

개인화된 신호 해석을 위한 맥락 기반 생체 신호의 모델링 기법

↓

Physiological signal Modeling for personalized analysis

↙↘

최아영*, Ahyoung Choi*, 우운택**, Woontack Woo**

요약 ~ 일상생활에서 활용 가능한 다양한 종류의 생체 신호 획득 및 분석 방법이 연구되고 있다. 기존의 생체 신호 분석 방법은 표준화된 임계치를 사용하여 해석한 결과를 제공하며 신호 측정 당시의 상황이 고려되지 않아 잡음 혹은 외부 환경의 영향을 받기 쉬운 단점이 있다. 본 논문에서는 생체 신호뿐만 아니라 기타 정황정보를 기반으로 하여 개인화된 신호를 분석하기 위한 모델(Personalized Decision Making method, PDM)을 제안한다. 개인화된 신호 해석 모델은 사용자의 맥락 정보, 사용자의 맥락 정보, 사용자의 나이, 성별, 현재의 몸 및 정신 상태, 음식 및 카페인 섭취 여부, 측정 시간 및 측정 요일 등을 기반으로 각 맥락 간의 연관 관계를 나타내고, 이상적인 사용자의 생체 신호 예측치를 베이스 정리를 기반으로 획득한다. 개인화된 해석 모델(ACM)을 통해 표준 임계치를 적용한 해석에 비해 인식의 정확도를 높일 수 있으며, 다양한 측정시의 조건을 알면 현재 사용자의 건강상태를 개인화된 분석과 유사한 정확도로 예측이 가능하다. 제안한 방법은 현재 관측된 관측치의 분포를 모르더라도, 현재 사용자의 상태를 맥락정보를 기반으로 하여 예측할 수 있으므로, 일반적인 데이터 모델을 기반으로 개개인에 맞는 얼굴 표정을 인식하는 연구 등에 활용이 가능하다.

Abstract ~ With the advent of light-weight daily physiological signal monitoring sensors, intelligent inference and analysis method for physiological signal monitoring application, commercialized products and services are released. However, practical constraints still remain for daily physiological signal monitoring. Most devices provide rough health check function and analyze with randomly sampled measurements. In this work, we propose the probabilistic modeling of physiological signal analysis. This model represent the relationship between previous user measurement (history), other groups type, model and current observation. From the experiment, we found that the personalized analysis with long term regular data shows reliable result and reduces the analyzing errors. In addition, participants agree that the personalized analysis shows reliable and adaptive information than other standard analysis method.

↓

핵심어: *Physiological signal analysis, Personalized analysis, Context awareness*

본 연구는 문화체육관광부 및 한국문화콘텐츠진흥원의 문화콘텐츠기술연구소육성사업의 연구결과로 수행되었음

*주저자 : 광주과학기술원 정보통신공학과 박사과정; e-mail: achoi@gist.ac.kr

**교신저자 : 광주과학기술원 정보통신공학과 교수; e-mail: wwoo@gist.ac.kr

1. 서론

기존의 신호를 해석하는 연구에서는 신호의 잡음을 최소화 하고, 복잡도를 줄이기 위해 현재의 데이터가 과거 획득된 데이터 셋에 영향을 미친다는 가정아래 회귀분석

등의 통계적 방법이 제안되었다[1]. 또한, 시간 복잡성을 최소화 하기 위해 주파수 공간에서 분석하는 방법 등이 활용되기도 하였다[2]. 그러나 일상생활에서 모니터링 하는 센서의 경우, 생체 신호가 측정 환경과 사용자에 따라 신호의 변화가 크므로 신호 해석의 단계에서 신호가 획득된 정황 정보 및 사용자 정보도 복합적으로 고려가 되어야 한다[3]. 본 연구에서는 신호의 복잡성을 고려하기 위해 맥락을 기반으로 한 신호 해석 모델을 제안한다. 제안한 모델은 사용자의 맥락 정보, 사용자의 맥락 정보, 사용자의 나이, 성별, 현재의 몸 및 정신 상태, 음식 및 카페인 섭취 여부, 측정 시간 및 측정 요일 등의 맥락정보를 기반으로 현재 사용자의 데이터 분포를 모르더라도 신뢰성 있는 건강상태를 분석할 수 있도록 한다. 제안한 방법은 기존의 모델을 통해 현재 사용자의 상태를 예측할 수 있으며, 이를 통해 신호 해석을 위한 학습의 단계를 최소화 할 수 있는 장점이 있다.

2. 맥락 기반의 개인화된 신호 해석 모델

사용자의 이전 히스토리 정보와 다른 사용자의 모델에 대한 타입 등의 정보를 통해 현재 사용자의 이상적인 신호 패턴을 추출하여 신호의 정상 범위를 파악하는 방법(Personalized Decision Making method, PDM)을 제안한다. 그림 1 은 개인화된 분석을 위한 순서를 나타낸다. 사용자로부터 생체 신호 데이터를 획득할 때 획득 당시의 맥락정보를 함께 기록하여 사용자의 히스토리에 태그의 형태로 저장한다. 저장된 정보를 통해 기존의 사용자 모델이 있는 경우 파라미터를 초기화 하고, 초기화된 파라미터와 현재 입력된 맥락 정보를 기반으로 사용자의 확률밀도 함수를 구하게 된다. 구해진 확률 밀도 함수는 MAP criteria 에 따라 상태를 분석하는데 적용이 된다. 베이스 확률 이론을 기반으로 하여 신호 해석 모델을 제안하는 이유는 각 맥락 정보와 측정된 관측치 간에 모호성이 존재하기 때문이다[4].

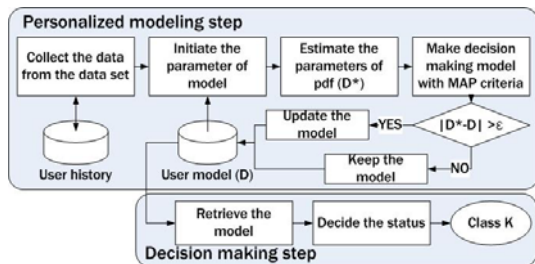


그림 1 개인화된 판단을 위한 프로시저

데이터의 분포는 정상 사람의 경우 정상치 값의 분포를 가질 확률이 가장 크므로 확률 분포 함수는 정상치 값이 최대이며, 비정상적으로 적거나 크게 나오는 확률이 대칭적이고 적다고 가정하여, 가우시안으로 정의한다. 사용자의 생체 신호 값의 확률을 최대화 할 수 있는 값을 추출하고 활용한다. 이를 위해서 본 논문에서는 UCAM 구조를 활용하여 정보를 저장하고, 맥락의 형태로 보존하는

프레임워크를 활용하여 개인화된 생체 신호 해석을 확률적으로 모델링 하는데 활용한다. 본 논문에서는 그림 2 와 같이 UCAM 프레임워크를 활용하여 이전 히스토리 정보와 사용자의 프로파일 정보를 생체 신호와 함께 저장하고, 맥락화 하여 개인화된 생체 신호 해석의 확률적으로 모델링에 적용한다[5].

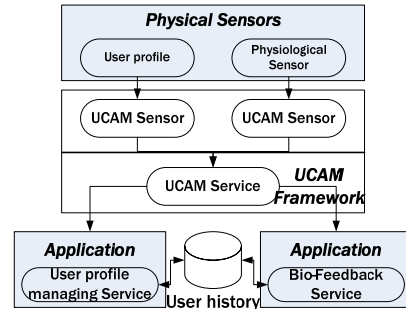


그림 2 생체신호 관련 데이터 획득을 위한 맥락 인식 프레임워크[2]

본 논문에서는 추정하고자 하는 신호로 맥파 신호를 사용하였다. 정상 사람의 생체 신호의 경우 정상 범위의 신호값이 관측될 확률이 가장 높고, 정상이 아닌 범위의 신호 값이 관측될 확률이 낮고, 정상 범위 이상 혹은 이하의 관측치가 관측될 확률이 비슷한 수준으로 낮게 관찰이 되므로 확률밀도 함수는 가우시안으로 가정한다. 개인화된 해석을 위한 첫 단계로, 관측치로부터 이상적인 데이터를 획득하기 위해 신호의 모델을 통해 관측치로부터 추정치를 예측한다. 맥파 신호는 시간에 따른 변화율이 크지 않다. 데이터를 예측하기 위해서는 관측치가 주어졌을 때 예측한 값의 확률을 최대화 할 값을 선택한다.

$$\arg \max_{d^*} P(d^* | d_{obs}) \propto \arg \max_{d^*} \{P(d_{obs} | d^*) \times P(d^*)\} \quad (1)$$

d^* 는 추정하고자 하는 데이터 셋이며, d_{obs} 는 현재 관측된 데이터를 나타낸다. 데이터의 확률밀도 함수는 가우시안을 만족하고, 조건부 확률도 가우시안을 만족하므로, 에러를 최소화 할 수 있는 신호가 예측하고자 하는 신호가 된다. 즉, 예측치를 통해 관측치 값을 최대화 할 수 있는 확률은 에너지를 최소화 하는 경우로 표현이 가능하다. 따라서 이상적인 신호 모델의 정의에 따라 관측치와 예측치 간의 신호 값의 차이가 적으면서 예측치는 신호의 변화율이 급격하지 않은 값을 여러 후보 값 중에 선택하여 최종 예측값을 획득한다. 단, 예측치와 관측치는 유사하며, 노이즈는 신호에 영향을 미치지 않을 정도로 작다고 가정한다.

$$\arg \min_{d^*} \{e_1 + e_2\} = \arg \min_{d^*} \{|d_{obs} - d^*|^2 + \nabla d\} \quad (2)$$

맥락에 기반한 신호의 해석은 구하고자 하는 특징의 확률 분포를 기반으로 획득된다. 본 논문에서는 맥파신호를 통해 RR 간격 값의 평균값을 신호 해석의 기준으로 사용한다. 정상범위의 사용자에 대한 확률밀도 함수를 바탕으로 판단 하는 임계치를 결정하는데, 정상인 사람의 생체 맥파는 정상 범위에서 발견될 확률이 높으므로

가우시안으로 가정한다. 임계치는 평균과 표준편차의 함수로 나타낼 수 있으며, 이는 타입과 히스토리 값에 영향을 받는다.

3. 실험 및 분석

사용자의 상태는 크게 정상과 비 정상으로 구분하며 본 논문에서는 정상인 사용자의 상태를 획득하고 이를 확인하기 위한 방법과 절차를 적용하였다. 실험에서는 총 12 명의 심장과 관련한 질병이 없는 정상 사용자를 대상으로 생체 신호 센서를 BIOPAC 과 본 연구실에서 이전에 개발한 UMPC 타입의 센서를 사용하여 수집하였다. 12 명의 사용자 중 10 명은 남성, 2 명은 여성이며 모든 사용자가 심장과 관련한 병력이 없으며 정상 상태를 확인하였다. 하루에 같은 시간에 각 센서당 5 분씩, 총 10 분 동안 신호를 수집하였다. 신호 수집 시 외부 환경의 영향을 최소화 하기 위해 사용자에게 측정하는 동안 매번 같은 종류의 음악 혹은 영상매체를 시청하도록 하였다. 또한 음식에 의한 영향과 운동에 의한 영향을 최소화 하기 위해 음식을 섭취한 시간과 운동을 한 시간을 기록하였으며, 음식의 경우 카페인의 섭취가 심장 박동수에 영향을 미치므로 이에 대한 조사도 추가하였다. 사용자 조사로 측정 후 사용자의 나이, 성별, 현재 사용자가 느끼는 심장 박동수의 빠르기 정도와, 땀이 나는 정도, 체온의 정도를 기록하게 하였다.

실험에서 획득한 기본정보를 바탕으로 신호를 수집하였는데, 획득된 신호가 stationary 하며 대표성을 띄는지 확인하기 위해 일별 데이터를 분석하였다. 분석 전 신호의 outlier 를 Grubbs' test 를 통해 제거하였다. 또한 신호의 노이즈를 가우시안이라 가정하고 신호 노이즈를 제거하기 위하여 smoothing 방법과 칼만 필터를 그림 3 와 같이 획득된 신호에 적용하였다.

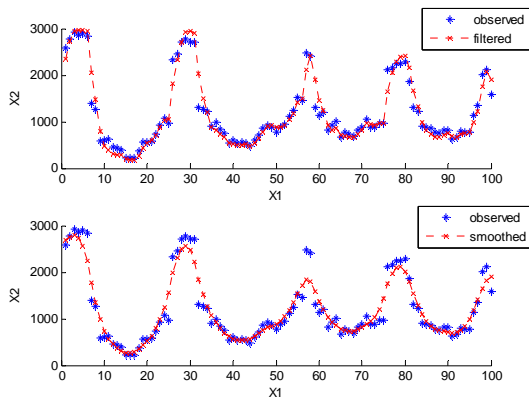


그림 3 노이즈 필터링 (위) 칼만필터 적용 (아래) smoothing 방법을 적용

신호를 estimation 한 결과, 칼만 필터를 적용한 경우가 예측하고자 하는 그래프와 신호 사이의 유클리디안 거리를 최소화 할 수 있는 방법이나, 실제 본 실험에서 사용하는 측정값인 심박수를 계산하기 위해서는 에러를 최소화 하는 방법보다 첨두치를 정확하게 예측하는 것이 필요하므로 본 논문에서는 smoothing 기법을 사용하여 노이즈를 필터링

하였다. Smoothing을 위한 적절한 창 크기를 결정하기 위해 본 논문에서는 크기의 종류를 다양화 하였고, smoothing window의 크기가 3인 경우가 가장 안정적으로 첨두치를 찾아내면서 신호의 왜곡도 최소화 할 수 있는 지점이므로 창 크기를 3으로 설정하여, 신호를 전처리 하였다. 창 크기에 따른 신호의 차이는 그림 4와 같다.

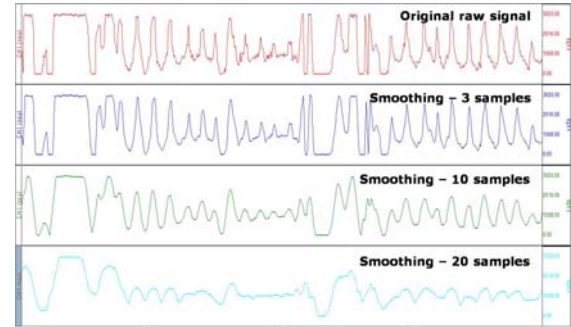


그림 4 다양한 smoothing 창을 적용한 데이터

전처리 된 신호는 특징 추출하는 단에서 특징을 추출하게 된다. 맥파에서 추출한 신호는 심장 박동수, 한 주기 내 맥파의 최대/최소 변화율, 최대값, 최소값, 심장 박동수의 변화율 등이다. 개인의 심장의 상태를 판단하는 판단 모듈에서 추출한 특징의 분포를 정규 분포로 가정하였으므로, 본 논문에서는 추출한 신호 중 Kolmogorov-Smirnov test 를 적용하여 추출한 특징이 정규 분포를 만족하는지 확인하였다. 가능한 특징 벡터 중 가우시안 분포를 나타내는 특징으로 대부분의 피험자의 경우 심장박동수의 데이터에서 정규성을 만족하였으므로, 특징으로 심장 박동수를 선택하였다. 심장 박동수는 피크간의 간격을 사용하여 계산하였으며, 피크는 5 분간의 데이터 중 호흡이나 움직임에 대한 신호의 변화량이 적으며 안정적으로 신호의 변화가 유지되는 구간의 100 개 샘플 신호를 사용하였다. 본 논문에 참여한 피험자로부터 추출된 특징 값은 그림 5 와 같이 나타내었다.

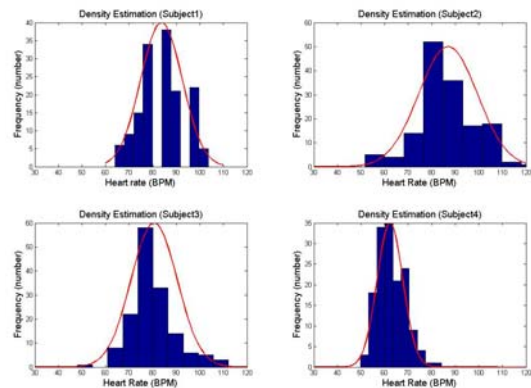


그림 5 관측치로부터 추출한 생체 신호 분포

추출한 신호는 판단 경계를 그룹별, 개인별 조사된 판단 임계치를 활용하여 정상인지 비 정상인지를 구분하게 된다. 그림 6 은 개인화, 그룹화 및 표준의 정상 분포를 판단하기

위한 임계치를 나타낸다. x 축은 심장박동수의 분포를 나타내며, y 축은 획득된 데이터 들의 확률 분포를 나타낸다. 그림에서 r 검은색의 분포는 심장 박동수의 정상 여부를 판단할 때 일반적으로 사용하는 표준 범위를 나타내며 60bpm~80bpm 까지를 나타낸다. 푸른색, 동그란 선으로 표시된 데이터 분포는 획득된 데이터를 남성으로 구분하여 추출한 데이터 분포를 나타내며 이 분포를 기반으로 한 판단 임계치는 48bpm ~93bpm 까지 이다. 붉은색 별로 표시된 그래프는 맥락을 기반으로 한 그룹의 분포 중 여성 그룹을 바탕으로 한 분포를 나타내며 붉은 실선 그래프는 여성인 개별 피험자의 분포를 나타낸다. 여성인 경우 59bpm 와 97bpm 의 임계치를 갖는다.

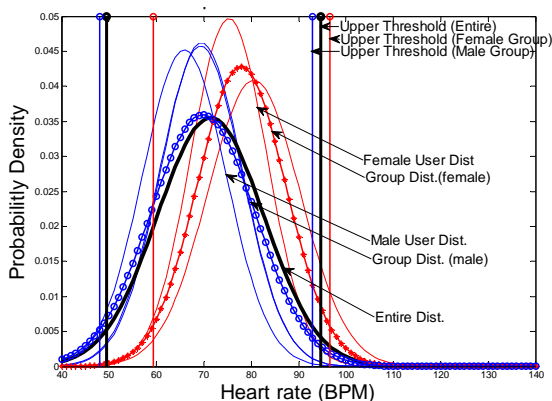


그림 6 생체신호 분석을 위한 그룹별 임계치

그림 7 은 생체 신호 판단의 분석 결과를 나타낸다. 정상과 비 정상 상태에 대한 분석 결과 그림 7 과 같이 나타난다. 표준 임계치 (Standard Threshold)는 심장 박동수의 표준 범위를 임계치를 나타내며, 개인화된 임계치 (Individual Threshold)는 개인의 분포를 바탕으로 하여 도출한 판단 임계치를 기반으로 한 분석을 나타낸다. 그룹 임계치는 아래 그림에서는 남성과 여성으로 나눈 그룹간의 임계치를 나타내었다.

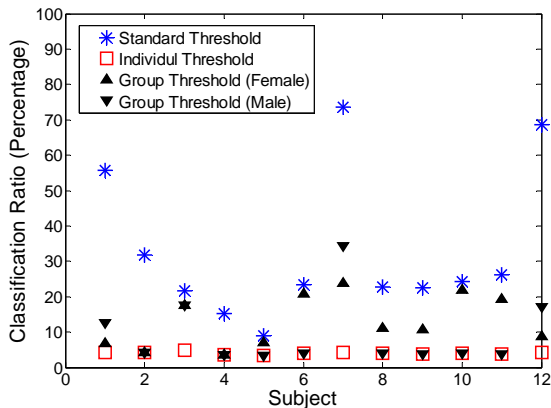


그림 7 생체신호 판단의 어려움

생체 신호 판단의 분석율은 표준화된 임계치를 적용하는 경우 정상인 사용자 임에도 불구하고 최대 74%에서 비정상이라 판단하는 어려가 발생함을 확인할 수 있었다. 사람에

따라 데이터의 분포가 변화하므로 실제 사용자가 정상인 경우에도 비정상의 데이터를 보이는 경우가 있음을 확인할 수 있다. 분석 결과 개인화된 해석의 경우 판단을 개인의 정상인 분포를 기반으로 가정하였으므로 최대의 인식률을 보였으며, 그룹을 나누어 분석한 경우 표준화된 분석보다는 정확한 분석을 나타냄을 확인할 수 있었다. 그림 8은 다양한 그룹핑을 위한 맥락정보의 종류를 나타낸다. 그룹을 나누는 기준은 신호 획득 시에 함께 기록한 다양한 맥락 정보들을 기반으로 한다. 이번 실험에서는 사용자의 성별, 나이, 측정요일, 사용자가 느끼는 몸의 상태, 음식 섭취나 카페인 섭취 여부를 맥락정보로 활용할 수 있다. 아래 그림에서 신호의 분석에 영향을 미치는 요소로는 피험자의 타입, 나이, 측정 시간, 피험자의 컨디션, 음식 및 카페인의 섭취 여부에 의해 영향을 받음을 알 수 있으며 이러한 정보를 기반으로 생체 신호를 해석하는 경우 더 개인화된 해석과 유사한 성능을 보이는 분석 결과를 확인할 수 있다.

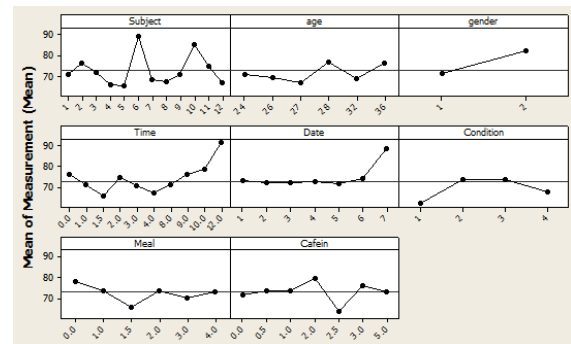


그림 8 개인화된 생체 신호 분석에 영향을 미치는 다양한 맥락 정보

4. 결론 및 향후 과제

본 논문에서는 맥락 정보를 기반으로 한 개인화된 생체 신호 분석 방법을 제안하였다. 제안한 개인화된 신호 해석 모델은 베이지 정리를 기반으로 각 맥락 요소와 생체 신호와의 관계를 모델링 하였다. 12 명 피험자를 대상으로 한 생체 신호 획득 및 분석 실험을 통해 개인화된 해석 모델(PDM)이 단일화된 해석에 비해 해석의 정확도를 높일 수 있으며, 다양한 조건(사용자의 맥락 정보, 사용자의 나이, 성별, 현재의 몸 및 정신 상태, 음식 및 카페인의 섭취 여부, 측정 시간 및 측정 요일 등)을 미리 고려하면 정확도 높은 신호 분석을 위한 예측이 가능함을 보였다. 제안한 방법은 맥락 정보를 활용하여, 현재 관측된 관측치의 모델을 예측할 수 있으므로, 기존의 참조 가능한 데이터를 이용하여 현재 데이터 모델을 예측하는 얼굴표정 인식 등의 연구에 응용이 가능하다.

참고문헌

[1] I.A. Rezek, S.J. Roberts, " Stochastic Complexity Measures for Physiological Signal Analysis," IEEE Trans. Biomedical Eng. vol. 44, no. 9, (1998).

- [2] W. A. Chaovaitwongse, O. A. Prokopyev, P. M. Pardalos, "Electroencephalogram (EEG) time series classification: Applications in epilepsy," *Annals of Operations Research*, vol.148, pp. 227–250, (2006).
- [3] W. H. Wu, M. A. Batalin, L. K. Au, A. Bui, and W. J. Kaiser, "Context-aware Sensing of Physiological Signals," In *Proc. of IEEE EMBS 2007*, pp.5271 – 5275, IEEE Press, New York (2007).
- [4] M. Öztürk, and A. Tsoukiàs, "Modeling uncertain positive and negative reasons in decision aiding," *Decision Support Systems*, vol.43, pp.1512–1526, (2007).
- [5] Y. Oh, W. Woo, "A Unified Application Service Model for ubiHome by Exploiting Intelligent Context-Awareness," In: Murakami, H., Nakashima, H., Tokuda, H., Yasumura, M. (eds.) *UCS 2004. LNCS*, vol. 3598, pp. 192–202. Springer, Heidelberg (2005).