

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 / Ubuntu
Language	Python / C++ / C / JavaScript
DL Frameworks	PyTorch / Keras / Tensorflow
Python Libraries	pandas / KoNLPy matplotlib / OpenCV
Systems and Tools	Git / MySQL Amazon Web Services / Google Cloud Platform Raspberry Pi / Arduino / Verilog

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 AI 프로젝트

#02 연구 경험

#03 대학원 입학 후 연구 계획

#04 주요 교과목 이수 현황

#05 비교과 활동

#06 Appendix

#01

2021 ~ 2022

AI 프로젝트

- (1) 컴퓨터 비전 기반 주차정보시스템 개발
- (2) TSP에서의 유전 알고리즘 성능 평가
- (3) 게임 닉네임 기반 티어 예측 모델 개발

(1) 컴퓨터 비전 기반 주차정보시스템 개발 (팀 프로젝트)

1) 소개

- 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](#)
- [Demo Video](#)
- [논문](#)

(2) TSP에서의 유전 알고리즘 성능 평가 (개인 프로젝트)

1) 소개

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

2) 사용 기술

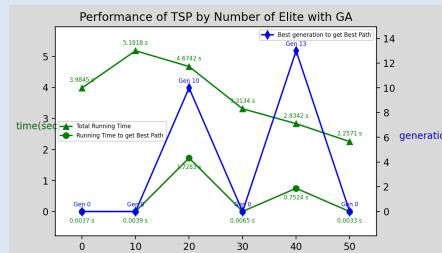
- 알고리즘 개발
Python
- 그래프 시각화
matplotlib

3) 관련 링크

- [GitHub Repository](#)

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



Full Image : [Appendix A](#)

(3) 게임 닉네임 기반 티어 예측 모델 개발 (팀 프로젝트)

1) 소개

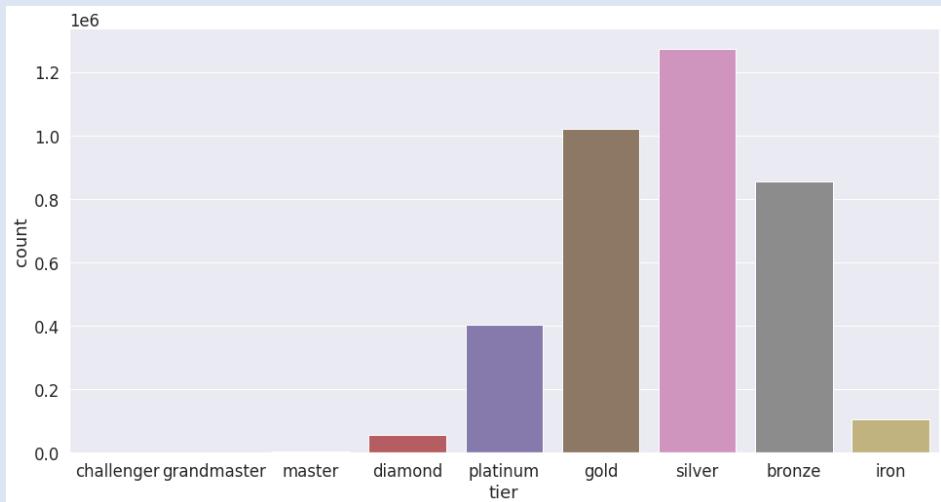
- '리그오브레전드' 게임 속 소환사 이름(닉네임)으로 게임 속 티어(등급)을 예측하는 토이 프로젝트
- **프로젝트 기간 :**
(2022.07. ~ 2022. 8.)

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](#)
- [모델 훈련 코드 \(Google Colab\)](#)
- [모델 사용 코드 \(Google Colab\)](#)



#02

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. ~ 현재)

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 ★

- 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 중

- 딥 러닝 기반의 자연어처리 기술 활용

학부연구생 활동 (1) : 인하대학교 의과대학 간호정보학 연구실

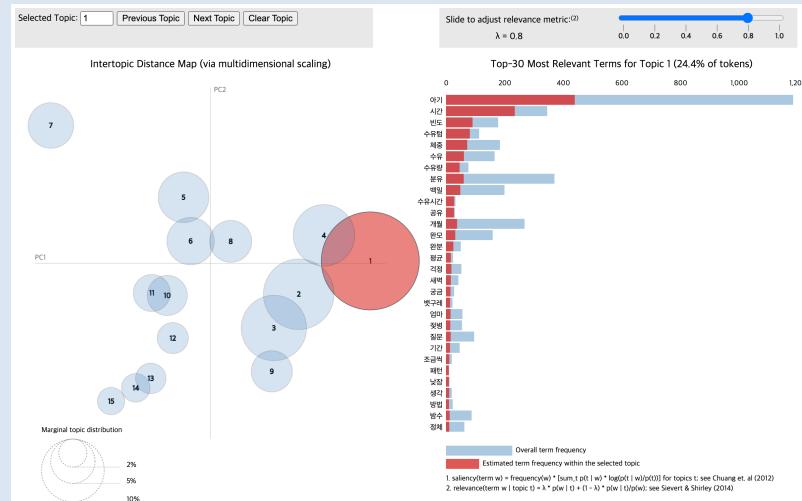
1. LDA 기반 토픽모델링 시계열 분석 및 시각화 (2021. 08. ~ 2021. 12.)

1) 소개

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
(아기의 개월 수로 구분한 시계열 텍스트 데이터)



Full Image : [Appendix B](#)

학부연구생 활동 (1) : 인하대학교 의과대학 간호정보학 연구실

2. BERT 기반의 낙상 진술문 탐지 모델 개발 (2022. 01. ~ 2022. 12.)

1) 소개

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
진료기록 텍스트 데이터 (EMR)

3) 성과

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

학부연구생 활동 (1) : 인하대학교 의과대학 간호정보학 연구실

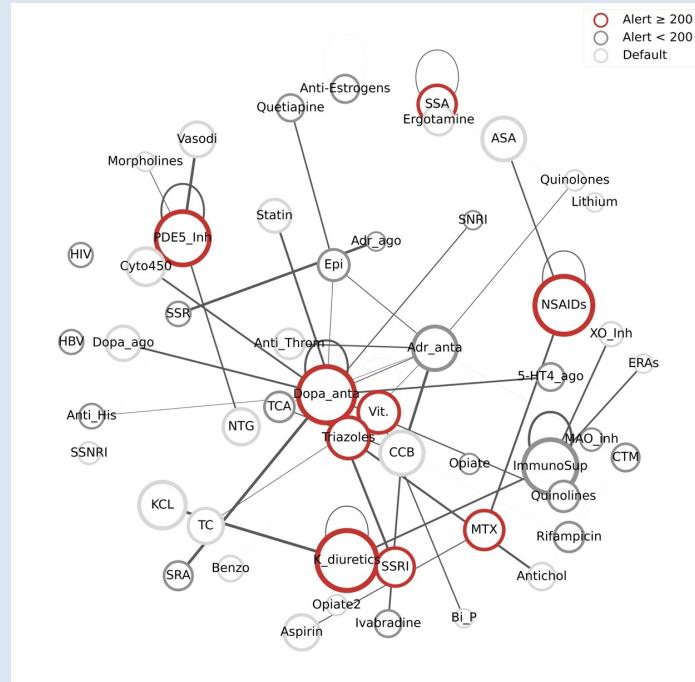
3. 약물간 상호작용 네트워크 분석 및 시각화 (2022. 08. ~ 2022. 10.)

1) 소개

병원의 의사들이 환자들에게 약을 처방할 때,
어떤 약물들 간에 상호작용 경고가 발생하는지
경고 발생률을 중심으로 네트워크 분석

2) 사용 기술 및 데이터

- 데이터 전처리
Python, pandas
- 네트워크 분석 및 시각화
NetworkX, matplotlib
- 데이터
약물 처방 및 약물간 상호작용 경고 데이터



학부연구생 활동 (1) : 인하대학교 의과대학 간호정보학 연구실

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

1) 소개

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. 자연어처리 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

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

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

논문 작성 및 발표 : 학술대회 발표

1. Changhun Koo, Yoonjoo Jung, **Dong-geon Lee**, "Through deep learning-based video processing, Design and implementation of Smart Port Parking Information System", *Annual Conference of KIPS 2021*, Yeosu, Korea (Nov. 2021) - Oral Presentation
 - 발표 영상(외부 링크) : <https://youtu.be/9NI3AhgallA>

논문 작성 및 발표 : 학술대회 발표 예정 (2023)

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) - Accepted as Oral Presentation
3. Insook Cho, EunJu Lee, Dong-geon Lee, "Effects of Language Differences on Inpatient Fall Detection Using Deep Learning", *Proceedings of the 19th World Congress on Medical and Health Informatics (MedInfo 2023)*, Sydney, Australia (July, 2023) - Accepted as Poster
 - [본문 내 관련 페이지 이동 \(“BERT 기반의 낙상 진술문 탐지 모델 개발”\)](#)

#03

2024 ~

대학원 입학 후 연구 계획

- 연구 주제
- 연구 배경
- 연구 방법

연구 계획 (1) : 제한된 데이터 환경에서의 문서 요약 및 이상치 탐지

1. 연구 주제

- 제한된 데이터 환경에서의 문서 요약 및 이상치 탐지

2. 연구 배경

- 문서 요약 및 이상치 탐지는 자연어 처리 및 기계 학습 분야에서 중요한 작업이다.
- 해당 작업에서 필요한 대부분의 실제 산업 데이터에는 레이블이 지정되지 않았거나 부분적으로 레이블이 지정되었기 때문에 효과적인 모델을 개발하기가 어려울 수 있다.
- 경우에 따라서는 비정상 데이터의 수가 정상 데이터의 수보다 훨씬 적은 불균형으로 비지도 학습 및 준지도 학습의 새로운 방법론 연구가 필요하다.

연구 계획 (1) : 제한된 데이터 환경에서의 문서 요약 및 이상치 탐지

3. 연구 방법

1) 데이터 수집 및 전처리

- 데이터의 특성에 맞는 전처리 기술을 사용하여 데이터 정제, 토큰화, 불용어 제거 등을 수행

2) 비지도 및 준지도 학습 기반 모델 개발

- 문서 요약 기법
 - 추출적, 추상적 요약 기법을 사용하여 중요한 정보를 보존하면서도 문서를 요약하는 방법 연구
- 이상치 탐지 기법 개발
 - 이상치 탐지에 대한 정의와 목적에 맞는 평가 척도를 고려한 이상치 탐지 알고리즘 개발

3) 성능 평가 및 결과 분석

- 다양한 평가 지표를 사용하여 모델의 성능을 분석하고, 개선할 수 있는 방법을 연구

연구 계획 (2) : 설명 가능한 멀티모달 학습 모델 개발

1. 연구 주제

- 설명 가능한 멀티모달 학습 모델 개발 (자연어+시계열 / 자연어+이미지)

2. 연구 배경

- 멀티모달 데이터 (예: 자연어, 이미지 및 시계열 데이터)의 가용성이 증가함에 따라 멀티모달 데이터를 기반으로 학습하고 의사 결정을 내릴 수 있는 모델 개발에 수요가 증가하고 있다.
- 그러나 이러한 모델은 의료진단이나 금융위험평가와 같은 중요한 작업에서 사용될 때, 모델이 어떻게 결정을 하는지 설명하기 어렵다는 것이 여전히 큰 과제로 남아있다.
- 따라서 의사 결정 과정에 대한 명확하고 간결한 설명을 제공할 수 있는 설명 가능한 멀티모달 학습 모델을 개발할 필요가 있다.

연구 계획 (2) : 설명 가능한 멀티모달 학습 모델 개발

3. 연구 방법

1) 데이터 수집 및 전처리

- 멀티모달 데이터를 수집하고 전처리하여, 모델에 적용할 수 있는 형식으로 변환
- 이때, 자연어, 이미지, 시계열 데이터의 특성을 고려하여 적절한 전처리 기법을 적용

2) 멀티모달 학습 모델 개발

- 자연어와 시계열 또는 자연어와 이미지를 결합한 멀티모달 학습 모델 개발
- 딥러닝 알고리즘을 활용하여 구현하며, 설명 가능한 모델 (eXplainable AI, XAI)을 적용

3) 성능 평가, 개선 및 분석

- 설명 가능성을 평가하기 위해 SHAP, LIME 등의 기법을 적용
- 모델의 성능이 낮을 경우, 하이퍼 파라미터 튜닝 및 모델 구조 개선을 통해 성능 개선

#04

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
- 일반수학 2
- 공업수학 1
- 공업수학 2
- 선형대수
- 신호 및 시스템
- 확률변수론
- ** 인공지능을 위한 수학 1

*** 컴퓨터 과학 관련

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

* : Python, SageMath 사용

** : 2023학년도 1학기 현재 수강 중인 교과목

*** : 해당 교과목들의 개념들과 관련 코드들을 깃헙에 마크다운으로 정리하였음 ([외부 링크](#))

주요 교과목 이수 현황 : 기초공학 및 정보통신공학

기초공학 관련

- 물리학 1
- 물리학 2
- 창의적 정보통신공학설계
- 공학 윤리와 토론
- 세상을 바꾸는 스타트업 이야기

정보통신공학 관련

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

#05

2021 ~ 2023

비교과 활동

- 자격증 및 외부 강의 수료
- 튜터링 및 강의 활동

자격증

및

외부 강의 수료

Appendix C

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)

튜터링

및

강의 경험

2021

- “프로그래밍 기초” 과목 튜터링 튜터 (2021. 03. ~ 2021. 06.)
- “(주) 잼코딩” 소프트웨어 교육 강사 (2021. 03. ~ 현재)
 - 데이터 분석 및 시각화 수업
 - 코드 및 결과물 Repository ([외부 링크](#))
 - Python / C 프로그래밍 및 알고리즘 수업

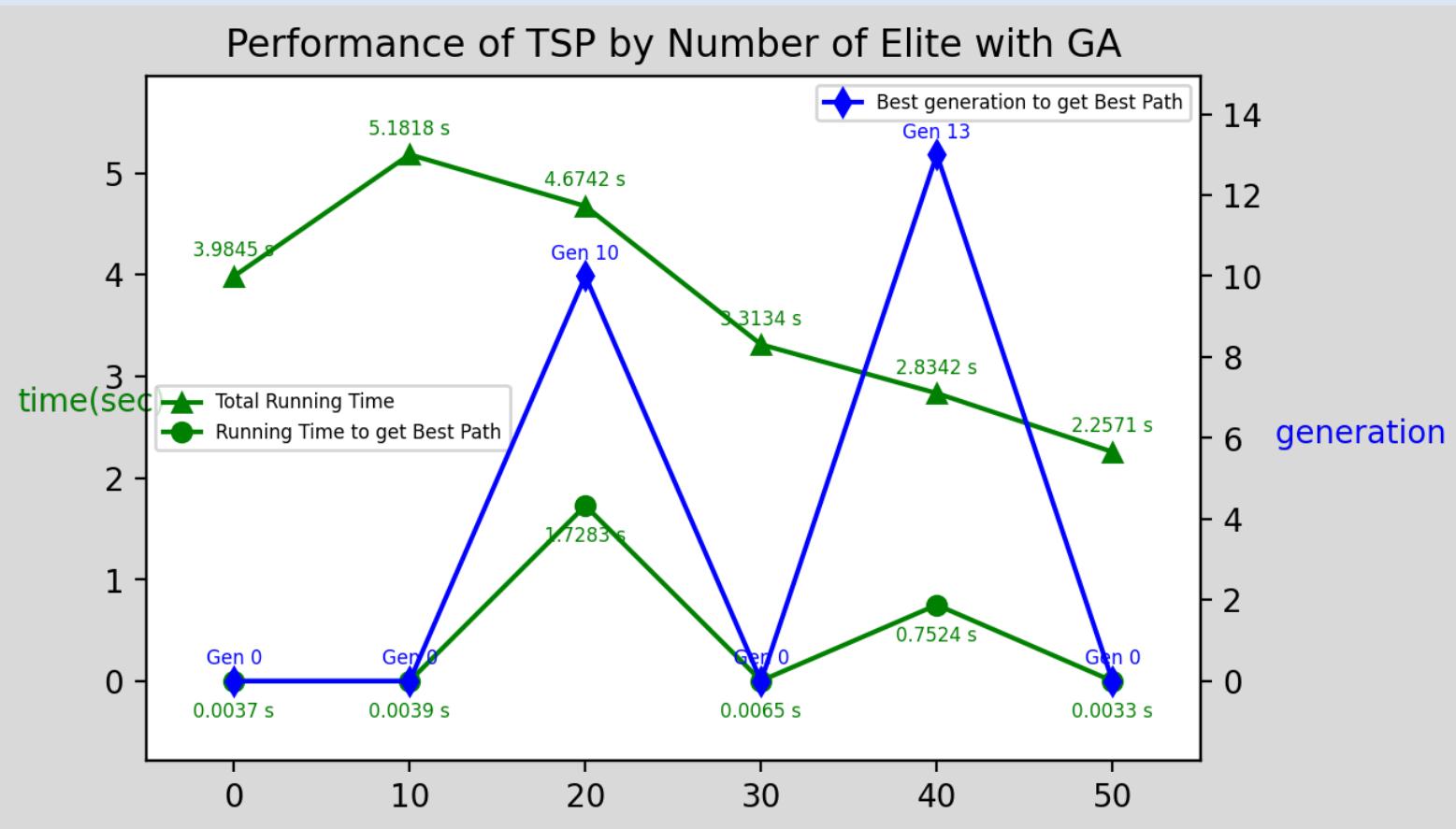
2023

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

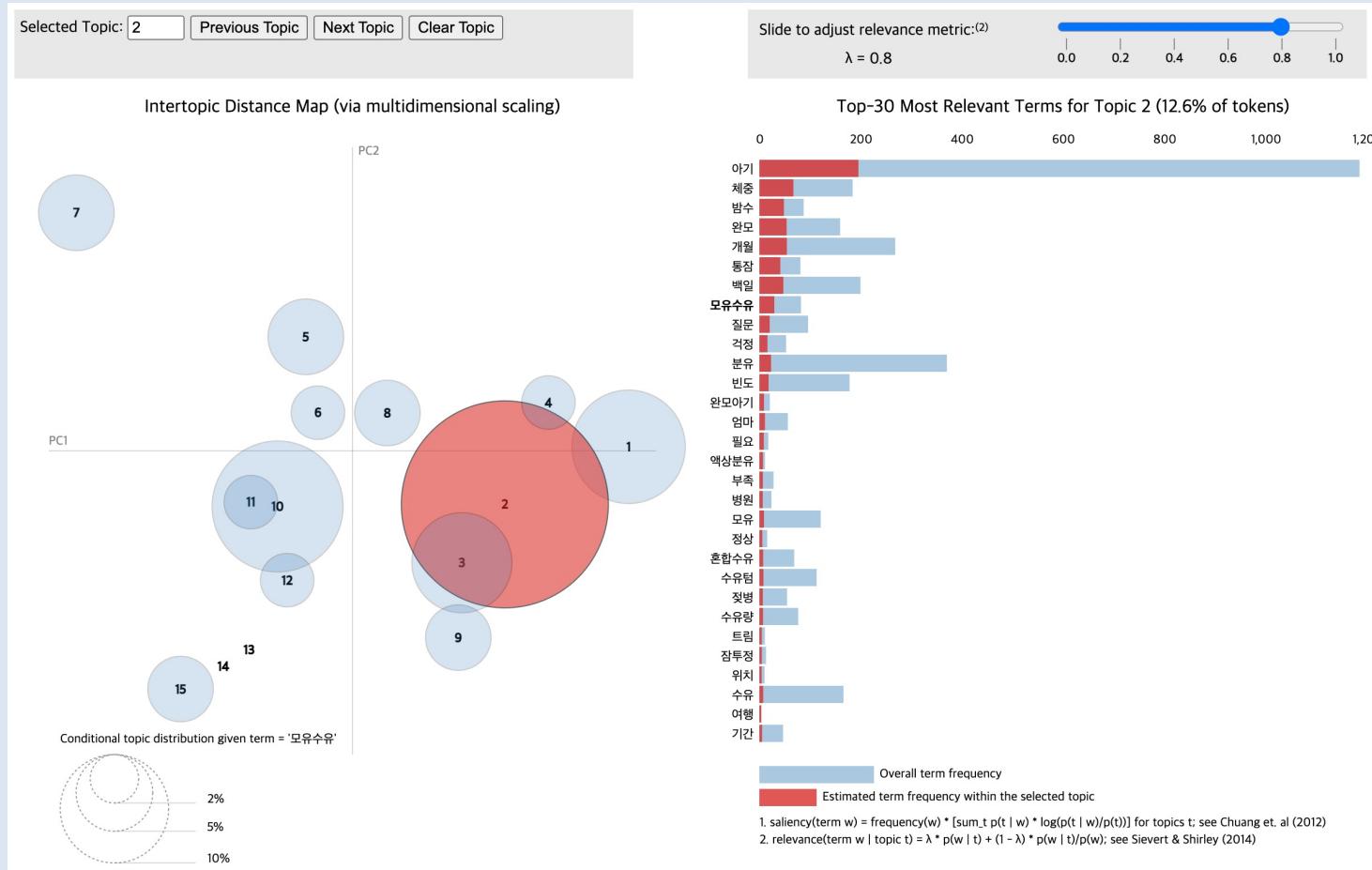
Dong-geon Lee

lee.dg.125@gmail.com

Appendix A : 유전 알고리즘을 통한 TSP 경로탐색



Appendix B : 토픽모델링 시각화



Appendix C : 자격증 및 외부 강의 수료 (1. ~ 2.)

1.

 한컴아카데미

22-HAAC-0223020호

수료증

성명 : 이동건
소속 : 인하대학교 정보통신공학과
교육과정명 : 딥러닝 이해
교육일정 : 2022. 2. 22. ~ 2022. 2. 23. (2일 16시간)

위 사람은 한컴아카데미에서 주관하는
위의 과정을 수료하였으므로 이 증서를 수여합니다.

2022. 02. 23.

(주)한컴아카데미



2.

미래자동차 혁신공유대학사업단 2021-038호



수료증

교육과정명: 2021 신기술 Skill-Up 비교과 단기 프로그램
Amazon Web Services (AWS) 머신러닝 교육
교육 일정: 2022.02.24. ~ 2022.02.25. (2일 8시간)
학과: 정보통신공학과
성명: 이동건

위 학생은 인하대학교 미래자동차 혁신공유대학
사업단에서 주관하는 위의 과정을 수료하였으므로
이 증서를 수여합니다.

2022년 02월 25일

인하대학교
미래자동차 혁신공유대학 사업단장



본문 이동

Appendix C : 자격증 및 외부 강의 수료 (3. ~ 4.)

3.



CERTIFICATE OF COMPETENCY

This NVIDIA DLI Certificate has been awarded to

Lee Donggeon

for the successful completion of
딥러닝의 기초

A handwritten signature of Will Ramey.

Will Ramey
Senior Director, Developer Programs

August 22,2022

4.



CERTIFICATE OF COMPETENCY

This NVIDIA DLI Certificate has been awarded to

Lee Donggeon

for the successful completion of
트랜스포머 기반 자연어 처리 애플리케이션 구축

A handwritten signature of Will Ramey.

Will Ramey
Senior Director, Developer Programs

August 23,2022

[본문 이동](#)

Appendix C : 자격증 및 외부 강의 수료 (5.)

5.



교육부



NRF
한국연구재단

디지털 혁신공유대학

미래자동차-2022-인하대-256호

수료증

프로그램명: 2022 동계방학 신기술 Skill-up 비교과 교육프로그램

- 딥러닝 교육(고급)

교육기간: 2023.1.27 (금) 8시간

학과: 정보통신공학과

성명: 이동건

위 학생은 인하대학교 미래자동차 혁신공유대학 사업단에서 주관하는 '2022 동계방학 신기술 Skill-up 비교과 교육프로그램 - 딥러닝 교육(고급)' 과정을 성실히 이수하여 이 수료증을 수여합니다.



2023년 1월 27일

인하대학교
미래자동차 혁신공유대학 사업단장



본문 이동