

# 기계학습 기반 저 복잡도 긴장 상태 분류 모델

## Design of Low Complexity Human Anxiety Classification Model based on Machine Learning

홍은재\* · 박형곤†  
(Eunjae Hong · Hyunggon Park)

**Abstract** - Recently, services for personal biometric data analysis based on real-time monitoring systems has been increasing and many of them have focused on recognition of emotions. In this paper, we propose a classification model to classify anxiety emotion using biometric data actually collected from people. We propose to deploy the support vector machine to build a classification model. In order to improve the classification accuracy, we propose two data pre-processing procedures, which are normalization and data deletion. The proposed algorithms are actually implemented based on Real-time Traffic Flow Measurement structure, which consists of data collection module, data preprocessing module, and creating classification model module. Our experiment results show that the proposed classification model can infer anxiety emotions of people with the accuracy of 65.18%. Moreover, the proposed model with the proposed pre-processing techniques shows the improved accuracy, which is 78.77%. Therefore, we can conclude that the proposed classification model based on the pre-processing process can improve the classification accuracy with lower computation complexity.

**Key Words** : Classification model, Human anxiety, Complexity, Machine learning, Support vector machine

### 1. 서 론

최근 개인의 생체 데이터를 수집하고 실시간으로 모니터링하여 데이터를 분석하는 서비스가 크게 증가하고 있다. 이러한 서비스는 단순히 개인이 자신의 생활 흐름을 알고 간단히 건강을 보조하는 것에서부터 만성 환자의 생체 정보를 실시간으로 수집하고 의사가 알맞은 진단을 내려주는 것까지 전문화되고 있으며, 생체 데이터만을 이용해 사람의 감정을 추론하는 연구까지 진행되고 있다. 이와 같은 서비스를 위해서는 24시간 모니터링 시스템과 지속적으로 사용자의 디바이스와 연결하여 실시간으로 데이터를 전송하는 기술, 그리고 정확하고 빠른 자동 데이터 분석 기술 등이 필요하다. 그러나 데이터를 이용한 분석 기술은 주로 정확도에 초점을 맞추기 때문에 속도는 크게 고려가 되지 않거나 혹은 실시간으로 측정하여 전송하기가 어려운 생체 데이터인 경우가 대부분이다.

[1]은 긴장 상태인 사람의 심전도(electrocardiogram, ECG)를 측정하여 특징(feature)을 추출하고 다른 데이터를 통해 이 특징의 정확도를 테스트하는 데이터 분석 기술을 제안한다. 하지만 자동화된 실시간 데이터 모니터링 및 분석은 이루어지지 않고 있

며 ECG는 측정을 위해 6개의 센서를 몸에 부착한 채로 측정해야 한다는 단점이 있어 실제 생활에서 이를 이용하기에는 적합하지 않다는 문제가 있다. [2]에서는 뇌전도(electroencephalography, EEG) 데이터, 즉 뇌파 데이터를 이용하여 기쁨이나 공포, 슬픔과 같은 감정을 감지하였다. 이를 분석하는 데에는 서포트벡터머신(Support Vector Machine, SVM) 방식을 이용하였는데 흔히 알려진 감정일수록 분류 정확도가 매우 낮았고 또한 EEG 데이터의 측정이 쉽지 않아 실제로 이용하기는 어렵다. 따라서 실시간 모니터링을 통한 생체 데이터를 이용하여 높은 정확도와 빠른 연산 속도를 가진 감정 분류 모델이 필요하다.

본 연구에서는 다양한 사람의 생체 데이터를 실시간으로 수집하고 이 데이터를 바탕으로 긴장 감정 상태를 분류할 수 있는 기계 학습 기반 분류 모델을 제안한다. 핵심은 여러 대상에서 실시간으로 수집한 심박수 데이터를 이용해 시간을 단축하면서도 높은 정확도를 가지는 분류 모델을 설계하는 것이다. 이를 위해 기계학습 중 SVM [3-6] 알고리즘을 바탕으로 모델을 생성하였다. SVM은 기계 학습 알고리즘 중 하나로 간단하고 효율적인 계산을 이용하며, 패턴 인식(pattern recognition) 및 분류 문제(classification problem)[11]와 생체 신호로부터 감정을 분류하기 위한 분야에서 널리 사용된다[12]. 이때, 단순히 SVM을 적용해 얻은 모델은 심박수를 기반으로 분류하는 데 제한된 성능만 보이기 때문에 데이터 전처리 과정을 통해 성능을 높였다. 이를 통해 SVM 기반 모델의 정확성과 속도가 향상될 수 있음을 실험 결과를 통해 확인했다. 실시간으로 데이터를 수집하고 분류할 수 있는 환경과 유사하게 실험을 진행하기 위해 본 논문에서는 시스

† Corresponding Author : Dept. of Electronic and Electrical Engineering, Ewha Womans University, Korea.  
E-mail: hyunggon.park@ewha.ac.kr

\* Dept. of Electronic and Electrical Engineering, Ewha Womans University, Korea.

Received : June 28, 2017; Accepted : August 23, 2017

템을 구현하여 분류 모델을 생성하고 사람들의 불안 상태를 분류하였다.

본 논문은 다음과 같이 구성된다. 2장에서는 본 논문에서 해결하고자 하는 문제 정의를 다루고 3장에서는 수집한 데이터의 특징에 대해 알아보며, 4장에서는 데이터의 특징에 따라 문제를 해결하기 위한 방안을 제안한다. 5장에서는 본 연구에서 제안하는 방안을 적용하여 구축한 시스템에 대해 살펴보고, 이를 이용한 실험 결과에 대해 알아본다. 6장에서는 결론을 맺고 마무리한다.

## 2. 문제 정의

본 논문에서는 수집한 심박수 학습 데이터 집합  $x$ 를 통해 긴장 상태를 분류하는 모델  $f(x)$ 를 생성하는 해법으로 SVM을 사용한다. 이 때 테스트 집합  $T$ 의 데이터 중 하나인  $x_t$  데이터에 대한  $f(x)$ 의 출력 값  $y_t$ 는 -1 혹은 1이다. 즉,

$$y_t \in \{-1, 1\} \quad (1)$$

이며 이때 출력 값 '1'과 '-1'은 각각 긴장한 상태일 때와 긴장하지 않은 상태를 나타낸다.

분류 모델  $f(x)$ 의 분류 성능을 측정하기 위해, 본 논문에서는 아래와 같이

$$TA(f(x)) = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad (2)$$

분류 정확도(Total Accuracy)를 정의한다. 이때  $FN$ 은 실제 레이블이 '-1'인 데이터를 '1'로 잘못 분류한 데이터 수를,  $FP$ 는 실제 레이블이 '1'인 데이터를 '-1'로 잘못 분류한 데이터의 수를 나타낸다.  $TP$ 와  $TN$ 은 각각 실제 레이블이 '1', '-1'인 데이터를 옳게 분류한 데이터 개수이다. 그러므로,  $TN+TP$  값은 분류 모델이 옳게 분류한 데이터의 총 개수를 나타내고,  $TA$ 는 전체 데이터 중 옳게 분류한 데이터의 백분율 값이다.

분류 모델을 형성할 때, 심박수 데이터의 특징을 고려하여 분류 모델의 성능을 더 높일 수 있다. 본 논문에서는 2단계의 데이터 전처리 과정을 거친 데이터를 이용해 분류 모델을 생성하는 방안을 제안하며, 이때 연산 시간 대비 데이터의 분류 정확도가 높은 모델  $g(x)$ 를 구하고자 한다. 아래의 식 (3)은 모델  $f(x)$ 의 연산 시간 대비 분류 모델의 분류 정확도를 나타내며

$$TO(f(x)) = \frac{TA(f(x))}{Computation\ Time} \quad (3)$$

2단계의 데이터 전처리 과정 중 첫 번째 과정만을 거친 데이터를 이용해 생성한 분류 모델을  $h(x)$ 라 했을 때 본 논문에서는 식 (4)를 만족하는  $g(x)$ 를 구하고자 한다.

$$TO(g(x)) \geq TO(h(x)) \quad (4)$$

이때  $TO(g(x))$ 는  $g(x)$ 를 이용하는  $TO$  값이고,  $TO(h(x))$ 는  $h(x)$ 를 이용하였을 때의  $TO$  값이다.

## 3. 심박수 데이터의 특징

심박수는 단위 시간당 심장 박동 횟수로 분당 박동수 (bpm)로 표시한다. 심박수는 손목, 팔꿈치, 목 등 신체의 맥박에서 측정하며 심장병 등의 증세가 없는 성인의 심박수는 60-100 bpm이다.

심박수는 두 가지 특성이 있다. 첫째, 심박수의 히스토그램은 평균을 중심으로 정규분포와 같이 종모양의 형태를 가진다[14]. 즉, 평균을 중심으로 밀집된 데이터가 측정된다. 둘째, 사람이 불안을 느끼면 심장 박동이 증가한다. [15]에서 볼 수 있듯이, 20대 남성 그룹이 걱정 또는 불안을 느낄 때와 느끼지 않을 때 심박수를 측정 한 결과, 평균은 각각 75.9 bpm과 57.6 bpm이었다.

이 논문에서 분류하는 불안 상태는 시험 불안 상태로 개인이 시험을 위협적으로 느끼고 평가 상황에서 실패를 두려워하여 느끼는 정서반응이고[16], 이러한 불안은 자율 생리 반응, 즉 걱정과 공포와 같은 감정으로 정의 된다[17]. 이와 같이 불안은 자율 신경계에 의한 반응인데 자율 신경계(ANS)는 흥분성 교감 신경계(SNS) 및 억제성 부교감 신경계(PNS)로 세분되며, 둘의 상호 작용으로 다양한 생리적 흥분이 생성된다. 이 중 SNS는 불안한 상태에서 활동적이며, 생리적 각성을 유발하여 환경 상황이나 조건을 신속하게 조정한다. 심박수의 증가, 발한 및 빠르고 얇은 호흡은 이러한 각성 상태의 특징이다.

본 연구에서는 피실험자에게 시험을 치르게 하여 불안감을 유도한 상태에서 심박수 데이터를 수집하고, 이와 같은 불안한 상태를 식별할 수 있는 분류 모델을 생성하고자 하며, 심박수 데이터의 특성을 고려한 전처리 과정을 제안한다.

## 4. 제안 방안

### 4.1 SVM을 이용한 분류 모델

$x = \{x_1, \dots, x_i, x_j, \dots, x_p\}$ 는 수집한 데이터 중 학습을 위해 사용하는 데이터 집합이다. 이때,  $x$ 는 다음의 두 가지 조건을 따른다.

$$x, T \subseteq C \text{ and } x \cap T = \emptyset$$

여기서  $C$ 는 수집한 전체 데이터 집합이고  $T$ 는 테스트 데이터 집합이다. 수집한 학습 데이터  $x_i$ 가 어느 클래스에 속해 있는지를 나타내는 레이블은  $y_i$ 이고,  $y_i \in \{-1, 1\}$ 이다.  $x_i$ 가 긴장 상태에 수집한 데이터일 때에  $y_i = 1$ 을,  $x_i$ 가 비긴장 상태에 수집한 데이터일 때에  $y_i = -1$  레이블을 붙인다.

학습 데이터  $(x_i, y_i), i = 1, \dots, p$ 에 대해 SVM은 최적화된 분

류 면, 즉 분류 모델인 초평면 (hyperplane)  $f(x)$ 와  $(w, b)$ 를 결정한다.  $(w, b)$ 는 아래의 식에서 기반한다.

$$w \cdot x + b = 0. \tag{5}$$

위 식의  $b \in \mathbb{R}$ 이며,  $\mathbb{R}$ 은 모든 실수의 집합을 나타내고,  $w \cdot x$ 는  $w \in \mathbb{R}^p$ 와  $x$  사이의 내적을 나타낸다. 최적화된 초평면은 항상 하나로 결정되고,  $w$ 와  $b$  값에 대해 입력 데이터인 학습 데이터는 다음을 만족한다.

$$y_i = 1 \quad \text{if} \quad w \cdot x_i + b \geq 1 \tag{6}$$

와

$$y_i = -1 \quad \text{if} \quad w \cdot x_i + b \leq -1. \tag{7}$$

이때 초평면과 가장 가까운 데이터 간의 거리인 최대 마진 값은  $\frac{2}{\|w\|}$ 이다.

(5)에서의 SVM은 선형 방식이므로 모델의 분류 성능을 향상시키기 위하여 본 논문에서는 비선형 방식을 이용하는 것을 제안한다. 비선형 방식은 커널 함수를 이용해 비선형 초평면을 생성하는 방식이고, 커널 함수는 비선형 매핑( $\Phi$ )을 이용해 학습 데이터를 높은 차원으로 변환하여 높은 차원에서 선형 초평면을 생성함으로써, 실제 데이터의 차원에서는 비선형 초평면을 생성한다. 커널 함수는 아래와 같이 표현된다.

$$K(x_i, x_j) = \Phi(x_i)\Phi(x_j)^T. \tag{8}$$

본 논문에서는 가장 일반적으로 사용되며 효율적이고 정확도 높은 학습 기능을 제공하는 가우시안 커널을 사용하였다. 가우시안 커널은 다음과 같이 표현되며,

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \tag{9}$$

여기서  $\sigma$ 는 SVM의 일반화 가능성 (generalizability)과 밀접한 관련이 있는 분산 매개 변수이다.

위의 방식으로 생성한 모델의 성능을 평가할 때는  $k$  묶음 교차검증( $k$  fold cross-validation)을 이용한다.  $k$  묶음 교차검증은 데이터 세트를 무작위로 혼합하고 크기가 동일한  $k$ 개의 묶음으로 나누어 한 묶음은 테스트를 위해 사용하는 반면  $k-1$ 개의 묶음은 분류 모델의 학습을 위해 사용한다. 이를  $k$ 번 반복하여  $k$ 개의 오류율(error rate)을 구하는데 모델의 최종 분류 정확도는 이 오류율들의 평균값을 이용하여 구하여[13] 분류 정확도에 대한 신뢰도를 높인다. 일반적으로  $k$ 를 5 또는 10으로 설정하며 [18], 본 논문에서는 10으로 설정하였다.

## 4.2 데이터 전처리

분류 모델은 여러 수집 대상으로부터 수집 된 데이터에 기초

하여 구성되기 때문에, 각 개인의 심박수 특성에 영향을 받을 수 있다. 또한, SVM은 아웃라이어(outlier) 데이터에 민감한 알고리즘으로, SVM에 기반한 모델의 성능은 소수의 아웃라이어 데이터에 의해 크게 영향을 받을 수 있다. 이 문제를 해결하고 SVM만을 사용하여 생성한 분류 모델  $f(x)$ 보다 성능이 향상된 모델  $g(x)$ 를 구하기 위하여 본 논문에서는 정규화(normalization)[6]와 데이터 삭제 (data deletion)[7]를 데이터 전처리 방안으로 제안한다.

### 4.2.1 데이터 정규화

정규화는 입력 값의 크기와 상관없이 입력 데이터들이 분류 모델을 생성할 때 미치는 영향을 동일하게 한다[6]. 식 (9)에 나타난 것과 같이 비선형 분류 모델을 위해 가우시안 커널을 사용할 때  $\|x_i - x_j\|^2$  값에 영향을 받게 된다. 따라서 일부 거리 값이 큰 경우에 의해 편향성(bias)을 가지게 되고 절대 값이 큰  $x_i$ 가 분류 모델의 생성에 다른 데이터보다 큰 영향을 미친다. 따라서 정규화를 통해 데이터의 범위를  $[0, 1]$ 로 바꿔  $\|x_i - x_j\|$ 의 크기를 줄일 수 있고 편향성을 없앨 수 있다.

또한, 제안 모델을 생성 시 각기 다른 평균, 표준 편차 및 특징을 갖는 여러 대상의 데이터를 하나의 데이터 집합으로 묶어 학습 데이터로 이용한다. 이때 정규화를 통해 각 개인의 데이터 특성이 전체 데이터에 미치는 영향을 줄인다. 데이터는 다음과 같이 정규화 하였다.

$$x_n = \frac{x_c - x_{\min}}{x_{\max} - x_{\min}} \tag{10}$$

여기서  $x_c$ 는 수집 데이터,  $x_n$ 은 정규화 된 데이터,  $x_{\min}$ 과  $x_{\max}$ 는 수집 데이터 중에 최솟값, 최댓값을 각각 나타낸다. 또한,  $x_c, x_{\min}, x_{\max} \in C$ 이다.

### 4.2.2 데이터 삭제

데이터 삭제 방법[7]은 SVM의 성능에 영향을 미칠 수 있는 아웃라이어를 줄임으로써 SVM의 성능을 향상시키는 방식이다. 심박수 데이터는 평균을 중심으로 데이터 밀도가 높다. 그러므로, 밀도가 낮은 영역에 있는 데이터를 아웃라이어로 규정하고 삭제하여 아웃라이어의 영향력을 줄일 수 있다. 또한 SVM 알고리즘의 복잡도는  $O(n^3)$ 이며 데이터의 수에 영향을 받기 때문에[8] 데이터 삭제 방식은 SVM의 복잡성을 줄여 알고리즘의 실행 속도를 향상시킬 수 있다.

$x_{ni} = \{x_{ni,1}, x_{ni,2}, \dots, x_{ni,i}, x_{ni,j}, \dots, x_{ni,q}\}$ 는 정규화 과정을 거친 학습 데이터 중에서 동일한 레이블을 가진 데이터 집합이다. 먼저, 식 (11)과 (12)를 통해 모든 데이터 간 거리의 합의 평균 값  $\hat{D}$ 을 구한다.

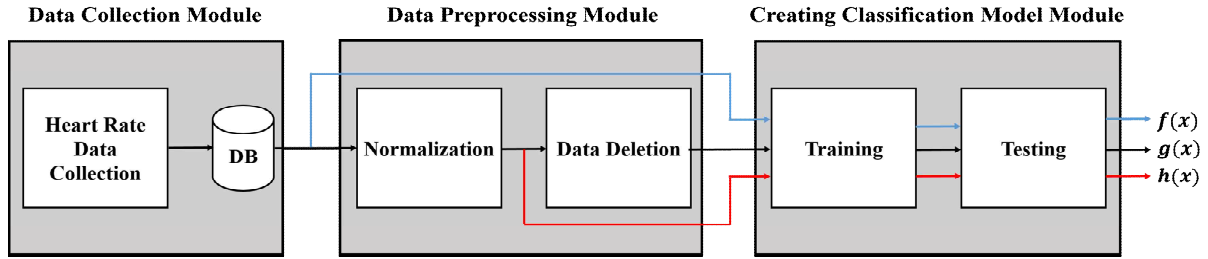


그림 1 구현 시스템 개요  
 Fig. 1 System Overview

$$D(x_{ni,i}) = \sum_j \sqrt{(x_{ni,i} - x_{ni,j})^2} \quad (11)$$

$$\hat{D} = \frac{1}{q} \sum_{i=1}^q D(x_{ni,i}). \quad (12)$$

이때 아래의 식 (13)을 만족하는  $x_{ni,i}$ 를 외부 범주로 분류하고 데이터에서 제외한다.

$$D(x_{ni,i}) > \hat{D} \quad (13)$$

내부 범주로 분류된 데이터  $x_{ni,j}$ 의 집합  $x_d$ 는 데이터 정규화와 데이터 삭제 과정을 거친 데이터이며, 이 데이터를 SVM을 학습 시키는데 이용해 분류 모델  $g(x)$ 를 생성한다.

## 5. 실험 및 결과

### 5.1 시스템 구축

본 논문에서는 실험을 진행할 때 효율적으로 데이터를 수집하고 체계적으로 분류 모델을 생성하기 위한 시스템을 구현하였다. 시스템은 Real-time Traffic Flow Measurement (RTFM)[9]를 기반으로 구성하였으며, 그림 1과 같이 데이터 수집, 데이터 전처리 및 분류 모델 생성의 세 가지 모듈로 구성하였다. 이 시스템에서 데이터 수집 모듈과 데이터 분석 모듈만을 사용하는 분류 모델은  $f(x)$ 이고, 데이터 전처리 모듈의 정규화 과정과 데이터 수집 모듈, 데이터 분석 모듈을 사용하여 만든 분류 모델은  $h(x)$ 이고, 세 모듈 모두를 사용하는 분류 모델은  $g(x)$ 이다.

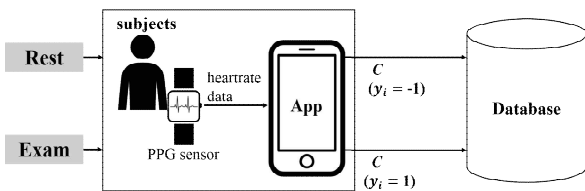


그림 2 데이터 수집 프로세스  
 Fig. 2 Data collection process

### 5.2 RTFM

본 연구에서는 네트워크 환경에서 데이터 트래픽 흐름을 측정하고 보고하는 데 사용되는 구조인 RTFM을 기반으로 시스템을 설계하였다. RTFM 구조는 네트워크의 트래픽 흐름 측정을 위해 설계되었지만 필요에 맞는 설정이 가능하고 모듈화에 적합하다는 특징을 가지고 있어, 모든 유형의 데이터 수집 및 분석 시스템에 응용하기 적합한 구조이다. 구현한 시스템에서 데이터 수집 모듈은 RTFM 구조의 측정기 및 측정 판독기와 같은 기능을 하고, 다른 모듈은 분석 응용 프로그램과 동일한 기능을 한다.

### 5.3 데이터 수집 모듈

이 모듈에서는 그림 2와 같이 3단계로 심박수 데이터 집합  $C$ 를 수집한다. 피실험자의 심박수의 측정은 애플 위치에 내장된 광 혈류량 검사(photoplethysmography, PPG) 센서를 이용하였다. 광 혈류량 검사 센서는 흡수된 녹색 빛의 변화를 통해 심혈관 맥박을 탐지하여 심박수를 측정하는 비 침습적 생체 인식(noninvasive biometric) 방법이다[10]. 실험이 시작되는 순간부터 애플 위치를 통해 지속적으로 데이터를 측정 및 수집한다.

먼저, 피실험자의 심박수를 안정시키고 통계 강의를 보기 전에 휴식을 취하게 하였다. 이때 수집한 데이터는 긴장하지 않은 상태의 데이터이다. 통계 강의를 청강한 후 피실험자의 긴장 상태를 유도하기 위해 해당 내용과 관련된 시험을 보도록 한다. 이때 수집한 데이터는 긴장 상태의 데이터이다. 다음으로, 실험이 진행되는 동안 애플 위치에서 수집한 모든 데이터를 블루투스 4.0을 사용하여 아이폰(iPhone)의 건강 어플리케이션으로 자동으로 전송한다. 마지막으로, 아이폰에 수집한 데이터를 본 저자가 개발한 아이폰용 어플리케이션을 통하여 서버의 데이터베이스에 전송한다. 이때 휴식 상태에서 수집한 데이터는 레이블  $y_i = -1$ 을 지정하고 긴장 상태에서 수집한 데이터는 레이블 값  $y_i = 1$ 을 지정한다. 데이터베이스는 날짜, 심박수 및 사용자 ID, 레이블 등 4개의 열로 구성한다. 이렇게 수집한 데이터는 그림 3과 같다.

### 5.4 데이터 전처리 모듈

이 모듈에서는 수집한 데이터에 데이터 정규화 및 데이터 삭

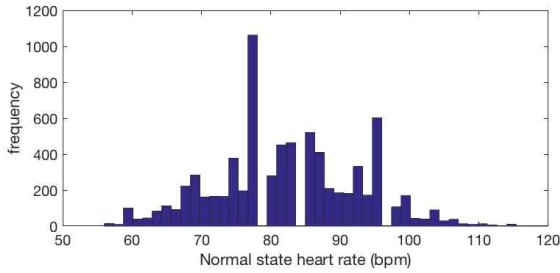
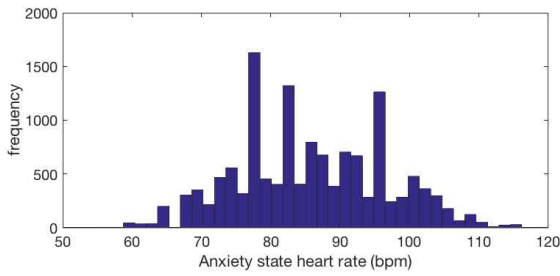


그림 3 수집 데이터  $C$   
 Fig. 3 Collected data  $C$

제 방법을 적용한다. 이 모듈을 거쳐 학습 데이터  $x$ 로부터 데이터 집합  $x_d$ 를 생성한다.

5.5 분류 모델 생성 모듈

이 모듈에서는 학습 데이터로 SVM을 학습해 긴장 상태를 분류하는 모델을 만든다. 그리고 테스트 데이터를 이용해 생성한 모델의 분류 정확도와  $TO$  값을 구한다.

5.6 실험 결과

5.6.1 데이터 수집

총 52명의 실험자를 대상으로 수집한 심박수 데이터를 사용하였으며 이중 27명은 남성이고, 나머지는 여성이었다. 수집된 52명의 데이터 집합은 총 21,253개의 데이터 샘플이며, 데이터에서 가장 많은 비중을 차지하는 개인의 데이터는 645개로 전체 데이터의 3.2%, 가장 적은 비중을 차지하는 개인의 데이터는 25개로 전체 데이터의 1.26%이다.

5.6.2 데이터 전처리

그림 4는 정규화가 된 학습 데이터의 모습이다. 그림 3과 비교해봤을 때 데이터 범위가  $[0, 1]$ 로 변경되었다. 범위가 변경될 뿐 아니라 각 개인에 따른 데이터의 평균, 분산과 같은 특성 역시 변경되었다. 이때, 불안 상태의 데이터의 평균값은 0.520이고, 불안하지 않은 상태의 데이터의 평균값은 0.399이었다.

데이터 삭제를 한 후의 데이터 집합  $x_d$ 의 모습은 그림 5와 같다. 데이터의 삭제를 통해 데이터 밀도가 낮은 부분의 아웃라

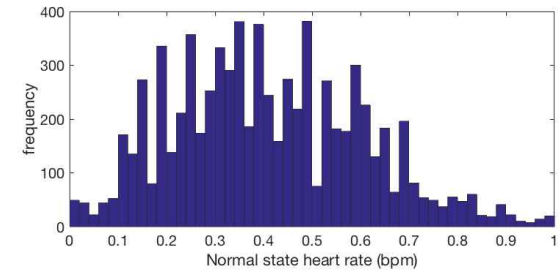
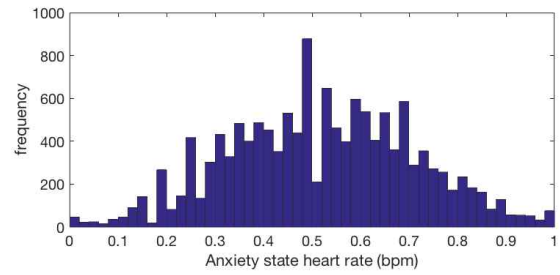


그림 4 정규화 된 데이터  $x_n$ 의 집합  
 Fig. 4 Normalized data set  $x_n$

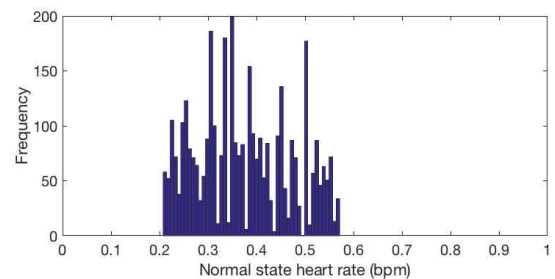
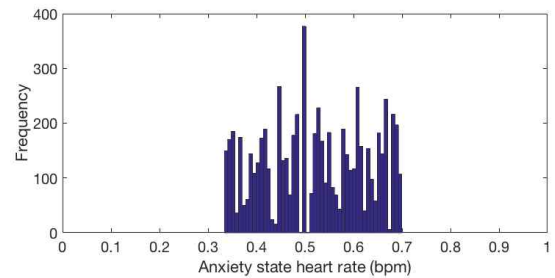


그림 5 데이터 삭제 후 데이터 집합  $x_d$   
 Fig. 5 Data set  $x_d$  after data deletion

이어 데이터 수를 감소시켰다. 데이터 삭제 후에 총 13,570 개의 데이터 샘플이 남게 되었다.

5.6.3 분류 모델 생성

학습 데이터  $x$ ,  $x_n$ 의 집합,  $x_d$ 를 이용해 SVM을 학습하여 분류 모델  $f(x), h(x), g(x)$ 를 생성하였다. 본 실험에 사용된 시

**표 1**  $f(x), h(x), g(x)$ 의 분류 정확도, 연산 속도 및  $TO$  값  
**Table 1** Classification accuracy, computation time and  $TO$  value of  $f(x), h(x)$  and  $g(x)$

Classification model	$f(x)$	$h(x)$	$g(x)$
Accuracy	65.18%	66.97%	78.77%
Computation Time (sec)	67.55	66.63	23.26
$TO$ 값	0.965	1.005	3.387

시스템은 6-Core Mac Pro (Intel Xeon E5 6-Core CPU 3.5GHz, 16GB RAM)의 mac OS Sierra이며 MATLAB R2016b를 이용해 실험하였다.

표 1은  $f(x), h(x), g(x)$ 의 분류 정확도와 연산 시간,  $TO$  값을 보여주며, 분류 정확도는 각각 65.18%, 66.97%, 78.77%이다.  $f(x)$ 는 각 피실험자들의 데이터 특성과 아웃라이어의 영향을 받기 때문에 정확도가 가장 낮은 것을 볼 수 있다.  $h(x)$ 는 정규화 과정을 거치며 각 피실험자들의 특성을 줄이고, 불안 상태의 심박수 데이터의 공통된 특징을 보다 분명하게 함으로써,  $f(x)$ 보다 높은 분류 정확도를 보인다. 하지만 두 모델의 데이터 개수는 동일하기 때문에 연산 시간에는 크게 영향을 미치지 않는다. 반면에  $g(x)$ 는 데이터 삭제 과정까지 거치며 아웃라이어의 영향을 줄여 가장 높은 분류 성능을 보일 뿐 아니라 데이터 개수가 적어지기 때문에 다른 두 모델보다 연산 시간이 짧은 것을 볼 수 있다.

분류 모델을 검증하면  $TO(g(x))$ 의 값은 3.387이고  $TO(h(x))$ 의 값은 1.005로  $g(x)$ 의 경우가 더 커, 식 (4)를 만족하는 분류모델인 것을 볼 수 있다.

## 6. 결 론

본 논문에서는 SVM을 이용하여 피실험자의 심박수 데이터를 바탕으로 불안 감정 상태를 분류할 수 있는 분류 모델을 생성하는 방안을 제안하였다. 또한 모델의 성능을 향상시키기 위해 정규화와 데이터 삭제를 이용하는 데이터 전처리 방안을 제안하였다. 이를 통해 각 피실험자의 데이터 특성을 줄이고 데이터의 중요도를 동일하게 할 뿐만 아니라 아웃라이어가 분류 모델에 미치는 영향을 줄여 보다 정확도가 높은 분류 모델을 생성하였다. 이를 실험적으로 검증하기 위해 체계적으로 분류 모델을 생성하고 데이터를 관리하는 시스템을 RTFM을 기반으로 구현하였다. 시스템에 따라 실제 피실험자로부터 수집한 데이터를 이용해 만든 분류 모델  $f(x)$ 와 데이터 전처리를 거친 후 만든 분류 모델  $g(x)$ 의 정확도는 각각 65.18%와 78.77%로 제안하는 분류 모델의 정확도가 더 높았다. 뿐만 아니라 모델 생성을 위한 연산 시간이 각각 67.66초와 23.26초로, 짧아지는 것을 확인하였다.

## 감사의 글

이 논문은 2015년도 및 2017년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구이며 (NRF-2015S1A5B6036244, No. NRF-2017R1A2B4005041), 과학기술정보통신부 및 정보통신기술진흥센터의 대학ICT연구센터육성 지원사업(IITP-2017-2012-0-00559)의 연구결과로 수행되었습니다.

## References

- [1] S. A. Israel, J. M. Irvine, A. Cheng, M. D. Wiederhold, and B. K. Wiederhold, "ECG to identify individuals," *Pattern recognition*, vol. 38, no. 1, pp. 133-142, 2005.
- [2] K. Takahashi, "Remarks on SVM-based emotion recognition from multi-modal bio-potential signals," *Proceedings of IEEE International Workshop on Robot and Human Interactive Communication(ROMAN4)*, pp. 95-100, Sep. 2004.
- [3] V. N. Vapnik, and V. Vapnik, *Statistical learning theory*, New York: Wiley, 1998.
- [4] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," *Proceedings of the fifth annual workshop on Computational learning theory*, pp. 144-152, July. 1992.
- [5] E. Alpaydin, *Introduction to machine learning*, MIT Press, 2014.
- [6] C. W. Hsu, C. C. Chang, and C. J. Lin, "A practical guide to support vector classification," Technical report, Department of Computer Science, National Taiwan University, Taipei, 2003.
- [7] Z. Nazari and D. Kang, "Density Based Support Vector Machines for Classification," *International Journal of Advanced Research in Artificial Intelligence(IJARAI)*, vol. 1, no. 4, pp. 69-76, 2015.
- [8] L. Bottou and C. J. Lin, "Support vector machine solvers," *Large scale kernel machines*, vol. 3, no. 1, pp. 301-320, 2007.
- [9] N. Brownlee, C. Mills, and G. Ruth, "Traffic flow measurement: Architecture," RFC 2063, *Traffic*, Oct. 1999.
- [10] M. Poh, D. J. McDuff, and R. W. Picard, "Advancements in noncontact, multiparameter physiological measurements using a webcam," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 1, pp. 7-11, 2011.
- [11] P. Shen, Z. Changjun, and X. Chen, "Automatic speech emotion recognition using support vector machine." *Proceedings of the International Conference on Electronic*

and *Mechanical Engineering and Information Technology (EMEIT)*, vol. 2, pp. 621-625, Aug. 2011.

- [12] M. Murugappan, S. Murugappan, and B. S. Zheng. "Frequency band analysis of electrocardiogram (ECG) signals for human emotional state classification using discrete wavelet transform (DWT)." *Journal of physical therapy science*, vol. 25, no. 7, pp. 753-759, 2013.
- [13] C. Guger, H. Ramoser, and G. Pfurtscheller, "Real-time EEG analysis with subject-specific spatial patterns for a brain-computer interface (BCI)," *IEEE transactions on rehabilitation engineering*, vol. 8, no. 4, pp. 447-456, 2000.
- [14] D. F. Polit, *Statistics and Data Analysis for Nursing Research 2e*, Pearson, 2009.
- [15] J.M. Notterman, W. N. Schoenfeld, and P. J. Bersh, "Conditioned heart rate response in human beings during experimental anxiety," *Journal of Comparative and Physiological Psychology*, vol. 45, no. 1, 1952.
- [16] S. Tobias, "Anxiety research in educational psychology." *Journal of Educational Psychology*, vol. 71, no. 5, pp. 573-582, 1979.
- [17] R. M. Liebert, and L. W. Morris. "Cognitive and emotional components of test anxiety: A distinction and some initial data," *Psychological reports*, vol. 20, No. 3 pp. 975-978, 1967.
- [18] D. Anguita, A. Ghio, S. Ridella, and D. Sterpi, "K-Fold Cross Validation for Error Rate Estimate in Support Vector Machines," *DMIN*, pp. 291-297, 2009.

## 저 자 소 개



### 홍 은 재 (Eunjae Hong)

2015년 2월 이화여자대학교 공학사  
2017년 2월 이화여자대학교 공학석사  
2017년 3월~현재 이화여자대학교 전자전기  
공학과 멀티미디어 통신 및 네트워크 연구실  
연구원

Email : eunjae.hong@ewhain.net



### 박 형 곤 (Hyunggon Park)

2004년 2월 포항공과대학교 공학사  
2006년 3월 미국 University of California,  
Los Angeles (UCLA) M.S.  
2008년 12월 미국 University of California,  
Los Angeles (UCLA) Ph.D.

2010년 2월 스위스 Swiss Federal Institute  
of Technology, Lausanne (EPFL) 신호처리  
연구실 선임연구원

2010년 3월~현재 이화여자대학교 전자전기  
공학과 부교수

Email : hyunggon.park@ewha.ac.kr