(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2021-0004058
(43) 공개일자 2021년01월13일

(51) 국제특허분류(Int. Cl.)
G16H 10/60 (2018.01) *G06F 16/35* (2019.01)
G06F 16/36 (2019.01) *G16H 50/70* (2018.01)
(52) CPC특허분류
G16H 10/60 (2018.01)
G06F 16/35 (2019.01)
(21) 출원번호 10-2019-0079878
(22) 출원일자 2019년07월03일
심사청구일자 2019년07월03일

(71) 출원인
인하대학교 산학협력단
인천광역시 미추홀구 인하로 100(용현동, 인하대학교)
(72) 발명자
곽경섭
인천광역시 연수구 해돋이로84번길 10(송도동, 송도풍림아이원6단지아파트)
샤커
인천광역시 미추홀구 인하로 100, 하이테크 400(용현동, 인하대학교)
(뒷면에 계속)
(74) 대리인
양성보

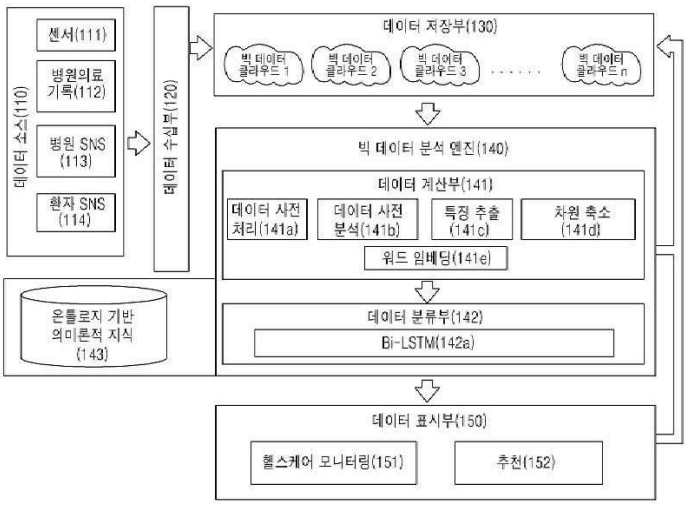
전체 청구항 수 : 총 8 항

(54) 발명의 명칭 **웨어러블 센서와 소셜네트워크 데이터를 활용한 헬스케어 모니터링 방법 및 장치**

(57) 요 약

웨어러블 센서와 소셜네트워크 데이터를 활용한 헬스케어 모니터링 방법 및 장치가 제시된다. 본 발명에서 제안하는 웨어러블 센서와 소셜네트워크 데이터를 활용한 헬스케어 모니터링 장치는 웨어러블 기기로부터 생리적 데이터, 스마트 단말로부터 약물 및 증상 데이터, 어플리케이션 프로그래밍 인터페이스(Application Programming Interfaces; API)를 통한 환자들 및 의사들의 소셜 네트워크 상에서의 논의 데이터를 수집하는 데이터 수집부, 개인 클라우드 서버와 연결되어 수집된 생리적 데이터, 약물 및 증상 데이터 및 논의 데이터를 저장하는 데이터 저장부 및 저장된 데이터에 대한 특징 극성 식별, 워드 임베딩 및 온톨로지 기반 특징 추출, 웨어러블 센서 데이터에서 추출, 데이터 구성 요소 분석, Bi-LSTM 기반 당뇨병과 BP 분류 및 약물 부작용 예측을 포함하는 데이터의 계산 및 분류를 수행하는 빅데이터 분석 엔진을 포함한다.

대 표 도



(52) CPC특허분류
G06F 16/367 (2019.01)
G16H 50/70 (2018.01)
(72) 발명자
리아즈
인천광역시 미추홀구 인하로 100(용현동, 인하대학교)
알리펄만
인천광역시 미추홀구 인하로 100(용현동, 인하대학교)

명 세 서

청구범위

청구항 1

웨어러블 기기로부터 생리적 데이터, 스마트 단말로부터 약물 및 증상 데이터, 어플리케이션 프로그래밍 인터페이스(Application Programming Interfaces; API)를 통한 환자들 및 의사들의 소셜 네트워크 상에서의 논의 데이터를 수집하는 데이터 수집부;

개인 클라우드 서버와 연결되어 수집된 생리적 데이터, 약물 및 증상 데이터 및 논의 데이터를 저장하는 데이터 저장부; 및

저장된 데이터에 대한 특징 극성 식별, 워드 임베딩 및 온톨로지 기반 특징 추출, 웨어러블 센서 데이터에서 추출, 데이터 구성 요소 분석, Bi-LSTM 기반 당뇨병과 BP 분류 및 약물 부작용 예측을 포함하는 데이터의 계산 및 분류를 수행하는 빅데이터 분석 엔진

을 포함하는 헬스케어 모니터링 장치.

청구항 2

제1항에 있어서,

빅데이터 분석 엔진은,

저장된 데이터는 센서 데이터, 의료 기록 및 소셜 네트워크 콘텐츠를 포함하고, 센서 데이터의 사전 분석, 센서 데이터의 사전 처리 및 필터링, 의료 기록의 사전 처리, 소셜 네트워크 콘텐츠의 사전 처리 및 불일치와 노이즈를 제거하기 위한 데이터 필터링을 수행한 후 데이터를 구조화된 형태로 변환하는

헬스케어 모니터링 장치.

청구항 3

제1항에 있어서,

빅데이터 분석 엔진은,

단어의 의미를 이해하는 신경망 모델에 대하여 추가 정보를 제공하기 위해 온톨로지 기반 특징 추출을 수행하여 모든 온톨로지 정보를 검색하고, 분류를 위한 워드 임베딩과 통합하며, 온톨로지로부터 정보를 추출하기 위해, 용어 주파수(Term Frequency; TF), 용어 빈도 및 TF-IDF(Inverse Document Frequency)를 포함하는 통계 방법을 사용하여 수학적으로 정의하는

헬스케어 모니터링 장치.

청구항 4

제1항에 있어서,

빅데이터 분석 엔진은,

WEKA(Waikato Environment for Knowledge Analysis)를 통해 계산된 정보 이득(Information Gain; IG) 방법을 사용하여 웨어러블 센서 데이터의 특징을 추출하고 속성의 가치를 찾아, 불확실성을 측정하기 위해 엔트로피를 이용하고 이전의 엔트로피와 이후의 엔트로피 차이를 정의하는

헬스케어 모니터링 장치.

청구항 5

데이터 수집부를 통해 웨어러블 기기로부터 생리적 데이터, 스마트 단말로부터 약물 및 증상 데이터, 어플리케이션 프로그래밍 인터페이스(Application Programming Interfaces; API)를 통한 환자들 및 의사들의 소셜 네트

워크 상에서의 논의 데이터를 수집하는 단계;

데이터 저장부는 개인 클라우드 서버와 연결되어 수집된 생리적 데이터, 약물 및 증상 데이터 및 논의 데이터를 저장하는 단계; 및

저장된 데이터는 빅데이터 분석 엔진을 통해 특징 극성 식별, 워드 임베딩 및 온톨로지 기반 특징 추출, 웨어러블 센서 데이터에서 추출, 데이터 구성 요소 분석, Bi-LSTM 기반 당뇨병과 BP 분류 및 약물 부작용 예측을 포함하는 데이터의 계산 및 분류를 수행하는 단계

를 포함하는 헬스케어 모니터링 방법.

청구항 6

제5항에 있어서,

저장된 데이터는 빅데이터 분석 엔진을 통해 특징 극성 식별, 워드 임베딩 및 온톨로지 기반 특징 추출, 웨어러블 센서 데이터에서 추출, 데이터 구성 요소 분석, Bi-LSTM 기반 당뇨병과 BP 분류 및 약물 부작용 예측을 포함하는 데이터의 계산 및 분류를 수행하는 단계는,

저장된 데이터는 센서 데이터, 의료 기록 및 소셜 네트워크 콘텐츠를 포함하고, 센서 데이터의 사전 분석, 센서 데이터의 사전 처리 및 필터링, 의료 기록의 사전 처리, 소셜 네트워크 콘텐츠의 사전 처리 및 불일치와 노이즈를 제거하기 위한 데이터 필터링을 수행한 후 데이터를 구조화된 형태로 변환하는

헬스케어 모니터링 방법.

청구항 7

제5항에 있어서,

저장된 데이터는 빅데이터 분석 엔진을 통해 특징 극성 식별, 워드 임베딩 및 온톨로지 기반 특징 추출, 웨어러블 센서 데이터에서 추출, 데이터 구성 요소 분석, Bi-LSTM 기반 당뇨병과 BP 분류 및 약물 부작용 예측을 포함하는 데이터의 계산 및 분류를 수행하는 단계는,

단어의 의미를 이해하는 신경망 모델에 대하여 추가 정보를 제공하기 위해 온톨로지 기반 특징 추출을 수행하여 모든 온톨로지 정보를 검색하고, 분류를 위한 워드 임베딩과 통합하며, 온톨로지로부터 정보를 추출하기 위해, 용어 주파수(Term Frequency; TF), 용어 빈도 및 TF-IDF(Inverse Document Frequency)를 포함하는 통계 방법을 사용하여 수학적으로 정의하는

헬스케어 모니터링 방법.

청구항 8

제5항에 있어서,

저장된 데이터는 빅데이터 분석 엔진을 통해 특징 극성 식별, 워드 임베딩 및 온톨로지 기반 특징 추출, 웨어러블 센서 데이터에서 추출, 데이터 구성 요소 분석, Bi-LSTM 기반 당뇨병과 BP 분류 및 약물 부작용 예측을 포함하는 데이터의 계산 및 분류를 수행하는 단계는,

WEKA(Waikato Environment for Knowledge Analysis)를 통해 계산된 정보 이득(Information Gain; IG) 방법을 사용하여 웨어러블 센서 데이터의 특징을 추출하고 속성의 가치를 찾아, 불확실성을 측정하기 위해 엔트로피를 이용하고 이전의 엔트로피와 이후의 엔트로피 차이를 정의하는

헬스케어 모니터링 방법.

발명의 설명

기 술 분 야

[0001] 본 발명은 웨어러블 센서와 소셜네트워크 데이터를 활용한 헬스케어 모니터링 방법 및 장치에 관한 것이다.

배 경 기 술

[0002] 현재 의료 산업의 기술은 개별 환자에 대한 생리적 정보를 수집하는 효율적인 방법을 제공하는 데 핵심적인 역할을 한다. 또한 스마트폰과 웨어러블 센서를 활용하여 환자 데이터를 수집함으로써 효율적인 의료 모니터링을 할 수 있다. 사람들은 또한 특정 약물에 대한 감정과 의견을 공유하기 위해 소셜 네트워킹을 사용하는데, 이것은 정서적 장애를 관찰하고 약물 부작용을 예측하는 데 유용하다. 그러나 지속적인 환자 모니터링은 대량의 의료 데이터를 생성하며, 소셜 네트워킹 사이트의 사용자가 생성한 의료 데이터는 대량 제공되고 구조화되지 않으며, 예기치 않은 것일 수 있다. 따라서 현재의 추세는 스마트폰, 웨어러블 센서, 의료기록, 소셜 네트워크(다시 말해, 빅데이터) 등 다양한 출처에서 추출한 방대한 양의 의료 데이터를 처리할 수 있는 고급 접근법이 필요하다. 현재의 의료 모니터링 시스템은 기기 및 소셜 네트워킹 데이터에서 중요한 정보를 추출하는 데 효율적이지 않으며, 효과적으로 분석하는 데 어려움을 겪고 있다. 또한, 기존의 기계 학습 기법은 비정상적인 예측을 위해 추출된 의료 데이터를 다루고 처리할 수 없다. 따라서, 의료 데이터를 정밀하게 저장 및 분석하고 분류 정확도를 개선하기 위해 클라우드 환경 및 빅데이터 분석 엔진을 기반으로 하는 새로운 헬스케어 모니터링 아키텍처를 제안한다.

발명의 내용

해결하려는 과제

[0003] 본 발명이 이루고자 하는 기술적 과제는 의료 데이터를 정밀하게 저장 및 분석하고 분류 정확도를 개선하기 위해 클라우드 환경 및 빅데이터 분석 엔진을 기반으로 하는 새로운 헬스케어 모니터링 방법 및 장치를 제공하는 데 있다.

과제의 해결 수단

[0004] 일 측면에 있어서, 본 발명에서 제안하는 웨어러블 센서와 소셜네트워크 데이터를 활용한 헬스케어 모니터링 장치는 웨어러블 기기로부터 생리적 데이터, 스마트 단말로부터 약물 및 증상 데이터, 어플리케이션 프로그래밍 인터페이스(Application Programming Interfaces; API)를 통한 환자들 및 의사들의 소셜 네트워크 상에서의 논의 데이터를 수집하는 데이터 수집부, 개인 클라우드 서버와 연결되어 수집된 생리적 데이터, 약물 및 증상 데이터 및 논의 데이터를 저장하는 데이터 저장부, 및 저장된 데이터에 대한 특징 극성 식별, 워드 임베딩 및 온톨로지 기반 특징 추출, 웨어러블 센서 데이터에서 추출, 데이터 구성 요소 분석, Bi-LSTM 기반 당뇨병과 BP 분류 및 약물 부작용 예측을 포함하는 데이터의 계산 및 분류를 수행하는 빅데이터 분석 엔진을 포함한다.

[0005] 빅데이터 분석 엔진은 저장된 데이터는 센서 데이터, 의료 기록 및 소셜 네트워크 콘텐츠를 포함하고, 센서 데이터의 사전 분석, 센서 데이터의 사전 처리 및 필터링, 의료 기록의 사전 처리, 소셜 네트워크 콘텐츠의 사전 처리 및 불일치와 노이즈를 제거하기 위한 데이터 필터링을 수행한 후 데이터를 구조화된 형태로 변환한다.

[0006] 빅데이터 분석 엔진은 단어의 의미를 이해하는 신경망 모델에 대하여 추가 정보를 제공하기 위해 온톨로지 기반 특징 추출을 수행하여 모든 온톨로지 정보를 검색하고, 분류를 위한 워드 임베딩과 통합하며, 온톨로지로부터 정보를 추출하기 위해, 용어 주파수(Term Frequency; TF), 용어 빈도 및 TF-IDF(Inverse Document Frequency)를 포함하는 통계 방법을 사용하여 수학적으로 정의한다.

[0007] 빅데이터 분석 엔진은 WEKA(Waikato Environment for Knowledge Analysis)를 통해 계산된 정보 이득(Information Gain; IG) 방법을 사용하여 웨어러블 센서 데이터의 특징을 추출하고 속성의 가치를 찾아, 불확실성을 측정하기 위해 엔트로피를 이용하고 이전의 엔트로피와 이후의 엔트로피 차이를 정의한다.

[0008] 또 다른 일 측면에 있어서, 본 발명에서 제안하는 웨어러블 센서와 소셜네트워크 데이터를 활용한 헬스케어 모니터링 방법은 데이터 수집부를 통해 웨어러블 기기로부터 생리적 데이터, 스마트 단말로부터 약물 및 증상 데이터, 어플리케이션 프로그래밍 인터페이스(Application Programming Interfaces; API)를 통한 환자들 및 의사들의 소셜 네트워크 상에서의 논의 데이터를 수집하는 단계, 데이터 저장부는 개인 클라우드 서버와 연결되어 수집된 생리적 데이터, 약물 및 증상 데이터 및 논의 데이터를 저장하는 단계, 및 저장된 데이터는 빅데이터 분석 엔진을 통해 특징 극성 식별, 워드 임베딩 및 온톨로지 기반 특징 추출, 웨어러블 센서 데이터에서 추출, 데이터 구성 요소 분석, Bi-LSTM 기반 당뇨병과 BP 분류 및 약물 부작용 예측을 포함하는 데이터의 계산 및 분류를 수행하는 단계를 포함한다.

발명의 효과

[0009] 본 발명의 실시예들에 따르면 제안된 빅데이터 분석 엔진을 통해 데이터 마이닝 기술, 온톨로지 및 양방향 장단

기 메모리(Bi-LSTM)를 기반하여, 데이터 마이닝 기술을 통해 의료 데이터를 효율적으로 사전 처리하고 유용한 기능을 추출하며 데이터의 차원수를 줄일 수 있다. 또한, 제안된 온톨로지에서는 실체 및 측면에 대한 의미론적 지식과 당뇨병 및 혈압 영역에서의 관계를 제공하며, Bi-LSTM는 의료 데이터를 정확하게 분류하여 환자의 약물 부작용과 비정상 상태를 예측할 수 있다. 또한, 제안된 장치는 당뇨병, BP, 정신 건강 및 약물 검토와 관련된 의료 데이터를 사용하여 환자 분류에 활용되어, 제안된 모델을 통해 이질적인 데이터를 정확하게 처리하고 건강 상태 분류와 약물 부작용 예측의 정확성을 향상시킬 수 있다.

도면의 간단한 설명

[0010] 도 1은 본 발명의 일 실시예에 따른 웨어러블 센서와 소셜네트워크 데이터를 활용한 헬스케어 모니터링 장치의 구성을 나타내는 도면이다.

도 2는 본 발명의 일 실시예에 따른 빅데이터 클라우드 서버 및 HDFS를 설명하기 위한 도면이다.

도 3은 본 발명의 일 실시예에 따른 빅데이터 분석 엔진을 설명하기 위한 도면이다.

도 4는 본 발명의 일 실시예에 따른 임베딩 및 온톨로지 기반 특징 추출을 설명하기 위한 도면이다.

도 5는 본 발명의 일 실시예에 따른 LSTM 기반 분류 및 예측을 설명하기 위한 도면이다.

도 6은 본 발명의 일 실시예에 따른 웨어러블 센서와 소셜네트워크 데이터를 활용한 헬스케어 모니터링 방법을 설명하기 위한 흐름도이다.

도 7은 본 발명의 일 실시예에 따른 온톨로지 사용에 따른 LSTM 기반 분류의 정확성과 MAE를 나타내는 그래프이다.

도 8은 본 발명의 일 실시예에 따른 PCA 및 IG와 함께 분류 모델을 사용하여 얻은 결과를 나타내는 그래프이다.

발명을 실시하기 위한 구체적인 내용

[0011] 이하, 본 발명의 실시 예를 첨부된 도면을 참조하여 상세하게 설명한다.

[0013] 도 1은 본 발명의 일 실시예에 따른 웨어러블 센서와 소셜네트워크 데이터를 활용한 헬스케어 모니터링 장치의 구성을 나타내는 도면이다.

[0014] 제안하는 웨어러블 센서와 소셜네트워크 데이터를 활용한 헬스케어 모니터링 장치는 데이터 수집부(120), 데이터 저장부(130), 빅데이터 분석 엔진(140)을 포함한다.

[0015] 데이터 수집부(120)는 웨어러블 기기로부터 생리적 데이터, 스마트 단말로부터 약물 및 증상 데이터, 어플리케이션 프로그래밍 인터페이스(Application Programming Interfaces; API)를 통한 환자들 및 의사들의 소셜 네트워크 상에서의 논의 데이터를 수집한다.

[0016] 예를 들어, 데이터 수집부(120)는 센서(111), 병원의료 기록(1120), 병원 SNS(113) 및 환자 SNS(114)를 포함하는 데이터 소스(110)로부터 데이터를 수집할 수 있다. 다시 말해, 웨어러블 기기, 스마트폰에서의 약물 및 증상 정보, 애플리케이션 프로그래밍 인터페이스(API)를 통한 환자 및 의사들의 소셜 네트워크 논의에서 생리적 정보를 수집한다.

[0017] 웨어러블 기기는 혈압 모니터, 글루코미터 센서, 스마트워치, 맥박산소미터, 가속도계, 온도 센서, 체중계 등 스마트 센서와 웨어러블 장치를 사용하여 혈압, 혈당 레벨, 심장 박동수, 스트레스 비율, 산소 포화도, 온도 및 체중 등 환자의 생리적 징후를 실시간으로 모니터링할 수 있다.

[0018] 스마트폰은 당뇨병과 BP 환자로부터 식이요법, 운동, 기타 활동 정보를 수집하고, 개인 정보(예를 들어, 나이, 성별, 키, 기타 정보)를 수집하는 데 사용된다.

[0019] 의료기록은 환자의 병력(예를 들어, 치료, 실험실 테스트, 약물 섭취)을 기재한 환자의 의료기록을 포함한다.

[0020] 소셜 네트워킹 플랫폼 및 웹 페이지의 데이터는 약품 리뷰나 환자의 감정적 게시물 같은 소셜 미디어 데이터를 수집하여 스트레스와 우울증 정도를 예측하고, 식생활과 생활양식의 맥락에서 당뇨병 약물의 부작용을 파악하고, 환자의 치료와 지식을 향상시킨다.

[0021] 데이터 저장부(130)는 이와 같이 개인 클라우드 서버와 연결되어 수집된 생리적 데이터, 약물 및 증상 데이터 및 논의 데이터를 저장한다.

[0022] 앞서 설명된 바와 같이, 다양한 출처로부터 수집된 대량의 의료 데이터는 저장 및 처리가 어렵다. 따라서 제안된 장치는 환자 정보를 저장하기 위해 확장성이 뛰어나고 보안이 뛰어난 Amazon S3라는 개인 클라우드 서버와 연결된다. 도 2를 참조하여 데이터 저장 과정을 더욱 상세히 설명한다.

[0024] 도 2는 본 발명의 일 실시예에 따른 빅데이터 클라우드 서버 및 HDFS를 설명하기 위한 도면이다.

[0025] 도 2와 같이 웨어러블 센서와 소셜 네트워킹 데이터는 클라우드 서버(210)에서 HDFS(Hadoop Distributed File System)(232)으로 전송되고 있다. HDFS(232)는 데이터를 HBase(231) 클러스터로 전송하기 위해 높은 대역폭을 제공하는 분산 파일 스토리지 시스템이다. HBase(231)는 HDFS(232) 위에서 실행된다. Flume(220)은 클라우드 서버에서 Hadoop 에코시스템으로 데이터를 전송한다.

[0026] 본발명의 실시예에 따르면, 웨어러블 기기에 의해 추출된 데이터를 MapReduce(242)로 전송하기 위해 Apache Pig(242)를 사용한다. MapReduce(242)는 두 가지 주요 기능을 가지고 있다. 즉, 맵(Map)과 축소(Reduce)이다. 맵(Map) 태스크는 한 쌍으로 된 키 값의 빅데이터에서 값을 수집한다. 축소(Reduce) 기능은 특정 키에 대해 설정된 값을 저장한다. 이후, 저장된 데이터는 분석 엔진(Analytics Engine)(250)으로 전달된다.

[0028] 빅데이터 분석 엔진(140)은 저장된 데이터에 대한 특징 극성 식별, 온톨로지 기반 의미론적 지식(1430으로부터 워드 임베딩 및 온톨로지 기반 특징 추출, 웨어러블 센서 데이터에서 추출, 데이터 구성 요소 분석, Bi-LSTM 기반 당뇨병과 BP 분류 및 약물 부작용 예측을 포함하는 데이터의 계산 및 분류를 수행한다. 저장된 데이터는 센서 데이터, 의료 기록 및 소셜 네트워킹 데이터로 구성된다. 그러나 비일관성, 누락값, 노이즈, 다른 형식, 큰 크기, 높은 차원성 때문에 실제 빅데이터를 처리하기는 극히 어렵다.

[0029] 빅데이터 분석 엔진(140)은 제안된 프레임워크의 가장 중요한 구성이다. 빅데이터 분석 엔진(140)은 데이터 계산부(141)와 데이터 분류부(142)의 두 하위 계층으로 나뉜다. 데이터 계산부(141)에는 데이터 사전 처리(141a) 데이터 사전 분석(141b), 특징 추출(141c), 워드 차원 축소(141d) 및 워드 임베딩(141e)과 같은 하위 모듈이 있다. 여기에서, 온톨로지 기반의 의미론적 지식(143)은 소프트 컴퓨팅 접근법과 함께, 필요한 정보의 추출에 필요한 데이터를 처리하고 분석하는 데 이용된다. 데이터 분류부(142)에서는 온톨로지가 있는 Bi-LSTM(142a)을 당뇨병, BP, 정신건강, 약물 부작용의 분류에 활용한다. 데이터 분류부(142)는 당뇨병과 BP에 관한 다차원 빅데이터를 지능적으로 분석하여 데이터로부터 의사 결정에 대한 통찰력을 얻고, 환자에게 맞춤형 당뇨병과 BP 의료 시스템을 제공한다. 도 3을 참조하여 데이터 분석 과정에 대해 더욱 상세히 설명한다.

[0031] 도 3은 본 발명의 일 실시예에 따른 빅데이터 분석 엔진을 설명하기 위한 도면이다.

[0032] 도 3과 같이 빅데이터 분석 엔진은 센서 데이터의 사전 분석, 센서 데이터의 사전 처리 및 필터링, 의료 기록의 사전 처리, 소셜 네트워크 콘텐츠의 사전 처리를 수행한다.

표 1

[0033]

리소스	ID	파라미터	설명
센서	S1	혈당	혈당 (mg / dL)
	S2	체온	환자의 현재 체온
	S3	혈압	수축기 혈압 (mmHg)
			이완기 혈압 (mmHg)
	S4	산소 포화도	환자의 SpO_2 소비량 (mmHg)
	S5	심박수	환자의 심박수 (bpm)
	S6	ECG	ECG 센서를 이용한 환자의 심전도
	S7	EEG	EEG 센서를 이용한 환자 뇌파 검사
	S8	스트레스	ECG + EEG 패턴을 이용한 환자의 스트레스 계산

스마트폰	SP1	나이	생년월일 환자의 나이
	SP2	키	환자의 신장을 찾기 위한 검시계
	SP3	BMI	체질량 지수 (kg / m^2)
	SP4	성별	환자의 성별 (0/1)
	SP6	활동	라이프 스타일: 좌식, 가볍게 활동, 보통 활동, 활성 또는 매우 활동적인
병원 의료 기록	MR1	지단백질 수준	저밀도 지단백질 수준 (LDL 콜레스테롤)
			고밀도 지단백질 수준 (HDL 콜레스테롤)
	MR 2	헤모글로빈	환자의 글리코 헤모글로빈 (A1c) (%)
	MR 3	혈당	환자의 혈액 검사
	MR 4	혈청 크레아티닌	환자의 혈액 검사
	MR5	중성 지방	환자의 혈액 검사
	MR6	콜레스테롤	환자의 혈액 검사
	MR7	AST (SGOT)	간 손상을 검사하는 혈액 검사
	MR8	ALT (SGPT)	간 손상을 검사하는 혈액 검사
	MR9	마약 복용	처방전 목록에서 추출한
	MR10	흡연	예 / 아니오 (자체 시험)
	MR11	음주	예 / 아니오 (자체 시험)
	MR12	소화불량	예 / 아니오 (자체 시험)
	MR13	가족 병력	예 / 아니오 (자체 시험)
	MR14	환자의 병력	예 / 아니오 (자체 시험)

[0034] 센서 데이터의 사전 분석은 표 1과 같이 센서와 스마트폰을 사용하는 당뇨병 및 BP 환자들로부터 서로 다른 파라미터가 감지된다. 또한 환자의 몸에서 다른 파라미터들을 추출한다. 첫째, 데이터 세트는 쉽게 구문 분석할 수 있도록 쉼표로 구분된 값(CSV) 파일로 변환된다. 이름은 각 속성(파라미터)에 할당되며, 숫자 값과 함께 열의 형태로 표시된다. 그런 다음 센서 데이터의 ID를 실제 센서 이름으로 표시한다(예를 들어, S1은 혈당 ID). 연령과 활동에 각각 명시적인 연도와 시간을 사용하는 대신 속성연령을 일, 월, 년으로 나누고, 활동시간을 아침, 오후, 저녁으로 나눈다. 이렇게 생성된 기능은 월의 시작인지 끝인지 또는 시작 월과 종료 월 중에 심각한 상태가 발생하는지 여부에 따라 정상적 및 비정상적인 상태의 패턴과 같이 환자 건강에 대한 통찰력을 제공한다. 이후 추가 처리를 위해 최종 데이터 세트를 Hadoop 클라우드 환경으로 업로드한다.

[0035] 센서 데이터는 분석 전에 미리 처리되고 필터링된다. 불일치와 노이즈를 제거하기 위해 데이터를 필터링한다. ASCII 문자를 제거하여 데이터를 정리한다. 데이터에서 이러한 노이즈를 제거하기 위해 칼만(Kalman) 필터링이라는 잘 알려진 필터링 접근방식을 사용한다. 또한, ReplaceMissingValues라는 비지도(unsupervised) 필터는 데이터 세트에 누락된 모든 숫자 값을 사용 가능한 데이터의 수단과 모드로 대체하기 위해 사용된다. 쓸모 없는 속성은 RemoveUseless 필터라는 비지도 필터를 사용하여 최대 90%의 분산으로 제거된다. 그런 다음 숫자 값은 정규화 필터를 사용하여 정규화되며, 모든 분류에 대해 0과 1 사이로 제한된다. 이러한 단계가 끝나면 EmEditor를 활용하여 데이터 세트를 n개의 데이터 파일로 나누고, 추가 처리를 위해 Hadoop 클라우드 환경으로 업로드한다.

[0036] 의료기록(Medical Record; MR)의 사전 처리에 있어서, 진단을 위해 처방된 의약품 데이터에 포함되어 있는 약물에 대한 정보를 이용한다. MR의 특성 중 일부는 환자 분류에 사용될 수 있다. 따라서 데이터 분석을 위해 MR의 각 속성에 ID와 기준 값(0, 1, 2)을 할당한다. 기준치 0, 1, 2는 환자의 건강 상태(정상, 당뇨병 전, 당뇨병)를 나타낸다. 또한 센서 기반 생성 데이터의 한계를 극복하기 위해 MR 데이터를 사용하기도 한다. 예를 들어, 누락된 값을 데이터 세트의 현재 MR 속성 값으로 대체한다.

[0037] 소셜 네트워크 콘텐츠의 사전 처리에 있어서, 다음 단계를 적용하여 데이터를 구조화된 형태로 변환한다. 또한 이러한 단계는 유용한 데이터를 제거하므로 특징과 의견 단어를 쉽게 추출하는 데 도움이 된다.

[0038] 먼저, 단어 제거 중지를 수행한다. 전치사(to, in, and of), 모든 관사(a, an, and the), 기호(#, @ 등), 코퍼스(corpus, 말뭉치)데이터의 URL과 같은 단어는 문서의 의미를 방해하지 않는다. 그러나 텍스트 분류의 정확성은 감소한다. 본 발명의 실시예에 따른 레인보우(Rainbow)라는 잘 알려진 방법을 사용하여 이 내용을 삭제한다.

[0039] 이후, 토큰화 단계에서 흰 공간과 구분자를 제거하여 코퍼스의 복잡한 텍스트를 작은 용어나 토큰으로 구분한다. n-gram 토큰라이저를 적용하여 흰 공간과 구분자를 삭제한다. 그런 다음 출력은 추출된 단어의 음성

(PoS) 태깅 및 표제어와 같은 추가 분석을 위해 저장된다.

[0040] PoS 태깅 단계에서, 코퍼스 텍스트를 문장으로 나눈 다음, POS 태깅에 CoreNLP(Stanford Core Natural Language Processing)를 사용한다.

[0041] 어간 추출(stemming) 및 표제어 추출(lemmatization) 단계에서, 본 발명은 어간 추출을 위해 접미사 드로핑(dropping) 알고리즘을 적용한다. 표제어 추출은 본문에서 사용되는 단어들의 표제어를 표현한다. 표제어 추출 후, 각 단어의 어휘 정보를 쉽게 얻는다. 예를 들어, 혈당(blood sugar)은 혈당(blood glucose)와 관련이 있다. 따라서, 어간과 표제어 단어는 추가 처리를 위해 활용된다.

[0042] 특징 변환 단계에서, 환자는 분류기 결과에 영향을 미치는 SNS에서 특이한 단어(예를 들어, depressssssed)를 사용한다. 따라서 두 번 이상 나타나는 일련의 문자를 일반적인 단어로 변환한다(예를 들어, depressssed는 depressed(우울해지다)가 된다).

[0043] 다음으로, 특징 극성 식별을 수행한다. 제안된 장치는 소셜 네트워크에 게시된 내용을 사용하여 환자의 스트레스와 우울증을 감지한다. 또한, 당뇨병 치료제의 효율성과 부작용에 대한 그들의 의견을 파악하기 위해 약물 검토에 관한 여러 가지 업무를 수행한다. 앞에서 언급한 두 가지 과제에 대해 정서 분석 접근법을 사용한다. 따라서 정서 구분을 위한 특징 극성과 문서 라벨링을 찾는 것이 중요하다. 소셜 네트워크와 웹 페이지 콘텐츠를 사전 처리한 후 SWN(SentiWordNet)을 사용하여 특징 의견 단어 극성을 파악한다. 그런 다음 전체 문서의 극성을 찾기 위해 특징 극성의 결과가 누적된다. 이후, 워드 임베딩(embedding) 및 온톨로지 기반 특징 추출을 수행한다. 도 4를 참조하여 임베딩 및 온톨로지 기반 특징 추출 과정에 대하여 더욱 상세히 설명한다.

[0045] 도 4는 본 발명의 일 실시예에 따른 임베딩 및 온톨로지 기반 특징 추출을 설명하기 위한 도면이다.

[0046] 도 4와 같이 데이터 세트에 200차원의 벡터가 있는 Word2vec의 건너뛰기 프로그램 skip-gram 모델을 훈련시켰다. 이 Word2vec 모델은 형상들 사이의 연관성을 탐지하기 위해 ML 분류기를 훈련시키는데 이용된다. 일반적으로 ML 분류기는 특정 도메인에 따라 특징의 기본 의미론을 놓치게 된다. 온톨로지는 경우에 따라 주어진 도메인의 의미론을 나타낼 수 있다. 온라인에서 이용할 수 있는 생물의학 온톨로지는 당뇨병, 우울증, 고혈압, 약물, 그리고 음식에 관한 다양한 주제를 다룬다.

표 2

[0047]

온톨로지 및 약어	정의	클래스 수	특성 수	개인 수
영양학 온톨로지 (ONS)	복합 영양 연구에 대한 설명을 제공합니다.	3442	66	104
BioMedBridges 당뇨병 온톨로지 (DIAB)	텍스트 마이닝을위한 당뇨병 표현형 간의 관계를 나타냅니다.	375	4	0
당뇨병 치료 온톨로지 (DMTO)	당뇨병 치료를 위한 상호 운용성을 제공합니다.	10700	315	63
인간 질병 온톨로지 (DOID)	그것은 드문 질병의 개념을 나타냅니다.	12694	15	0
약물 목표 온톨로지 (DTO)	그것은 약물 목표 데이터의 분류에 대한 정보를 제공합니다.	10075	0	0
FHIR 및 SSN 기반 유형 1 당뇨병 온톨로지 (FASTO)	당뇨병 환자를 위한 인슐린 관리 정보를 제공합니다.	9577	822	460

[0049] 표 2와 같이 개념과 관계가 많은 영양학 연구용 온톨로지(ONS), BioMedBridges 당뇨 온톨로지(DIAB), 당뇨병 Mellitus 치료 온톨로지(DMTO), 인간 질병 온톨로지(DOID), 약물 표적 온톨로지(DTO), 패스트 헬스케어 상호운용성 자원(FHIR), 의미 센서 네트워크(SSN) 기반 타입1의 당뇨병 온톨로지(FASTO)의 최신판을 활용한다. 이러한 온톨로지는 특이한 단어의 의미를 이해하는 신경망 모델에 대하여 추가 정보를 제공한다.

[0050] 이것은 모든 온톨로지 정보를 검색하고 분류기를 위한 단어 임베딩과 통합하려고 시도한다. 각각의 온톨로지를 BOW(Bag-of-Words)로 생각한다. 온톨로지로부터 정보를 추출하기 위해, 용어 주파수(TF), 용어 빈도 및 TF-IDF(Inverse Document Frequency)라고 하는 잘 알려진 통계 방법을 사용한다. 여기서는 온톨로지의 각 개념이나 특성을 용어로서, 온톨로지를 문서로서 고찰한다. 따라서 TF는 온톨로지(ontology)에서 발견되는 용어(주변어)

이다. TF는 다음 수학식에서 수학적으로 정의된다.

[0051] $$TF(Term,Onto) = 1 + \log(Feature_{term}, Onto) \quad (1)$$

[0052] TF(Term,Onto) 값이 0보다 클 경우 이 단계를 반복하여 특정 개념을 추출한다. TF-IDF는 정보 추출에 대한 예시 용어를 선택한다. IDF는 주로 모든 온톨로지(예를 들어, 환자, 질병, 병원 등)에 나타나는 단어의 의미를 감소시킨다. 용어가 더 많은 온톨로지 또는 단일 온톨로지의 더 많은 개념에서 발생하는 경우, 그것은 정규 용어라는 것을 의미하며 정보 추출에 필요한 용어가 아닐 수 있다. 따라서 입력 단어의 로그 함수의 결과는 0으로 감소할 것이다. 이는 TF-IDF의 가치가 이 용어에 비해 작다는 것을 보여준다. IDF의 통계적 설명은 다음 수학식에 나타나 있다.

[0053] $$IDF\ (Term,Ontos) = \log \frac{N_f}{1 + |\{Onto\ \in\ Ontos : Term\ \in\ Onto\}|} \quad (2)$$

[0054] 여기서 N_f 는 데이터베이스의 전체 온톨로지 수 또는 온톨로지 내의 총 개념 수를 표시한다(예를 들어, $N_f = |O|$ 및 |{Onto∈Ontos:Term ∈Onto}|는 용어가 나타나는 온톨로지 내의 온톨로지 또는 개념의 수이다). 다음 TF-IDF 수학식을 사용하여 공통 용어를 제거한다.

[0055] $$\mathrm{TF} - \mathrm{IDF} = TF(Term,Onto)\ .IDF(Term,Ontos) \quad (3)$$

[0056] TF-IDF 결과는 온톨로지 코퍼스에서 온톨로지 기능이 온톨로지에게 얼마나 필수적인지를 보여주거나 온톨로지에서의 개념에 온톨로지 기능이 얼마나 중요한지를 보여준다.

[0057] 다음으로, 웨어러블 센서 데이터에서 추출하는 특징에 대하여 설명한다. 특징 추출에 정보 이득(IG) 방법을 사용하는데, 이는 소음과 관련 없는 특징을 줄임으로써 분류기에 영향을 미친다. WEKA(Waikato Environment for Knowledge Analysis)를 활용하여 정보 이득을 계산한다. 당뇨병 및 BP 데이터 세트에 대한 평가자로 정보 이득 필터 "infoGainAtrributeVal"을 적용하여 결과를 얻는다. 속성 값이 식별되면, IG 측정은 훈련 데이터 세트의 엔트로피 감소와 연계된다. 이 접근법은 분류에 따라 IG를 계산하여 속성의 가치를 찾는다. 제안된 IG 방법은 시스템 불확실성을 측정하기 위해 엔트로피를 이용하고 이전의 엔트로피와 이후의 엔트로피 차이를 발견한다. 그것은 수학식 4와 같이 B가 제공하는 A에 대한 추가 정보의 양을 명시한다.

[0058] $$IG(A|B) = H(A) - H(A|B) \quad (4)$$

[0059] 여기서 A와 B는 별개의 변수이다. A는 특징이며, 이전의 엔트로피는 수학식 5를 사용하여 측정할 수 있다.

[0060] $$H(A) = -\sum_i P(A_i)\log_2 P(A_i) \quad (5)$$

[0061] 여기서 $P(A_i)$ 는 A_i 의 이산 값에 대한 사전 확률을 나타낸다. A의 조건부 엔트로피는, 이후의 엔트로피 B가 주어진 후에, 수학식 6과 7에서와 같이 정의될 수 있다.

[0062] $$H(A|B) = -\sum_j P(B_j)\, H(A|B_j) \quad (6)$$

[0063] $$= -\sum_j P(B_j) \sum_i \left(P\left(A_i|B_j\right) log_2 P\left(A_i|B_j\right)\right) \quad (7)$$

[0064] 정보 이득은 수학식 8과 같이 수학식 5와 7을 수학식 4에 넣어 계산할 수 있다.

[0065] $$IG(A|B) = -\sum_i P(A_i)\log_2 P(A_i) - \left(-\sum_j P(B_j) \sum_i \left(P\left(A_i|B_j\right)\log_2 P\left(A_i|B_j\right)\right)\right) \quad (8)$$

[0066] 다음으로, 주요 구성 요소를 분석한다. 데이터 세트의 차원수는 사전 처리 및 추가 속성 생성 후에 증가한다. 이는 분류 정확도, 적합도 및 시간 복잡성의 감소와 같은 문제를 야기한다. 따라서, 차원 감소에 대한 통계적 접근방식인 주성분 분석(PCA)을 사용한다. PCA는 p차원의 X 데이터를 가장 적은 손실로 q차원의 Y 데이터로 변

환한다.

[0067] 이후, Bi-LSTM 기반 당뇨병과 BP 분류 및 약물 부작용을 예측한다. Bi-LSTM는 환자의 당뇨병, BP, 정신건강, 당뇨병 약물의 부작용 등을 분류하는 데 쓰인다.

[0069] 도 5는 본 발명의 일 실시예에 따른 LSTM 기반 분류 및 예측을 설명하기 위한 도면이다.

[0070] 도 5는 당뇨병, BP, 정신건강, 약물 부작용의 분류를 위한 Bi-LSTM의 구조를 보여준다. 그리드 검색 최적화 알고리즘을 적용하여 LSTM 모델의 하이퍼 파라미터에 대한 최적의 값을 식별한다. 하이퍼 파라미터 드롭아웃 속도, 에폭(epochs), 배치 크기 및 학습 속도에 대해 선택된 최적 값은 각각 0.3, 30, 32, 0.001이다. 도 5와 같이, LSTM은 네 가지 다른 유형의 데이터 세트를 분류하는 데 사용된다. 당뇨병 분류를 위해, Pima Indians Diabaetes 데이터 세트는 UCI 기계 학습 저장소에서 수집되었다. 데이터 세트에는 8개의 입력 속성이 있다. 그러나 당뇨병 분류 모형을 훈련하는 데는 6가지 속성만 사용된다. 이러한 속성은 연령, BMI, BP, 혈장 포도당, 당뇨병 혈통 기능, 2시간 혈청 인슐린 등이다. 집중 치료 II(MIMIC-II) 데이터베이스의 PhysioNet 다중 파라미터 지능형 모니터링 데이터 세트를 사용하여 BP 분류 모델을 교육했다. 이 데이터 집합은 BP와 심박수(HR)를 포함한 다수의 활력징후 샘플로 구성된다. 약물 부작용 예측을 위한 데이터 세트는 UCI 저장소에서 획득했다. 이 데이터 집합은 6개의 속성으로 구성되어 있다. 그러나 제안된 작업에서는 두 가지 속성(약물과 환자 리뷰 이름)만 고려된다. 본 발명에서는 $LSTM^{a}$, $LSTM^{b}$, $LSTM^{c}$, $LSTM^{d}$ 의 네 가지 Bi-LSTM 기반 분류기 모델을 개발했다.

[0071] 당뇨병 분류에는 나이, 가족, 성별, 활동, BMI, 혈압, 혈당 기능이 사용된다. 각 타임스텝, t, 특성 Xt를 $LSTM^{a}$ 단위로 입력하여 당뇨병 등급 {정상, 당뇨병 전, 당뇨병}을 예측한다. 당뇨병 및 혈압 환자를 분류하는 의료규칙은 표 3에 제시되어 있다.

표 3

[0072]

	카테고리	나이(연도)	가족력(Y/N)	성별 (M/F)	활동 (단계)	BMI	혈당(MG/dL)	심박수(bpm)	수축기 혈압 (mmHg)	이완기 혈압 (mmHg)
당뇨병 분류	정상	20-50	N	M/F	5000+	18-25	≤ 100 (fasting)	60-80	-	-
	당뇨병 전	41-65	Y	M/F	3001-5000	25-30 or more	100 to 125 (fasting)	140-160	-	-
	당뇨병	60+	Y	M/F	≤ 3000	30+	≥126 (fasting)	140-160 or ≥160	-	-
혈압 분류	정상	-	-	-	≤ 3000	-	-	60-100	90-119	60-79
	저혈압 (저혈압)	-	-	-	3001-5000	-	-	-	≤ 90	≤ 60
	고혈압 (고혈압)	-	-	-	5000+	-	-	-	≥140	≥ 90

[0073] 혈압과 심전도 검사에서 웨어러블 센서는 각 환자의 BP와 심장 박수를 식별하는 데 사용된다. 추출된 정보는 특징(형상) 구성에 사용된다. 특징을 생성 및 사전 처리한 후, $LSTM^{b}$ 라는 이름의 Bi-LSTM 기반의 모델을 개발했는데, 이 모델은 과거 시퀀싱된 데이터에 존재하는 시간 패턴을 유지할 수 있다. BP의 분류를 위해 각 단계 t에서 생성된 모든 파라미터와 특징을 카테고리 벡터 {0, 1, 2}로 변환한다.

[0074] 정신 건강 모니터링에서 환자 메시지 및 게시물은 소셜 네트워크에서 추출되며 지도 및 비지도 접근방식을 사용하여 필터링된다. 200차원의 Word2vec 모델을 훈련시킨 다음, LSTMc 모델에 워드 시퀀스를 공급했다. 그런 다음 $LSTM^{c}$ 의 출력은 소프트맥스 함수에 할당되며, 소프트맥스 함수는 게시된 내용의 각 문장에 대한 정서 라벨(양성, 중성, 음성 또는 강한 음성)을 예측한다. 예측 극성을 기준으로 환자의 정신건강을 분류하는 규칙은 표 4에 나와 있다.

표 4

[0075]

	카테고리	환자 게시물 및 의견에 대한 감정적 인 극성 (양성/중성/음성/ 강한음성)	항 당뇨병 약물에 대한 감정 극성 (양성/중성/음성)
환자 정신 건강 분류	행복	양성	-
	정상	중성	-
	우울	음성	-
	스트레스	강한 음성	-
약물 부작용 예측	부작용 없음	-	양성
	중간 부작용	-	중성
	심각한 부작용	-	음성

[0077] 약물 부작용 예측에서는 현재 섭취하는 당뇨병과 BP 약물을 입력 질의로 사용하고 있으며, 다른 웹사이트에서 이에 대한 리뷰를 검색했다. 필터링과 사전처리 후에, 리뷰는 자동적으로 부작용이 없고/적당한 부작용이 있고/심각한 부작용이 있다고 라벨을 붙인다. 이 라벨링에 대해 각 문장의 정서 극성을 파악한다. 그런 다음, 표 4와 같이 양성, 중성, 음성의 정서 극성은 각각 부작용이 없고, 온건한 부작용과 심각한 부작용이 없는 것으로 간주한다. 200차원 Word2vec 모델은 LSTM[d] 모델에서 일련의 단어로 약물 리뷰를 나타내도록 훈련되었다. 그런 다음 LSTM[d] 의 출력을 약물 부작용을 정확하게 예측하는 소프트맥스 기능에 공급한다.

[0078] 다시 도 1을 참조하면, 헬스케어 모니터링(151) 및 추천(152)을 위한 데이터 표시부(150)을 통해 데이터의 계산 및 분류 결과에 따른 분석 데이터를 표시한다. 당뇨병 분류에서 얻은 결과는 표 5에 제시되어 있다.

표 5

[0079]

제안된 모델 및 기타 분류기	정도(P) (%)	리콜(R) (%)	기능 측정(FM) (%)	정확도(Ac) (%)	RMSE	MAE
CNN	62	66	63	66	58	34
MLP	67	67	67	68	51	34
SVM	67	70	68	70	54	30
퍼지 분류기	73	72	65	72	52	26
로지스틱 회귀	70	68	69	69	55	32
랜덤 포레스트	67	70	66	70	46	42
KNN	57	65	59	65	52	50
LSTM[a]	74	75	75	**75**	50	**26**

[0081] 다른 분류기(CNN, MLP, SVM, 퍼지 분류기(fuzzy logic), 로지스틱 회귀(logistic regression), 랜덤 포레스트(random forest) 및 KNN)는 피마 인디언(Pima Indians) 데이터 세트를 사용하여 제안된 LSTMa 모델과 비교되었다. 그 결과 제안된 LSTM[a] 는 다른 분류기와 비교하여 가장 높은 정확도(75%)와 가장 낮은 MAE(26%)를 얻었다. 이 높은 정확도는 LSTM[a] 가 당뇨병 분류를 위해 기억 세포에 더 중요한 정보를 저장하고 있음을 나타낸다. 또한 최저 MAE는 LSTM[a]모델의 더 나은 성능을 보여준다. 퍼지 분류기는 다른 분류기에 비해 가장 낮은 RMSE(46)를 얻었다. 이 실험을 바탕으로 다른 분류기들은 시간이 많이 소요되고, 적은 수의 특징에도 성능이 저하되는 것을 관찰한다. 그러나 LSTM은 당뇨병 예측에서 다른 분류기들을 능가했다.

[0082] 표 6은 제안된 모델과 기타 분류기가 BP 분류 측면에서 얻은 결과를 보여준다.

표 6

[0083]

제안된 모델 및 기타 분류기	정도(P) (%)	리콜(R) (%)	기능 측정(FM) (%)	정확도(Ac) (%)	RMSE	MAE
CNN	71	70	69	70	54	30
MLP	80	80	80	80	37	22
SVM	83	75	73	74	50	25
퍼지 분류기	84	83	83	83	40	16
로지스틱 회귀	72	72	71	72	51	29
랜덤 포레스트	78	70	67	70	45	42
KNN	60	58	56	58	64	42
LSTM[b]	89	87	87	88	34	12

[0085] LSTM[b] 가 CNN(70%), MLP(80%), SVM(73%), 퍼지 분류기(83%), 로지스틱 회귀(72%), 랜덤 포레스트(70%), KNN(58%)에 비해 가장 높은 정확도(88%)를 달성함을 관측했다. 그러나 제안된 모델의 MAE와 RMSE는 각각 12와 34로 다른 분류기보다 낮았다.

[0086] 표 7은 정신 건강 분류의 측면에서 분류기의 성과를 보여준다.

표 7

[0087]

제안된 모델 및 기타 분류기	정도(P) (%)	리콜(R) (%)	기능 측정(FM) (%)	정확도(Ac) (%)	RMSE	MAE
CNN	66	66	66	66	56	34
MLP	72	71	71	72	47	29
SVM	68	68	68	68	56	32
퍼지 분류기	60	58	56	59	67	45
로지스틱 회귀	70	75	72	71	44	34
랜덤 포레스트	60	60	59	60	58	40
KNN	64	63	64	64	60	36
LSTM[c]	87	90	87	89	35	15

[0089] 표 7에 따르면, 소프트맥스를 사용하는 LSTM[c] 는 다른 분류기와 비교했을 때 가장 높은 정확도(89%)를 보여준다. 또 LSTM[c] 의 MAE와 RMSE는 각각 15와 35로 다른 분류기보다 낮다.

표 8

[0090]

제안된 모델 및 기타 분류기	정도(P) (%)	리콜(R) (%)	기능 측정(FM) (%)	정확도(Ac) (%)	RMSE	MAE
CNN	76	68	65	68	62	39
MLP	82	81	81	81	39	20
SVM	82	82	82	82	36	18
퍼지 분류기	79	78	78	78	41	28
로지스틱 회귀	84	82	83	83	36	17
랜덤 포레스트	76	66	63	66	58	33
KNN	77	75	75	76	49	24
LSTM[d]	88	90	89	90	32	13

[0092] 제안된 모델의 약물 리뷰를 이용한 분류 결과를 평가하기 위해, 표 8에서와 같이 $LSTM^d$ 모델을 다른 분류기와 비교했고, 로지스틱 회귀 분석은 각각 90%와 83%에서 높은 정확도를 얻었다. 다른 분류기들은 $LSTM^d$ 및 로지스틱 회귀 분석과 비교하여 더 낮은 정확도를 달성했다. $LSTM^d$ 의 RMSE와 MAE는 각각 32와 13으로 다른 분류기보다 낮다. CNN의 정확도는 매우 낮으며, 다른 분류기들에 비해 RMSE가 매우 높다. 이는 $LSTM^d$ 가 단어 벡터의 크기가 증가함에 따라 단어 벡터의 긴 시퀀스를 처리할 수 있음을 나타낸다. 이와는 대조적으로, CNN은 단어 순서와 단어 임베딩의 차원이 증가하면 과적합하게 된다.

[0094] 도 6은 본 발명의 일 실시예에 따른 웨어러블 센서와 소셜네트워크 데이터를 활용한 헬스케어 모니터링 방법을 설명하기 위한 흐름도이다.

[0095] 제안하는 웨어러블 센서와 소셜네트워크 데이터를 활용한 헬스케어 모니터링 방법은 데이터 수집부를 통해 웨어러블 기기로부터 생리적 데이터, 스마트 단말로부터 약물 및 증상 데이터, 어플리케이션 프로그래밍 인터페이스(Application Programming Interfaces; API)를 통한 환자들 및 의사들의 소셜 네트워크 상에서의 논의 데이터를 수집하는 단계(610), 데이터 저장부는 개인 클라우드 서버와 연결되어 수집된 생리적 데이터, 약물 및 증상 데이터 및 논의 데이터를 저장하는 단계(620), 저장된 데이터는 빅데이터 분석 엔진을 통해 특징 극성 식별, 워드 임베딩 및 온톨로지 기반 특징 추출, 웨어러블 센서 데이터에서 추출, 데이터 구성 요소 분석, Bi-LSTM 기반 당뇨병과 BP 분류 및 약물 부작용 예측을 포함하는 데이터의 계산 및 분류를 수행하는 단계(630) 및 데이터의 계산 및 분류 결과에 따는 분석 데이터를 표시하는 단계(640)를 포함한다.

[0096] 단계(610)에서, 데이터 수집부를 통해 웨어러블 기기로부터 생리적 데이터, 스마트 단말로부터 약물 및 증상 데이터, 어플리케이션 프로그래밍 인터페이스(Application Programming Interfaces; API)를 통한 환자들 및 의사들의 소셜 네트워크 상에서의 논의 데이터를 수집한다.

[0097] 단계(620)에서, 데이터 저장부는 개인 클라우드 서버와 연결되어 수집된 생리적 데이터, 약물 및 증상 데이터 및 논의 데이터를 저장한다.

[0098] 저장된 데이터는 센서 데이터, 의료 기록 및 소셜 네트워크 콘텐츠를 포함하고, 센서 데이터의 사전 분석, 센서 데이터의 사전 처리 및 필터링, 의료 기록의 사전 처리, 소셜 네트워크 콘텐츠의 사전 처리 및 불일치와 노이즈를 제거하기 위한 데이터 필터링을 수행한 후 데이터를 구조화된 형태로 변환한다.

[0099] 단어의 의미를 이해하는 신경망 모델에 대하여 추가 정보를 제공하기 위해 온톨로지 기반 특징 추출을 수행하여 모든 온톨로지 정보를 검색하고, 분류를 위한 워드 임베딩과 통합하며, 온톨로지로부터 정보를 추출하기 위해, 용어 주파수(Term Frequency; TF), 용어 빈도 및 TF-IDF(Inverse Document Frequency)를 포함하는 통계 방법을 사용하여 수학적으로 정의한다.

[0100] WEKA(Waikato Environment for Knowledge Analysis)를 통해 계산된 정보 이득(Information Gain; IG) 방법을 사용하여 웨어러블 센서 데이터의 특징을 추출하고 속성의 가치를 찾아, 불확실성을 측정하기 위해 엔트로피를 이용하고 이전의 엔트로피와 이후의 엔트로피 차이를 정의한다.

[0101] 단계(630)에서, 저장된 데이터는 빅데이터 분석 엔진을 통해 특징 극성 식별, 워드 임베딩 및 온톨로지 기반 특징 추출, 웨어러블 센서 데이터에서 추출, 데이터 구성 요소 분석, Bi-LSTM 기반 당뇨병과 BP 분류 및 약물 부작용 예측을 포함하는 데이터의 계산 및 분류를 수행한다.

[0102] 단계(640)에서, 데이터의 계산 및 분류 결과에 따는 분석 데이터를 표시한다.

[0104] 도 7은 본 발명의 일 실시예에 따른 온톨로지 사용에 따른 LSTM 기반 분류의 정확성과 MAE를 나타내는 그래프이다.

[0105] 도 7은 온톨로지를 사용하지 않고 / 온톨로지를 사용하여 제안된 $LSTM^c$ 및 $LSTM^d$ 모델의 정확성과 MAE를 나타낸다. 볼 수 있듯이, 제안된 온톨로지 LSTM는 단순한 LSTM보다 상당히 개선된 것을 보여준다. 정서문자 분류에서는 $LSTM^c$ 의 정확도가 84%로 온톨로지 사용시 87%로 높아졌다. 하지만 MAE는 1개만 줄었다. 또한 약물 부작용 분류에서 $LSTM^d$ 의 정확도와 MAE는 각각 90%와 14였다. 그러나 $LSTM^d$와 함께 온톨로지를 사용할 경우 이 정확도

는 93%로 증가하였고, MAE는 1만큼 감소하였다.

[0107] 도 8은 본 발명의 일 실시예에 따른 PCA 및 IG와 함께 분류 모델을 사용하여 얻은 결과를 나타내는 그래프이다.

[0108] PCA 및 IG와 함께 분류 모델을 사용하여 얻은 결과는 도 8에 나타나 있다. 도 8에서 얻은 결과를 바탕으로, PCA 및 IG를 사용한 LSTM[a], LSTM[b], 온톨로지-LSTM[c] 및 온톨로지-LSTM[d] 의 정확도가 각각 4%, 1%, 3%, 1% 증가했음을 주목한다.

[0109] 본 발명의 실시예들에 따르면 제안된 빅데이터 분석 엔진을 통해 데이터 마이닝 기술, 온톨로지 및 양방향 장단기 메모리(Bi-LSTM)를 기반하여, 데이터 마이닝 기술을 통해 의료 데이터를 효율적으로 사전 처리하고 유용한 기능을 추출하며 데이터의 차원수를 줄일 수 있다. 또한, 제안된 온톨로지에서는 실체 및 측면에 대한 의미론적 지식과 당뇨병 및 혈압 영역에서의 관계를 제공하며, Bi-LSTM는 의료 데이터를 정확하게 분류하여 환자의 약물 부작용과 비정상 상태를 예측할 수 있다. 또한, 제안된 장치는 당뇨병, BP, 정신 건강 및 약물 검토와 관련된 의료 데이터를 사용하여 환자 분류에 활용되어, 제안된 모델을 통해 이질적인 데이터를 정확하게 처리하고 건강 상태 분류와 약물 부작용 예측의 정확성을 향상시킬 수 있다.

[0111] 이상에서 설명된 장치는 하드웨어 구성요소, 소프트웨어 구성요소, 및/또는 하드웨어 구성요소 및 소프트웨어 구성요소의 조합으로 구현될 수 있다. 예를 들어, 실시예들에서 설명된 장치 및 구성요소는, 예를 들어, 프로세서, 콘트롤러, ALU(arithmetic logic unit), 디지털 신호 프로세서(digital signal processor), 마이크로컴퓨터, FPA(field programmable array), PLU(programmable logic unit), 마이크로프로세서, 또는 명령(instruction)을 실행하고 응답할 수 있는 다른 어떠한 장치와 같이, 하나 이상의 범용 컴퓨터 또는 특수 목적 컴퓨터를 이용하여 구현될 수 있다. 처리 장치는 운영 체제(OS) 및 상기 운영 체제 상에서 수행되는 하나 이상의 소프트웨어 애플리케이션을 수행할 수 있다. 또한, 처리 장치는 소프트웨어의 실행에 응답하여, 데이터를 접근, 저장, 조작, 처리 및 생성할 수도 있다. 이해의 편의를 위하여, 처리 장치는 하나가 사용되는 것으로 설명된 경우도 있지만, 해당 기술분야에서 통상의 지식을 가진 자는, 처리 장치가 복수 개의 처리 요소(processing element) 및/또는 복수 유형의 처리 요소를 포함할 수 있음을 알 수 있다. 예를 들어, 처리 장치는 복수 개의 프로세서 또는 하나의 프로세서 및 하나의 콘트롤러를 포함할 수 있다. 또한, 병렬 프로세서(parallel processor)와 같은, 다른 처리 구성(processing configuration)도 가능하다.

[0112] 소프트웨어는 컴퓨터 프로그램(computer program), 코드(code), 명령(instruction), 또는 이들 중 하나 이상의 조합을 포함할 수 있으며, 원하는 대로 동작하도록 처리 장치를 구성하거나 독립적으로 또는 결합적으로(collectively) 처리 장치를 명령할 수 있다. 소프트웨어 및/또는 데이터는, 처리 장치에 의하여 해석되거나 처리 장치에 명령 또는 데이터를 제공하기 위하여, 어떤 유형의 기계, 구성요소(component), 물리적 장치, 가상 장치(virtual equipment), 컴퓨터 저장 매체 또는 장치에 구체화(embody)될 수 있다. 소프트웨어는 네트워크로 연결된 컴퓨터 시스템 상에 분산되어서, 분산된 방법으로 저장되거나 실행될 수도 있다. 소프트웨어 및 데이터는 하나 이상의 컴퓨터 판독 가능 기록 매체에 저장될 수 있다.

[0113] 실시예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 실시예를 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다.

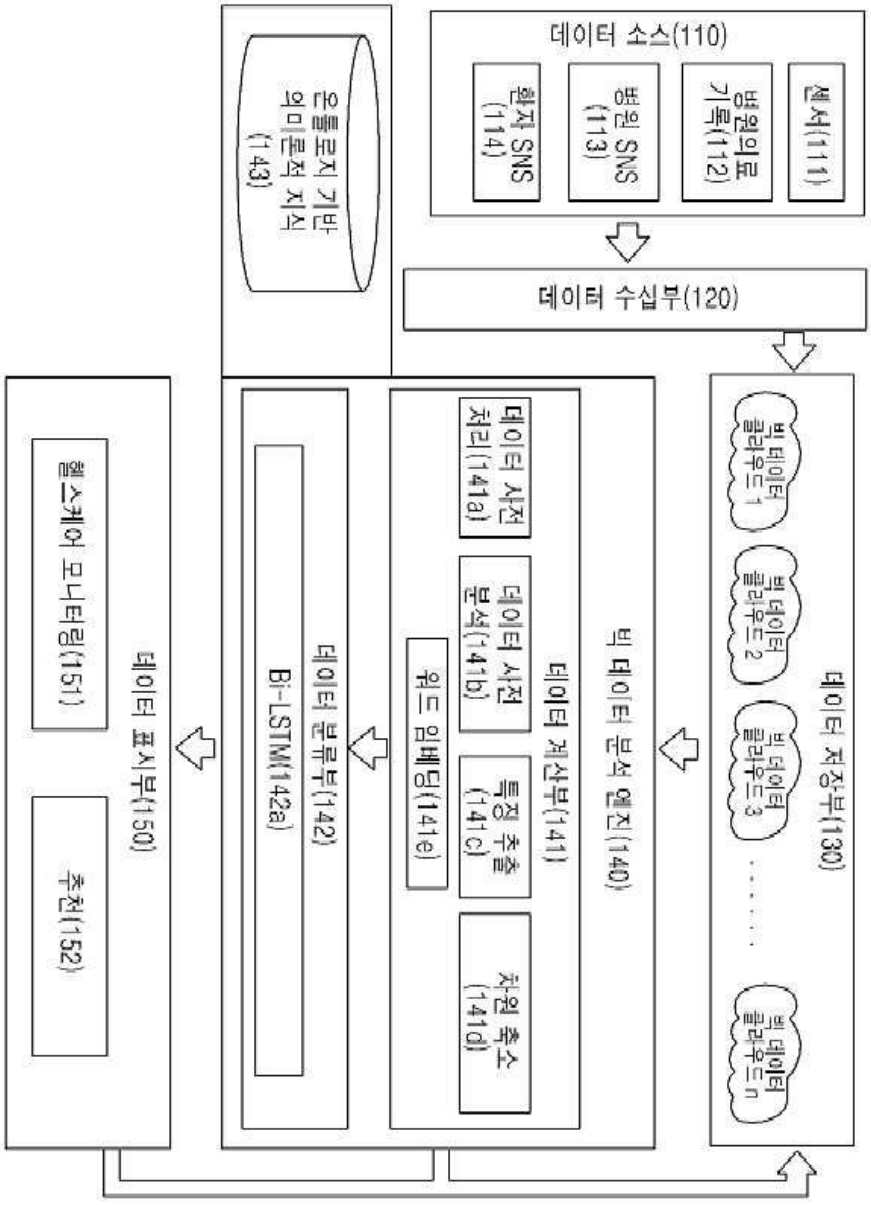
[0114] 이상과 같이 실시예들이 비록 한정된 실시예와 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 상기의 기재로부터 다양한 수정 및 변형이 가능하다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성

될 수 있다.

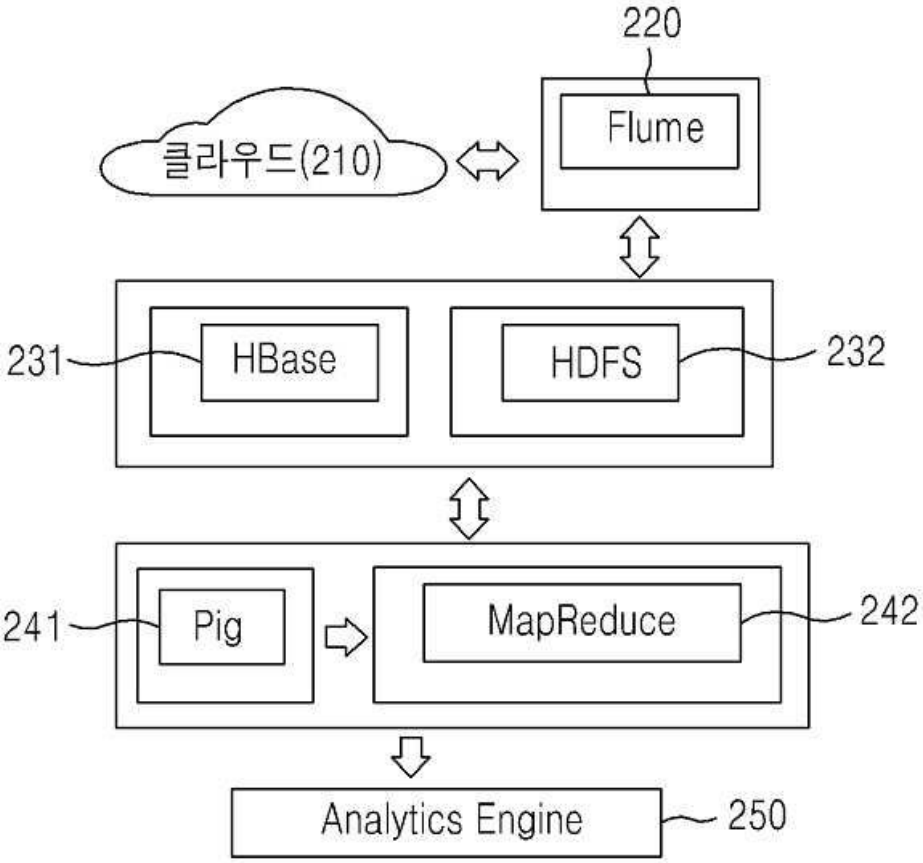
[0115] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

도면

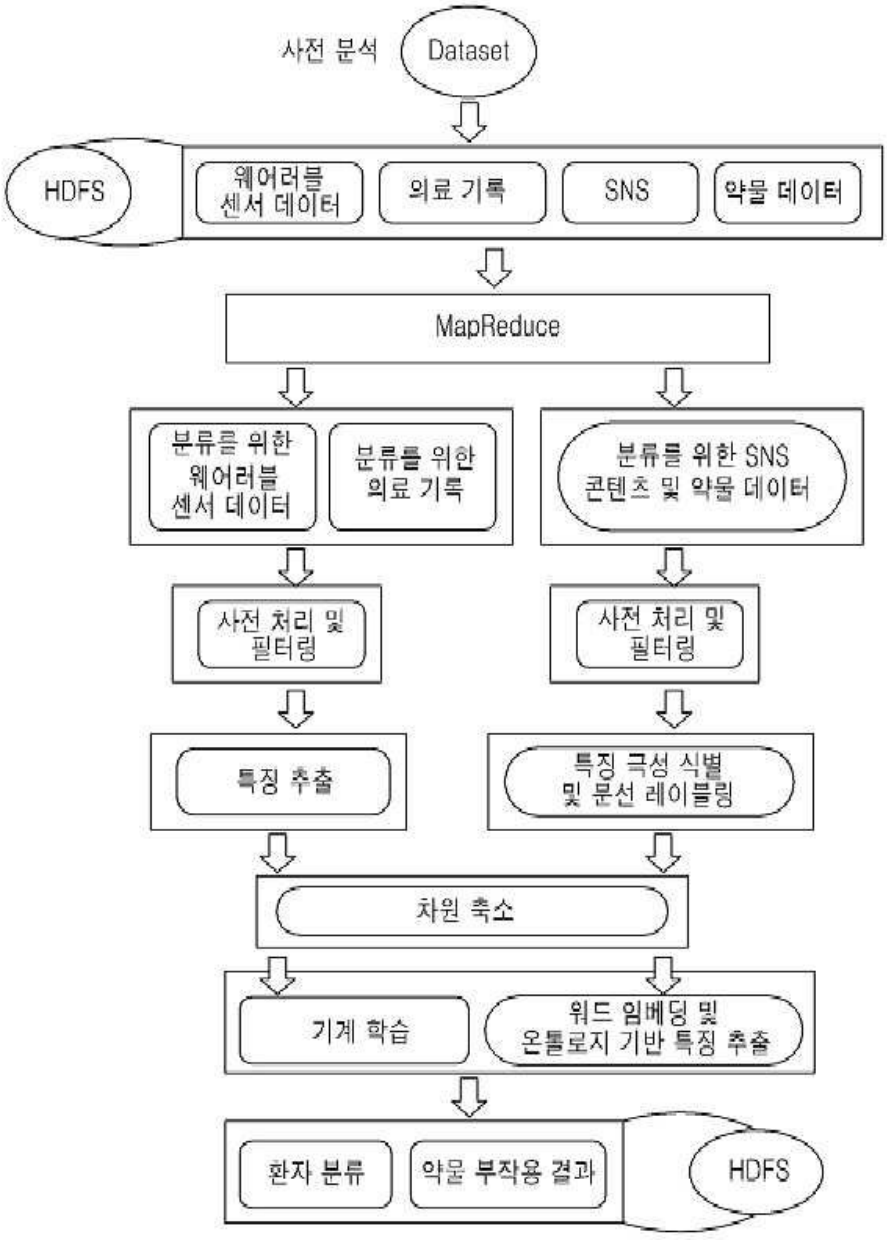
도면1



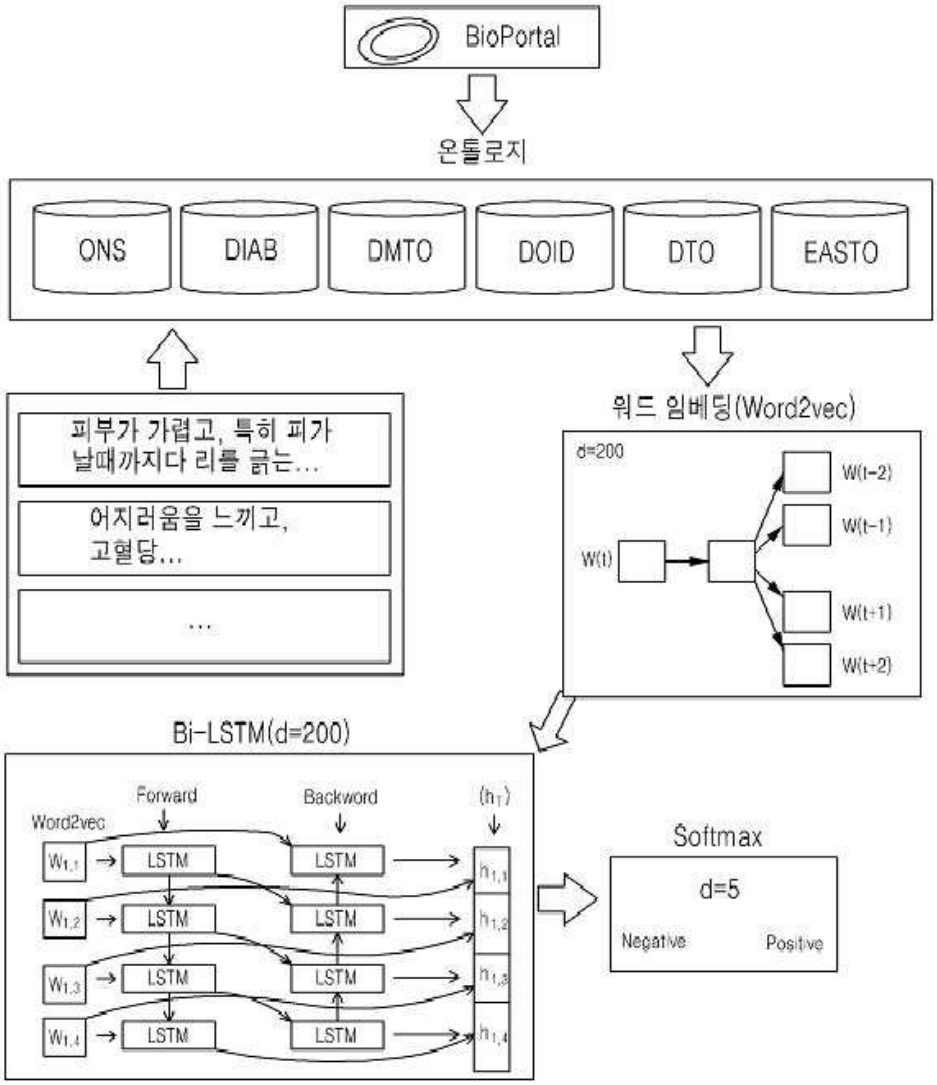
도면2


220
Flume
클라우드(210)
231
HBase
HDFS
232
241
Pig
MapReduce
242
Analytics Engine
250

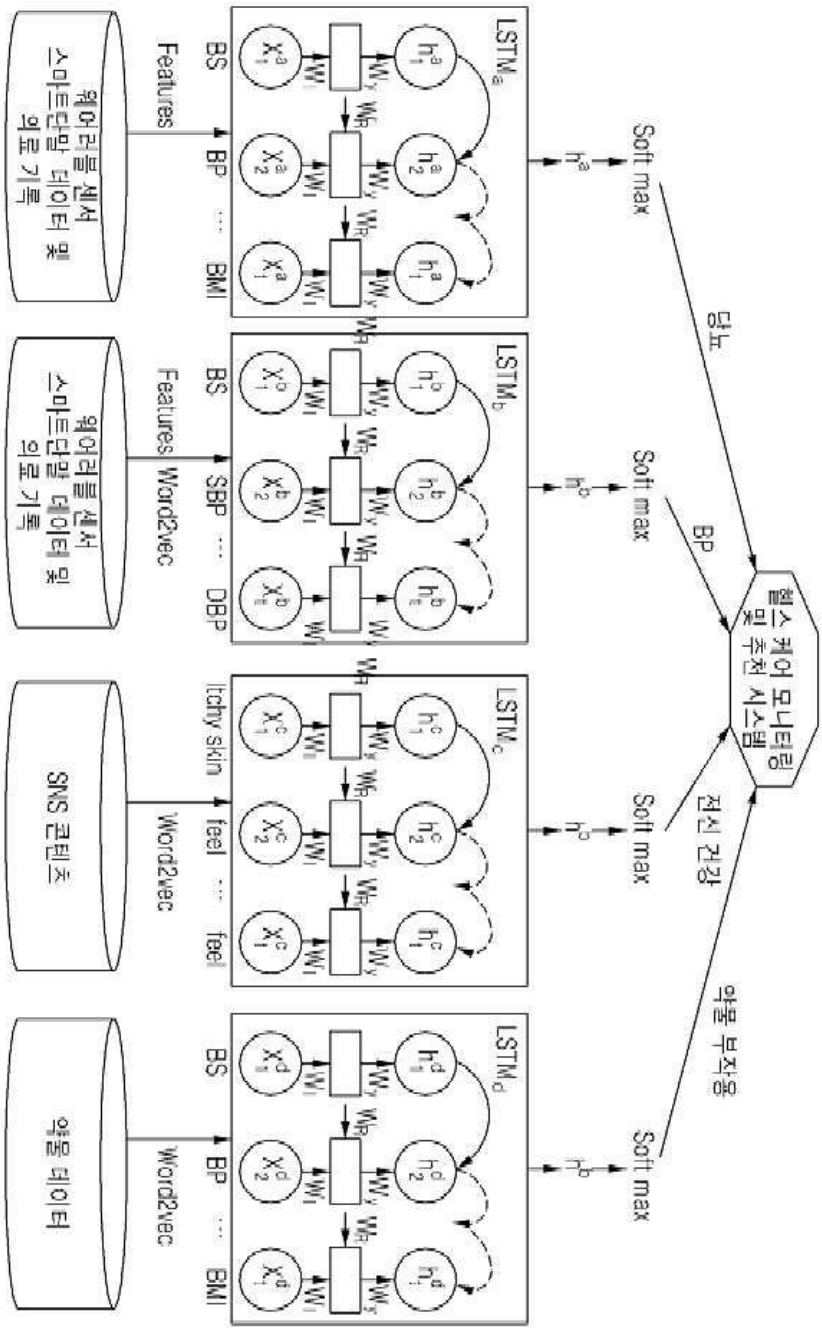
도면3


사전 분석
Dataset
HDFS
웨어러블 센서 데이터
의료 기록
SNS
약물 데이터
MapReduce
분류를 위한 웨어러블 센서 데이터
분류를 위한 의료 기록
분류를 위한 SNS 콘텐츠 및 약물 데이터
사전 처리 및 필터링
사전 처리 및 필터링
특징 추출
특징 극성 식별 및 문선 레이블링
차원 축소
기계 학습
워드 임베딩 및 온톨로지 기반 특징 추출
환자 분류
약물 부작용 결과
HDFS

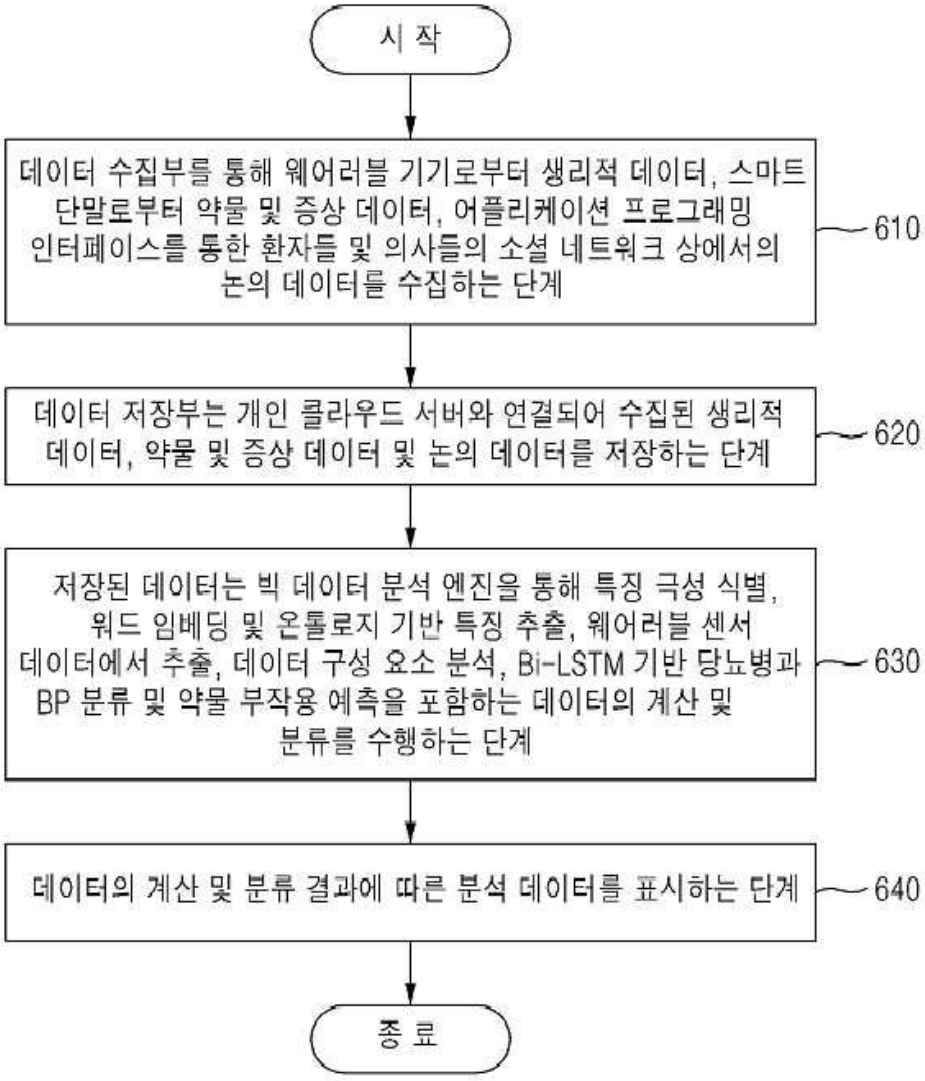
도면4


BioPortal
온톨로지
ONS
DIAB
DMTO
DOID
DTO
EASTO
피부가 가렵고, 특히 피가 날때까지다 리를 긁는…
어지러움을 느끼고, 고혈당…
…
워드 임베딩(Word2vec)
d=200
W(t)
W(t-2)
W(t-1)
W(t+1)
W(t+2)
Bi-LSTM(d=200)
Forward
Backword
Word2vec
LSTM
Softmax
d=5
Negative
Positive

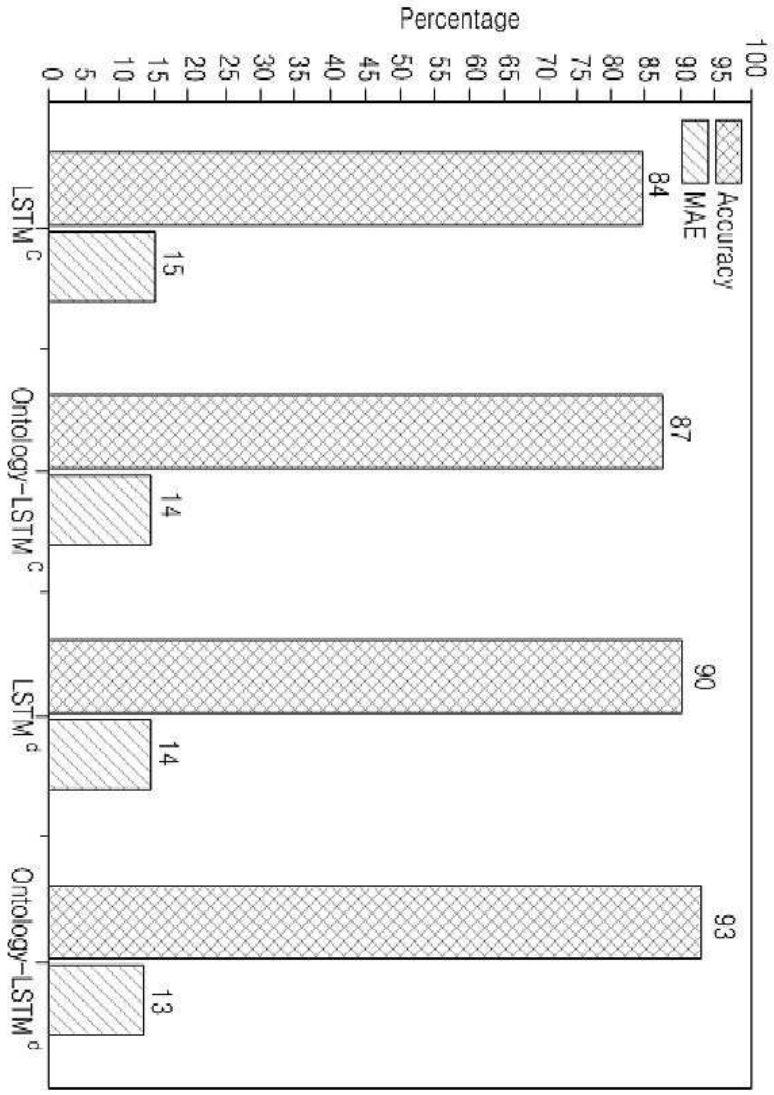
도면5


웨어러블 센서 스마트단말 데이터 및 의료 기록
Features
BS
BP
BMI
LSTMa
Soft max
당뇨
웨어러블 센서 스마트단말 데이터 및 의료 기록
Features
Word2vec
BS
SBP
DBP
LSTMb
Soft max
BP
헬스 케어 모니터링 및 추천 시스템
SNS 콘텐츠
Word2vec
Itchy skin
feel
feel
LSTMc
Soft max
전신 건강
약물 데이터
Word2vec
BS
BP
BMI
LSTMd
Soft max
약물 부작용

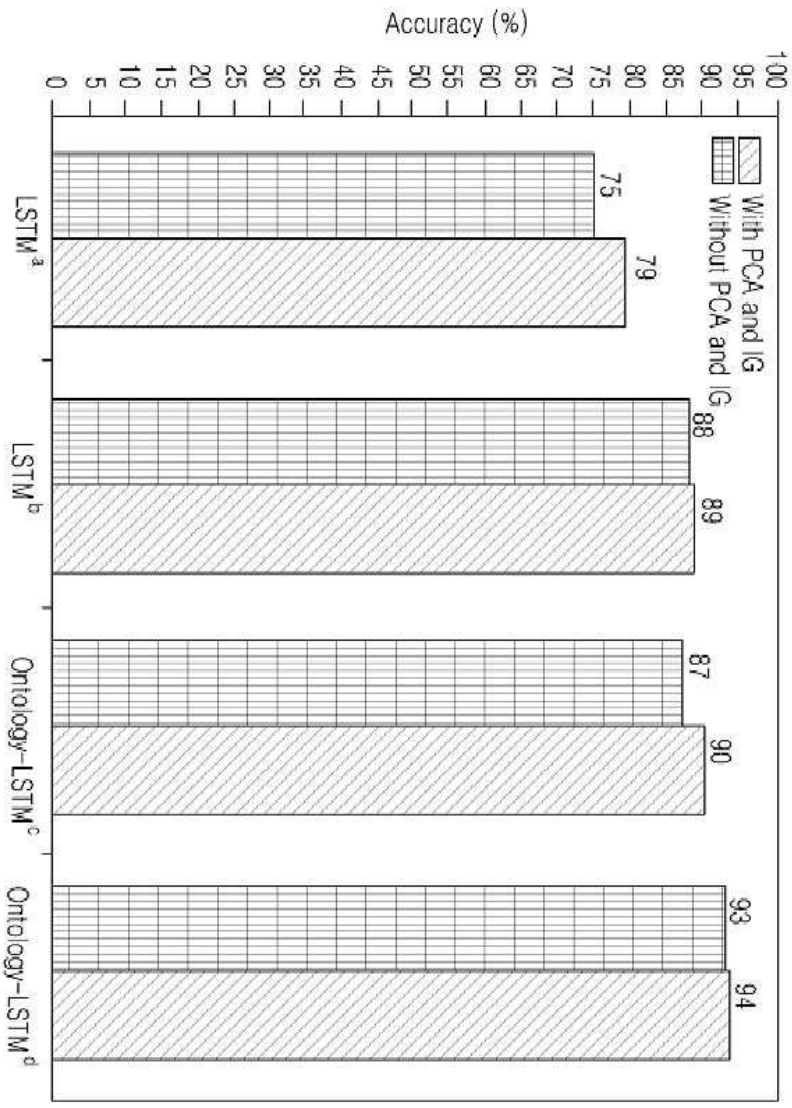
도면6


시 작
데이터 수집부를 통해 웨어러블 기기로부터 생리적 데이터, 스마트 단말로부터 약물 및 증상 데이터, 어플리케이션 프로그래밍 인터페이스를 통한 환자들 및 의사들의 소셜 네트워크 상에서의 논의 데이터를 수집하는 단계
610
데이터 저장부는 개인 클라우드 서버와 연결되어 수집된 생리적 데이터, 약물 및 증상 데이터 및 논의 데이터를 저장하는 단계
620
저장된 데이터는 빅 데이터 분석 엔진을 통해 특징 극성 식별, 워드 임베딩 및 온톨로지 기반 특징 추출, 웨어러블 센서 데이터에서 추출, 데이터 구성 요소 분석, Bi-LSTM 기반 당뇨병과 BP 분류 및 약물 부작용 예측을 포함하는 데이터의 계산 및 분류를 수행하는 단계
630
데이터의 계산 및 분류 결과에 따른 분석 데이터를 표시하는 단계
640
종 료

도면7


Percentage
0
5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100
Accuracy
MAE
84
15
87
14
90
14
93
13
LSTM c
Ontology-LSTM c
LSTM d
Ontology-LSTM d

도면8


Accuracy (%)
0
5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100
With PCA and IG
Without PCA and IG
LSTM[a]
LSTM[b]
Ontology-LSTM[c]
Ontology-LSTM[d]
75
79
88
89
87
90
93
94