

ARIMA 모델 기반 웨어러블 센서 데이터의 특징 벡터 추출 방법론

황보선¹, 박철수², 장병탁¹

서울대학교 컴퓨터공학부¹, 광운대학교 컴퓨터 공학과²
{bshwang, [btzhang](mailto:btzhang@bi.snu.ac.kr)}@bi.snu.ac.kr¹, parkcheolsoo@kw.ac.kr²

The ARIMA model based feature extraction methodology for wearable sensor data

Bosun Hwang¹, Cheolsoo Park², Byoung-Tak Zhang¹

School of Computer Science and Engineering, Seoul National University¹
Computer Engineering Department, KwangWoon University²

요 약

전자 기기의 SoC화와 저전력 설계, 그리고 이를 구동하는 모바일 운영체제의 발달 및 무선 기술의 발달은 현대 사회의 생활 양식에 많은 영향을 미치고 있다. 이러한 맥락에서 웨어러블 센서의 등장은 그 활용도에서 많은 기대를 나타내고 있는데, 이미 스마트폰의 보급이 일반화된 상황이며, 스마트 시계 역시 그 개발이 활발한 상황이다. 본 논문은 이러한 모바일 디바이스 상에서 측정이 가능한 시계열 웨어러블 센서 데이터의 통계 모델을 활용한 특징 벡터 추출 방법에 대해 논한다. 제안된 방법론은 가속도 센서 및 피부전도도 데이터의 특징 벡터 추출을 시계열 데이터의 통계 회귀 모델인 ARIMA 모델에 대해 적용하여 추출한다. 본문에서는 ARIMA 모델 적용을 위한 전처리 기법 및 모델 적합 검정에 대해 논하며, ARIMA 모델의 계수를 특징 벡터로서 최종 결과물을 도출한다. 본 논문의 실험 결과는 자체 측정된 일상 생활에서의 3축 가속도 센서 데이터와 동시에 측정된 피부전도도 데이터를 사용하였으며, 최종 검정 결과는 제안된 ARIMA 모델 기반의 특징 벡터 추출 방법이 웨어러블 센서 데이터를 대변하는 특징 벡터로서 적합하다는 결과를 보인다.

1. 서론.

모바일 디바이스의 대중적인 보급은 일반적인 생활 속에서 기존 전자기기의 활용과는 다른 패러다임 전환을 의미한다. 단순한 통신과 메시지 전달의 기능을 넘어서 이제는 현대인들의 생활 전반에 관여하며 그 응용 분야들이 폭발적으로 증가하고 있는 상황이다.

개인의 일상적인 생활 패턴 분석 역시 이러한 흐름 속에서의 커다란 응용 분야로서, 여러 인공지능을 활용할 수 있는 흥미로운 관심 대상이다.

본 논문은 이러한 인공지능 알고리즘 적용의 전 단계로서 대표적인 웨어러블 센서 데이터로서 가속도 센서와 피부전도도를 활용한 통계적 특징 벡터 추출 기법을 제안한다.

웨어러블 센서 데이터는 그 특성상 많은 잡음 상황에 노출 될 수 있으며, 이에 적합한 신호처리 및 추출 방법론이 요구 된다. 본 논문에서는 MEMD 알고리즘을 활용하여 센서 데이터의 주요 성분 신호를 추출하였으며, 기존의 신호 피크 추출과 같은 단순한 방법을 넘어서, 통계적인 시계열 회귀 모델인 ARIMA 모델을 적용하는 방법을 제안하였다[6].

실험 결과는 ARIMA 모델의 적합성 검정 실험 결과를 통해 제안된 특징 벡터 추출 기법이 웨어러블 센서 데이터에 대해 적합함을 보인다.

2. ARIMA 모델

ARIMA 모델은 정상성(Stationarity)을 보이는 시계열 데이터에 적합한 전통적인 통계 회귀 모델이다. 자기 회귀 모델(Auto Regressive: AR)과 이동 평균 모델(Moving Average: MA)의 결합 형태로 이루어지며, 경우에 따라 신호 정상성 특성을 얻기 위해, 원래 신호의 미분 및 log를 취한 신호를 사용하기도 한다[6]. 본 논문에서는 ARIMA 모델 사용을 위해 대표적인 통계 프로그램인 R을 활용하였으며, 'auto.arima()' 함수를 활용하여 신호마다 적합한 ARIMA 모델의 차수를 얻은 후 이를 활용하여 모델을 구축하였으며, 구축된 모델은 통계적인 적합 검정을 통해 그 적합성을 판정하였다.

3. 신호 전처리 기법

본 논문에서는 일상적인 생활 속에서 17분간 32Hz로 스마트 시계를 통해 측정된 4종의 웨어러블 센서

데이터(3축 가속 센서, 피부전도도: GSR)를 사용하였다. 이종의 센서 신호는 각기 다른 주파수 특성을 보이지만, ARIMA 적합에 따른 특징벡터 추출이라는 관점에서 피부전도도의 주요 주파수 특성이 1Hz 미만의 저주파 특성이라는 점은 적합성 실험을 통해 맞지 않는다. 이에 따라 두 종류의 센서 신호를 모두 웨어러블 장비에서 지원하는 최대 샘플링 주파수와 아티팩트 등을 고려하여 8Hz 저주파 필터로 MEMD 알고리즘을 적용하여 사용하게 된다. 가속도 센서의 경우도 피부전도도와 동일한 기본적인 MEMD 필터링과 전처리를 통해 함께 준비하여, 전체 측정된 데이터를 5초 윈도우로 나누어 총 213세트의 실험 데이터를 준비할 수 있었다[1].

측정된 신호들의 공통의 신호 특성을 지닌 신호를 추출하기 위하여 multivariate empirical mode decomposition(MEMD) 알고리즘을 사용하였다 [3]. MEMD 알고리즘은 시계열 데이터의 주요한 주파수 성분을 기존 디지털 필터 (FIR 또는 IIR) 나 wavelet 변환보다 더 효율적으로 추출해 낸다고 알려져 있다 [3]. 기존 디지털 필터나 wavelet 변환 알고리즘들은 코사인 함수나 mother wavelet 함수와 같은 basis 함수에 의존하여 비선형적인 뇌전도 신호를 선형적으로 분석을 하지만, MEMD 알고리즘은 입력 신호 자체에서 적응적으로 basis 함수를 추출하여 비선형적인 시계열 신호를 보다 효율적으로 분석할 수 있다. 또한 기존 분석 알고리즘들은 단채널 분석 방법인데 반해 MEMD 알고리즘은 다채널 데이터를 동시에 분석하여 여러 채널에 동시에 나타나는 성분들을 보다 정확히 추출해 낼 수 있다. MEMD 신호 처리 알고리즘은 MATLAB상에서 구현되어 수행되었다.

4. ARIMA 모델 적합성 검증

웨어러블 센서 데이터의 ARIMA 모델 적합성을 검증하기 위하여, 시계열 데이터의 정상성 판정 테스트(Ljung-Box, Augmented Dickey-Fuller test) 및 Q-test를 이용한 적합 검정을 수행하였다. 통계 프로그램인 R이 사용되었으며, 실험 결과에서와 같이 정상성 테스트 후 p-value 값의 유의성 여부를 확인하고, 적합 판정 자체는 회귀 모델의 잔차 특성이 백색 잡음 특징을 보이게 되면 생성된 모델이 적합하다는 결론을 내림으로써 적합성 유무를 판별하게 된다[6].

5. 실험 결과

5.1 신호 정상성 검증

ARIMA 모델 적용을 위해서는 먼저 시계열 데이터가 신호 정상성 특성을 유지해야 한다. ARIMA 모델 자체가 정상성 시계열 데이터에 대한 회귀 모형이기 때문에, 실험을 통해 정상성 유무를 판정하였다. 실제 많은 시계열 신호의 경우, 신호 자체가 정상성을 갖지 못하는 경우가 있으므로 신호 특성에 따라 가변적으로 차분된

신호나 log를 취하여 사용하게 된다[6]. 웨어러블 센서 데이터의 경우도 샘플별 정상성 유무에 따라 차분된 신호를 사용하였다.

본 논문에서는 대표적인 정상성 판정 테스트인 Ljung-Box 테스트와 Augmented Dickey-Fuller(ADF) 테스트를 통해 검증하였는데, 두 테스트 모두 p-Value가 낮을 때 해당 신호가 정상성을 나타낸다. p-Value 확인 결과 아래와 같이 통계적 유의 수준에서 거의 대부분의 신호가 정상성 신호 특성을 보이고 있는 것을 확인할 수 있었다.

Ljung-Box test	p-Value Mean	p-Value Median	# of > 0.05	% of > 0.05
X	0.000	0.0000000	0	100.00
Y	0.000	0.0000000	0	100.00
Z	0.000	0.0000000	0	100.00
GSR	0.000	0.0000000	0	100.00

표 1) 신호 정상성 판정 결과 (Ljung-Box test)

ADF test	p-Value Mean	p-Value Median	# of > 0.05	% of > 0.05
X	0.020	0.0100000	9	95.77
Y	0.027	0.0100000	16	92.49
Z	0.013	0.0100000	5	97.65
GSR	0.022	0.0100000	12	94.37

표 2) 신호 정상성 판정 결과 (ADF test)

5.2 ARIMA 모델 적합 검증

웨어러블 센서 신호의 전처리를 끝내고 나면, ARIMA 모델을 구축하게 되는데, 이때 적절한 AR 모델 차수와 MA 모델 차수를 얻어내어 이에 맞는 ARIMA 모델 계수를 얻어내게 된다. 이 때 구축된 ARIMA 모델이 적합하게 잘 구성이 되었는지 아래와 같이 원래 신호와 예측된 신호와의 잔차 특성과 Q-test를 통한 백색 잡음 여부를 확인하게 된다. 즉, 적합이 잘 되었을 때, 잔차 자체가 백색 잡음과 같은 형태로 보이며, 이때 ACF(Auto Correlation Function)은 Lag '0'에서만 상관성을 나타내게 된다. 맨 아래 그림의 Q-test 결과는 p-value가 유의 수준을 벗어나는지 여부를 확인하는데, p-value가 클수록 해당 신호가 백색잡음임을 보이게 되고, 이는 구축된 ARIMA 모델이 적합함을 의미한다[6]. 실험을 통해 4종의 신호 모두 적합하게 모델 구성이 되었음을 확인할 수 있었다. 이렇게 구축된 ARIMA 모델의 AR, MA 모델 계수가 해당 신호를 대변할 수 있으며, 이는 차후 머신 러닝 알고리즘을 적용한 응용에 주요한 특징벡터로 사용될 수 있을 것이다.

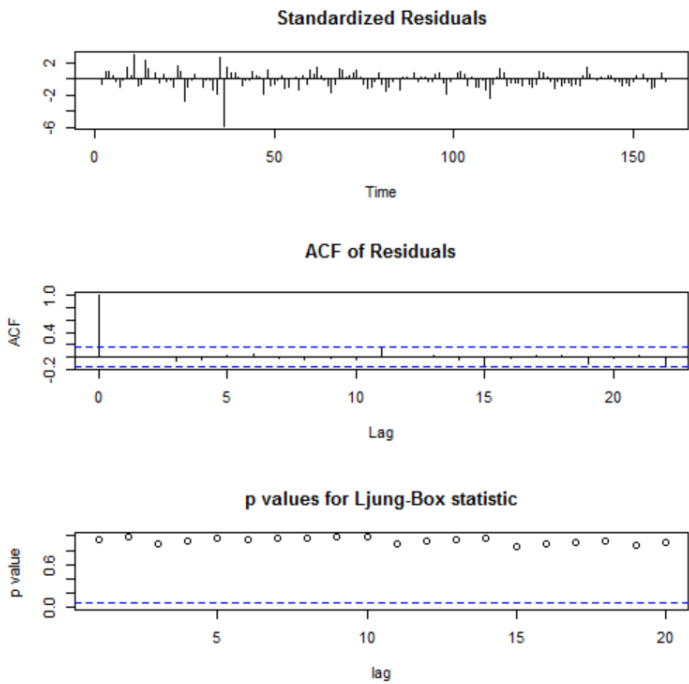


그림 1) ARIMA 모델 적합성 판정 결과 예제 (가속도 센서)

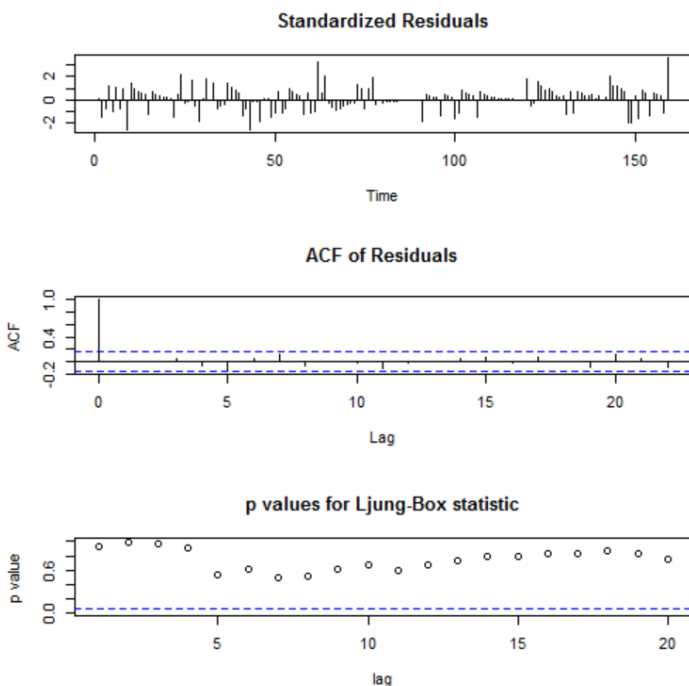


그림 2) ARIMA 모델 적합성 판정 결과 예제 (GSR 센서)

6. 결론 및 향후 연구 방향

본 논문에서는 웨어러블 센서 데이터의 특징 벡터 추출 문제에 대해, MEMD 알고리즘 적용 및 시계열 회귀 모델인 ARIMA 모델을 이용하는 방법론을 제안하였다. 제안된 방법론은 ARIMA 모델이 웨어러블 센서 데이터의 특징 벡터 추출에 적용될 수 있으며, 모델

적용을 위해 제안된 신호 전처리 기법이 ARIMA 모델 적합에 주요함을 실험을 통해 확인할 수 있었다.

향 후 연구 과제로는 현재 측정된 웨어러블 센서 데이터의 태깅을 통한 클래스 정보를 얻어내어, 일상 생활 속의 웨어러블 센서 데이터를 이용한 생활 패턴 학습 방법론까지의 확장 연구가 진행될 예정이다.

8. 참고 문헌

- [1] Alkane Sano, Rosalind W. Picard, "Recognition of Sleep Dependent Memory Consolidation with Multi-modal Sensor Data", IEEE 2013
- [2] J.T. Turner, A. Page, T. Mohsenin, and T. Oates, "Deep Belief Networks used on High Resolution Multichannel Electroencephalography Data for Seizure Detection," 2014 AAAI Spring Symposium Series, 2014
- [3] C. Park, D. Looney, N. Rehman, A. Ahrabian, and D. P. Mandic, "Classification of Motor Imagery BCI Using Multivariate Empirical Mode Decomposition", IEEE Transactions on neural systems and rehabilitation engineering, January 2013
- [4] Chung-Yeon Lee, Byoung-Tak Zhang, "Effective EEG Connectivity Analysis of Episodic Memory Retrieval", In Proceedings of Annual Meeting of the Cognitive Science Society (CogSci 2014), pp. 833-838, 2014
- [5] S. Oh, M.S. Lee, and B.-T. Zhang, "Ensemble learning with active example selection for imbalanced biomedical data classification", IEEE/ACM Transactions on Computational Biology and Bioinformatics, 8(2):316-325, 2011
- [6] Robert H. Shumway, David S. Stoffer, "Time Series Analysis and Its Applications With R Examples.", Second Edition, Springer, pp. 84-164
- [7] M. Chen, D. Mandic, P. Kidmose, and M. Ungstrup, "Qualitative Assessment of Intrinsic Mode Functions of Empirical Mode Decomposition," In Proceedings of ICASSP 2008, pp. 1905-1908, 2008