

---

# Matlab을 활용한 빅데이터 기반 분석 시스템 연구

주문일\* · 김희철\*\*

\*인제대학교

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

Moon-il Joo\* · Hee-cheol Kim\*\*

\*Inje University

E-mail : joomi@inje.ac.kr

### 요 약

최근 급속한 데이터의 생성으로 인하여 빅데이터 기술이 발전하고 있으며, 빅데이터를 분석하기 위한 다양한 빅데이터 분석 툴이 개발되어지고 있다. 대표적인 빅데이터 기반의 분석 툴은 R 프로그램, Hive, Tajo 등 다양한 분석 툴이 있다. 그러나, Matlab을 활용한 데이터 분석과 이를 위한 알고리즘 개발이 여전히 보편적이며, 빅데이터 분석에서도 Matlab이 광범위하게 사용되고 있다. 본 논문은 생체신호를 분석하는 Matlab을 활용한 빅데이터 기반 분석 시스템을 연구하고자 한다.

### ABSTRACT

Recently, big data technology develop due to the rapid data generation. Thus big data analysis tools for analyzing big data has been developed. Typical big data tools are the R program, Hive, Tajo and more. But data analysis based on Matlab is still common used. And it is still used in big data analysis. In this paper, it research into big data analysis system based on the Matlab for analyzing vital signals.

### 키워드

빅데이터, Matlab, 생체신호, 분석 툴

## 1. 서 론

최근 다양한 웨어러블 장비와 스마트폰 기기의 보급되면서 개인, 기업, 그룹 등 다양한 정보가 수집이 가능해졌으면, 방대한 양의 데이터가 급속하게 생성이 되었다. 급속한 데이터의 생성으로 인하여 빅데이터 기술이 발전하고 있으며 빅데이터를 분석하기 위한 다양한 빅데이터 분석 툴이 개발되어졌고, 개발이 진행중이다[1-2].

빅데이터란 디지털 환경에서 생성되는 데이터로 그 규모가 방대하고, 생성 주기도 짧고, 형태도 수치 데이터뿐만 아니라 문자와 영상 데이터를 포함하는 대규모 데이터이다. 빅데이터는 다양한 형태의 데이터로 연속적으로 유입되는 선형 구조, 데이터와 단순히 하나의 특징 값을 가지는 데이터, 텍스트 기반의 데이터 등 다양한 형태를 가지고 있다. 이러한 형태의 데이터를 분류하고

분석하여 의미 있는 데이터 가공이 필요하다[3]. 대표적인 빅데이터 기반의 분석 툴은 R 프로그램, Hive, Tajo 등의 분석 툴이 있으며, 이 중에서 Hive와 Tajo는 SQL과 유사한 HiveQL을 사용하여 검색 및 통계 분석이 가능하다. 이처럼 현재의 분석 툴은 통계 기반의 분석을 위한 서비스를 하기 위해 활용하고 있다. 그러나 빅데이터는 다양한 사물에서 측정된 데이터, 웹에서 수집한 데이터, 개인 및 기업에서 수집한 데이터 등 다양한 데이터가 있으며, 이러한 데이터를 분석하기 위한 다양한 방법이 존재한다. 현재까지 Matlab은 다양한 데이터를 분석하기 위해 보편적으로 사용되어지고 있으며, 또한 빅데이터 분석에서도 광범위하게 사용되고 있다.

본 논문은 Matlab 코드를 빅데이터에 적용하여 데이터를 분석을 하고자 한다. 많은 빅데이터 플랫폼이 예를 들면 하둡은 자바 기반의 플랫폼은

로 Matlab을 사용하여 알고리즘을 개발 시에 자바로 컨버팅 작업을 진행해야 하기 때문에 알고리즘 적용의 어려움이 있다. 그래서 본 논문은 Matlab코드 자체를 웹서비스를 이용하여 업로드하여 코드의 삽입, 삭제, 변경적용을 하는 시스템을 연구하고자 한다.

본 논문은 2장에서는 빅데이터 기반의 생체신호 분석 시스템에 대하여 기술하였으며, 3장에서는 Matlab 코드 삽입 기반의 생체신호 분석에 대하여 기술한다. 마지막으로 4장에서는 결론을 기술한다.

## II. 빅데이터 기반 생체신호 분석 시스템

생체신호 분석 플랫폼은 그림1과 같이 다양한 사용자들로부터 측정된 생체신호 원시데이터를 SOAP (Simple Object Access Protocol)메시지를 이용하여 서버로 전송하고 서버는 빅데이터를 저장하는 하둡 플랫폼의 저장소인 하둡 분산파일 시스템(Hadoop Distributed File System, HDFS)에 생체신호 원시데이터를 저장한다[4]. 저장된 생체신호 원시데이터는 선형의 구조를 가지고 있으며 데이터 자체만으로는 의미가 없는 비정형의 형태를 가진다. 따라서 의미를 가지는 가공된 데이터의 추출이 필요하다. 그래서 생체신호 분석 알고리즘을 적용시켜 생체신호 특징 값을 추출한다. 추출된 특징 값은 정형화된 데이터로 일반적인 데이터베이스로 관리하기에 부적합한 특징을 가지고 있다. 왜냐하면, 심전도에서 추출한 맥박은 하나의 컬럼에 방대한 데이터를 저장해야 한다. 이러한 데이터는 데이터베이스로 관리하기에는 한계가 있다. 그래서 빅데이터 기반의 분석 툴인 데이터웨어하우스에 저장을 하여 서비스를 제공한다. 서비스를 사용하기 위한 사용자가 서비스로 요청하면 데이터웨어하우스에서 요청한 데이터를 검색 및 추출하여 데이터 분석을 한 후에 사용자에게 요구하는 결과 값을 웹서비스 형태로 제공한다.

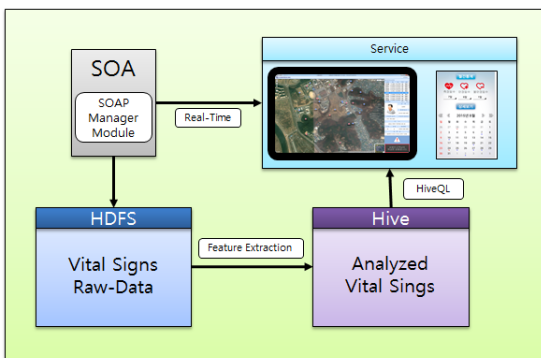


그림 1. 빅데이터 기반 생체신호 분석 시스템 워크플로우

## III. Matlab 코드 삽입 기반의 생체신호 분석

그림 2와 같이 Matlab 코드 삽입 기반의 생체신호 분석을 한다. 본 논문은 Matlab 기반의 생체신호 분석을 위해 구조는 Matlab 코드를 SOAP 메시지를 이용하여 서버에 전송하여 하둡 분산파일 시스템에 Matlab 코드를 저장하여 생체신호를 분석하고자 한다.

생체신호를 분석하기 위한 다양한 Matlab 코드의 정확한 명세화된 모듈 명세서에 입력값과 출력값을 정확히 명세화를 해야 한다. 알고리즘 명세화를 하기 위하여 명세서 엔진을 사용하여 정의하고자 한다. 명세서 엔진은 알고리즘의 정의, 입력값, 출력값 정보를 포함하고 있다. 입력값에는 알고리즘 적용을 입력 데이터가 정의 되어 있으며, 출력값에는 출력된 결과값을 정의하고 있다.

즉, 알고리즘 저장소에 저장된 알고리즘들을 활용하기 위하여 첫째, 명세서 엔진의 검색에서 클라이언트의 SOAP 메시지에서 포함된 찾고자하는 알고리즘 정보를 통해 원하는 알고리즘을 검색한다. 둘째, 검색된 알고리즘이 필요로 하는 입력값에 따라 생체신호 데이터 및 특징 값을 가져온다. 생체신호 데이터는 하둡에 저장된 데이터에서 값을 가져오고 특징 값은 데이터웨어하우스를 이용하여 값을 가져와서 알고리즘에 적용한다.

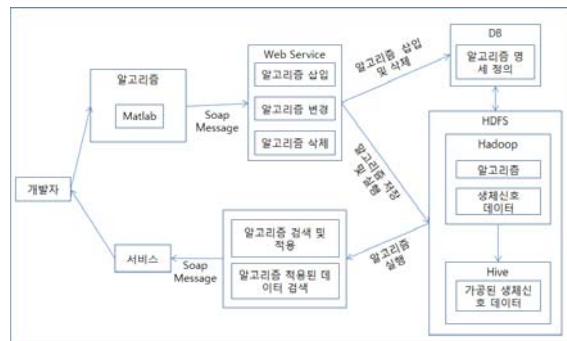


그림 2. Matlab 코드 삽입 기반의 생체신호 분석 워크플로우

## IV. 결론

IoT시대로 접어들면서 다양한 환경에서 방대한 데이터가 생성되는 시대이다. 이처럼 급속하게 늘어난 데이터를 처리하기 위하여 빅데이터 기술과 빅데이터를 분석하기 위한 기술이 발전하고 있다.

빅데이터를 분석하기 위한 분석 툴 기술이 개발 되어지고 있으며 또한, 지속적으로 연구를 하고 있다. 그러나, 현재까지 Matlab기반의 데이터 분석이 보편적으로 사용되어지고 있기 때문에 본 논문은 Matlab 코드를 빅데이터에 적용하여 데이터를 분석을 하고자 한다. 많은 빅데이터 플랫폼

이 예를 들면 하둡은 자바 기반의 플랫폼으로 Matlab을 사용하여 알고리즘을 개발 시에 자바로 컨버팅 작업을 진행해야 하기 때문에 알고리즘 적용의 어려움이 있다. 그래서 본 논문은 Matlab 코드 자체를 웹서비스를 이용하여 업로드하여 적용을 하는 시스템을 개발하면, 데이터마이닝, 생체신호, IoT 등 다양한 전문가들이 시스템을 사용하여 연구 및 개발에 도움이 될 것이라고 예상된다.

### 감사의 글

본 연구는 산업통상자원부의 창의산업융합 특성화인재양성사업(과제번호 201604840001)지원으로 수행하였음

### 참고문헌

- [1] M. Saecker and V. Markl, "Big Data Analytics on Modern Hardware Architectures: A Technology Survey," Proceeding of European Business Intelligence Summer School. pp. 125-149, 2012.
- [2] J.S. Ward and A. Barker, "Undefined By Data: A Survey of Big Data Definitions," The Computing Research Repository, 2013.
- [3] 박기진, "빅데이터 Eco System," ie 매거진, 제10권, 제3호, pp. 41-47, 2012.9.
- [4] 주문일, 고동희, 김희철, "생체신호 습득과 건강 모니터링을 위한 스마트 헬스케어 의복 개발," 멀티미디어학회논문지, 제19권, 제5호, pp. 808-817, 2016.5.