

데이터마이닝 연구실

Data Mining &
Computational Biology
Laboratory



이현주
교수

hyunjulee@gist.ac.kr

062-715-2213

<https://combio.gist.ac.kr>

Education

- 2002 ~ 2006** University of Southern California (Ph.D. – Computer Science)
- 1997 ~ 1999** Seoul National University (M.S. - Computer Engineering)
- 1993 ~ 1997** KAIST, South Korea (B.S. - Computer Science)

Experience

- 2007 ~** Gwangju Institute of Science and Technology, Professor
- 2006 ~ 2007** Brigham and Women's Hospital and Harvard Medical School, Postdoctoral Fellow
- 2002 ~ 2006** University of Southern California, Research Assistant
- 2001 ~ 2002** Korea Wisenut Inc., Senior Engineer
- 1998 ~ 2001** Intus Technology inc., Xinics Inc., Engineer

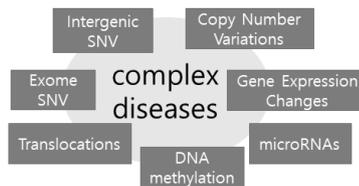
Fact Sheet

- 2018** GIST Best Research Award

연구실 소개

데이터마이닝 및 계산 생물학 연구실은 생명/의학/사회 분야에 인공지능 방법론을 적용하는 융합 연구를 수행한다. 인터넷에서 생명과학에 걸친 다양한 분야의 이질적 데이터의 통합/분석을 통하여, 보편적이고 조직적인 정보추출을 위한 데이터마이닝을 연구하고 있다. 현재 진행 중인 연구들은 다음과 같다. i) 암과 관련된 유전자 및 세포 신호 전달 경로들을 발굴하는 인공지능 기술을 개발하고, 암환자들의 개인화된 게놈 정보를 분석하여 표적 항암제를 추천함으로써, 암환자들의 진료에 도움을 주고자 한다. ii) 유전체 데이터를 기반으로 알츠하이머 병 등의 노화 관련 질병의 조기 발굴 및 치료를 위한 인공지능 방법론을 개발한다. iii) 자연언어처리 방법론을 바탕으로 질병 관련 검색엔진을 개발함으로써 질병 연구를 촉진시키는데 기여하고자 한다. iv) 대화형 인공지능 및 멀티모달 이해 시스템에 최신 대형 언어모델(LLM)을 적용하기 위한 자연언어처리 학습 방법론을 개발한다.

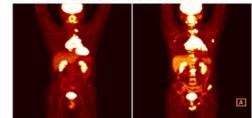
헬스케어 빅 데이터



알츠하이머 병



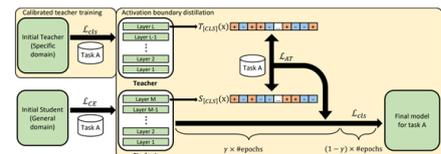
암 발병/진행/치료



질병-유전자-약물 검색엔진



자연언어처리 학습 방법론 개발



연구 성과

수행중인 주요 연구과제 (주요과제경력)

- 분자유전학적 원인 불명인 악성 종양 환자의 개인 맞춤 원인 규명 및 치료제 개발을 위한 지능형 SW 시스템 개발
- 빅데이터를 활용한 알츠하이머성 치매의 바이오마커 추천 및 진단을 위한 인공지능 시스템 개발 (한국연구재단)
- MCMT 분석을 위한 멀티패킷 텍스트 마이닝 시스템 개발 (한국연구재단)
- 사전 훈련된 언어 모델의 성능 향상 기법 (한국연구재단)

주요논문 (대표실적)

- Domain Knowledge Transferring for Pre-trained Language Model via Calibrated Activation Boundary Distillation. Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (2022)
- Integrative reconstruction of cancer genome karyotypes using InfoGenomeR. Nature Communications (2021)
- Multiresolution correction of GC bias and application to identification of copy number alterations. Bioinformatics (2019)
- Multitask learning approach for understanding the relationship between two sentences, Information Sciences (2019)
- DigChem: Identification of disease-gene-chemical relationships from Medline abstracts. PLoS Computational Biology (2019)

주요특허

- 전장 유전체 데이터를 이용한 유전체 복원 방법 (2022, 해외 PCT)
- 미분류 데이터를 이용하여 신경망의 학습을 조기 종료하는 방법 (2022, 해외 PCT)
- 암 연구를 위한 모듈 구성 방법 (2019)
- 암 연관 마이크로RNA의 우선순위화 방법 (2019)
- 유전자와 질병간의 관계를 포함하는 문장 검색 엔진 (2018)

주요연구시설

- CPU 서버 8대 이상 (총 120코어 이상, 메모리 1테라 이상, 하드 500테라 이상), GPU 70대 이상 보유

G.I.S.T. AI for X 융합연구

약물 데이터, 약물표적상호작용 데이터, 단백질 데이터들을 통합하여, 새로운 약물표적 상호작용을 예측하는 딥러닝 및 머신러닝 방법론 개발.

약물

복제수 변이 데이터, 유전자 발현 데이터, 마이크로 RNA 데이터 등의 생명정보 데이터를 통합하여, 개인 맞춤 치료를 위한 인공지능 시스템 개발.

생물학/의학

대화문 요약 모델 성능 향상을 위한 데이터 증강 기법 개발, 페르소나에 기반한 대화 응답 생성 기법 개발, 멀티모달 이해 능력 향상을 위한 지식 증류 기반의 학습 방법론 개발.

자연언어처리