



김인환 Kim Inhwan



010 3111 3734



moreih29@gmail.com



<https://github.com/moreih29>



<https://moreih29.github.io/>

ABOUT ME

분석하는 개발자 김인환입니다.

컴퓨터 과학을 전공했으며, 질병과 관련된 공공 데이터/웹 데이터 분석 및 시계열 예측 연구를 해왔습니다. 이 과정에서, 데이터 라벨링 및 데이터 정제 그리고 이상치 탐지 및 시계열 데이터 간의 상관관계 분석을 한 경험이 있습니다. 이러한 기술들을 바탕으로 자연어 처리 및 시계열 예측 분야의 논문들을 읽고 모델들을 설계했으며, AI와 관련된 주제로 SCIE급 해외 저널에 논문을 게재한 경험이 있습니다.

반복되는 작업들을 자동화하고 프로세스를 체계화 하는 것을 좋아합니다. 또한 연구 뿐만 아니라, AI를 라이브 서비스에 적용하는 것에 관심이 있습니다. 최근에는 사회적으로 문제가 되고 있는 코로나와 관련하여 현황과 예측 정보를 사용자들에게 제공하는 웹 서비스를 개발하여 운영하고 있습니다. 이러한 경험들을 바탕으로 현실의 다양한 문제들을 정의하고 AI를 활용하여 해결하는 일에 참여하고 싶습니다.

SKILLS

Programming Language

- Java, Python, JavaScript, HTML5

Framework / Library

- Spring, Django, Vue.js, Tensorflow, Numpy, Pandas, Scikit-learn

Server

- PostgreSQL, NginX, Jetty

DevOps

- Docker, Docker-compose, Git

Environment

- AWS, Linux, macOS, Windows

EDUCATION

2016.03 - 2019.08

상명대학교 컴퓨터과학과 졸업

2019.09 - 2021.08

상명대학원 컴퓨터과학과 졸업

PROJECTS

2021.07 - 현재

코로나 내일 <https://coronatomorrow.co.kr/>

- 국내 코로나 예측 및 현황 정보 제공
- 공공데이터 포털 API 및 질병관리청 크롤링을 통해 자동으로 일일 확진자, 사망자, 백신 접종자, 거리두기 단계 등 현황 데이터 수집
- PostgreSQL을 사용해 데이터베이스 구축
- Tensorflow를 사용해 예측 모델(Seq2Seq, Attention)을 설계하고 수집된 데이터를 통해 학습 및 평가
- Django와 nginx를 사용해 back-end 구축
- Vue.js를 사용해 front-end 개발
- 유지 보수를 위해 docker를 사용해 DB, 웹서버를 컨테이너화하고 docker-compose를 통해 컨테이너 관리
- AWS를 통해 웹 서비스 배포

2018.09 - 2021.09

전 세계 질병 정보 제공 시스템 <http://www.epidemic.co.kr/worldwide/news/>

- 웹 데이터(뉴스, 트위터) 기반 전 세계 질병 정보(일일 핵심 키워드, 뉴스 기사 및 트윗, 지역별 질병 키워드 노출 개수, 위험 질병 순위) 제공
- Bing news, Twitter API를 통해 하루 동안 생성된 웹 데이터 중 질병 키워드가 포함된 기사 및 트윗 수집
- 광고성 기사 및 트윗을 제거하기 위해 데이터 라벨링
- Tensorflow를 사용해 텍스트 분류 모델(Word2vec, BiLSTM)설계 후 라벨링 된 데이터 학습. 이후 광고성 기사 및 트윗 제거
- TF-IDF를 사용해 일일 핵심 질병 키워드 추출

PUBLICATIONS

- **Long-term Influenza Outbreak Forecast Using Time-precedence Correlation of Web Data**
Jang, Beakcheol, Inhwan Kim, and Jong Wook Kim
IEEE Transactions on Neural Networks and Learning Systems (2021), IF:8.793
 - ✓ 인플루엔자의 시계열 데이터 예측에서 예측 기간이 증가할수록 오차가 크게 증가하는 것을 문제로 정의
 - ✓ Cross-correlation을 사용해 인플루엔자 발생과 웹 데이터 발생 빈도 사이에 시간차 상관관계가 존재함을 확인
 - ✓ 웹 데이터의 발생 빈도와 과거 인플루엔자 발생 데이터를 기반으로 장기간 예측을 위해 다중 인코더 기반 Seq2seq with attention 모델을 설계 및 구현
 - ✓ RMSE, MAPE, Pearson correlation 평가 기준을 기반으로 2주 이상의 장기간 예측에서 비교 모델들에 비해 예측 오류가 감소한 결과를 보임.
- **Effective Training Data Extraction Method to Improve Influenza Outbreak Prediction from Online News Articles: Deep Learning Model Study**
Jang, Beakcheol, Inhwan Kim, and Jong Wook Kim
JMIR Medical Informatics 9.5 (2021): e23305., IF:2.96
 - ✓ 인플루엔자와 높은 상관관계를 갖는 웹 데이터 발생 빈도를 찾기 위한 키워드 탐색 비용이 큰 것을 문제로 정의
 - ✓ 단어 임베딩의 단어 벡터간 코사인 유사도를 사용해 적은 비용으로 연관 단어를 탐색하여 학습 데이터를 추출하는 방법 제안
 - ✓ 세 가지 단어 임베딩(Word2vec, GloVe, Fasttext)을 사용해 추출한 키워드를 기반으로 모델을 학습시키고 예측 성능을 비교
- **Word2vec convolutional neural networks for classification of news articles and tweets**
Jang, Beakcheol, Inhwan Kim, and Jong Wook Kim
PloS one 14.8 (2019): e0220976., IF:3.240
 - ✓ Word2vec의 학습 방법인 skip-gram과 CBOW 또는 CNN의 하이퍼 파라미터 튜닝에 따른 성능 변화 분석이 부족한 것을 문제로 정의
 - ✓ 뉴스와 트윗 데이터를 기반으로 하이퍼 파라미터 튜닝과 skip-gram, CBOW 방식에 따른 분류 성능 차이 비교 및 분석