패턴 인식 문제에 있어서는 LDA가 PCA보다 일반적으로 더 적합하다고 볼 수 있다. 그러나 PCA에 대한 LDA의 우위는 모든 패턴 인식 문제에 적용되지는 않는다. LDA는 클래스들의 평균 값들의 차이가 큰 문제에 대해서는 좋은 성능을 보이는 반면, 목적함수의 특성 상 LDA의 특징 공간의 차원이 분류하고자 하는 클래스의 개수에 제한을 받기 때문에 이 분류(two-class classification) 문제와 같이 클래스의 개수가 적은 문제에서는 분류 성능에 한계를 보인다[21]. 이런 경우, 데이터의 분포 속성에 따라 PCA가 LDA보다 좋은 분류 성능을 보이기도 한다. 마찬가지로, PCA와 LDA 이 외의 여러 종류의 방법들도 풀고자 하는 문제의 성격과 데이터의 분포에 따라 성능의 편차를 보인다[22].

한 종류의 특징이 갖는 한계를 극복하기 위해 두 종류 이상의 특징들을 함께 이용하는 하이브리드 특징 추출 방법들도 개발되어 왔다. [23]에서는 얼굴인식을 위해 전체 얼굴 영상으로부터 추출한 진역 특징과 눈, 코, 입의 지역 특징들을 함께 사용하여 얼굴 인식을 위한 결합 특징 공간을 구성하였고, [24]에서는 LBP (Local Binary Pattern) 특징과 DTDWT (Dual-Tree Discrete Wavelet Transform)를 함께 사용하였다. 또는 특징마다 분류기를 만들어 여러 종류의 혼합 규칙을(fusion rule) 통해 [25] 여러 종류의 특징을 함께 사용할 수도 있다.

본 논문에서는 서로 다른 속성을 가지는 방법들로부터 추출한 특징들을 기본특징 집합으로 구성하여, 이 중 샘플들을 분류

하는 데에 적합한 특징들을 선별적으로 결합하여 분류 성능을 향상 시키는 복합(Hybrid) 패턴 인식 방법을 제안한다. 서로 다른 특징 추출 방법들의 시너지(synergy) 효과를 얻기 위해서는 개별 방법들의 성질이 상호 보완적이어야 하므로, 본 논문에서는 앞서 언급한 대표적인 특징 추출 방법인 PCA와 LDA를 기본특징들의 추출하는 데에 사용하였다. 기본특징 집합이 구성되면, 대표적인 특징 선택 방법 가운데 하나인 ReliefF 방법을 [8] 사용하여 샘플을 분류하기 위한 각 특징들의 유용성을 측정된다. 그런 다음, 특징들에 대한 측정 결과를 바탕으로 유용한 정보가 많은 특징들만으로 결합 특징을 구성하여 인식을 위한 최종 분류기의 입력으로 사용한다. UCI repository[26]에서 제공하는 여러 종류의 데이터 셋에 대한 분류 실험을 수행한 결과, PCA와 LDA를 각각 단독으로 사용하였을 때에는 데이터 셋의 종류에 따라 성능의 우위가 바뀌었으나, 제안한 결합 특징은 모든 데이터 셋에서 가장 우수한 분류 성능을 보여 주었다.

본 논문은 다음과 같이 구성되어 있다. 2장에서는 특징 추출을 위한 대표적인 방법으로서 PCA와 LDA에 대해 간략히 설명하고 각각의 특성을 비교한다. 3장에서는 판별 분석에 기반하여 PCA와 LDA를 이용해 구성된 기본특징들의 분별력을 측정하여 결합 특징을 구성하는 방법을 제안하고, 4장에서는 다양한 데이터 셋에 대한 분류 성능을 평가한다. 마지막으로 5장에서 결론을 맺는다.

II. Feature Extraction Methods

1. PCA (Principal Component Analysis)

N 개의 샘플로 구성된 데이터 셋(set)이 주어졌을 때, 각각의 샘플은 n 차원 공간에서 한 점($\mathbf{x}_i \in R^n$)으로 표현될 수 있다. PCA는 투영된 샘플들의 분산이 최대가 되게 하는 선형변환(linear transformation)을 찾는데, PCA의 목적 함수는 다음과 같이 정의된다[12].

$$W_{PCA} = \operatorname{argmax}_W |W^T S_T W| \quad (1)$$

여기서 $S_T = \sum_{i=1}^N (\mathbf{x}_i - \mu)(\mathbf{x}_i - \mu)^T$ 는 전체 산란행렬(total scatter matrix)이고 μ 는 전체 샘플의 평균이다. 투영행렬(projection matrix) $W_{PCA} (= [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_m])$ 는 n 차원의 입력 공간(input space)를 $p (\ll n)$ 차원의 특징공간으로 맵핑(mapping)하는 선형변환이다. W_{PCA} 를 구성하고 있는 투영 벡터(\mathbf{w}_i)들은 PCA 특징공간의 기저벡터들로서 원래의 입력공간(input space)에서 최대 분산이 되는 방향에 해당하며, 전체 산란행렬 S_T 의 고유벡터(eigen vector)로 사용하기 때문에 결국 유니모달(unimodal) 가우시안 가정 하에서 데이터들의 독립 축들(independent axes)을 찾게 된다. 이 때 p 개의 고유벡터는

해당 고유값(eigen value)이 큰 순서대로 선택한다. S_T 의 고유 벡터를 구하기에 앞서, 모든 데이터 샘플들은 $\|\mathbf{x}_i\| = 1$ 이 되도록 정규화(normalization)하며, 각각의 정규화 된 샘플들은 전체 정규화 된 샘플들의 평균을 빼줌으로써, 가장 큰 고유값을 가진 고유벡터들이 상관관계(correlation)의 관점에서 샘플들의 분산이 최대가 되는 고유공간의 차원을 나타내도록 한다 [22]. PCA에 의해 추출된 특징벡터

$$\mathbf{y}_k^{PCA} = W_{PCA}^T \mathbf{x}_k, k = 1, \dots, N, \text{ 으로 표현된다.}$$

2. LDA (Linear Discriminant Analysis)

LDA는 클래스 정보를 이용하는 지도학습 방법으로서, 클래스들을 서로 가장 잘 구분할 수 특징공간을 찾는다. C 개의 클래스를 가진 데이터 셋이 주어졌을 때, 클래스 간 산란행렬 (between class scatter matrix) S_B 와 클래스 내 산란행렬 (within class scatter matrix) S_W 를 아래와 같이 정의한다 [12].

$$\begin{aligned} S_B &= \sum_{i=1}^C N_i (\mu_i - \mu)(\mu_i - \mu)^T \\ S_W &= \sum_{i=1}^C \sum_{\mathbf{x}_k \in c_i} (\mathbf{x}_k - \mu_i)(\mathbf{x}_k - \mu_i)^T \end{aligned} \quad (2)$$

여기서 N_i 는 i 번째 클래스 c_i 에 속한 샘플의 개수, μ_i 와 μ 는 c_i 의 평균과 전체 평균이다.

LDA는 다음과 같은 목적함수를 만족시키는 투영행렬 W_{LDA} 를 구함으로써, 같은 클래스 내의 샘플들의 분산은 최소로 하고 동시에 클래스들의 평균들의 분산은 최대로 하는 특징 공간을 찾는다.

$$W_{LDA} = \underset{W}{\operatorname{argmax}} \frac{|W^T S_B W|}{|W^T S_W W|} \quad (3)$$

(3)에서 W_{LDA} 는 S_B 와 S_W 의 일반화 된 고유값(generalized eigenvalue)이 큰 순서대로 q 개의 일반화 된 고유벡터 (generalized eigenvector)들의 집합으로서, $S_W^{-1} S_B$ 의 고유벡터들에 해당한다. PCA에서와 같이 추출된 특징벡터 $\mathbf{y}_k^{LDA} = W_{LDA}^T \mathbf{x}_k, k = 1, \dots, N$ 으로 표현된다.

3. PCA versus LDA

PCA는 고차원의 데이터의 분포를 2차원 또는 3차원 공간에 효과적으로 시각화 하는데 유용한 반면, PCA의 대표적인 한계점으로는 샘플들의 클래스 정보를 사용하지 않기 때문에 PCA의 목적함수를 만족하는 변환을 찾는 과정에서 클래스의 분리성에 대해 고려를 하지 않는다는 점이다. PCA는 변환된 축 (axes)들이 최대 분산의 방향이 되도록 단지 좌표의 회전을 수

행하기 때문에 분류를 위한 좋은 특징의 추출을 보장하지는 않는다[22]. 반면, LDA는 목적함수 자체가 클래스 정보를 기반으로 분류를 목적으로 하고 있다. 일반적으로, 클래스 정보를 이용하는 지도학습 기반의 LDA가 비지도학습(unsupervised learning)을 기반으로 하는 PCA보다 분류 문제에서 더 우수하다고 여겨진다. 그림 1(a)에서와 같이 PCA의 목적함수는 전체 데이터의 분산이 가장 큰 방향의 벡터를 찾지만, LDA의 경우, 두 클래스의 분류에 가장 적합한 방향의 벡터를 찾는 것을 볼 수 있다. 그러나 이는 모든 분류 문제에서 적용되지는 않으며, 문제의 성격에 따라 PCA가 LDA보다 더 우수한 분류 성능을 보이는 경우도 많다.

LDA는 같은 클래스 내의 샘플들의 분산이 작고 클래스들 간의 분산이 큰 경우에는 좋은 성능을 보이지만, 그렇지 않은 경우에는 좋은 성능을 보이지 못한다[19,22]. 예를 들어, 그림 1(b)와 같이 각 클래스의 평균들이 비슷한 경우에는 LDA보다 오히려 PCA가 더 좋은 분류 성능을 나타낸다. 또한 LDA의 목적함수는 샘플들이 각 클래스에서 정규 분포를 갖고 있다는 것을 가정하기 때문에 클래스 당 샘플의 개수가 적거나, 샘플의 분포가 정규 분포와 상이한 경우에는 성능이 떨어질 수 있다.

이와 같이, 데이터 분포의 성격에 따라 PCA와 LDA의 분류 성능상의 우위가 바뀔 수 있는데, 고차원 데이터의 실제 분포를 확인하기는 어렵기 때문에 주어진 문제에 대해 PCA와 LDA 중 어떤 방법을 사용해야 할지를 선택하는 것은 어려운 문제이다. 따라서 본 논문에서는 PCA 또는 LDA 중 어떤 한 방법만을 사용하는 것이 아니라, PCA와 LDA의 특징들을 모두 추출하고 각각의 특징에 대해 샘플의 분별성을 평가한다. 평가 결과를 바탕으로, 분별력이 높은 특징들을 선별하여 결합 특징을 구성함으로써 PCA와 LDA의 장점을 모두 활용할 수 있는 방법을 제안한다.

III. Hybrid Pattern Classification

먼저, 학습 셋으로부터 PCA와 LDA의 투영행렬 $W_{PCA} \in R^{n \times m}$ 과 $W_{LDA} \in R^{n \times m}$ 을 이용하여 각각 특징벡터 $\mathbf{y}^{PCA} = [y_1^{PCA}, \dots, y_p^{PCA}]^T$ 와 $\mathbf{y}^{LDA} = [y_1^{LDA}, \dots, y_q^{LDA}]^T$ 를 추출하여 결합특징으로 사용될 기본특징 집합 $\mathbf{Y}^{BoI} = [Y_1, \dots, Y_{pq}]^T = ([y_1^{PCA}, \dots, y_p^{PCA}, y_1^{LDA}, \dots, y_q^{LDA}]^T)$ 을 구성한다.

각각의 기본특징들이 패턴을 분류하는데 얼마나 유용한지를 측정하기 위해 대표적인 특징 선택 방법인 Relief를 [8] 사용하였다. Relief 방법은 이 분류 문제를 위해 제안된 Relief 방법 [27]에서 확률 근사의 신뢰도를 개선하고 다 분류 (multi-class classification) 문제로 확장한 방법이다. Relief 계열의 방법들은 잡음의 영향을 크게 받지 않고, 특징들의 상호 작용들에 강인하다. 또한 문맥적인 정보들을 이용하기 때문에 특징들 사이에 의존성이 있는 문제들에 특히 유용하다[27]. 이

에, 각각의 기본특징 Y^{pool} 들에 대해 ReliefF 방법을 이용하여 같은 클래스에 속한 샘플들의 특징 값과 다른 클래스에 속한 샘플들의 특징 값들 간의 관계로부터 특징 값들이 샘플이 속한 클래스와 어떤 연관성이 있는지를 수치화 하고, 이로써 해당 특징이 분류에 얼마나 적합한지를 측정하는데 사용하였다.

각 특징의 값이 서로 가까이 있는 샘플들을 얼마나 잘 구별할 수 있는지를 평가하기 위해, 먼저 기본특징 Y^{pool} 들에 대한

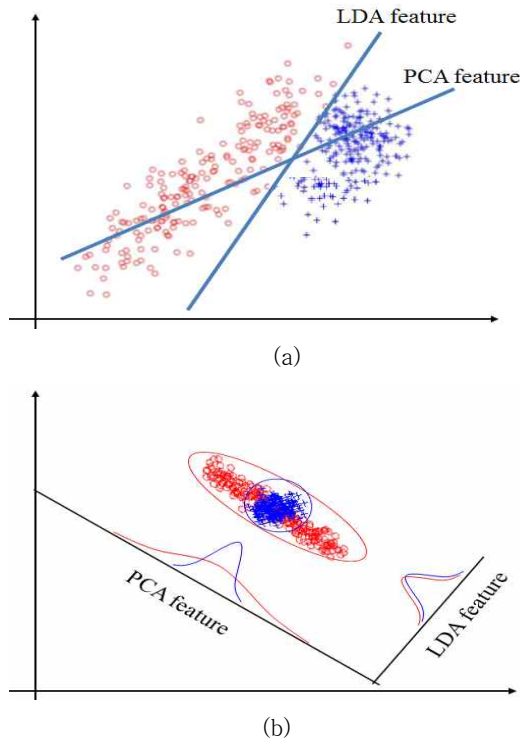


Fig. 1. Comparison of PCA and LDA

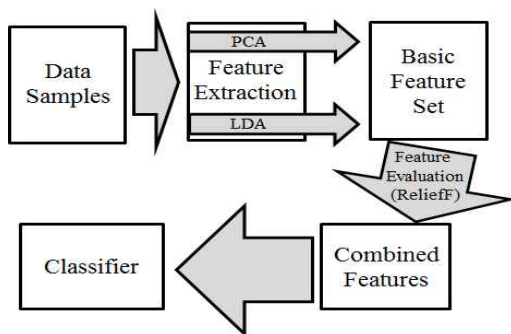


Fig. 2. Overall procedure of the proposed method

평가 결과를 기록할 측정 벡터 $S = [S_1, \dots, S_{pq}]^T$ 를 정의한다. 클래스 C_l 에 속한 샘플들의 개수를 N_{C_l} , 전체 샘플의 개수를 $N (= \sum_{l=1}^C N_{C_l})$ 이라 한다. 그런 다음, 임의로 선택한 k 번째 샘플

$\mathbf{x}_k = [x_{k1}, \dots, x_{k(pq)}]^T$ 에 대해 \mathbf{x}_k 와 같은 클래스(C_l)에 속한 샘플들 중에 가장 거리가 가까운 m 개의 샘플(\mathbf{x}_j^S)들의 집합을 $N = \{\mathbf{x}_1^S, \dots, \mathbf{x}_m^S\}$, $C-1$ 개의 다른 클래스(C_l)에 속한 샘플들 중에 \mathbf{x}_k 와 거리가 가까운 m 개의 샘플(\mathbf{x}_j^D)들의 집합을 $D_l = \{\mathbf{x}_1^{D_l}, \dots, \mathbf{x}_m^{D_l}\} (l \neq 1)$ 을 구성한다. 만약 \mathbf{x}_k 의 i 번째 특징 x_{ki} ($i = 1, \dots, pq$)가 N 에 속한 샘플들 $\mathbf{x}_j^S (j = 1, \dots, m)$ 의 특징 x_{ji}^S 와 다른 값을 갖는다면 i 번째 특징은 같은 클래스에 속한 두 샘플을 서로 분리하는 셈이 된다. 이는 클래스 분류에 적합하지 않으므로 i 번째 특징에 대한 측정 벡터의 성분 S_i 의 값을 작게 해야 한다. 반면 \mathbf{x}_k 의 i 번째 특징 x_{ki} ($i = 1, \dots, pq$)가 D_l 에 속한 샘플들 $\mathbf{x}_j^{D_l} (j = 1, \dots, m)$ 의 특징 $x_{ji}^{D_l}$ 와 다른 값을 갖는다면 i 번째 특징은 다른 클래스에 속한 두 샘플을 서로 분리하는 분류에 적합한 특징이므로 S_i 의 값을 증가시켜야 한다. 이러한 개념을 바탕으로 C 개의 클래스와 N 개의 샘플들로 구성된 학습 셋으로부터, 각 특징에 해당하는 측정치 $S_i, i = 1, \dots, pq$, 의 값은 다음과 같이 계산한다[27].

1) 측정 벡터의 모든 성분을 0으로 초기화 한다 ($S = \mathbf{0}$).

2) for $k = 1 : N$

학습 영상 \mathbf{x}_k 에 대해, N 과 $C-1$ 개의 D_l 를 각각 구성한다.

3) for $i = 1 : pq$

$$S_i = S_i - \sum_{j=1}^m \text{Diff}(Y_i, \mathbf{x}_k, \mathbf{x}_j^S) / (N \cdot m) + \sum_{l=1}^{C-1} \left[\frac{1 - P(C_l)}{1 - P(C_1)} \sum_{j=1}^m \text{Diff}(Y_i, \mathbf{x}_k, \mathbf{x}_j^{D_l}) \right] / (N \cdot m)$$

end

end

여기에서 $P(C_l)$ 은 클래스 C_l 의 사전확률(prior probability)이고, $\text{Diff}(Y_i, \mathbf{x}^A, \mathbf{x}^B)$ 는 \mathbf{x}^A 와 \mathbf{x}^B 의 i 번째 특징들의 차이($|x_i^A - x_i^B|$)이다.

4) 2),3) 과정을 N 번 반복하여 계산된 최종 S 를 가지고 S_i 의 값이 큰 특징들만을 선택하여 결합 특징으로 사용한다. 제안한 결합 특징 구성 방법의 전체 흐름은 Fig. 2에 나타나 있다.

IV. Experimental Results

제안한 방법의 효과를 확인하기 위해 UCI machine learning repository의 여러 데이터 셋에 대해 인식률을 실험해 보았다. 각 데이터 셋의 특성은 표 1에 나타나 있다. 각각의 데이터 셋에 대해 PCA와 LDA를 각각 사용하여 추출한 특징들을 단독으로 사용했을 때의 인식률과 특징 선택 과정 없이 PCA 특징과 LDA 특징을 연결하여 결합 특징을 이용했을 때 (Com(P+L)) [23], 그리고 제안한 방법을 이용하여 PCA, LDA

특징들 중 유용한 특징들만 선별하여 구성된 결합 특징을 사용하였을 때의 인식률을 비교해 보았다. PCA와 LDA를 이용하여 특징을 추출하기에 앞서, 학습 데이터는 학습 데이터 셋의 평균과 표준 편차를 이용하여 각각의 데이터 샘플을 구성하고 있는 입력특징들이 영 평균(zero mean), 단위 표준 편차(unit standard deviation)이 되도록 정규화(normalization) 하였고, 테스트 샘플에 대해서도 동일한 평균과 표준 편차 값으로 정규화하여 실험하였다. 분류기로는 최근접 이웃 분류기(nearest neighborhood rule)을 사용하였으며, 두 샘플 간의 거리의 측정자로는 l_2 노름(norm)을 사용하였다[1]. 인식률 산출 방법은 10-fold cross validation을 적용하였다[28]. 이러한 과정을 5차례 수행하여 그 평균 인식률을 표 2에 나타내었다.

표 2의 PCA와 LDA의 결과를 보면, 데이터 셋에 따라 성능의 우위가 바뀌는 것을 볼 수 있다. 분류 문제임에도 불구하고 Ionosphere, Sonar, Bupa 데이터 셋의 경우는 PCA가 LDA보다 작게는 3.1%부터 크게는 14.1%까지 높은 인식률을 보여주었다. 반면, Monk3, Heart disease, Wine, Iris 데이터 셋의 경우는 LDA가 PCA보다 작게는 0.7%부터 크게는 18.2%까지 높은 인식률을 나타내었다. 이와 같은 PCA와 LDA의 성능편차는 각 데이터 셋의 클래스별 샘플의 분포 및 클래스 개수 등의 데이터 셋의 구조와 관련이 있다. 그림 3과 그림 4는 그림 3은 PCA와 LDA 특징 공간에서의 Wine 데이터 샘플들과 Sonar 데이터 샘플들의 분포를 보여준다.

PCA 특징 공간은 Wine과 Sonar 데이터 셋의 경우 각각 13차원 60차원의 특징 공간을 구성할 수 있으나, 시각화를 위해 2차원, 3차원의 특징 공간에 샘플들을 도시하였다. 그림 3을 보면, Wine 데이터 샘플의 경우는 2차원의 LDA 특징 공간에서 클래스별로 클러스터링이 비교적 잘 되고 클래스 간에 겹침이 거의 발생하지 않는 것을 볼 수 있다.

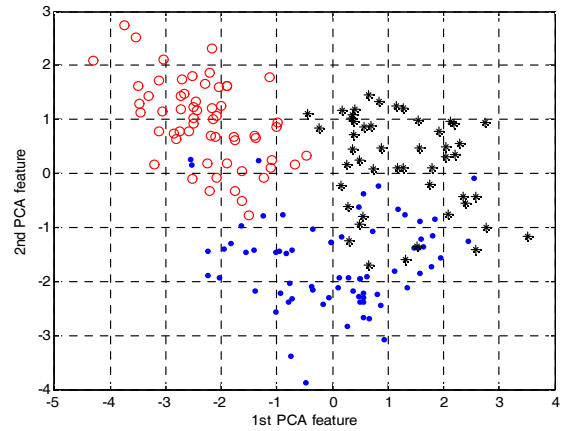
반면, PCA 특징 공간의 경우, 서로 다른 클래스에 속한 일부 샘플들이 서로 겹치는 것을 볼 수 있다. 한편, Sonar 데이터 셋의 클래스 개수가 2개 이므로 LDA 특징 공간의 차원은 1차원에 제한된다. 그림 4(b)와 같이 각 클래스의 샘플들은 1차원 LDA 특징 공간에서 서로 겹침이 발생하는 것을 볼 수 있다.

Table 1. Data sets from UCI machine learning repository used in the experiments

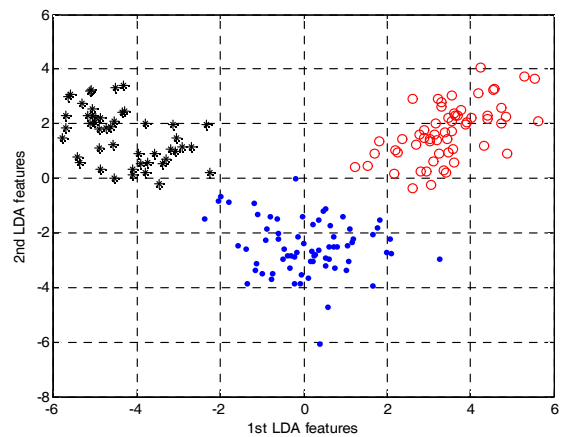
Data set	# of Input Features	# of Instances	# of classes
Ionosphere	33	351	2
Sonar	60	208	2
Bupa	6	345	2
Monk3	6	432	2
Heart disease	13	297	2
Wine	13	178	3
Iris	4	150	3

Table 2. Recognition rates for UCI data sets

Data set	PCA	LDA	Com(P+L)	Proposed Method
Ionosphere	86.6	83.5	88.7	93.8
Sonar	87.2	73.1	84.0	90.8
Bupa	63.3	60.6	66.6	71.0
Monk3	68.1	86.2	73.0	96.7
Heart disease	75.0	75.7	74.4	84.2
Wine	95.4	99.0	98.3	99.8
Iris	94.9	95.7	95.7	96.9



(a)



(b)

Fig. 3. Distribution of Wine data samples in the feature space: (a) PCA feature space. (b) LDA feature space.

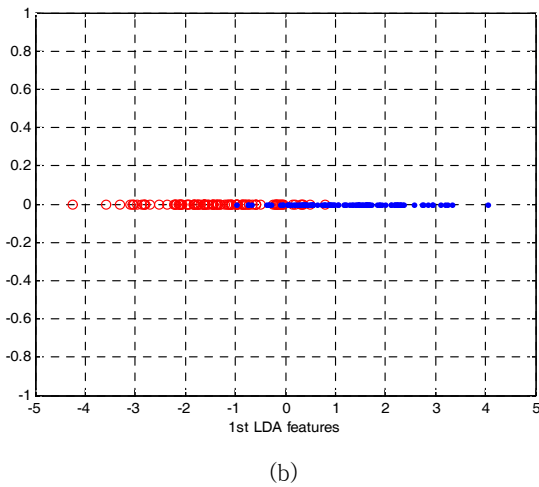
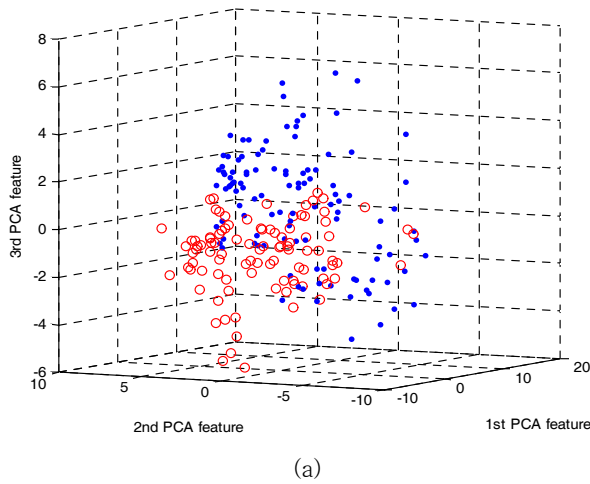


Fig. 4. Distribution of Sonar data samples in the feature space: (a) PCA feature space. (b) LDA feature space.

반면, 그림 4(a)의 고차원의 PCA 특징 공간에서는 데이터 샘플들이 서로 겹침이 없이 분포할 수 있음을 볼 수 있다. 이와 같은 특성은 표 2의 결과들에서도 확인할 수 있는데, PCA의 경우는 주로 이 분류 문제(Ionosphere, Sonar, Bupa 데이터 셋)에서 LDA보다 좋은 성능을 보이는 경향을 보여 주었다. 이는 LDA의 경우 1차원의 특징 공간만을 만들 수 있는 반면, PCA는 고차원의 특징 공간을 구성할 수 있기 때문으로 분석된다(Cover's rule[29]).

그러나 Cover's rule만으로 PCA와 LDA의 성능 차이에 대해 분석을 하는 데에는 한계가 있다. Monk3와 Heart disease 데이터 셋의 경우는 이 분류 문제임에도 불구하고 LDA가 더 좋은 성능을 나타내는 것을 볼 수 있는데, 이는 Monk3와 Heart disease 데이터 셋의 경우, 클래스 별로 입력 특징들의 평균 값이 큰 차이가 나는 등의 다른 샘플 분포의 특성들이 특징 공간의 차원보다 더 큰 영향을 미쳤기 때문이다. 그러나 데이터들의 구조가 어떠한지를 정확하게 파악하는 것은 매우 어려운 일이다. 따라서 주어진 문제(데이터 셋의 종류)에 대해 어

떤 한 가지 방법을 사용하는 것은 효율적이지 못하다.

표 2에서, 제안한 결합 특징을 사용하였을 때 모든 데이터 셋에서 PCA 또는 LDA 특징을 단독으로 사용했을 때보다 작게는 Wine 데이터 셋에서 0.8%(Wine 데이터 셋의 경우 이미 LDA의 성능이 포화성능(saturated)을 나타내었다)에서 많게는 Monk 3 데이터 셋에서 10.5%까지 더 높은 인식률을 보여 주었다. 주목할 점은 PCA 특징과 LDA 특징을 단순히 결합했을 때는 PCA와 LDA의 성능의 중간 성능을 보이거나 소폭의 성능(Ionosphere, Bupa 데이터 셋) 향상을 보였다는 점이다. 반면, 각 특징들의 분별력을 측정하여 이를 바탕으로 유용한 특징들만을 선택하여 결합한 경우(제안한 방법) PCA 특징과 LDA 특징의 장점들이 서로 시너지(synergy) 효과를 보여 결과적으로 인식 성능을 크게 향상 시킨 것을 확인할 수 있다.

V. Conclusions

본 논문에서는, 서로 다른 성질을 가지고 있는 특징들을 효과적으로 결합하여 데이터 분류의 성능을 높이는 하이브리드 패턴 인식 방법을 제안하였다. 패턴인식을 위한 여러 종류의 특징 추출 방법들은 각각 나름대로의 특성이 있기 때문에 데이터의 구조와 해결하고자 하는 문제에 따라 성능의 차이가 발생할 수 있다. 이에 제안한 방법은 어느 한 방법만을 사용하는 것이 아니라 다른 종류의 특징들로부터 각각 유용한 정보들만을 선별적으로 사용함으로써 각 특징 추출 방법의 효과를 극대화하였다. 이를 위해, 패턴 인식에서 대표적으로 널리 사용되는 PCA와 LDA의 특징들로 기본특징 집합을 구성하고 각각의 특징들의 분별력을 측정하였으며, 측정 결과를 바탕으로 분류에 유용한 특징들만으로 결합 특징을 구성하여 분류기에 사용하였다. 또한 제안한 방법에서는 PCA와 LDA가 병렬로 진행되기 때문에 한 가지 방법을 사용하는 경우와 비교했을 때 총 연산 시간은 ReliefF 수행 시간만 증가하는데, 차원 축소 방법인 PCA와 LDA는 저차원의 특성을 추출하므로 ReliefF는 실시간 수준으로 수행될 수 있다. 여러 종류의 UCI machine learning repository 데이터 셋에 대한 실험 결과, 제안한 결합 특징은 PCA와 LDA를 단독으로 사용했을 때와 두 종류의 특징을 단순히 결합하여 사용했을 때보다 우수한 인식 성능을 보여 주었다.

향후, PCA와 LDA 이외에도 여러 종류의 특징들을 효과적으로 결합하여 사용한다면 보다 다양한 패턴 인식 문제들을 해결하는 데 도움이 될 것으로 기대한다.

REFERENCES

[1] S.-I. Choi, J. Oh, C.-H. Choi, and C. Kim, "Input variable selection for feature extraction in classification problems," *Signal Processing*, Vol. 92, No. 3, pp. 636-648, March 2011.

- [2] S.T. Jung, "Robust Extraction of Facial Features under Illumination Variations," *Journal of the Korea Society of Computer and Information*, Vol. 10, No. 6, pp. 1-8, December 2005.
- [3] C. Kim and C.-H. Choi, "A discriminant analysis using composite features for classification problems," *Pattern Recognition*, Vol. 40, No. 7, pp. 2118-2125, November 2007.
- [4] G.Y. Heo, H. Choi, and J.S. Youn, "Supervised Rank Normalization with Training Sample Selection," *Journal of the Korea Society of Computer and Information*, Vol. 20, No. 1, pp. 21-28, January 2015.
- [5] N. Kwak and C.-H. Choi, "Input feature selection by mutual information based on parzen window," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 12, pp. 1667-1671, December 2002.
- [6] S.-I. Choi, C.-H. Choi, G.-M. Jeong, and N. Kwak, "Pixel selection based on discriminant features with application to face recognition," *Pattern Recognition Letters*, Vol. 33, No. 9, pp. 1083-1092, July 2012.
- [7] J. Liang, S. Yang, and A. Winstanley, "Invariant optimal feature selection: a distance discriminant and feature ranking based solution," *Pattern Recognition*, Vol. 41, No. 5, pp. 1429-1439, May 2008.
- [8] I. Kononenko, E. Šimec, and M. Robnik-Šikonja, "Overcoming the myopia of inductive learning algorithms with RELIEFF," *Applied Intelligence*, Vol. 7, No. 1, pp. 39-55, January 1997.
- [9] X. He, D. Cai, and P. Niyogi, "Laplacian score for feature selection," *Advances in Neural Information Processing Systems*, pp. 507-514, 2005.
- [10] G.Y. Heo, C.S. Park, and C.W. Lee, "Training Sample and Feature Selection Methods for Pseudo Sample Neural Networks," *Journal of the Korea Society of Computer and Information*, Vol. 18, No. 4, pp. 19-26, March 2013.
- [11] M. Turk, and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, Vol. 3, No. 1, pp. 71-86, 1991.
- [12] P.N. Belhumeur, J.P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, pp. 711-720, July 1997.
- [13] C. Kim, S.-I. Choi, M. Turk, and C.-H. Choi, "A New Biased Discriminant Analysis Using Composite Vectors for Eye Detection," *IEEE Trans. on Systems, Man, and Cybernetics, Part B: Cybernetics*, Vol. 42 No. 4, pp. 1095-1106, August 2012.
- [14] J. Yang, D. Zhang, A.F. Frangi, and J.Y. Yang, "Two-dimensional PCA: a new approach to appearance-based face representation and recognition," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 26, No. 1, pp. 131-137, January 2004.
- [15] H. Cevikalp, M. Neamtu, M. Wilkes, and A. Barkana, "Discriminative common vectors for face recognition," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 27, No. 1, pp. 4-13, January 2005.
- [16] H. Yu and J. Yang, "A direct LDA algorithm for high-dimensional data — with application to face recognition," *Pattern Recognition*, Vol. 34, No. 10, pp. 2067-2070, October 2001.
- [17] J. Doak, "An evaluation of feature selection methods and their application to computer security," *CSE Technical Report 92-18*, University of California, Davis, 1992.
- [18] N. Kwak, and C.-H. Choi, "Input feature selection for classification problems," *IEEE Trans. on Neural Networks*, Vol. 13, No. 1, pp. 143-159, January 2002.
- [19] S.-I. Choi, "Face Recognition Based on 2D Images Under Various Conditions," Ph.D. Thesis, Seoul National University, 2010.
- [20] N. Kwak, "Feature selection and extraction based on mutual information for classification," Ph.D. Thesis, Seoul National University, 2003.
- [21] N. Kwak, and J. Oh, "Feature extraction for one-class classification problems: Enhancements to biased discriminant analysis," *Pattern Recognition*, Vol. 42, No. 1, pp. 17-26, January 2009.
- [22] A.M. Martínez, and A.C. Kak, "PCA versus LDA," *Pattern Recognition*, Vol. 23, No. 2, pp. 228-233, February 2001.
- [23] C. Kim, J. Oh, and C.-H. Choi, "Combined subspace method using global and local features for face recognition," in *Proc. of International Joint Conference on Neural Networks*, pp. 2030-2035, 2005.
- [24] J. Ravi, and K.B. Raja, "Hybrid domain based face recognition system," *International Journal of Advanced Networking and Applications*, Vol. 3, No.

- 6, pp. 1402-1408, 2012.
- [25] J. Kittler, M. Hatef, R.P. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 3, pp. 226-239, March 1998.
- [26] Repository of Machine Learning Databases, <http://www.ics.uci.edu/mlearn/MLRepository.html>
- [27] M. R. Sikonja and I. Kononenko, "Theoretical and Empirical Analysis of ReliefF and RReliefF," *Machine Learning*, Vol. 53, No. 1, pp. 23-69, October 2003.
- [28] S.-I. Choi, and G.-M. Jeong, "A discriminant distance based composite vector selection method for odor classification," *Sensors*, Vol. 14, No. 4, pp. 6938-6951, 2014.
- [29] T.M. Cover, "Geometrical and Statistical properties of systems of linear inequalities with applications in pattern recognition," *IEEE Trans. on Electronic Computers*, Vol. EC-14, No. 3, pp. 326-334, June 1965.

Authors



Sang-Il Choi received the B.S. degree in the division of electronic engineering from Sogang University, Seoul, Korea, in 2005 and the Ph.D. degree from the School of Electrical Engineering and Computer Science, Seoul National University, Seoul, in 2010.

He was a Postdoctoral Researcher in the BK21 Information Technology, Seoul National University, in 2010 and in the Institute for Robotics and Intelligent Systems of Computer Science Department, University of Southern California, Los Angeles, until August of 2011. He is currently an Assistant Professor with the Department of Computer Science and Engineering, Dankook University, Gyeonggi-do, Korea. His research interests include pattern recognition, feature extraction and selection, machine learning, computer vision, and their applications.