

SpPCA와 MLP에 기반을 둔 융합법칙에 의한 MRS 패턴분류

정회원 송창규*, 이대종*, 전병석**, 유정웅*

MRS Pattern Classification Using Fusion Method based on SpPCA and MLP

Chang kyu Song*, Dae jong Lee*, Byeong seok Jeon**, Jeong woong Ryu* *Regular Members*

요 약

본 논문에서는 SpPCA와 MLP에 기반을 둔 융합법칙에 의한 MRS 패턴분류기법을 제안한다. 차원축소를 위해 사용되는 기존의 PCA 기법은 입력데이터가 비선형 특성을 갖는 경우 최적의 변환행렬을 구할 수 없다는 문제점을 가지고 있다. 따라서, 본 논문에서는 구간별로 입력데이터를 분할한 후 PCA에 의해 특징을 추출하는 SpPCA 기법을 이용하여 입력패턴의 차원을 축소한다. 다음 단계인 분류단계에서는 MLP 비선형분류기를 이용하여 구간마다 추출된 특징벡터를 이용하여 기준패턴과의 유사도를 산출한다. 최종 분류단계에서는 MLP에 의해서 산출된 유사도에 기반을 둔 융합법칙에 의하여 MRS 패턴을 분류한다. 제안된 방법의 유용성을 보이기 위한 실험결과에서 기존의 방법들에 비해서 향상된 인식결과를 보임을 확인하였다.

Key Words : PCA, SpPCA, MLP, MRS pattern classification

ABSTRACT

In this paper, we propose the MRS pattern classification techniques by the fusion scheme based on the SpPCA and MLP. A conventional PCA technique for the dimension reduction has the problem that it can't find a optimal transformation matrix if the property of input data is nonlinear. To overcome this drawback, we extract features by the SpPCA technique which use the local patterns rather than whole patterns. In a next classification step, individual classifier based on MLP calculates the similarity of each class for local features. Finally, MRS patterns is classified by the fusion scheme to effectively combine the individual information. As the simulation results to verify the effectiveness, the proposed method showed more improved classification results than conventional methods.

I. 서 론

자기공명 스펙트럼 분석기에 의해서 취득된 MRS (magnetic resonance spectra) 생체신호는 높은 입력 차원을 가지고 있을 뿐만 아니라, 데이터들간 유사성이 매우 높기 때문에 패턴을 분류하는데 있어서

많은 어려움을 가지고 있다^[1]. 또한, 입력차원이 높은 신호를 데이터의 축소 없이 패턴을 분류할 경우 데이터 저장용량의 증가 및 많은 계산처리에 의한 속도 저하를 초래한다. 이러한 문제점으로 인하여 입력차원이 높을 경우 중요한 특성만을 추출하는 다양한 특징선택 또는 특징추출기법이 제안되고 있다^{[2],[3]}.

* 충북대학교 전기전자컴퓨터공학부 (sckyu@chungbuk.ac.kr, leebigbell@hotmail.com, jwryu@vod.chungbuk.ac.kr)

** (주)세화폴리텍 (koamma88@chungbuk.ac.kr)

논문번호 : KICS2005-07-302, 접수일자 : 2005년 07월 25일

특징추출기법으로는 영상처리에서 많이 적용되는 주성분분석기법(principal component analysis; PCA)과 선형판별분석기법(linear discriminant analysis; LDA) 등이 있다^{[4]-[7]}. 선형판별분석기법은 분류를 위해 최적의 투영행렬을 찾는 방법 중의 하나로 투영된 분산행렬을 최대화하는 투영행렬을 찾기보다는 클래스간 분산(between-class scatter matrix) 행렬과 클래스내 분산(within-class scatter matrix) 행렬의 비가 최대가 되는 최적의 투영행렬을 찾는 것이다. 또한, 클래스간 데이터가 유사한 분산행렬을 갖지만 평균값이 상이한 경우에 대해서 최적의 해를 구할 수 있다. 그러나, 본 논문에서 고려되는 MRS 패턴과 같은 고차원의 데이터들이 위의 조건을 만족하는 경우는 극히 한정되어 있다.

PCA는 다변수 선형 데이터 분석에서 잘 알려진 대표적인 기법이며 주된 개념은 데이터의 최대 분산 방향을 나타내는 상호직교 기저벡터의 집합을 찾는 것을 목적으로 한다. 구해진 기저벡터행렬을 이용하여 입력공간을 저차원의 특징공간으로 선형적으로 투영시킨다. 이 방법은 차원축소에는 효과적으로 사용되고 있으나 투영된 데이터간의 겹치는 현상으로 인하여 분류문제에서는 효과적이지 못하다. 또한, 기존의 주성분분석기법은 모든 입력차원의 선형성에 기반을 두고 있기 때문에 입력데이터가 구간적으로 선형 특성을 보이지만 모든 차원에 대해서 비선형 특성을 갖는 경우 최적의 변환행렬을 구할 수 없는 문제점을 지니고 있다^[8].

이러한 문제점으로 인해서 전구역 데이터를 구간마다 분할하여 특징을 추출하는 SpPCA(sub-pattern based on PCA)와 Modular PCA 등이 제안되었다^{[9]-[11]}. 이 방법들은 기존의 PCA와 같이 데이터의 모든 차원을 이용하기 보다는 데이터를 균등하게 몇 구간으로 분할한 후, 구간별로 PCA에 의해 특징을 추출하는 방법이다. 이와 같이 구간별로 구분하여 특징을 추출하기 때문에 최종 분류를 위해선 구간별 특징벡터를 효과적으로 융합하여 인식률을 향상시킬 수 있는 융합분류기의 설계가 요구된다. 기존연구에서는 구간별로 추출한 특징벡터를 기준벡터로 설정한 후 검증하고자 하는 입력벡터와의 비교를 위해 k-NN(k-nearest neighbor) 선형분류기를 적용하였다. 그러나, k-NN 분류기의 경우 동일한 클래스 내의 특징벡터가 선형성을 갖는 경우 우수한 인식성능을 보이지만 비선형 특성을 갖는 경우에는 k-NN 분류기와 같은 선형분류기는 효과적이지 못하다.

따라서 본 논문에서는 비선형 분류기로서 널리 사용되고 우수한 특성을 보이는 다층신경망을 적용한다. 기존의 신경망은 차원의 저주 특성으로 인하여 입력차원이 높은 패턴에는 사용하는데 문제점이 있다. 이를 위해서 본 논문에서는 SpPCA 기법을 이용하여 차원을 축소한 후 최종 분류를 위해 MLP(multi-layer perceptron) 비선형분류기를 사용한다. 또한 SpPCA를 이용하여 구한 특징벡터를 융합하는 문제에 있어서 기존 연구는 전구역 특징벡터(global feature)에 기반을 둔 방법을 채택한 반면에 본 논문에서는 가중치합법칙(sum rule), 우수후보선출법칙(best candidate vote rule), 다수선출법칙(majority vote rule) 등 다각적인 융합법칙을 적용하여 패턴 특성에 맞는 결정법칙을 제안한다.

본 논문의 구성은 다음과 같다. 2장에서는 제안된 SpPCA와 MLP에 기반을 둔 융합법칙에 의한 MRS 패턴기법에 대하여 기술하고, 3장에서는 Wine^[9]과 MRS^[12] 데이터베이스에 각각 제안된 방법을 이용함으로써 적용된 방법의 우수성을 보인다. 마지막으로 4장에서는 결론을 제시한다.

II. SpPCA와 MLP에 기반을 둔 융합 분류기

PCA 기법은 다변수 선형 데이터 해석에서 잘 알려진 기법이다. PCA 기법은 데이터의 분산이 최대화 되는 변환행렬을 구하고 이 변환행렬에 의해 고차원의 입력공간을 저차원의 공간으로 축소하는 기법으로 선형 특징추출과 데이터 압축 등에 널리 사용되고 있다.

상세한 설명을 위해서 다음과 같은 N 개의 샘플을 갖는 학습데이터를 고려하자.

$$Z = (z_1, z_2, \dots, z_N) \tag{1}$$

여기서, $z_i = [z_i^1, z_i^2, \dots, z_i^d]^T$ 는 d 차원을 갖는 데이터이며 공분산행렬은 다음과 같다.

$$R = \frac{1}{N} \sum_{i=1}^N (z_i - \bar{m})(z_i - \bar{m})^T \tag{2}$$

$$\bar{m} = \frac{1}{N} \sum_{i=1}^N z_i \tag{3}$$

다음으로 공분산 R 의 고유값과 고유벡터를 계산한다. $W = (w_1, w_2, \dots, w_r)$ 를 r 개의 가장 큰 고유값에 대응하는 r 개의 고유벡터라 하자. 그러면 원래의

샘플 Z 의 집합에 대해서 그에 대응하는 특징벡터 $Y = (y_1, y_2, \dots, y_N)$ 는 다음과 같이 Z 를 PCA-변환 공간으로 투영함으로써 얻어진다.

$$y_i = W^T(z_i - \bar{m}) \quad (4)$$

여기서, W 가 $p \times r$ 행렬이고 z_i 가 i 번째 $p \times 1$ 벡터 이면 y_i 는 i 번째 $r \times 1$ 벡터이다.

SpPCA는 전구역 패턴보다는 구간마다 나누어진 분할패턴을 사용한다. SpPCA를 적용하기 위해서 원래의 데이터 $Z = (z_1, z_2, \dots, z_N)$ 는 다음과 같이 동일한 크기의 부패턴들의 집합으로 분할된다.

$$Z_j = (z_{j1}, z_{j2}, \dots, z_{jN}) \quad (5)$$

여기서, j 는 패턴지수이고, $j = 1, 2, 3, \dots, L$ (L 은 분할의 수)이다. 다음으로 식 (6)과 (7)에 의해서 j 번째 분할에 대한 공분산을 계산한다.

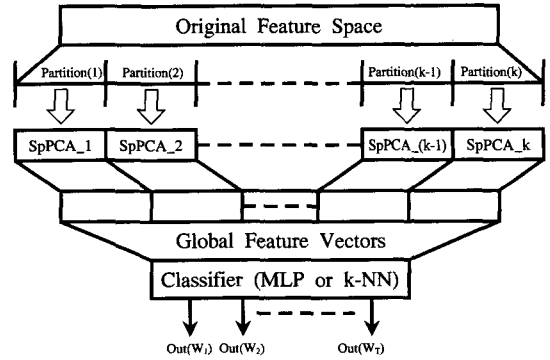
$$R_j = \frac{1}{N} \sum_{i=1}^N (z_{ji} - \bar{m}_j)(z_{ji} - \bar{m}_j)^T \quad (6)$$

$$\bar{m}_j = \frac{1}{N} \sum_{i=1}^N z_{ji} \quad (7)$$

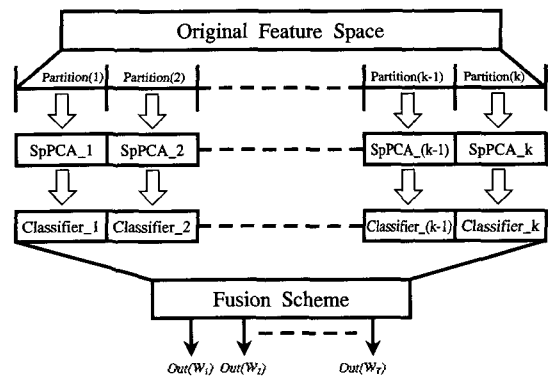
일반적인 PCA기법을 이용한 특징추출 과정과 마찬가지로 j 번째 분할에 대한 r 개의 가장 큰 고유값에 대응하는 r 개의 고유벡터 $W_j = (w_{j1}, w_{j2}, \dots, w_{jr})$ 를 결정한다. 마지막으로 식 (8)에서 나타낸 바와 같이 구간마다 나누어진 분할패턴에서 특징벡터를 구할 수 있다.

$$y_{ji} = W_j^T(z_{ji} - \bar{m}_j) \quad (8)$$

따라서, j 번째 분할 $Y_j = (y_{j1}, y_{j2}, \dots, y_{jN})$ 은 부패턴 데이터에서 얻어질 수 있다. 패턴 분류를 위해 전구역 특징 그리고 구간 특징과 같은 두 형태의 특징 조합을 고려할 수 있다. 여기서 전구역 특징은 $Y = (Y_1, Y_2, \dots, Y_L)^T$ 와 같이 각각의 부패턴의 특징을 조합함으로써 얻어진다. 구간 특징은 단지 식 (8)에 의해 얻어진다. 그림 1은 특징 조합에 따른 의사결정 구조를 나타낸다. 여기서 선형 분류기로 k-NN을 비선형 분류기로 MLP를 적용한다. k-NN 선형분류기를 위한 학습과 검증데이터 사이의 유사도를 계산하기 위해서 표 1의 거리척도를 사용한다. 또한 MLP 비선형분류기는 입력층, 은닉층 그리고



(a) 전구역 특징을 이용한 결정법칙



(b) 구간특징을 이용한 결정법칙

그림 1. 전구역 특징 및 구간특징에 기반을 둔 결정법칙.

표 1. 유사도를 위한 거리척도 방법.

Distance measure	Mathematical description
Euclidean	$\ x - y\ = \sqrt{\sum_{i=1}^n x_i - y_i ^2}$
Hamming	$\ x - y\ = \sum_{i=1}^n x_i - y_i $
Mahanobis	$\ x - y\ = (x - y)^T \Sigma^{-1} (x - y),$ Σ^{-1} : Inverse covariance

출력층의 3층으로 설계된다. 여기서 입력수는 PCA에 의해 계산된 특징의 차원과 같다. 또한 전달함수로 기호함수(logistic function)를 사용한다.

본 논문에서 사용된 의사결정구조는 다음과 같다. $\Omega = \{w_1, w_2, \dots, w_c\}$ 는 c 개의 클래스 라벨의 집합이라 하고, $D_i = [d_1(O_i), d_2(O_i), \dots, d_c(O_i)]$ 는 c -차원 벡터를 갖는 분류기 출력이라 하자. 여기서, Q_i 는 i 번째 입력 특징이다. 전구역 특징에 기반을 둔 의사결정에 따라서 각 클래스에 대한 훈련데이

터와 입력된 검증데이터 사이의 가장 우수한 유사도를 계산할 수 있다. 마지막으로, 분류는 식 (9)에 의해서 클래스들 사이의 가장 우수한 유사도를 갖는 클래스를 선택함으로써 얻어진다.

$$i_o = \begin{cases} \operatorname{argmin}_i (d_i) & \text{for } k\text{-NN}, \quad \forall i, i = 1, 2, \dots, c \\ \operatorname{argmax}_i (d_i) & \text{for } k\text{-NN}, \quad \forall i, i = 1, 2, \dots, c \end{cases} \quad (9)$$

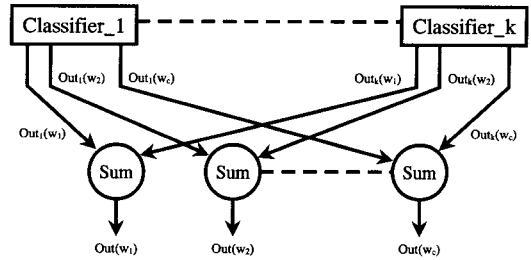
전구역 특징에 기반을 둔 의사결정과 달리 각각의 개별 분류기는 분할영역에 대해서 각 클래스의 유사도를 계산하기 때문에 구간 특징에 기반을 둔 분류기는 정보의 다중소스를 갖는다. 각각의 정보를 효과적으로 조합할 수 있는 많은 융합구조가 있지만, 본 논문에서는 가중치합법칙, 우수후보선출법칙과 다수선출법칙의 세 가지 융합법칙을 적용한다. 가중치합 법칙에 기반을 둔 의사결정은 모든 분할 영역에 대한 각 클래스의 유사도의 합으로 계산된 출력값들 중에서 가장 우수한 유사도를 갖는 클래스를 선택함으로써 수행된다. 그리고 우수후보 선출 방법은 모든 분할영역에 대한 각 클래스의 유사도들 사이의 가장 우수한 유사도를 갖는 클래스를 선택함으로써 수행된다. 또한 다수선출법칙은 각각의 분할영역에 대해서 가장 우수한 유사도를 갖는 클래스의 개수의 합으로 계산된 지수들 중에서 최대 지수 값을 갖는 클래스를 선택함으로써 수행된다.

III. 실험 결과

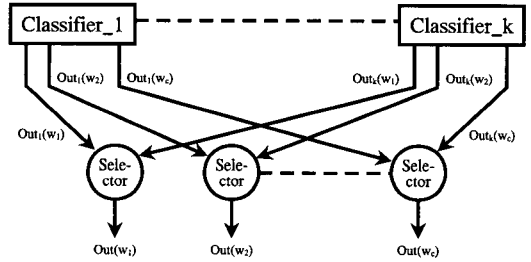
3.1 Wine 데이터

제안된 방법의 타당성을 보이기 위해 논문에서 자주 언급되는 벤치마크 데이터중 하나인 wine 데이터를 이용하였다. 이 데이터는 UCI 머신 러닝 센터에서 구축된 데이터베이스이며 관련 사이트에서 쉽게 얻을 수 있다(<http://www.sgi.com/tech/mlc/db>). Wine 데이터는 3개의 클래스를 가지며 각각의 클래스당 59, 71, 48개의 패턴들로 구성되어 있고 각각의 데이터는 20의 차원을 갖는다. 이중 훈련에 사용된 데이터는 클래스당 각각 30, 35, 24개이며 나머지 데이터는 제안된 방법의 성능을 분석하는데 이용하였다. 본 실험은 보다 정확한 결과의 해석을 위해서 학습데이터와 검증데이터를 임의로 10번씩 선택하여 총 10회의 실험을 하였다.

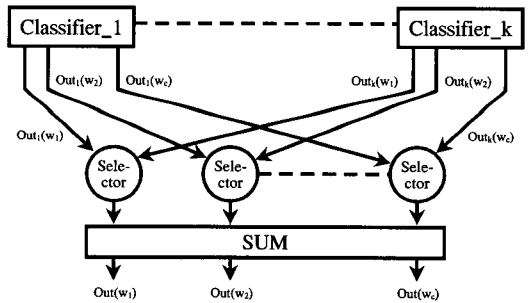
분류기를 적용하기 전에 주성분분석기법을 이용하여 특징벡터를 산출하고 이 과정 중에 미리 축소될 특징벡터의 개수를 설정하였다. 이러한 개수 결



(a) 가중치합법칙 (sum rule)



(b) 우수후보선출법칙 (best candidate vote rule)



(c) 다수선출법칙 (majority vote rule)

그림 2. 개별 분류기에 기반을 둔 융합법칙

정은 개수를 변경시키면서 가장 우수한 결과를 보이는 개수를 축소된 특징벡터의 차원으로 결정하였다. 분류기로는 기존의 k-NN 선형분류기 외에 비교를 위해서 MLP 비선형분류기를 사용하였다. 여기서, k-NN 분류기는 선택된 k의 값에 따라서 영향을 받기 때문에 k 변화에 따른 선형분류기의 특성을 살펴볼 필요가 있다. 이를 위해서 k의 값을 1, 3, 5 등으로 변경하여 성능을 평가하였다. 비선형 분류기인 MLP의 구조에서 입력차원은 PCA 기법에 의해서 축소된 특징벡터의 차원과 동일하다. 또한, 은닉층의 개수는 30, 학습률은 0.01, 모멘텀율은 0.2로 설정하였다. 이러한 파라미터들은 다양한 실험을 통해서 우수한 성능을 보이는 값들로 설정하였다.

표 2에 적용된 기법에 따른 인식결과를 나타냈다. k-NN 분류기를 사용하기 위한 거리척도 방법으로는 표 1에 나타낸 Euclidean, Hamming, Mahanabobis

표 2. 적용기법별 Wine 데이터에 대한 인식률.

[단위 : %]

Method	Training	Test	특징벡터의 수
PCA+kNN	100±0.0	85.5±4.3	6
PCA+MLP	99.4±0.8	93.1±2.4	6
SpPCA+kNN	100±0.0	91.8±2.6	분할영역당 4
SpPCA+MLP	98.9±1.2	94.7±2.5	분할영역당 4

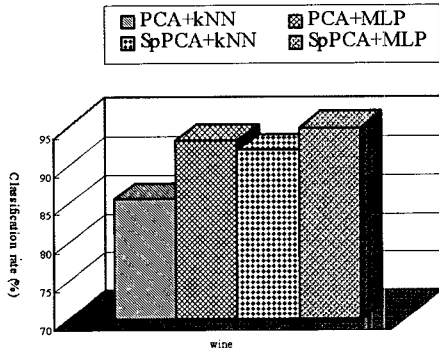


그림 3. Wine 데이터에 대한 인식률.

방법을 적용하였다. 이 중에서 Euclidean 거리척도 방법이 가장 좋은 결과를 나타냈다. 이러한 기존의 PCA에 의한 특징추출과 k-NN 선형분류기에 의해 wine 데이터를 분류한 결과 인식률은 85.5%를 나타냈다. 여기서 인식률은 총 사용데이터에 대하여 정확하게 인식한 데이터의 비를 구하여 산출하였다. 한편 본 논문에서 적용된 비선형 분류기인 MLP를 적용한 결과 인식률은 93.1%를 나타냈으며 기존 분류기에 비하여 약 8%의 높은 인식률 상승을 보였다. 이 결과로부터 기존의 방법인 k-NN 선형분류기에 비해 MLP 비선형분류기가 보다 효과적임을 알 수 있었다.

기존의 PCA에 의한 데이터의 차원을 축소하는 대신에 SpPCA에 의해서 구간별 차원을 축소한 후에 k-NN 선형분류기를 적용한 결과는 인식률 91.8%

로 기존의 방법인 PCA에 비해서 약 6%의 인식률 상승을 보였다. 이때 융합법칙으로는 보다 우수한 결과를 보인 기중치합을 이용하였고 Euclidean 거리 척도방법으로 유사도를 산출하였다. 마지막으로 본 논문에서 제안한 SpPCA와 MLP를 이용하여 wine 데이터를 분류한 결과 인식률은 94.7%로 가장 우수한 성능을 나타냈다. 이때 융합법칙으로는 기중치합을 이용하였고 분할 개수는 2개로 설정하였다.

Wine 데이터를 이용하여 실험한 결과 그림 3에서 보는 바와 같이 SpPCA+MLP, PCA+MLP, SpPCA+kNN, PCA+kNN 순으로 인식률이 우수하였다, 이 결과로부터 k-NN 선형분류기에 비해서 MLP 비선형 분류기가 보다 효과적이고, 또한 기존의 PCA 기법에 비해 SpPCA 기법이 우수한 결과를 나타냄을 확인하였다.

3.2 MRS 데이터

사용된 MRS 데이터는 수막종 M(menigiomas), 별아교세포종 A(astorcytomas) 및 간질환자의 뇌조직으로부터 추출한 비종양 E(non-tumourous) 샘플로부터 취득하였다^[12]. 3개의 클래스를 갖는 MRS 데이터의 수는 95(M), 74(A), 37(E)이며 하나의 데이터 당 550의 차원을 갖는다. 그림 4에 각각의 클래스별 패턴 파형을 나타냈다.

데이터 중에서 훈련에 사용된 데이터는 각각 47(M), 37(A), 19(E)로 103개이며 나머지 103개 데이터를 이용하여 제안된 방법의 성능을 분석하였다. 이전 실험과 마찬가지로 보다 정확한 결과의 해석을 위해서 훈련과 검증 패턴을 무작위로 10번씩 선택하여 총 10회의 실험을 하였다. MLP의 구조에서 은닉층의 개수는 20, 학습률은 0.005, 모멘트률은 0.008로 설정하였다. 기타 파라미터는 wine 실험에서와 동일한 조건하에서 실험하였다. PCA에 의해서 선택될 특징벡터의 수는 식 (10)과 같이 크기순으로 구해진 고유값의 합계와 선택하고자하는 p개의 개

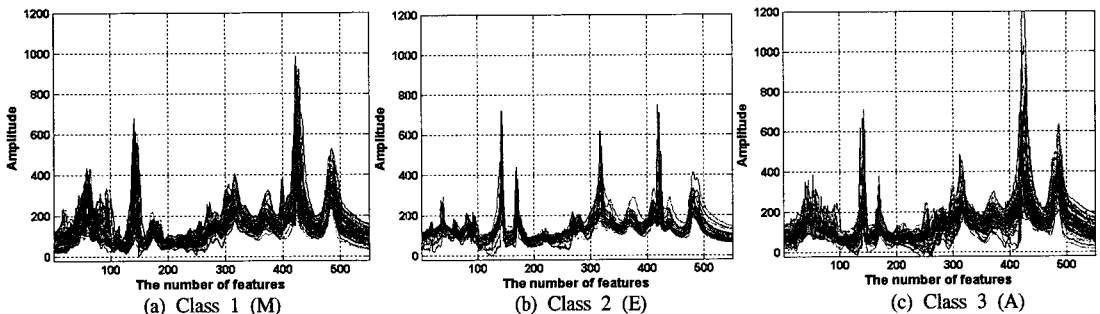


그림 4. 클래스별 MRS 파형.

표 3. PCA에 의한 인식결과.

[단위 : %]

Classifier		thr = 0.99		thr = 0.995		thr = 0.998	
k-NN	Distance	k = 1	k = 3	k = 1	k = 3	k = 1	k = 3
	Euclidean	79.7±3.5	77.7±4.2	79.4±3.8	77.3±4.4	79.3±4.0	77.4±4.0
	Hamming	82.6±4.1	81.0±4.3	81.8±4.8	80.8±4.6	82.1±4.7	81.3±4.1
	Maha	75.1±4.4	72.5±4.4	75.3±4.9	71.8±4.4	75.1±4.8	72.2±4.8
MLP	Training		Test	Training	Test	Training	Test
		97.0±1.7	74.1±4.9	96.7±2.0	75.5±3.1	96.3±1.2	75.3±5.8

표 4. SpPCA에 의한 인식결과.

[단위 : %]

Number of Partition	Fusion Method	k-NN			MLP	
		k = 1	k = 3	k = 5	Training	Test
3	Global Feature	83.7±2.8	83.0±3.6	79.8±3.9	97.1±1.5	75.1±2.5
	Sum Rule	83.7±2.8	83.0±3.6	79.8±3.9	99.8±0.4	81.4±3.4
	Best Vvote	75.3±2.2			99.7±0.7	79.5±3.1
	Majority Vote	79.9±2.3			100±0.0	81.9±2.8
5	Global Feature	84.3±3.3	81.8±2.6	81.0±3.6	98.3±1.5	77.2±5.1
	Sum Rule	84.3±3.3	81.8±2.6	81.0±3.6	99.5±0.5	81.4±3.6
	Best Vote	63.0±2.3			99.9±0.3	82.6±1.8
	Majority Vote	80.1±2.4			99.9±0.3	84.2±2.6
7	Global Feature	83.6±3.3	81.6±3.1	80.7±3.2	97.7±1.8	76.7±4.8
	Sum Rule	83.6±3.3	81.6±3.1	80.7±3.2	99.5±0.7	84.6±3.5
	Best Vote	67.1±4.2			100±0.0	86.7±2.7
	Majority Vote	80.2±2.4			100±0.0	88.2±2.0

수만큼 선택된 고유값의 비(thr)가 0.99, 0.95, 0.998을 만족하는 개수 p 를 특징벡터의 수로 설정하였다.

$$thr \geq \frac{\lambda_1 + \lambda_2 + \dots + \lambda_p}{\sum_{i=1}^p \lambda_i} \quad (10)$$

기존의 PCA를 이용하여 특징을 추출한 후 k-NN과 MLP 분류기를 적용한 경우의 인식률을 표 3에 나타냈다. 실험결과 k-NN 선형분류기의 경우 82.6%의 인식률을 나타냈다. 여기서 Hamming 거리척도 방법이 다른 거리척도 방법에 비해서 보다 좋은 결과를 보였다. 반면에 MLP 비선형분류기를 적용한 결과 인식률은 75.5%를 나타냈으며 k-NN 선형분류기에 비해서 약 7%만큼 인식률이 낮아졌다. 이 결과로부터 기존의 PCA 기법에 의해 추출된 특징벡터의 분류는 비선형분류기보다 선형분류기가 보다 우수한 인식률을 보임을 알 수 있다.

SpPCA에 의해서 특징을 추출한 후에 적용된 분류별 인식률을 표 4에 나타냈다. 표 4에서 보는 바와 같이 k-NN 분류기의 경우 분할수가 5이고 k=1일 때 인식률은 84.3%를 나타냈다. 또한 융합법칙으로는 전구역 특징벡터와 가중치합을 이용한 경우

가 동일하게 우수한 결과를 나타냈다. 기준벡터와 입력패턴과의 유사도는 성능이 우수한 Hamming 거리척도 방법으로 산출하였다. MLP 비선형분류기를 적용한 결과 분할수는 7이고 다수결정법칙에 의해서 융합한 겨우 인식률은 88.2%로 가장 좋은 인식률을 보였다. 그림 5에는 10개의 랜덤 검증데이터에 대한 인식률을 나타냈다. 이 결과 7번째의 검증데이터를 제외한 모든 데이터에 대하여 제안된 방법이 다른 기법에 비해 우수한 결과를 보임을 알 수 있다.

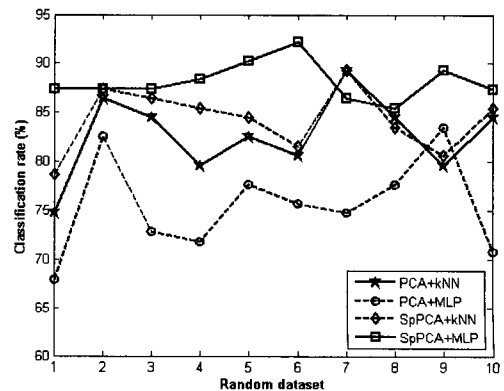


그림 5. 랜덤 검증 데이터에 대한 인식률.

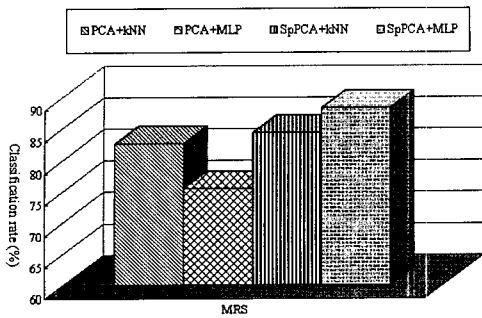


그림 6. MRS 데이터에 대한 인식률.

그림 6에 적용 기법별 인식결과를 나타냈다. 그림 6에서 보는 바와 같이 제안된 방법을 적용한 결과 인식률이 기존의 PCA+kNN 기법에 비해서 약 6% 높게 나타내어 제안된 방법의 우수성을 검증하였다. 또한, 고차원 입력패턴을 분할 없이 전차원에 대해서 PCA를 이용하여 특징을 추출한 후, MLP 비선형분류기에 의해서 패턴을 분류한 결과 인식률이 매우 낮게 나타났다. 반면에 구간별로 분할한 후 각각의 구간에 대해서 특징을 산출하고, MLP 비선형분류기를 적용한 결과 인식률이 매우 향상됨을 알 수 있었다. 이것은 다중결정법칙에 의해 구간별로 산출된 특징들을 효과적으로 융합하였기 때문이다.

IV. 결 론

본 논문에서는 SpPCA와 MLP에 기반을 둔 융합 기법을 이용한 MRS 패턴분류 기법을 제안하였다. 제안된 방법은 우선 고차원의 입력 데이터를 균등 분할한 후 PCA를 이용하여 분할영역별로 특징을 추출했다. 다음 단계에서는 추출된 특징벡터를 MLP 비선형분류기를 이용하여 구간별 유사도를 산출했다. 최종 분류단계에서는 효과적인 융합법칙에 의해서 패턴을 분류하였다.

제안된 방법의 타당성을 보이기 위해 저차원의 wine 데이터와 고차원의 MRS 데이터를 대상으로 실험하였다. 저차원의 wine 데이터를 대상으로 실험한 결과 기존의 방법인 PCA+kNN의 경우 85.5%의 인식률을 나타냈으며 제안된 SpPCA+MLP 기법을 적용한 결과 94.7±2.5%의 인식률을 나타내어 약 9%의 인식률 향상을 보였다. 또한, 고차원의 MRS 데이터를 대상으로 실험한 결과 기존의 PCA+kNN 방법의 경우 82.6%의 인식률을 나타냈으며 제안된 SpPCA+MLP 기법을 적용한 결과 88.2%의 인식률을 나타내어 약 6%의 인식률 향상을 보였다.

실험결과로부터 k-NN 비선형분류기에 비해서 MLP 비선형분류기가 분류기로서 우수함을 알 수 있다. 향후 SpPCA를 적용하기 위해서는 최적의 분할 개수의 선택과 분할영역의 설정 문제에 대한 체계적인 연구가 요구된다.

참 고 문 헌

- [1] R. Baumgartner, R. Somorjai, C. Bowman, T.C. Sorrell, C.E. Mountford, U. Himmelreich, Unsupervised feature dimension reduction for classification of MR spectra, *Magnetic Resonance Imaging* 22, pp. 251-256, 2004.
- [2] Y.Y.B. Lee, Y. Huang, W. El-Deredy, P.J.G. Lisboa, C. Arus, P. Harris, Robust methodology for the discrimination of brain tumours from in vivo magnetic resonance spectra: Science, Measurement and Technology, *IEE Proc.*, Vol.147, Issue 6, pp.309-314, 2000.
- [3] Yu Wang, S. Van Huffel, E. Heyvaert, L. Vanhamme, N. Mastrorardi, P. Van Hecke, Magnetic resonance spectroscopic quantitation via complex principal component analysis, *WCCC-ICSP 2000*, Vol.3, pp.2074-2077, 2000.
- [4] M. Turk and A. Pentland, Face recognition using eigenfaces, *IEEE Conf. on Computer Vision and Pattern Recognition*, pp.586-591, 1991.
- [5] Hong Yang, Joseph Irudayaraj and Manish M. Paradkar, Discriminant analysis of edible oils and fats by FTIR, FT-NIR and FT-Raman spectroscopy, *Food Chemistry*, Vol.93, Issue1, pp.25-32, 2005.
- [6] Marina Vannucci, Naijun Sha and Philip J. Brown, NIR and mass spectra classification: Bayesian methods for wavelet-based feature selection, *Chemometrics and Intelligent Laboratory Systems*, In Press, Corrected Proof, Available online 4 March 2005.
- [7] E.K. Tang, P.N. Suganthan, X. Yao and A.K. Qin, Linear dimensionality reduction using relevance weighted LDA, *Pattern Re-*

- cognition, Vol.38, Issue4, pp.485-493, 2005.
- [8] N.P. Hughes, S.J. Roberts, L. Tarassenko, Semi-supervised learning of probabilistic models for ECG segmentation, Engineering in Medicine and Biology Society, EMBC 2004, Vol.1, pp.434-437, 2004.
 - [9] Songcan Chen, Yulian Zhu, Subpattern-based principal component analysis, Pattern Recognition 37, pp.1081-1083, 2004.
 - [10] Rajkiran Gottumukkal, Vijayan K. Asari, An improved face recognition technique based on modular PCA approach, Pattern Recognition Letters 25, pp.429-436, 2004.
 - [11] Yang-Lang Chang, Chin-Chuan Han, Fan-Di Jou, Kuo-Chin Fan, K.S. Chen, Jeng-Horng Chang, A modular eigen subspace scheme for high-dimensional data classification, Future Generation Computer System 20, pp.1131-1143, 2004.
 - [12] Nicolino J. Pizzi, Fuzzy pre-processing of gold standards as applied to biomedical spectra classification, Artificial Intelligence in Medicine

송창규 (Chang Kyu Song)

정회원



1995년 2월 충북대학교 전기공학과 졸업(공학사)
 1997년 2월 충북대학교 대학원 전기공학과 졸업(공학석사)
 1997년 3월~현재 충북대학교 대학원 전기공학과 박사과정
 <관심분야> 영상처리, 영상압축,

패턴인식

이대종 (Dae Jong Lee)

정회원



1995년 2월 충북대학교 전기공학과 졸업(공학사)
 1997년 2월 충북대학교 대학원 전기공학과 졸업(공학석사)
 2002년 8월 충북대학교 대학원 전기공학과 졸업(공학박사)
 2003년~현재 충북대학교 컴퓨터정보통신연구소 연구원

2004년~2005년 7월 University of Alberta, Postdoctor.

<관심분야> 영상처리, 영상압축, 패턴인식

전병석 (Byeong Seok Jeon)

정회원



1995년 2월 충북대학교 전기공학과 졸업(공학사)
 1997년 2월 충북대학교 대학원 졸업(공학석사)
 2005년 2월 충북대학교 대학원 졸업(공학박사)
 2000년 3월~현재 (주)세화폴리텍

<관심분야> 영상처리, 영상압축, 패턴인식

유정웅 (Jeong Woong Ryu)

정회원



1965년 2월 한양대학교 전기공학과 졸업(공학사)
 1976년 8월 단국대학교 대학원 졸업(공학석사)
 1987년 8월 단국대학교 대학원 졸업(공학박사)
 1969년~1979년 대전산업대학교

부교수

1979년 3월~현재 충북대학교 전기전자컴퓨터공학부 교수

<관심분야> 강인제어, PID, QFT, 지능시스템, 인공지능, 퍼지이론