

# 생체 신호의 특징 추출 및 SVM을 이용한 분류

김만선\*, 이상용\*\*

{mansun, sylee}@kongju.ac.kr

## Feature Extraction and Classification using SVM for Biomedical Signal

### 요 약

최근 대용량의 데이터베이스로부터 유용한 정보를 발견하고 데이터간에 존재하는 연관성을 탐색하고 분석하는 데이터 마이닝에 관한 많은 연구들이 진행되고 있다. 다양한 생체 신호를 분석하기 위하여 데이터 마이닝 기법을 이용할 수 있다. 본 논문에서는 심전도 신호의 패턴을 분류하기 위하여 신경망 기법을 적용하였다. 최근 패턴분류에 있어서 각광을 받고 있는 SVM 모델은 학습과정에서 얻어진 확률분포를 이용하여 의사결정함수를 추정한 후 이 함수에 따라 새로운 데이터를 이원분류 하는 것으로 분류 문제에 있어서 일반화 기능이 매우 높다. 기존에 많이 이용되던 BP 모델과 비교평가 하였다.

### 1. 서 론

뇌파(EEG), 심전도(ECG), 근전도(EMG), 안전도(EOG), 체온(ST), 피부전기저항(GSR), 맥파(BVP) 등 다양한 생체 신호를 분석하기 위해 여러 가지 방법들이 연구되고 있다. 그 중에서 심전도 데이터는 심장의 전기적인 신호의 다양한 파형으로 이루어진다. 이와 같은 파형을 분석하고 분류하기 위해서 데이터마이닝의 기법을 이용할 수 있다.

데이터마이닝은 수많은 데이터들로부터 유용한 정보를 얻어내는 과정으로 대용량의 데이터 내에 존재하는 관계, 패턴, 규칙 등을 탐색하고 찾아내어 모형화 함으로써 지식을 추출하는 일련의 과정이라고 정의할 수 있다. 이러한 데이터마이닝은 KDD(Knowledge Discovery in Database)를 구성하는 여러 단계 중에서 가장 핵심적인 역할을 하며, 데이터마이닝의 여러 가지 기법중 분류는 생체공학 분야에서 진단을 목적으로 널리 사용되고 있다. 분류는 미리 정의된 범주를 가진 클래스들의 집합이 주어졌을 때, 특정한 데이터 항목이 이런 클래스들의 어디에 속하는 지를 결정하는 것이다.

분류의 단계는 전처리 및 특징 추출, 학습으로 구성된

다. 전처리 및 특징 추출 과정에서는 입력패턴에 대한 잡음 제거 및 효과적인 특징들을 추출하여 입력벡터를 구성함으로써 패턴 분류기의 입력 벡터를 만들어주고, 학습 과정을 통해서 입력 벡터들을 유사한 특징에 따라 결정된다.

신경망은 사람 뇌의 동작에 가깝게 만든 프로그램 또는 데이터 구조 시스템을 말한다. 학습 방법으로는 보통 감독 학습 방법과 비감독 학습 방법으로 나뉘어지며, 전자는 목적 패턴이 있는 것을 말하며, 후자는 목적 패턴이 존재하지 않는 신경망을 말한다. 여기에서 목적 패턴이란 주어진 입력 패턴에 대해 신경망이 출력해 주기 원하는 출력 패턴을 말하며, 외부로부터 주어진 패턴이다.

본 논문에서는 심전도 신호의 패턴을 분류하기 위하여 심전도 패턴의 특징으로부터 추출된 파라미터들을 신경망의 입력으로 하는 BP, SVM모델로 신경망이 패턴 분류를 효과적으로 수행하는지 특성을 평가하였다.

### 2. 관련 연구

심전도(ECG : electrocardiogram)는 심장의 상태를 비관혈적(non-invasive)으로 진단하는 매우 중요한 수단으로 활용되며, 진폭은 수 mV이고 주파수는 250Hz이내의 생체전위 신호 중 하나이다.

\* 공주대학교 컴퓨터공학과

\*\* 공주대학교 컴퓨터공학과 부교수

돌연사를 일으키는 대표적 심장질환은 허혈성 심장질환, 확장성 심근증, 비후성 심근증이 대부분을 차지하며 특히 허혈성 심장 질환이 돌연사의 80%를 차지하므로 이 질환의 예방 및 조기진단이 중요하다. 허혈성 심장질환의 증세로서 협심증과 심근경색증이 있는데, 심전도의 ST 분절이 elevation 또는 depression 되는 episode를 띄게 된다. ST 세그먼트는 1Hz미만의 주파수 대역을 가지고 있으므로 저주파신호의 기저선(baseline) 변동 잡음, 전 주파수 대역에 존재하는 근잡음(muscle artifact)과 같은 주파수 대역에 존재하므로 정확한 잡음을 제거하지 못하면 신호의 왜곡이 발생하므로 오진을 하게 된다.

ECG 데이터를 분류하기 위한 기존의 여러 가지 방법들[1][2][3]이 있었으나, 오분류률의 감소 및 학습시간의 단축에서 성능이 우수한 BP와 SVM을 적용하여 연구할 필요가 있다.

### 3. 특징 추출 및 신경망 시스템의 적용

#### 3.1 특징 추출

원신호의 왜곡을 최소화하여 기저선을 제거 할 수 있는 웨이브렛 모함수를 결정하기 위하여, 다양한 웨이브렛 모함수를 적용하여 기저선을 제거하였다.

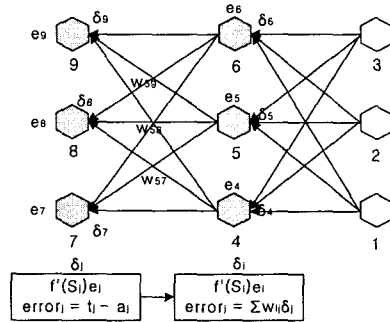
ST 분절 구간 설정은 RR 간격이 600ms보다 클 경우는 R-peak에서 60ms 후부터, 그렇지 않은 경우는 40ms 후부터 160ms 동안으로 하였다. ST 분절의 특징을 추출하기 위하여 RR 간격이 600ms보다 클 경우 또는 그렇지 않을 경우에 대하여, ST0[ST 분절이 시작되는 지점 (R+60 또는 R+40 ms)에서의 진폭], ST80[R+140 또는 R+120 ms 지점에서의 진폭], ST 분절의 slope, ST 분절의 면적(ST 분절의 구간에서 ST 분절과 isoelectric level로 둘러싸여진 면적)을 ST 분절의 특징을 나타내는 파라미터의 후보로 설정하였다.

#### 3.2 BP(BackPropagation)

다층 퍼셉트론(Multi Layer Perceptron)에 일반화된 델타 규칙(Generalization Delta Rule)을 학습 규칙으로 사용한 알고리즘을 오류 역전파 학습 알고리즘(BP)이라 한다.

“만일, 어떤 신경세포의 활성이 다른 신경세포의 잘못된 출력에 공헌을 하였다면, 두 신경세포 간의 연결 가중치를 그것에 비례하여 조절해 주어야 한다. 그리고, 그러한 과정은 그 아래에 있는 신경세포들까지 계속된다는

점이다. 이렇게 출력 층에서 발생한 에러를 위층으로 역전파시키므로, 오류 역전파 알고리즘이라 한다. [그림 1]은 일반적인 BP알고리즘의 모형이다. 복잡함을 피하기 위해 1개의 은닉층에 5개의 은닉노드가 있는 경우를 고려하였다.

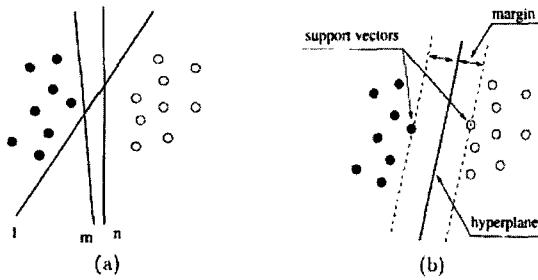


[그림 1] 오류 역전파 학습 알고리즘 모형

목적 패턴에서 출력 신경세포의 활성값을 뺀 값이 바로 해당 출력 신경세포의 에러가 된다. [그림 1]에서  $e_7, e_8, e_9$ 로 나타낸 것이 각 출력 신경세포들의 에러이다. 에러를 여러번 가공하여 출력층 각각의 신경세포에 대해 델타( $\delta$ )라는 것을 구하게 된다. 그리고,  $f'(S)$ 는 활성화함수의 미분 값을 말하며, 여기에서는 시그모이드 함수를 사용하였다.

#### 3.3 SVM

최근에 패턴분류에 있어서 각광을 받고 있는 SVM(Support Vector Machine) 모델은 1998년 V.N.Vapnik[4][5]에 의해 개발된 통계적 학습이론으로서 학습데이터와 범주 정보의 학습진단을 대상으로 학습과정에서 얻어진 확률분포를 이용하여 의사결정함수를 추정 한 후 이 함수에 따라 새로운 데이터를 이원 분류하는 것으로 VC(Vapnik Chervonenkis) 이론이라고도 한다. 기본 원리는 훈련 데이터들을 고차원의 특징공간으로 사상(mapping)시킨 후 두 분류 사이의 여백(margin)을 최대화 시키는 결정함수(hyperplane)를 찾는 것이다. SVM은 분류 문제에 있어서 일반화 능력이 높기 때문에 많은 분야에서 응용되고 있다. [그림 2]은 hyperplane을 사용하여 두 클래스 사이를 분류한 것이다. (a)는 1, m, n으로 두 클래스를 단순히 분류한 것이며, (b)는 support vectors에 의해 최대 margin을 주어 결정함수를 찾아 분류한 것이다.



[그림 2] 2 class에서 결정함수를 사용한 분류

4. 실험

4.1 실험 데이터

European ST-T database의 e0103, e0105, e0111 을 선택하였으며, European ST-T database로부터 ST 분절에서 임상소견 중 정상(normal)과 비정상(ST episode를 갖는 데이터)으로 분류하였다.

4.2 Classifier

SVM Classifier의 Kernel은 polynomial(차수 : 3)로 사용하였으며 BP의 학습 알고리즘 중에서 Conjugate gradient backpropagation 방법으로 사용하였다.

[표 1] 분류 성능 비교(Hit Rate(%))

Classifier \ Data	BP	SVM
e0103	66%	55%
e0105	35%	75%
e0111	53%	76%

5. 결론 및 향후 과제

심전도 신호의 패턴을 분류하기 위하여, 심전도 패턴의 특징으로부터 추출된 파라미터들을 신경망의 입력으로 하였다. BP 모델보다는 SVM 모델이 더 좋은 분류 성능을 보였으나, 90%이상의 Hit Rate를 보이려면 더 많은 훈련데이터로 학습을 시켜야하며, 일반화 성능도 높아질 것이다. 특징 추출된 파라미터들이 일반화의 성능을 좌우하므로, 최적의 입력 벡터들을 결정하는 것이 필요하다.

향후 과제로는 여러 가지 커널함수를 이용하여 SVM의 분류 성능을 비교 평가할 것이다.

6. 참고 문헌

[1] Zumray Dokur and Tamer Olmez, "ECG beat classification by a novel hybrid neural network", Computer Methods and Programs in Biomedicine, Volume 66, Issues 2-3, September 2001, Pages 167-181.

[2] Pentti M. Rautaharju, Sophia H. Zhou, E. William Hancock, B. Milan Hor[acacute][ccaron]jek, Dirk Q. Feild, James M. Lindauer, Galen S. Wagner, Olle Pahlm and Charles L. Feldman, "Comparability of 12-lead ECGs derived from EASI leads with standard 12-lead ECGs in the classification of acute myocardial ischemia and old myocardial infarction", Journal of Electrocardiology, Volume 35, Issue 4, Part 2, October 2002, Pages 35-39

[3] U. Rajendra Acharya, P. Subbanna Bhat, S. S. Iyengar, Ashok Rao and Sumeet Dua, "Classification of heart rate data using artificial neural network and fuzzy equivalence relation, Pattern Recognition", Volume 36, Issue 1, January 2003, Pages 61-68.

[4] N. Cristianini, "An Introduction to Support Vector Machines", Cambridge University Press, 2000.

[5] Z. Weida, "Linear programming support vector machines", J. Pattern Recognition Society, pp:1-10, 2001.