
Portfolio

김태형

고려대학교 산업경영공학부
인공지능 및 데이터 분석 연구실

Contents

1. Profile
2. Publications
3. Patents
4. Projects
5. Awards and Honors
6. Research in progress

1. Profile



김태형

Taehyeong Kim

M.S. / Ph.D. Student

Artificial Intelligence & Data Analysis
(AIDA) Lab.

School of Industrial and Management Engineering,
Korea University.

GitHub :

<https://github.com/GNOEYHEAT>

LinkedIn :

<https://www.linkedin.com/in/gnoeyheat/>

▪ Education

- Korea University, School of Industrial and Management Engineering, Ph.D. Student, Advisor: Sung Won Han (2019.03 ~ Present)
- Korea University, Department of Information Statistics, B.S. in Information Statistics (2012.03 ~ 2019.08)

▪ Research Interests

- Vision-Language Model
- Large Language Model
- Multimodal Learning
- Time Series Analysis
- Anomaly Detection

▪ Publications (SCI/SCIE)

- **Taehyeong Kim***, Sung Won Han*, Minji Kang, Se Ha Lee, Jong Ho Kim, Hyung Joon Joo, Jang Wook Sohn, "Similarity-Based Unsupervised Spelling Correction Using BioWordVec: Development and Usability Study of Bacterial Culture and Antimicrobial Susceptibility Reports", JMIR MEDICAL INFORMATICS, (2021)
- Jun-Guel Lee, **Taehyeong Kim**, Ki Woo Sung, Sung Won Han, "Automobile parts reliability prediction based on claim data: The comparison of predictive effects with deep learning", ENGINEERING FAILURE ANALYSIS, (2021)
- Hyun Joon Park*, **Taehyeong Kim***, Young Seok Kim, Jinhong Min, Ki Woo Sung, Sung Won Han, "CRFormer: Complementary Reliability Perspective Transformer for Automotive Components Reliability Prediction Based on Claim Data", IEEE ACCESS, (2022)
- Jae-Dong Kim*, **Tae-Hyeong Kim***, Sung Won Han, "Demand Forecasting of Spare Parts using Artificial Intelligence: A Case Study of K-X Tanks", MATHEMATICS, (2023)

1. Profile



김태형
Taehyeong Kim

M.S. / Ph.D. Student

Artificial Intelligence & Data Analysis
(AIDA) Lab.

School of Industrial and Management
Engineering, Korea University.

GitHub :
<https://github.com/GNOEYHEAT>

LinkedIn :
<https://www.linkedin.com/in/gnoeyheat/>

Conferences

- 김태형, 강민지, 주형준, 한성원, "BioWordVec을 활용한 유사도 기반의 철자 오류 교정", 한국데이터마이닝학회, Seoul, Korea (Nov. 2020) – Oral
- 이준걸, 박성훈, 김태형, 김창기, 한성원, "LSTM기반 딥러닝을 이용한 일사량 이미지 생성", 한국데이터마이닝학회, Seoul, Korea (Nov. 2020) – Oral
- 김태형, 임재성, 한성원, "디지털 워터마크를 이용한 위·변조 사진 Out-of-Distribution 탐지", 대한산업공학회, Seogwipo-si, Korea (Jun. 2021) – Oral
- 김태형, 박원준, 성기우, 한성원, "클레임 데이터를 활용한 준지도 학습 기반 자동차 부품 고장 예측", 한국신뢰성학회, Jeju-si, Korea (Jun. 2021) – Oral
- Jae-Dong Kim, **Taehyeong Kim**, Young-Seok Kim, Sung Won Han, "Demand Forecasting of Spare Parts via Stacked Generalization: Case Study of K-X Tanks", Asia Simulation Conference 2021, Virtual Conference, Korea (Dec. 2021) – Oral
- 박현준, 김태형, 김영석, 민진홍, 성기우, 한성원, "자동차 신뢰성 예측을 위한 상호보완적 신뢰성 관점의 트랜스포머 기반 예측 모델 개발", 대한산업공학회, Jeju-si, Korea (Jun. 2022) – Poster
- 김태형, 허준호, 박세홍, 한성원, "기침 소리 데이터를 이용한 멀티모달 기반의 COVID-19 검출", 한국데이터마이닝학회, Busan, Korea (Aug. 2022) – Oral
- 김태형, 강지수, 함유진, 한성원, "위성영상 기반 일사량을 활용한 전력거래량 스택킹 앙상블 예측 모형 개발", 한국태양에너지학회, Jeju-si, Korea (Oct. 2022) – Oral

Patents

- 손장욱, 한성원, 주형준, 김태형, 김종호, 강민지, 장준호, 김종훈, "온톨로지 데이터베이스 기반의 룰 기반 알고리즘을 이용한 세균배양검사결과보고서 정제 시스템", KR-Registration No. 10-2512095
- 김태형, 임재성, 김성원, "노이즈 부가를 이용한 위변조 이미지 판정 방법 및 그 장치", KR-Registration No. 10-2592718

1. Profile



김태형
Taehyeong Kim

M.S. / Ph.D. Student

Artificial Intelligence & Data Analysis
(AIDA) Lab.

School of Industrial and Management
Engineering, Korea University.

GitHub :
<https://github.com/GNOEYHEAT>

LinkedIn :
<https://www.linkedin.com/in/gnoeyheat/>

■ Projects

- 천리안 위성영상 기반 구름이동벡터 생산 알고리즘 구축 용역, 한국에너지기술연구원, Korea (Apr. 2019 ~ Jul. 2019)
- 감염성질환 데이터의 CDM 확장모델 개발 및 활용을 위한 Tool 개발, 한국보건산업진흥원, Korea (Apr. 2019 ~ Dec. 2021)
- 클레임 데이터 기반 고장발생예측 기술개발, (주)현대엔지비, Korea (Aug. 2019 ~ Feb. 2020)
- 클레임데이터를 활용한 내구열화 부품 발체 및 미래 고장 패턴 예측 기술 개발, (주)현대엔지비, Korea (May. 2020 ~ Aug. 2021)
- 양자 기계학습 시뮬레이터, 한국연구재단, Korea (Jul. 2020 ~ Feb. 2022)
- AI를 활용한 상품정보 자동 생성 알고리즘 연구, (주)프림로즈힐파트너스, Korea (Oct. 2022 ~ Dec. 2022)

■ Teaching Experiences

- 데이터분석 역량체계 구축, (주)LG이노텍, Korea (Sep. 2019 ~ Dec. 2019)
- Big Data 분석가 양성과정, 에스케이하이닉스 주식회사, Korea (Feb. 2020 ~ Jun. 2020)
- DX Expert 양성 과정, (주)엘지화학, Korea (Jun. 2020 ~ Nov. 2020)
- AI 전문가 양성 과정, 현대제철(주), Korea (Jun. 2021 ~ Aug. 2021)
- Big Data 분석가 양성과정, 에스케이하이닉스 주식회사, Korea (Jul. 2021 ~ Jul. 2022)

■ Certificates

- 빅데이터 분석기사 (Jul. 2021)
- 정보처리기사 (Nov. 2018)
- 데이터분석 준전문가(ADsP) (Dec. 2018)
- 사회조사분석사 2급 (Aug. 2017)
- AICE Associate (Nov. 2022)
- AICE Basic (Nov. 2022)

1. Profile



김태형
Taehyeong Kim

M.S. / Ph.D. Student

Artificial Intelligence & Data Analysis
(AIDA) Lab.

School of Industrial and Management
Engineering, Korea University.

GitHub :
<https://github.com/GNOEYHEAT>

LinkedIn :
<https://www.linkedin.com/in/gnoeyheat/>

▪ Awards and Honors

- 코드 유사성 판단 시즌2 AI 경진대회, 1st place, 데이콘, Korea (Apr. 2024)
- 반도체 소자 이상 탐지 AI 경진대회, 1st place, 데이콘, Korea (Mar. 2024)
- HD현대 AI Challenge, 우수상, 주최: HD한국조선해양 AI Center, 운영: 데이콘, Korea (Nov. 2023)
- AI vs Human 텍스트 판별 해커톤, 1st place, 데이콘, Korea (Nov. 2023)
- 2023 Samsung AI Challenge : Camera-Invariant Domain Adaptation, 장려상, 주최: 삼성전자 SAIT, 운영: 데이콘, Korea (Oct. 2023)
- 이미지 기반 질의 응답 AI 경진대회, 3st place, 데이콘, Korea (Aug. 2023)
- 법원 판결 예측 AI 경진대회, 4st place, 데이콘, Korea (Jul. 2023)
- 음향 데이터 COVID-19 검출 AI 경진대회, 3st place, 데이콘, Korea (Jul. 2022)
- 2021년 산업제어시스템 보안위협 탐지 AI 경진대회(HAIcon 2021), 입선, 주최·주관: 국가정보원, 국가보안기술연구소, 후원: 한국정보보호학회, 운영: 데이콘, Korea (Nov. 2021)
- SAS 논문경진대회, 우수 논문상, 주최: 한국쌔스소프트웨어(유), 주관: 한국데이터마이닝학회, Korea (Nov. 2020)
- FactoryHack Korea 2020, 대상, 주최: 산업통상자원부, 주관: 한국전자부품연구원(KETI), Korea (Jan. 2020)
- KU-AI 4.0 바이오 의료 빅데이터 분석 및 활용 과정, 최우수상, 주최: 고용노동부, 주관: 고려대학교 대학교육개발원, Korea (Nov. 2018)

▪ Training Courses

- 산업인공지능 전문인력 양성, 한국산업기술진흥원, Korea (Sep. 2019 ~ Mar. 2020)
- 4단계 BK21 빅데이터 애널리틱스 기반 문제 정의형 인재 양성, 교육부, Korea (Sep. 2020 ~ Present)
- LG Aimers / Data Intelligence, LG AI연구원, Korea (Jul. 2022 ~ Aug. 2022)
- LG Aimers / Data Intelligence, LG AI연구원, Korea (Jul. 2023 ~ Aug. 2023)

1. Profile



김태형
Taehyeong Kim

M.S. / Ph.D. Student

Artificial Intelligence & Data Analysis
(AIDA) Lab.

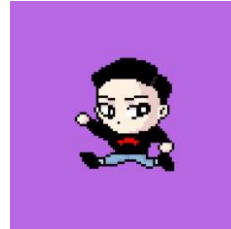
School of Industrial and Management
Engineering, Korea University.

GitHub :
<https://github.com/GNOEYHEAT>

LinkedIn :
<https://www.linkedin.com/in/gnoeyheat/>

AI Challenges

- Competition Challenger (Rank 4 of 104,309), 수상 8회, 1% 3회, 4% 9회, 10% 6회, 데이콘 (Mar. 2024)



GNOEYHEAT

<https://dacon.io/myprofile/189223/home>

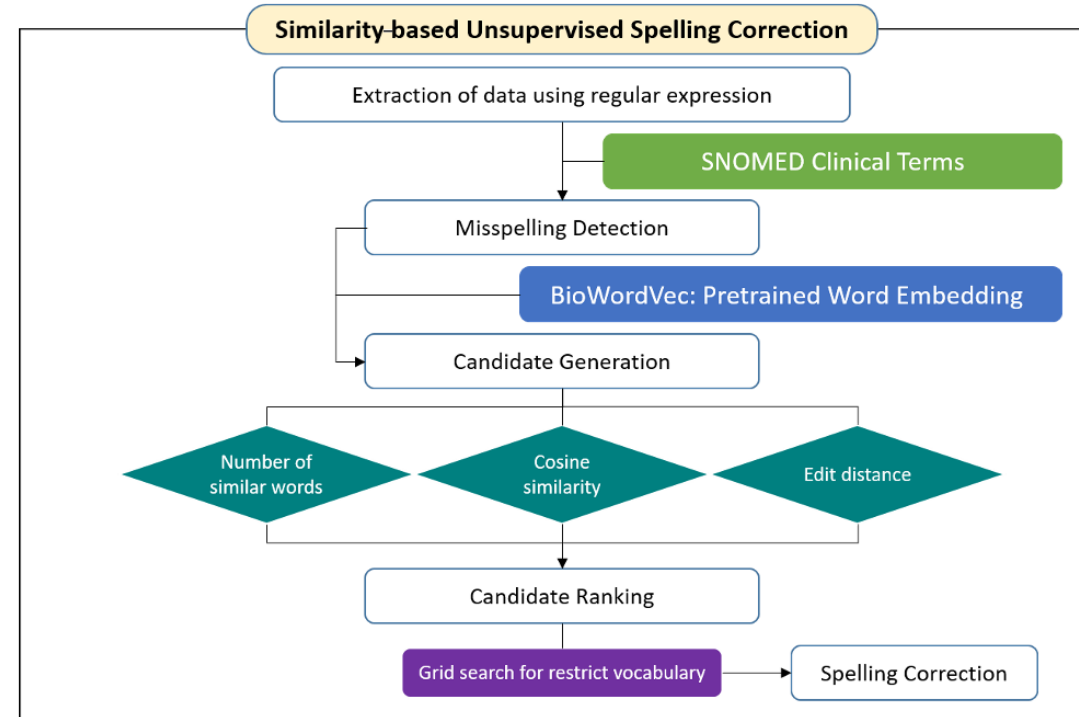
 Competition Challenger	3 of 105,293 현재 랭킹	3 최고 랭킹	수상 9회 1% 4회 4% 9회 10% 6회	개인 22회 단체 16회 전체 38회
--	-----------------------------	---------------	-----------------------------------	----------------------------

- 정형 및 시계열 도메인 경진대회 2회 수상.
(HD현대 AI Challenge, HAICon2021 산업제어시스템 보안위협 탐지 AI 경진대회)
- 비전 도메인 경진대회 2회 수상.
(반도체 소자 이상 탐지 AI 경진대회, 2023 Samsung AI Challenge : Camera-Invariant Domain Adaptation)
- 자연어처리 경진대회 3회 수상.
(코드 유사성 판단 시즌2 AI 경진대회, AI vs Human 텍스트 판별 해커톤, 법원 판결 예측 AI 경진대회)
- 음성처리 경진대회 1회 수상. (음향 데이터 COVID-19 검출 AI 경진대회)
- 멀티모달 도메인 경진대회 1회 수상. (이미지 기반 질의 응답 AI 경진대회)

2. Publications

2.1. Similarity-Based Unsupervised Spelling Correction Using BioWordVec: Development and Usability Study of Bacterial Culture and Antimicrobial Susceptibility Reports (JMIR Medical Informatics, 2021)

Keywords : spelling correction, natural language processing, bacteria, electronic health record

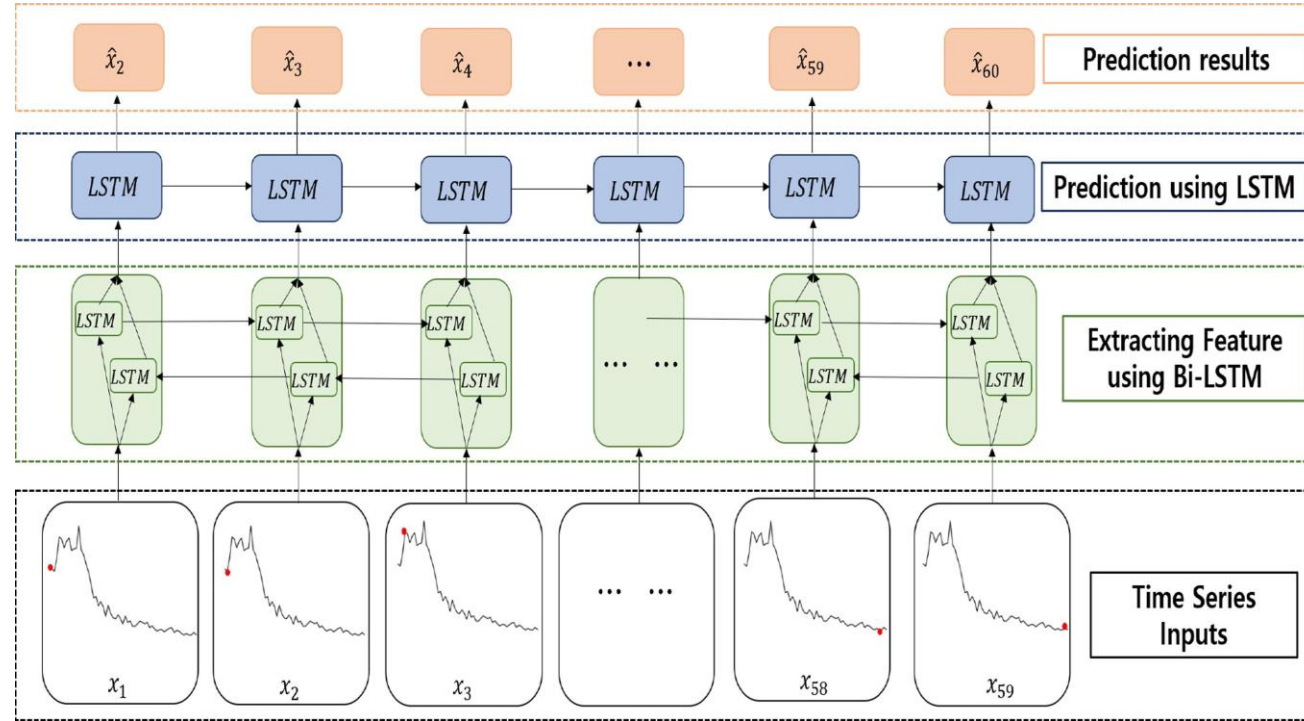


- 본 연구에서는 감염성 질환 관련 비정형 데이터 용어를 추출하기 위하여 BioWordVec을 사용한 유사도 기반의 맞춤법 검사 알고리즘을 제안함.
- 제안하는 알고리즘은 올바른 철자의 단어 사전이 구축되어 있지 않은 상황에서 맞춤법 검사 및 교정이 가능한 장점을 가짐.
- 또한, 오타에 강건한 장점을 가짐으로써 의료 용어와 같은 전문성이 높으나 통일되지 않는 자연어 처리에 적합함.
- 이를 통해, 대규모 의료 텍스트 데이터를 효율적으로 정제하고 처리할 수 있음.

2. Publications

2.2. Automobile parts reliability prediction based on claim data: The comparison of predictive effects with deep learning (Engineering Failure Analysis, 2021)

Keywords : automobile claim data, reliability prediction, convolutional neural network, recurrent neural network, deep learning

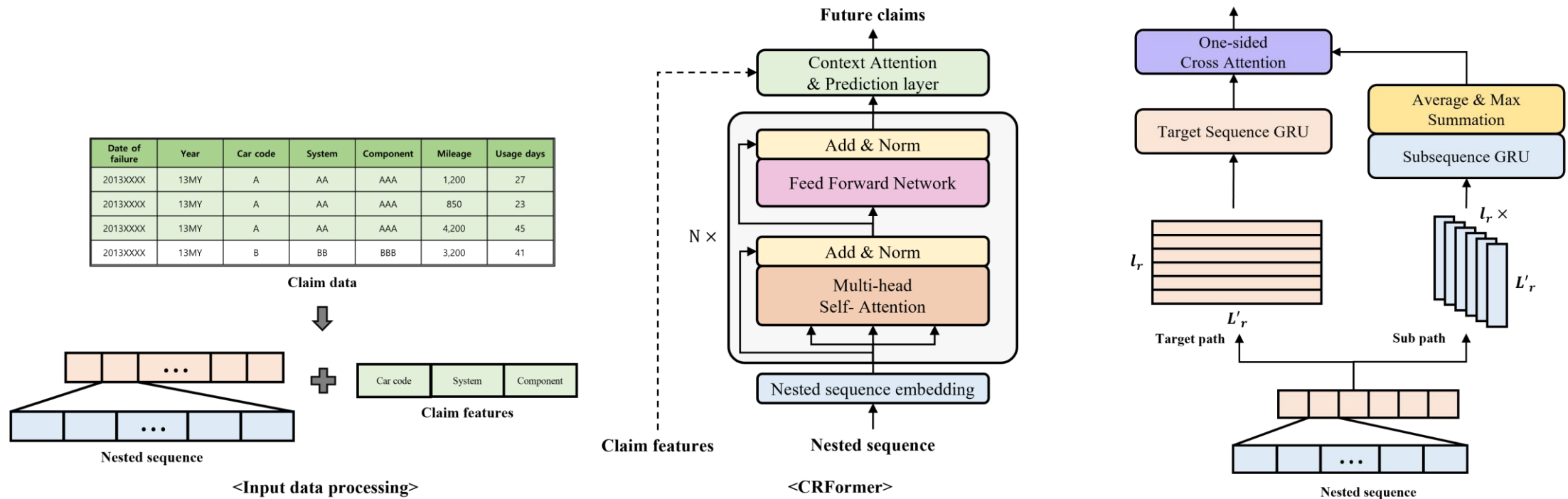


- 본 연구에서는 초기 6개월의 필드 클레임 데이터를 이용하여 자동차 부품의 고장 및 신뢰성 예측 모델을 개발함.
- 다양한 딥러닝 방법론을 제안하고 통계적 방법, 시계열 분석 방법, 머신러닝 등 기존의 선행 연구와 비교 실험함.
- Kaplan-Meier 추정 방법을 활용한 신뢰성 예측 평가 지표 RCA(Reliability Curve Accuracy)를 제안함.
- 고장 개수 예측과 신뢰성 예측 모두 RNN (many-to-many) 모델이 가장 우수한 성능을 보임.

2. Publications

2.3. CRFormer: Complementary Reliability Perspective Transformer for Automotive Components Reliability Prediction Based on Claim Data (IEEE Access, 2022)

Keywords : reliability, automotive engineering, predictive models, transformers, warranties, data models, costs



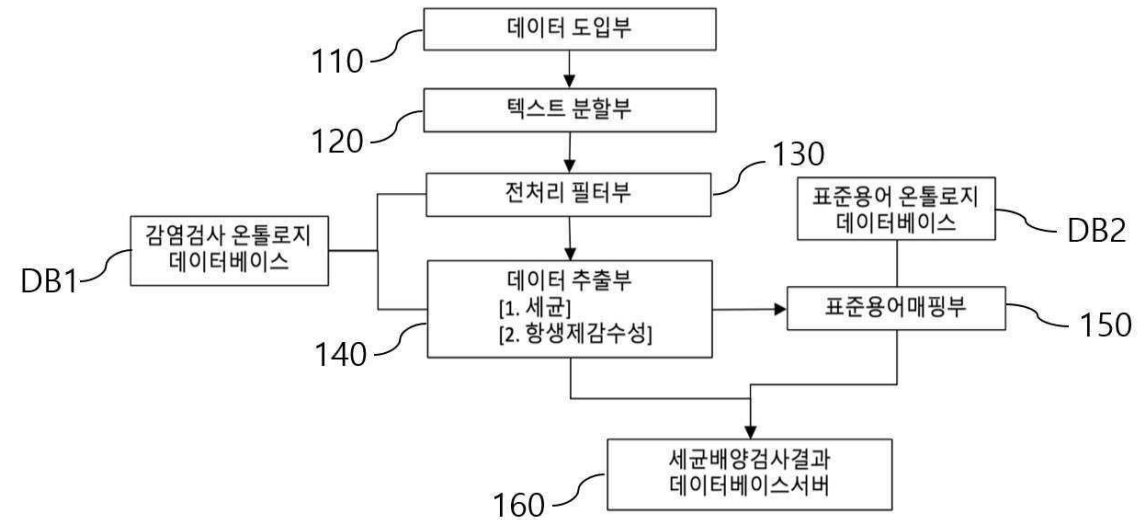
- 본 연구에서는 자동차 부품 신뢰성을 예측하기 위한 transformer encoder 기반의 CRFormer를 제안함.
- 서로 다른 신뢰성 관점(시간, 주행거리)을 상호보완적으로 융합하는 Nested sequence embedding을 통해 짧은 시퀀스에서도 풍부한 특징을 추출함.
- 부품 정보를 학습하는 Context Attention과 DTW(Dynamic Time Warping)를 활용한 pattern similarity 평가지표를 제안함.
- CRFormer는 불규칙 또는 감소 패턴에만 과대적합되는 기존 모델과는 다르게 증가 패턴에 대해 우수한 예측 성능을 도출함.

3. Patents

3.1. 온톨로지 데이터베이스 기반의 룰 기반 알고리즘을 이용한 세균배양검사결과보고서 정제 시스템

KR-Registration No. 10-2512095 (2023.03.15)

100

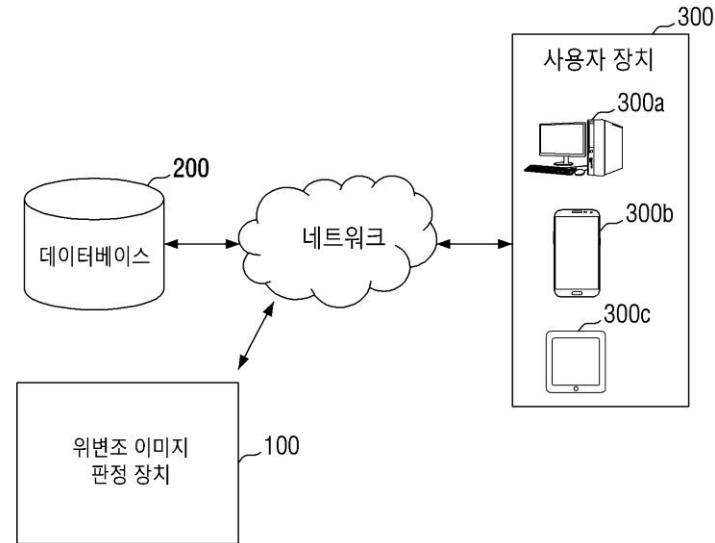


- 본 특허는 온톨로지 데이터베이스 기반의 룰 기반 알고리즘을 이용한 세균배양검사결과보고서를 정제 시스템에 관한 것임.
- 다기관 상호운용성을 높이고 데이터 분석을 용이하기 위해 세균배양검사결과보고서를 사용자가 원하는 용어집의 용어로 분석함.
- 이를 룰베이스 알고리즘을 통해 정제하고 사용자가 원하는 용어로 변환할 수 있는 시스템 및 그 방법을 개발함.
- 각 의료기관이 보유한 다른 구조의 의료데이터에 적용 가능한 동일한 구조와 규격의 공통 데이터 모델을 구축할 수 있음.

3. Patents

3.2. 노이즈 부가를 이용한 위변조 이미지 판정 방법 및 그 장치

KR-Registration No. 10-2592718 (2023.10.18)



- 본 특허는 노이즈 부가를 이용한 위변조 이미지 판정 방법 및 그 장치에 관한 것임.
- 입력 이미지의 오브젝트 영역에 기초하여, 입력 이미지의 적어도 일부에 노이즈를 추가함.
- 노이즈가 추가된 입력 이미지에 대한 OOD(Out-Of-Distribution) 탐지를 수행함.
- 위변조 이미지 판정의 성능을 향상시킬 수 있는 방법 및 그 장치를 개발함.

4. Projects

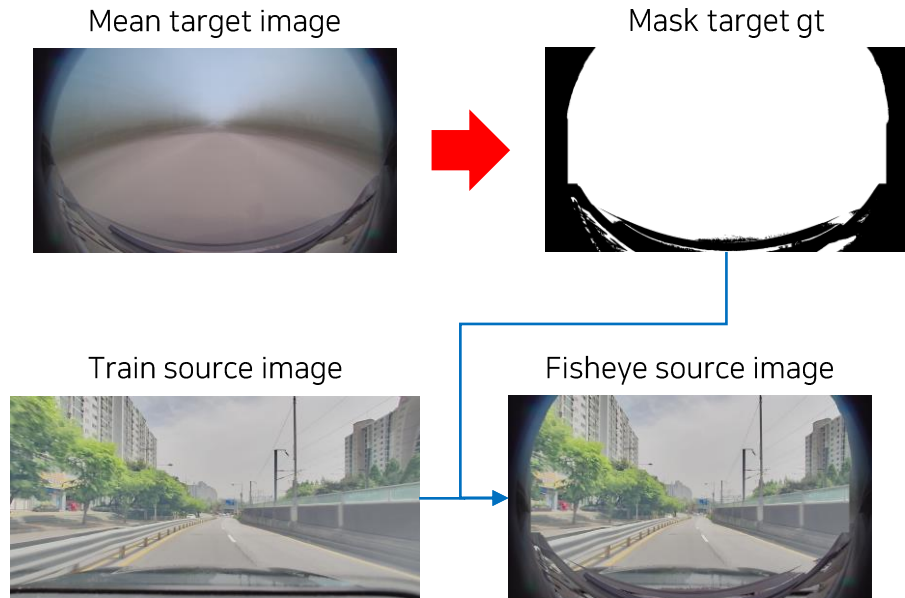
지원기관명	연구 기간	연구 제목
(주)프림로즈힐파트너스	2022.10 ~ 2022.12	AI를 활용한 상품정보 자동 생성 알고리즘 연구
<ul style="list-style-type: none"> • 딥러닝 기술을 활용하여 상품 이미지, 상품 설명 등의 상품정보를 분석하고 계층적인 상품 상세 속성 정보를 자동으로 태깅하는 기술임. • 상품 이미지와 상품 설명 같은 이미지와 텍스트 데이터를 활용하여 멀티모달 기반의 태깅 예측 모델을 개발함. • Dual Encoder와 Fusion Encoder 구조를 통합한 Encoder-Decoder 구조의 Unified Vision-Language 모델을 구축함. • 제품 이미지, 제품 상세 이미지, 전체 착용 이미지 등 복수의 이미지로 구성된 제품에 대해 상품의 범주를 가장 잘 표현하는 대표 이미지를 선정하는 알고리즘을 개발함. 		
지원기관명	연구 기간	연구 제목
한국연구재단	2020.06 ~ 2022.02	양자 기계학습 시뮬레이터
<ul style="list-style-type: none"> • 50 ~ 100개의 qubit로 구성된 Noisy Intermediate-Scale Quantum(NISQ)에서 실현 가능한 양자 기계학습 알고리즘을 중점으로 연구함. • Quantum Convolutional Neural Network(QCNN) 기반의 새로운 양자 기계학습 아키텍처를 연구함. • 양자 신경망의 Calibration 분석을 위해 매개변수화된 양자 회로의 비교 연구를 진행함. 		
지원기관명	연구 기간	연구 제목
(주)현대엔지비	2019.08 ~ 2021.08	클레임데이터를 활용한 내구열화 부품 발체 및 미래 고장 패턴 예측 기술 개발
<ul style="list-style-type: none"> • 부품 단위의 고장 발생을 예측하여 클레임 비용을 절감하고 차량 내구 개발 시 사전에 문제를 예지하여 반영하고자 함. • 단기간의 고장 패턴을 바탕으로 장기간의 고장 발생을 예측하고자 함. 패턴분석 및 기계학습(AI)을 통한 예측 기술을 개발함. • 고장발생 영향 변수 기반 다변량 신뢰도 예측 기술을 개발함. 클레임 데이터 기반 내구 및 품질 문제 분류 알고리즘을 개발함. • 본 연구를 통해 'CRFormer: Complementary Reliability Perspective Transformer for Automotive Components Reliability Prediction Based on Claim Data'라는 제목의 SCI급 논문을 게재함. 		
지원기관명	연구 기간	연구 제목
한국보건산업진흥원	2019.04 ~ 2021.12	감염성질환 데이터의 CDM 확장모델 개발 및 활용을 위한 Tool 개발
<ul style="list-style-type: none"> • 감염성 질환 관련 검사결과 및 환자이동정보 데이터에 대한 Common Data Model을 개발하고 Extract, Transform, Load 도구를 개발함. • BioWordVec 기법으로 사전 훈련된 단어 임베딩을 활용한 유사도 기반의 맞춤법 검사 알고리즘을 제안함. • 제안하는 모형은 사전이 구축되지 않은 상태에서도 철자를 교정할 수 있으며 앞으로 발생할 수 있는 오타에 대해서도 강건하다는 장점을 가짐. • 본 연구를 통해 'Similarity-Based Unsupervised Spelling Correction Using BioWordVec: Development and Usability Study of Bacterial Culture and Antimicrobial Susceptibility Reports'라는 제목의 SCI급 논문을 게재함. 		

5. Awards and Honors

5.1. 2023 Samsung AI Challenge : Camera-Invariant Domain Adaptation (장려상, 삼성전자 SAIT)

- Fisheye 변환을 활용한 InternImage 기반의 도메인 적응 이미지 분할.

Keywords : semantic segmentation, domain adaptation, computer vision, fisheye camera calibration

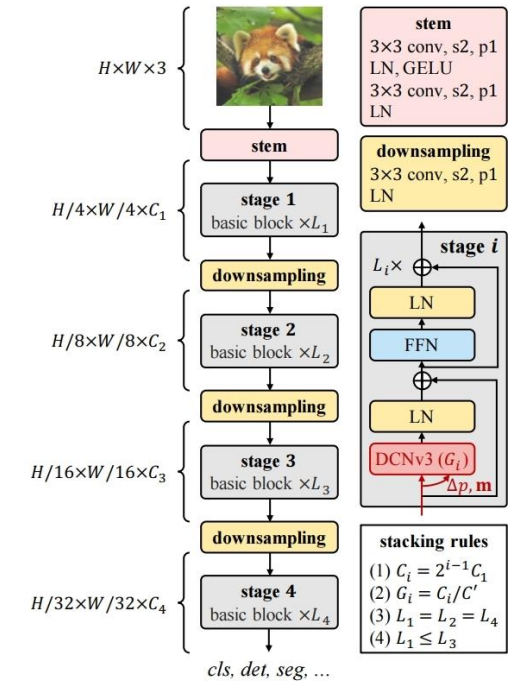


<Fisheye Transformation>

Fisheye Transformation Pipeline

- ① Train target image sample를 평균한 Mean target image를 생성함.
- ② threshold 기준으로 Mask target gt를 생성함.
- ③ Mean target image를 source image와 결합한 Fisheye source image와 background class가 추가된 Fisheye source gt를 생성함.

Fisheye Ground Truth Visualization



<Overall Architecture of InternImage>

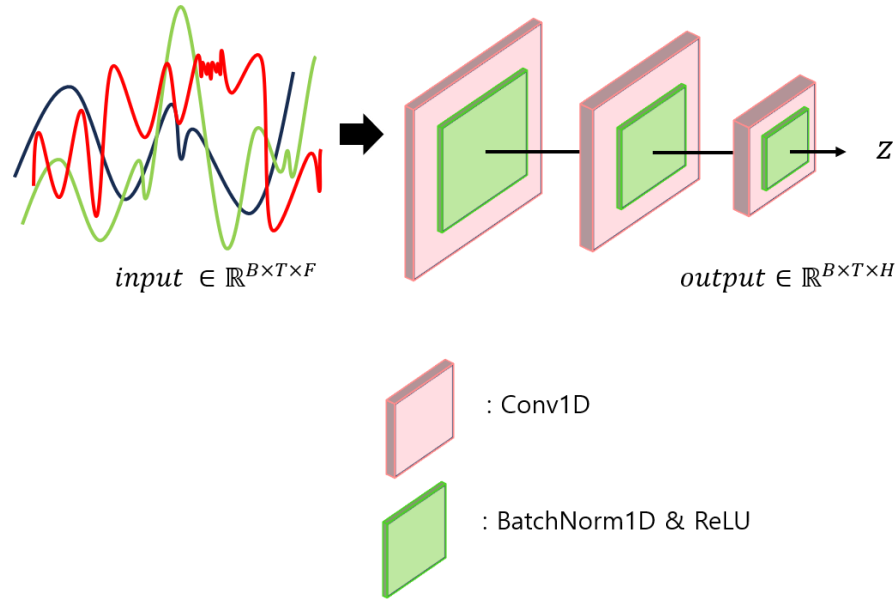
- 본 경진대회에서는 왜곡이 존재하지 않는 이미지(Source Domain)와 레이블을 활용하여 왜곡된 이미지(Fisheye Target Domain)에 대해서도 고성능의 이미지 분할(Semantic Segmentation)을 수행하는 AI 알고리즘을 개발하고자 함.
- 레이블이 존재하지 않는 Train target image를 활용하여 Fisheye data의 background를 정의하는 Fisheye Transformation을 개발함.
- Deformable Convolution v3를 사용하는 CNN 기반의 SOTA 모델인 InternImage를 backbone으로 사용함.
- Dice Loss, Multi-scale Test Time Augmentation(TTA)를 적용하여 소수 범주에 대한 다양한 크기의 예측 성능(mIoU)을 향상시킴.

5. Awards and Honors

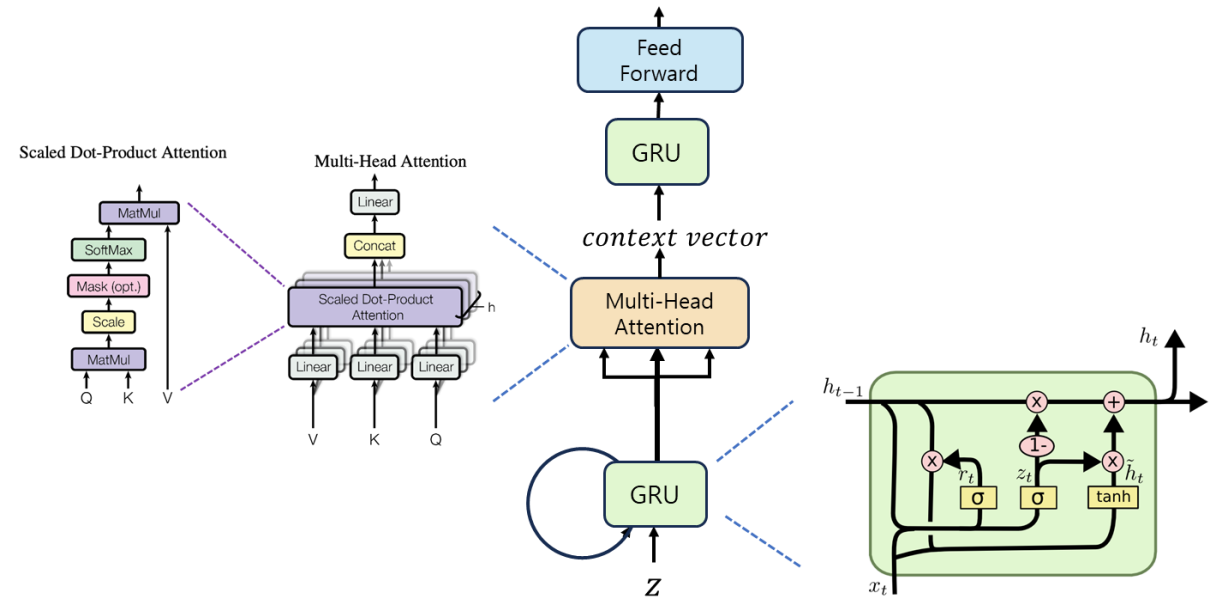
5.2. 2023 HD현대 AI Challenge (우수상, HD한국조선해양 AI Center)

- 어텐션 순환신경망을 활용한 센서 기반의 작업 중량 예측.

Keywords : time series regression, temporal context modeling, attention mechanism, overpredicted mean absolute error



<Temporal context modeling - Step 1>



<Temporal context modeling - Step 2>

- 본 경진대회에서는 건설기계 센서데이터를 활용하여 작업자의 운전 성향과 숙련도에 영향을 받지 않는 건설장비의 작업 중량 예측 모델을 개발하고자 함.
- 신호 및 센서의 노이즈를 줄이기 위해 지수가중평균 기법을 사용하여 smoothing함 .
- 점진적으로 kernel의 크기가 감소하고 channel이 증가하는 CNN 구조를 사용하여 신호 및 센서의 노이즈를 줄이면서 다양한 특징을 추출하고, Attention 기반의 GRU를 활용하여 시간적 정보를 고려하는 모델을 제안함.
- Overpred MAE Loss를 구현하여 검증에 활용함.

Thank you 😊

taehyeong93@korea.ac.kr