

의료인공지능 인터페이스를 갖춘 의료정보시스템 연구

이정훈¹, 김영옥¹, 김상우¹, 최소운¹, 김진우¹, 엄태현¹, 이혁준¹
1 광운대학교

leejh9226@kw.ac.kr, rumia0601@kw.ac.kr, anna7a@kw.ac.kr, tidhtmc1wm@kw.ac.kr, fkrk56@kw.ac.kr, crackscendo@kw.ac.kr, hlee@kw.ac.kr

A Study on the Health Information System with Medical AI Interface

JeongHoon Lee¹, YoungWook Kim¹, SoYun Choi¹, SangWoo Kim¹, JinWoo Kim¹,
Tae-Hyun Eom¹, Hyukjoon Lee¹
1 Kwangwoon Univ.

요 약

의료 인공지능 개발 및 사용에는 방대한 의료 데이터가 요구된다. 따라서 의료 인공지능과 의료정보시스템 사이의 데이터를 중개할 시스템이 요구된다. 하지만, 의료 데이터에는 개인정보 등 반출이 어려운 정보가 포함되어 그 개발과 도입에 걸림돌이 되고 있다. 이 논문에서는, 의료정보시스템과 인공지능 모듈 사이에 이벤트 허브를 사용하는 이벤트 기반 데이터 처리 구조(Event Driven Architecture)를 제안한다.

I. 서 론

지난 수년간 전자의무기록의 전산화, 수집을 수행하는 데이터 웨어하우스의 실용화가 이루어졌다.[1] 하지만 축적한 임상데이터에 대한 인공지능 서비스의 실제 활용사례는 찾아보기가 어렵다. 이러한 배경에는 데이터 비식별화의 어려움을 들 수 있다. 대표적인 예시로 EMR(Electronic Medical Record)에는 환자 개인정보와 환자의 진찰기록이 병기되어, 반출 및 활용에 어려움이 있다.[2]

이러한 어려움 속에 의료인공지능의 도입은 상급종합병원 기준 17.5%, 종합병원(300병상 이상) 기준 0.9%로 더딘 편이지만, 대형병원 중심으로 IBM Watson 등의 의료 인공지능 도입 사례가 점차 늘고 있다.[3] 또한, 의료인공지능은 의료인의 의사결정을 지원하여 의료서비스의 질을 크게 향상할 것으로 기대된다.[4] 따라서, 본 논문에서는 의료인공지능 도입 및 활용에 특화된 의료정보 시스템을 제시하여 의료인공지능 도입 환경을 개선하려 한다.

이러한 시스템의 기술 실증으로, 1) 텍스트 기반의 의료데이터를 생산 및 수집하는 EMR 서비스, 2) 의료인공지능 분석 대상 데이터를 외부 인공지능 모듈로 전달하는 이벤트 허브, 3) 의료 데이터에서 핵심 키워드(증상, 의약품, 검사)를 추출하는 인공지능 모델을 연계하여 하나의 시스템으로 구현하였다.

II. 의료정보시스템

시스템 구현을 위한 구성요소를 세가지로 구분하였다. 사용자와 직접 상호작용을 하여 EMR / Authentication 등의 의료정보시스템 기능을 제공하는 웹서비스, EMR을 통해 생산된 데이터를 분석하여 결과물을 제공 하는 의료인공지능 모듈 그리고 구성요소 간의 이벤트를 중개하는 이벤트 허브이다. 이러한 구조는 그림 1을 통해 살펴볼 수 있다.

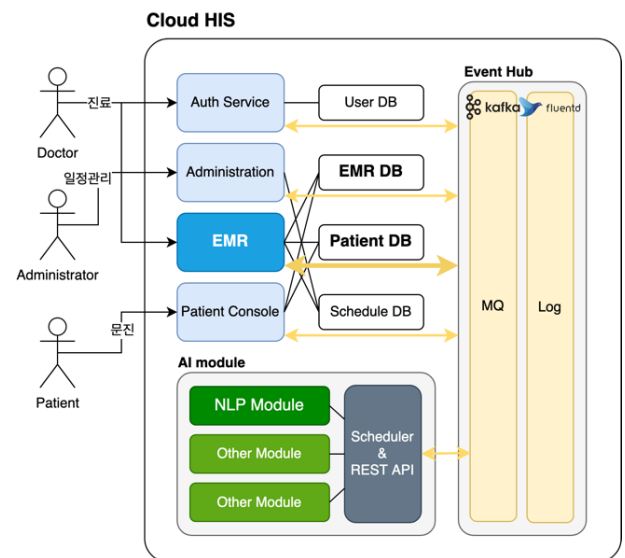


그림 1. 의료정보시스템 구성도

2-1) 웹서비스

이 논문에서는 의료정보시스템의 각 구성요소를 마이크로 서비스 단위의 웹 서비스로 설계하였다. 이들 중, 의료인공지능과 연계된 일부 서비스는 이벤트 허브를 통해 해당 인공지능과 상호작용하여 동작한다.

이 논문에서는 기술 실증을 위해 최소한의 인증 서비스와 EMR 서비스를 구현하였다.

인증 서비스는 HIS를 사용하려는 사용자가 충분한 권한을 갖고 있는지 검증하여 승인하며, 승인된 사용자는 EMR 서비스를 통해 진료예약, 진단, 처방 등의 의료행위를 할 수 있다.

2-2) 이벤트 허브

이벤트 허브는 시스템에서 발생한 의료행위이벤트(이하 이벤트)를 관리한다. 이벤트는 의료행위 과정에서 발생한 의학정보들로 구성되며, 해당 이벤트에 대한

분석을 담당하는 특정 의료인공지능 모듈은 이벤트 허브를 통해 해당 데이터를 전달받는다.

이 논문에서는 이벤트 허브 구현을 위해 아파치 카프카를 사용하였다. 카프카는 이벤트 허브에 연결된 서비스를 생산자와 소비자로 구분하여 필요한 API를 제공한다. 이 시스템에서는 EMR 서비스를 의료정보에 대한 생산자, 의료인공지능 모듈을 소비자로서 정의하여 설계하였다.

카프카를 통한 이벤트 중개방식은 전통적인 통신기법과 달리 고가용성을 보장하여 안정적인 데이터 전달을 가능케 한다.[5]

또한, 데이터를 직접 전달하지 않고 이벤트 허브에 이벤트를 등록하고 구독하는 방식으로 동작하여 어느 한쪽의 시스템이 증설/변경되는 경우에도 추가 개발소요가 적다는 장점을 갖는다.

2-3) 의료인공지능 모듈

의료 인공지능 모듈은 시스템을 도입하려는 의료기관에서 필요로 하는 분야(영상분석, 예후)에 대한 분석을 수행한다.

이 논문에서는 의료정보시스템에서 생산된 임의의 텍스트(문진, 소견)를 대상으로 개체명인식을 수행하는, 의학 개체명인식 인공지능 모델을 구현하였다. 해당 모델은 텍스트에서 관련 키워드(증상, 의약품, 검사)를 예측하고, 의료정보시스템이 사용 가능한 형태로 가공하여 해당 시스템으로 전달한다.

개체명인식 모델을 구현하기 위해 Google이 개발한 언어 사전학습 모델, BERT의 파생모델인 DistilBERT를 사용하였다.[6] 해당 모델은 기존 BERT 대비 97%의 언어이해도를 유지하면서 40%가량 경량화된 모델이다. 해당 모델에 개체명 인식을 위한 토큰 분류 레이어를 얹은 파인 튜닝을 통하여 구현하였다.

이러한 모델의 실제 구현을 위해 Hugging Face에서 제공하는 transformers 라이브러리 중 DistilBERT 패키지를 사용하였다. 해당 패키지 내의 TokenClassification 클래스를 통해 EMR 텍스트를 몇 가지 키워드 라벨로 구분해 낼 수 있었다.

모델 학습에는 n2c2 NLP Research Data Sets 2010 Relations Challenge 데이터셋을 사용하였다.[7] 해당 데이터셋은 426 개의 영문 EMR 문서내의 키워드정보를 라벨(label) 데이터로 갖고 있다. 라벨의 종류는 상병명, 검사명, 치료명 세가지이며, 학습된 모델은 추출한 키워드를 위의 세가지 라벨로 구분한다.

모델학습을 위한 하이퍼파라미터는 Learning rate 0.01, momentum 0.9, SGD optimizer, batch size 16, epochs 5로 설정하였다. 해당 모델을 RTX3080 단일 GPU 환경에서 50분간 학습시켜 평가지표를 도출한 결과, 모델의 정확도는 0.923, F1-score는 0.825에 달하는 것을 확인할 수 있었다.

구현결과

EMR 서비스의 진단차트 기능을 통해, 진료예약시 입력 받은 문진내역을 확인해 볼 수 있다. 또한, 연계된 의학 텍스트인식 인공지능을 통해 핵심 키워드를 추출하여 문진내역에 하이라이팅을 수행하고, EMR 시스템에 기록할 수 있다.

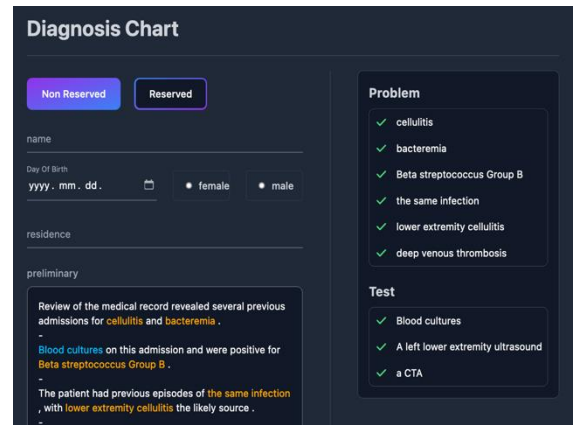


그림 2. 의학 개체명인식 인공지능 모델과 연계된 EMR

III. 결론

본 논문에서는 독립적인 Micro Service 들로 구성된 의료정보시스템을 구현하고, 해당 시스템에서 발생한 의료행위를 이벤트 단위로 추상화 하여 그에 따라 적합한 의료인공지능이 동작하는 구조를 제시하였다.

후속연구에서는 Micro Service의 이점을 활용하여 클라우드 서비스기반의 의료정보시스템으로 확장할 것이다. 또한, 의료정보시스템은 의료데이터를 생산하는 유일한 시스템이므로, 의료인공지능을 위한 학습데이터를 생산하고 처리할 수 있는 시스템을 연구 개발할 것이다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기술진흥센터의 SW 중심대학지원사업의 연구결과로 수행되었음. (2017-0-00096)

참고 문헌

- [1] 한국보건 의료정보원, “2020년 보건의료정보화 실태조사 결과보고서”, 2020
- [2] 손병은, “의료인공지능 연구/개발 및 실용화를 위한 지능형 병원정보시스템 모델”, 2022
- [3] 정원준, “국내 인공지능(AI) 의료기기 현황 및 규제 이슈”, 2018
- [4] 한국보건산업진흥원, “인공지능(AI) 기반 의료기기 현황 및 이슈(1)”, 2018
- [5] Jay Kreps, “Kafka: a Distributed Messaging System for Log Processing”, 2011
- [6] Victor Sanh, et al. “DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter” 2019
- [7] Ozlem Uzuner, et al. “2010 i2b2VA challenge on concepts, assertions, and relations in clinical text” 2011