
빅데이터 기반의 생체신호 수집 및 저장소 설계

주문일 · 서영우 · 김희철

인제대학교

Research on the Analysis System based on the Big Data for Matlab

Moon-il Joo · Young-woo Seo · Hee-cheol Kim

Inje University

E-mail : joomi@inje.ac.kr

요 약

최근 급속한 데이터의 생성으로 인하여 빅데이터 기술이 발전하고 있다. 특히 생체신호를 측정하는 웨어러블 디바이스의 발전으로 인하여 다양한 생체신호가 기하급수적으로 증가하고 있다. 그래서 기하급수적으로 증가한 생체신호의 특징을 파악하여 체계적으로 저장하는 저장소 기술이 필요하다. 본 논문은 생체신호의 특징을 파악하여 생체신호를 저장하는 저장소 설계와 생체신호를 수집하는 기술을 연구하고자 한다.

ABSTRACT

Recent rapid creation of data has resulted in the development of big data technologies. In particular, with the development of wearable devices that measure biological signals, a variety of biological signals are growing exponentially. Thus, storage technologies are required to identify and systematically store characteristics of exponential increase in biological signals. In this paper, we will study the storage design that stores the biometrics by identifying the characteristics of the biometrics and the techniques to collect the biometrics.

키워드

빅데이터, 생체신호, 저장소, 데이터 수집

I. 서 론

4차 산업 혁명의 시대에서는 데이터가 큰 자산으로 인정받는 시대로 통합된 생체데이터 수집, 관리, 분석 기술이 중요하다[1]. 특히, 웨어러블 기술의 발전으로 인해 일상생활 속에서 생체신호를 측정하는 다양한 디바이스가 개발 보급되었다[2]. 급속한 웨어러블 디바이스의 발전으로 인하여 생체신호를 활용한 개인맞춤형 건강관리서비스의 제공이 가능 할 것으로 전망한다[3].

건강관리서비스를 하기 위해서는 다양한 생체신호를 수집하는 수집 기술, 수집된 생체신호를 관리하기 위한 체계적인 저장소 설계와 생체신호 분석 알고리즘이 중요하다. 생체신호 분석 알고리즘을 개발하기 위해서는 다양한 데이터가 필요하다. 그래서, 최근에는 생체신호를 공유하기 위한 다양한 연구가 진행이 되었으며, 대표적으로 MIT-BIH의 PhysioBank ECG Database[4]가 있다.

본 논문은 생체신호의 특징을 파악하여 생체신호를 수집하기 위한 수집 기술과 수집된 생체신

호를 저장하기 위한 생체신호 저장소를 설계하고자 한다.

II. 생체신호 수집 및 저장 시나리오

생체신호 기반의 건강관리서비스를 하기 위해서는 생체신호 분석 알고리즘의 개발이 중요하다. 그러나 생체신호 분석 알고리즘을 개발하기 위해서는 다양하게 수집 및 저장된 생체신호 데이터가 필요하다. 그래서, 생체신호를 수집 및 저장을 위한 다양한 연구가 진행이 되고 있으며, 대표적으로 MIT-BIH[44]에서 개발한 Physionet[45] 데이터베이스가 있다. 이처럼 생체신호 분석 알고리즘의 개발에서는 생체신호의 수집을 통한 체계적인 저장소 구축이 필수적이다.

본 논문은 그림 6과 같은 3단계의 절차를 통하여 생체신호 수집과 체계적인 저장소 설계를 하고자 한다.

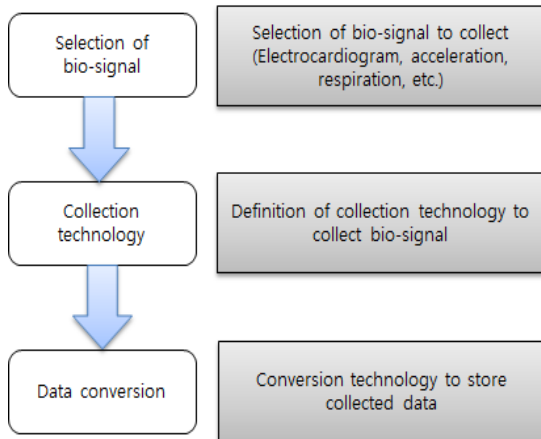


그림 1. 생체신호 수집 및 저장 시나리오

생체신호 저장소를 설계하기 위해서는 생체신호 특징에 대한 이해가 필요하다. 생체신호는 크게 3분류로 나눌 수 있다.

첫째, 생체신호를 측정할 수 있는 웨어러블 디바이스에 측정된 생체신호 Raw Data

둘째, 생체신호 Raw Data에서부터 생체신호 분석 알고리즘을 이용하여 추출한 특징값

셋째, 특징값에서 추출한 파라미터

생체신호 Raw Data는 심전도, 호흡, 가속도 신호 등과 같이 센서로부터 측정된 연속적 데이터로 데이터만으로는 아무런 의미를 가지고 있지 않은 신호로 인체에서 발생하는 가장 기본적인 생체신호이다.

생체신호 특징값은 심전도, 호흡, 가속도 신호 등 생체신호 Raw Data를 분석하여 추출한 데이터로 일차적으로 가공된 데이터를 의미한다. 예를 들면, 심전도에서 추출한 R-R Interval과 가속도 신호에서 추출한 Step Interval 등이 있다. 이 추출한 데이터는 데이터 자체적으로 의미를 가지고 있지만 의료 및 건강관리서비스를 제공하기에는 한계를 가지고 있다.

특징값에서 추출한 파라미터는 R-R Interval 및 Step Interval 데이터로부터 추출한 평균 맥박 및 스텝수, 칼로리, 거리 등과 같은 데이터는 서비스를 제공하기 위한 데이터를 의미한다.

III. 생체신호 선정

생체신호를 측정할 수 있는 다양한 웨어러블 디바이스의 발전으로 인하여 다양한 생체신호를 측정할 수 있다. 대표적으로 맥박 및 가속도신호를 측정할 수 있는 스마트워치, 심전도를 측정할 수 있는 심전도 측정 벨트 및 스마트의류, 가속도신호를 측정할 수 있는 스마트폰 등 다양한 웨어러블 디바이스가 있다. 또한, 국내 및 전세계적으로 생체신호의 확보 편리성을 높이기 위해 Physionet에서는 인간의 다양한 생체신호를 데이

터베이스로 구축을 진행하고 있다.

본 논문은 최근에 연구가 활발히 진행되고 있으며, 획득하기 비교적 쉬운 심전도 데이터와 가속도 데이터를 선정하였다. 심전도 데이터는 Physionet ECG Database에서 다양한 포맷으로 구축이 되어 있다. 또한 심전도 측정 벨트나 스마트 의류 등을 이용하여 측정할 수 있다. 가속도 데이터는 스마트폰, 단말기 및 스마트워치 등에서 신호 측정 및 저장을 할 수 있는 다양한 웨어러블 디바이스가 보편적으로 출시되었기 때문이다.

IV. 생체신호 수집 기술

최근에는 생체신호를 측정하는 착용형 웨어러블 디바이스를 이용하여 생체신호를 측정하는 것이 보편화되고 있는 추세이다. 대부분의 웨어러블 기기는 블루투스 기반의 무선통신을 활용하여 스마트폰이나 단말기에 블루투스로 연결하여 생체신호를 측정한다. 생체신호 측정기기를 이용하여 측정된 생체신호 데이터와 Physionet와 같이 생체신호 데이터를 구축한 데이터베이스로부터 원하는 생체신호 데이터를 검색하여 획득 할 수 있다.

일반적으로 데이터를 수집하기 위해서는 다양한 수집 기술이 사용되어 진다. 생체신호는 심전도와 같이 하나의 데이터가 대량으로 발생하거나, 가속도신호와 같이 3축(X, Y, Z축)을 데이터가 대량으로 발생한다. 그래서 Crawling, Streaming, Log Aggregator, RDB Aggregator 등과 같은 기술을 이용하여 뉴스, 웹 정보, 로그 수집 등과 같이 웹문서나 실시간으로 정보가 수집되는 형태에는 적합하지 않다. 생체신호는 대량으로 축적된 생체신호를 짧게는 몇분 길게는 몇일의 데이터를 분석하여 사용자의 건강상태를 파악한다. 그래서 축적된 생체신호를 파일에 저장하여 파일을 전송하는 기술이 효과적이다. 파일의 전송에 효과적인 FTP(File Transfer protocol) 및 Open API(Application Programming Interface)를 활용한 데이터 전송 기술이 적합하다.

V. 데이터 변환 및 통합

MIT-BIH의 PhysioBank ECG Database에서 검색하여 획득한 데이터이다. PhysioBank ECG Database의 심전도 데이터는 다양한 포맷을 가지고 있다. PhysioBank ECG Database의 심전도 데이터의 포맷은 측정시간과 여러 개의 리드를 포함하는 데이터를 가지고 있는 구조이다.

본 논문은 그림 2와 같이 MIT-BIH의 심전도 신호에서 필요한 리드값을 추출하기 위한 변환 작업을 한다.

Elapsed time hh:mm:ss.mmm	MLII (mV)	V5 (mV)
0:00.000	-0.145	-0.065
0:00.003	-0.145	-0.065
0:00.006	-0.145	-0.065
0:00.009	-0.145	-0.065
0:00.011	-0.145	-0.065
0:00.014	-0.145	-0.065
0:00.017	-0.145	-0.065
0:00.019	-0.145	-0.065
0:00.022	-0.120	-0.080
0:00.025	-0.135	-0.080
0:00.028	-0.145	-0.085
0:00.031	-0.150	-0.085
0:00.033	-0.160	-0.075
0:00.036	-0.155	-0.070
0:00.039	-0.160	-0.070
0:00.042	-0.175	-0.065
0:00.044	-0.180	-0.055
0:00.047	-0.185	-0.050
0:00.050	-0.170	-0.050
0:00.053	-0.155	-0.040
0:00.056	-0.175	-0.040
0:00.058	-0.180	-0.055
0:00.061	-0.190	-0.075

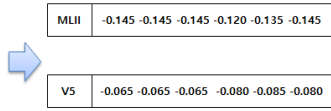


그림 2 MIT-BIH 심전도 데이터 추출

VI. 생체신호 저장소 설계

심전도, 호흡, 가속도 신호 등과 같이 센서로부터 측정된 데이터는 데이터 자체로 아무런 의미를 가지고 있지 않으며, 저장할 하기 위해서는 생체데이터 외에 생체데이터에 대한 정보를 나타내는 측정 일자, 측정자, Hz, 생체신호 종류와 같은 다양한 정보를 가지고 있어야 한다. 표5와 같이 HBase를 활용한 생체신호를 저장할 수 있는 데이터베이스를 설계한다. 본 논문은 Data와 User 컬럼을 생성한다. Data 컬럼은 Raw Data, Hz, Type, Overall Time을 가진다. User 컬럼은 Age와 Gender를 가진다.

표 1. HBase 기반 저장소

Row-key (Measurement Time_ID)	Data				User	
	Raw Data	Hz	Type	Overall Time	Age	Gender
2017.03.06.17. 45.07_joo	100 110 112 120 ...	200	ecg	125	38	man
2017.03.06.56. 10.07_joo	111 121 125 112 ...	200	ecg	100	38	man

VII. 결 론

본 논문은 다양한 웨어러블 디바이스에서 측정된 생체신호를 수집하기 위한 수집 기술 및 체계적으로 저장하기 위한 저장소 설계를 제안했다. 체계적인 생체신호 저장소 설계는 생체신호를 분석하고자 하는 개발자들에게 도움을 주고자 한다.

생체신호를 이용하여 건강관리서비스를 위해 분석 알고리즘을 개발하는 개발자들에게는 생체신호 데이터가 중요하다. 그래서, 체계적인 생체신호 저장소 설계를 통하여 다양한 생체신호 분석 알고리즘 개발에 필수적이다.

생체신호 수집 및 저장소 설계는 생체신호를 활용하여 건강관리서비스를 하고자 하는 개발자들을 위한 초석이 될 것으로 예상된다.

감사의 글

이 논문은 2017년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(NRF-2017R1D1A3B04032905)

참고문헌

- [1] 이희주, "4차 산업 혁명시대의 의료 환경 변화와 웰니스의 전망," 한국웰니스학회지, 12(4), 215-223, 2017.
- [2] 주정봉, 한상국, "IoT 와 IoE 의 웨어러블 디바이스가 미치는 영향과 그에 따른 사회변화," 한국통신학회 학술대회논문집, 65-66, 2014.
- [3] 박정훈, "헬스케어 웨어러블 디바이스의 기술 및 시장 동향," 전자공학회지, 45(3), 48-56, 2018.
- [4] Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., Stanley, H. E., "Physiobank, physiotookit, and physionet," Circulation, 101(23), e215-e220, 2000.