

PORTFOLIO



Dong-geon Lee

lee.dg.125@gmail.com



이동건

Dong-geon Lee



인하대학교 정보통신공학과 4학년

lee.dg.125@gmail.com

010-3317-8739

Tech Stacks

OS

Windows / MacOS / Linux

Language

Python / C++ / C / JavaScript

DL Frameworks

PyTorch / Keras / Tensorflow

Python Libraries

KoNLPy / pandas

OpenCV / matplotlib

Systems and Tools

Git / MySQL / Raspberry Pi / Arduino / Verilog

Amazon Web Services / Google Cloud Platform

대표 프로젝트

BERT 기반의 낙상 진술문 탐지 모델 개발 ([페이지 이동](#))

GitHub

<https://github.com/oneonlee>

CV

<https://github.com/oneonlee/my/CV.pdf>

목차

#01 연구 경험

#02 AI 프로젝트

#03 주요 교과목 이수 현황

#04 비교과 활동

#01

2021 ~ 2023

연구 경험

- 학부연구생 활동 (1) : 의과대학 간호정보학 연구실
- 학부연구생 활동 (2) : 디지털 인텔리전스 연구실
- 논문 작성 및 발표 : 학술대회 발표

학부연구생 활동 (1) : 인하대학교 의과대학 간호정보학 연구실

- 활동 (2021. 08. ~ 2023. 06.)
 1. LDA 기반 토픽모델링 시계열 분석 및 시각화
 - 통계 기반의 자연어처리 기술 활용
 - 한글 자연어처리를 위한 전처리 기술 함양
 2. BERT 기반의 낙상 진술문 탐지 모델 개발 및 배포 ★
 - 딥 러닝 기반의 자연어처리 기술 활용
 3. 약물간 상호작용 네트워크 분석 및 시각화
 - 그래프 기반의 데이터 분석 및 시각화 기술 함양
 4. 개체명 인식을 통한 낙상 보고서 자동 생성 모델 개발
 - Bi-LSTM-CRF 및 BERT 활용

학부연구생 활동 (1) : 인하대학교 의과대학 간호정보학 연구실

1. LDA 기반 토픽모델링 시계열 분석 및 시각화 (2021. 08. ~ 2021. 12.)

1) 개요

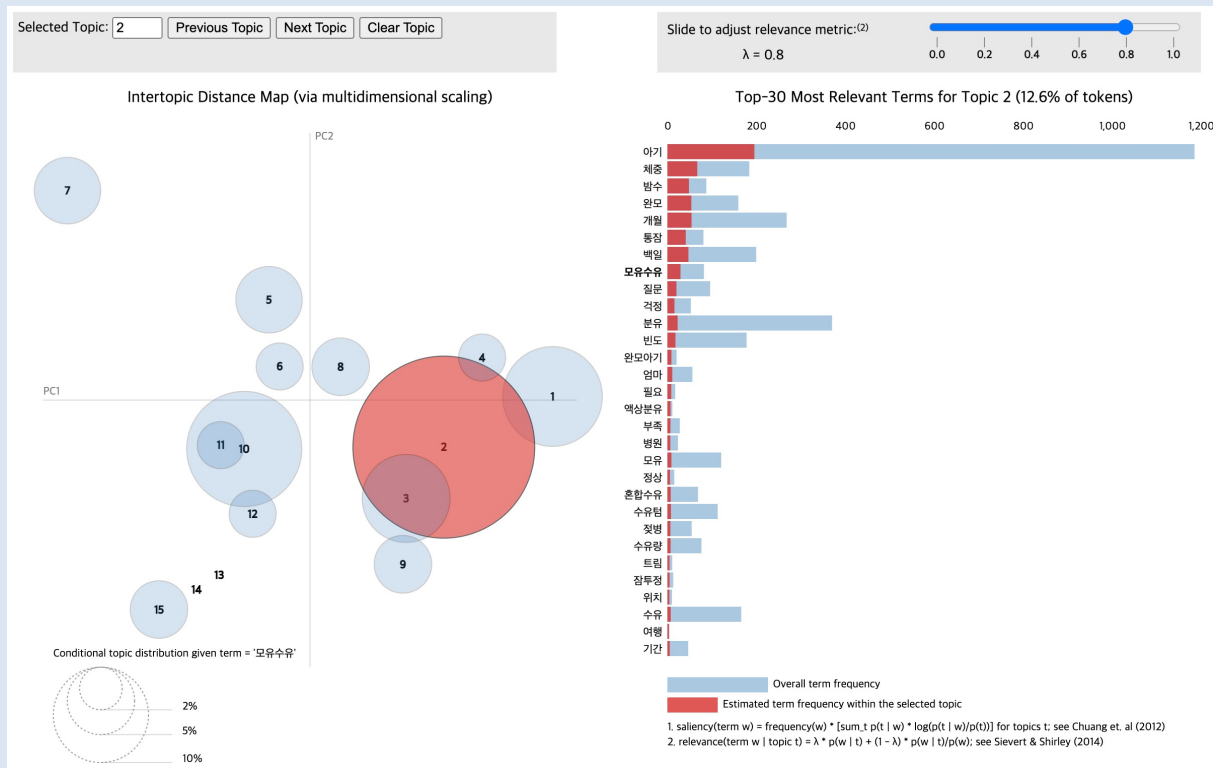
출산 후 부모의 모유수유 관련 관심사를
LDA 기반의 토픽모델링 기법을 이용하여
신생아의 개월 별로 토픽모델링 분석

2) 사용 기술 및 데이터

- 한글 및 한국어 전처리
Python, KoNLPy, hanspell, pandas
- 토픽 모델링 분석 및 시각화
genism, LDAvis, matplotlib
- 데이터
모유수유 관련 인터넷 게시글
(신생아의 개월 수로 구분한 시계열 텍스트 데이터)

학부연구생 활동 (1) : 인하대학교 의과대학 간호정보학 연구실

1. LDA 기반 토픽모델링 시계열 분석 및 시각화 (2021. 08. ~ 2021. 12.)



학부연구생 활동 (1) : 인하대학교 의과대학 간호정보학 연구실

2. BERT 기반의 낙상 진술문 탐지 모델 개발 및 배포 (2022. 01. ~ 2023. 05.)

1) 개요

병원 내 낙상 사건 보고 누락율을 낮추기 위해
딥 러닝 기반의 낙상 진술문 탐지 모델 개발

2) 사용 기술 및 데이터

- 한글 및 한국어 전처리
Python, KoNLPy, pandas
- 딥 러닝 모델 개발
TensorFlow, Keras, transformers
FastText, BERT (F1-score : 0.9890)
- 모델 배포 및 시스템 구축
PyMySQL, APScheduler
- 데이터
수도권 3개 병원 및 환자안전보고시스템의
의료기록 텍스트 데이터 (EMR)

3) 성과

- 수도권 공공종합병원에 모델 배포 및
낙상 진술문 자동 탐지 시스템 구축
- 해당 연구 성과를 국제의료정보학회 및
미국 임상정보학회에서 학술논문 발표
 - [본문 내 관련 페이지 이동](#)

4) 관련 링크 (GitHub Repository)

- 딥러닝 모델 관련 파트
<https://bit.ly/Fall-Model-GitHub>
- 시스템 배포 관련 파트
<https://bit.ly/Fall-System-GitHub>

학부연구생 활동 (1) : 인하대학교 의과대학 간호정보학 연구실

4. 개체명 인식을 통한 낙상 보고서 자동 생성 모델 개발 (2023. 01. ~ 2023. 06.)

1) 개요

병원 내 현저하게 낮은 낙상 사건 보고 누락율을 보완하기 위해
낙상 보고서를 자동으로 생성하는 딥 러닝 기반의 모델 개발

2) 사용 기술 및 데이터

- 한글 및 한국어 전처리
Python, KoNLPy, pandas
- 딥 러닝 기반 개체명 인식 (Named Entity Recognition) 모델 개발
TensorFlow, Keras, transformers
Bi-LSTM-CRF, BERT
- 데이터
수도권 3개 병원 및 환자안전보고시스템의 진료기록 텍스트 데이터 (EMR)

학부연구생 활동 (2) : 인하대학교 데이터 인텔리전스 연구실

- 활동 (2022. 11. ~ 현재)

1. 딥러닝 모델 관련 논문 리뷰 및 코드 구현 세미나

- 최신 딥 러닝 기반 자연어처리 기술의 토대가 되는 기초 논문 발표
- 해당 논문의 모델을 PyTorch로 직접 구현
- 구현한 모델로 시계열 데이터 혹은 자연어 데이터로 실험 비교
- GitHub Repository

<https://bit.ly/DL-GitHub>

2. Semi-Supervised Learning을 통한 Keyphrase 추출 모델 개발 중

- 목표 : 레이블된 데이터가 부족한 환경에서의 도메인 특화 모델 개발
- 딥 러닝 기반의 자연어처리 기술 활용

논문 작성 및 발표 : 학술대회 발표

1. 구창훈, 정윤주, 이동건, "딥 러닝 기반 영상처리를 통한 스마트 항만 주차정보 시스템 설계 및 구현", 한국정보처리학회 ACK 2021 (추계 학술발표대회) (Nov. 2021) - Oral Presentation (공동 1저자)
 - 발표 영상 (외부 링크) : <https://bit.ly/ack2021presentation>
2. **Dong-geon Lee**, EunJu Lee, Insook Cho, "Bridging the Reporting Gap of Inpatient Falls to Improve Safety Practices Using Deep-Learning-Based Language Models and Multisite Data", *AMIA 2023 Clinical Informatics Conference*, Chicago, United States (May, 2023) - Oral Presentation (Peer Reviewed)
 - 발표 영상 (외부 링크) : <https://bit.ly/cic23presentation>
3. Insook Cho, EunJu Lee, Dong-geon Lee, "Effects of Language Differences on Inpatient Fall Detection Using Deep Learning", The 19th World Congress on Medical and Health Informatics (MedInfo 2023), Sydney, Australia (July, 2023)
 - Poster (외부 링크) : <https://bit.ly/medinfo2023poster>
 - [본문 내 관련 페이지 이동](#) ("BERT 기반의 낙상 진술문 탐지 모델 개발")

#02

2021 ~ 2023

AI 프로젝트

- (1) BERT를 이용한 소설 작가 분류 모델 개발
- (2) 게임 닉네임 기반 티어 예측 모델 개발
- (3) TSP에서의 유전 알고리즘 성능 평가
- (4) 컴퓨터 비전 기반 주차정보시스템 개발

(1) BERT를 이용한 소설 작가 분류 모델 개발 (팀 프로젝트)

1) 개요

- 영문 소설 데이터의 문맥을 파악하여 각 소설 작가들이 보여주는 특징적인 문체를 판별
- 프로젝트 기간 :
(2023.05. ~ 2023. 06.)

2) 담당 파트 및 사용 기술

- EDA (탐색적 데이터 분석)
pandas, matplotlib
- 모델 훈련 및 하이퍼 파라미터 튜닝
PyTorch, scikit-learn

3) 관련 링크

- 발표 소개 영상
<https://bit.ly/BERT-presentation>
- 프로젝트 보고서
<https://bit.ly/BERT-report>
- GitHub Repository
<https://bit.ly/BERT-GitHub>

(2) 게임 닉네임 기반 티어 예측 모델 개발 (팀 프로젝트)

1) 개요

- '리그오브레전드' 게임 속 소환사 이름 (닉네임)으로 게임 속 티어 (등급)을 예측
- 프로젝트 기간 : (2022.07. ~ 2022. 08.)

2) 담당 파트 및 사용 기술

- EDA (탐색적 데이터 분석)
pandas, matplotlib, seaborn
- 한글 및 한국어 데이터 전처리
KoNLPy, hgtk
- RNN 계열 모델 훈련, 성능 비교
TensorFlow, scikit-learn

3) 관련 링크

- GitHub Repository
<https://bit.ly/Lol-GitHub>
- 모델 사용 코드 (Google Colab)
<https://bit.ly/Lol-Colab>

```
1 tier_predict('플래티넘 문지기')  
( '플래티넘', 0.215069)  
  
1 tier_predict('아몬드가 죽으면')  
( '다이아몬드', 0.28879693)  
  
1 tier_predict('아이언맨')  
( '아이언', 0.18356499)
```

(3) TSP에서의 유전 알고리즘 성능 평가 (개인 프로젝트)

1) 개요

- TSP (외판원 문제)를 자연계의 진화 과정을 모방한 '유전 알고리즘' 기법으로 구현 후 성능을 확인
- 프로젝트 기간 :
(2022.06. ~ 2022. 07.)

2) 사용 기술

- 알고리즘 개발
Python
- 그래프 시각화
matplotlib

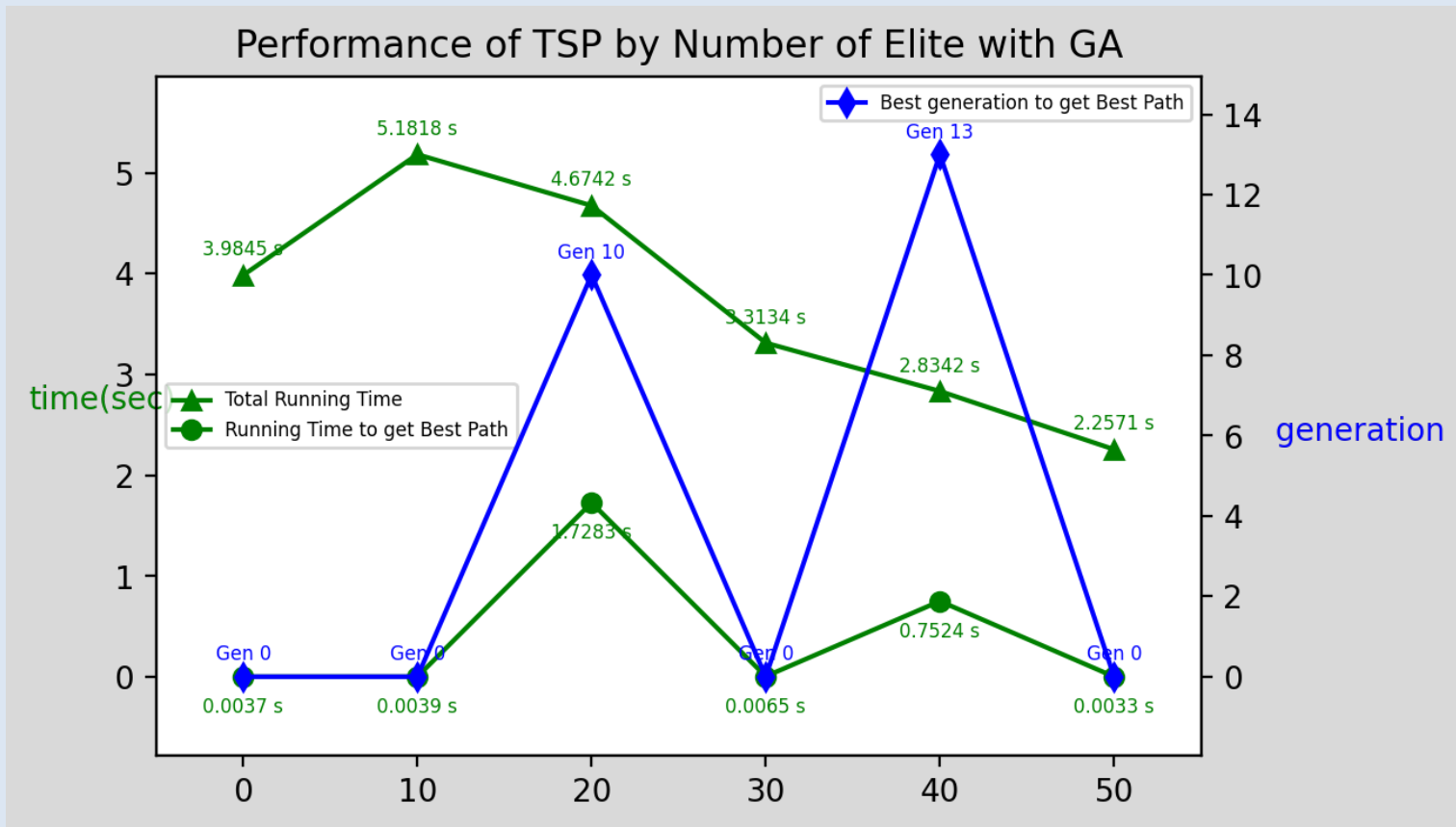
3) 관련 링크

- GitHub Repository
<https://bit.ly/GA-GitHub>

4) 특징

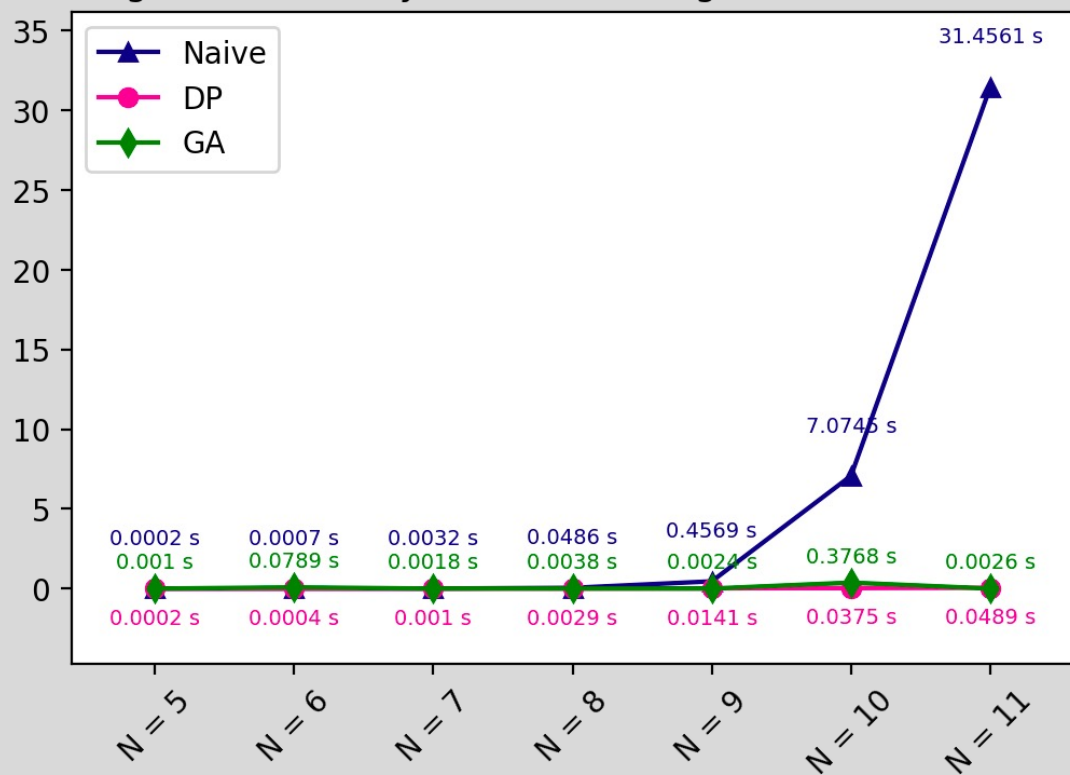
- 세대(generation)를 거쳐 적응도(fitness)가 높은 것들이 선택되는 유전 알고리즘의 원리가 epoch과 loss 같은 기계학습의 기본 구조와 유사하다는 것을 확인
- 디지털 염색체들의 교배율과 돌연변이 발생 비율 값 등에 따른 성능 차이를 확인하며, 인공지능의 학습 과정 속 하이퍼 파라미터 튜닝의 원리와 필요성을 확인

(3) TSP에서의 유전 알고리즘 성능 평가 (개인 프로젝트)



(3) TSP에서의 유전 알고리즘 성능 평가 (개인 프로젝트)

Running time of TSP by N Value and Algorithms (Naive, DP, GA)



(4) 컴퓨터 비전 기반 주차정보시스템 개발 (팀 프로젝트)

1) 개요

- 영상처리 및 컴퓨터 비전 기술로 주차장의 실시간 주차 현황을 분석하여 제공하는 웹 서비스
- 프로젝트 기간 :
(2021.03. ~ 2021. 11.)

2) 담당 파트 및 사용 기술

- 주차장의 주차선 검출 알고리즘 개발
Python, OpenCV
- YOLOv4를 이용한 차량 탐지
TensorFlow, OpenCV
- 프론트엔드 및 백엔드 개발
JavaScript, Django

3) 성과

- 학술논문 게재 및 학술대회 발표
(2021 한국정보처리학회 추계학술발표대회)
- 특허 출원
특허명 : “주차 정보를 제공하는 시스템 및 그 제어 방법”
출원번호 : 10-2021-0178090

4) 관련 링크

- GitHub Repository
<https://bit.ly/cargo-GitHub>
- Demo Video
<https://bit.ly/cargo-demo>
- 학술논문
<https://bit.ly/cargo-paper>

#03

2018 ~ 2023

주요 교과목 이수 현황

- 수학 및 컴퓨터 과학 관련
- 기초공학 및 정보통신공학 관련

주요 교과목 이수 현황 : 수학 및 컴퓨터 과학

수학 관련

* 계산수학
일반수학 1, 2
공업수학 1, 2
선형대수
신호 및 시스템
확률변수론
** 수치해석

*** 컴퓨터 과학 관련

객체지향프로그래밍 1, 2
자료구조론
시스템 프로그래밍
오퍼레이팅 시스템
정보보호론
컴퓨터 네트워크
알고리즘 설계
데이터베이스 설계
인공지능 응용
디지털 영상처리 설계

* : Python, SageMath 사용

** : 2023학년도 2학기 수강 중

*** : 해당 교과목들의 개념들과 관련 코드들을 깃헙에 마크다운으로 정리하였음 (<https://bit.ly/CS-lectures>)

주요 교과목 이수 현황 : 기초공학 및 정보통신공학

기초공학 관련

물리학 1, 2
물리학실험 1, 2
창의적 정보통신공학설계
공학 윤리와 토론
세상을 바꾸는 스타트업 이야기
컴퓨팅 사고와 데이터 분석

정보통신공학 관련

정보통신입문
정보통신기초설계/실습 1, 2
회로이론
전자기학 1
전자회로 1
디지털 논리회로

#04

2021 ~ 2023

비교과 활동

- 자격증 및 외부 강의 수료
- 튜터링 및 강의 활동
- 공인 어학 성적

자격증

및

외부 강의 수료

2022

1. “딥러닝의 이해” (한컴 아카데미)
2. “Amazon Web Services (AWS) 머신러닝 교육”
(인하대학교 미래자동차 혁신공유대학 사업단)
3. “딥러닝의 기초”
(NVIDIA Deep Learning Institute)
4. “트랜스포머 기반 자연어 처리 애플리케이션 구축”
(NVIDIA Deep Learning Institute)

2023

5. “딥러닝 교육 (고급)”
(인하대학교 미래자동차 혁신공유대학 사업단)
6. “LG Aimers” (LG AI 연구원)

튜터링

및

강의 경험

2021

- “프로그래밍 기초” 과목 튜터링 튜터 (2021. 03. ~ 2021. 06.)
- “(주) 잼코딩” 소프트웨어 교육 강사 (2021. 03. ~ 2023. 05.)
 - 데이터 분석 및 시각화 수업
 - GitHub Repository: <https://bit.ly/analysis-GitHub>
 - Python / C 프로그래밍 및 알고리즘 수업

2023

- “인공지능 프로그래밍 기초” 과목 TA (2023학년도 1학기)
- “알고리즘 설계” 과목 TA (2023학년도 1학기)
- “자료구조론” 과목 TA (2023학년도 2학기)
- “데이터베이스 설계” 과목 TA (2023학년도 2학기)

공인

어학 성적

English

- TOEIC 805점 (2023. 02.)
- OPIc IH(Intermediate High) 등급 (2023. 08.)

Dong-geon Lee

lee.dg.125@gmail.com