

강의개요

Single Cell, Pre-training, and Foundation Model

Single-cell RNA sequencing(scRNA-seq)은 다양한 생물학적 환경에서 세포 이질성과 조직 기능을 규명하기에 유용하지만, scRNA-seq 데이터의 희소성(sparsity) 및 라벨을 가진 주석 데이터(annotation data)의 부족에 의해서, 전사체 분석의 정확도에 한계를 가진다. 이러한 문제를 해결하기 위해 단일세포를 위한 사전 학습 (pre-training) 방법론 및 파운데이션 모델 (foundation model)이 활발히 연구되고 있다. 사전 학습은 대규모 데이터를 활용해 자가 지도 학습 (self-supervised) 방식으로 일반적인 표현(representation)을 학습함으로써 라벨이 없는 scRNA-seq 데이터에도 효과적으로 적용될 수 있다. 최근 발표된 단일 세포 기반 파운데이션 모델은 대형 scRNA-seq 데이터셋을 바탕으로 gene embedding과 cell embedding을 학습해 다양한 downstream task의 성능을 크게 향상시켰다. 이러한 모델들은 dropout으로 인해 결측된 발현값을 더 의미 있는 표현 공간에서 보정하고, 세포 타입 주석, 배치 보정, 희귀 세포 탐지 등에서 기존 방법보다 높은 정확도를 보여준다.

본 강의에서는 단일 세포 파운데이션 모델에서 사용되는 자가 지도 학습 방법론과 주요 모델 및 다양한 downstream task를 소개한다. 본 강의를 통해서 파운데이션 모델을 구축하기 위한 데이터 셋과 학습 방법론들과 이를 생물학 지식으로 변환하는 연구들을 이해하는 것을 목표로 한다.

- Single-cell RNA sequencing 소개
- 사전 학습 방법론 및 파운데이션 모델 구축 방법론
- scRNA-seq 기반 파운데이션 모델 사례
- Downstream task 및 성능
- 질병 관련 연구에의 응용

* 강의 난이도: 중급

* 강의: 이현주 교수 (광주과학기술원 AI융합학과)

Curriculum Vitae

Speaker Name: **Hyunju Lee, Ph.D.**



► Personal Info

Name Hyunju Lee
Title Professor
Affiliation Gwangju Institute of Science and Technology

► Contact Information

Address 123 Cheomdangwagi-ro, Buk-gu, Gwangju, 61005
Email hyunjlee@gist.ac.kr
Phone Number 062-715-2213

Research interest : Bioinformatics, Machine learning, and Text Mining

Educational Experience

1997 B.S. in Computer Science, KAIST, South Korea
1999 M.A. in Computer Engineering, Seoul National University, South Korea
2006 Ph.D. in Computer Science, University of Southern California, USA

Professional Experience

2006-2007 Post-doc Researcher, Brigham and Women's Hospital and Harvard Medical School, USA
2007- Professor, Dept of AI Convergence, Electrical Engineering and Computer Science, Gwangju Institute of Science and Technology

Selected Publications (3 maximum)

1. Yeonghun Lee, Sung-Hye Park, Hyunju Lee. Prediction of the 3D cancer genome from whole-genome sequencing using InfoHiC. *Molecular Systems Biology*, 20(11):1156-1172, 2024
2. Sejin Park, Hyunju Lee. Robust self-supervised learning strategy to tackle the inherent sparsity in single-cell RNA-seq data. *Briefings in Bioinformatics*, 25(6): bbae586, 2024.
3. Yeonghun Lee and Hyunju Lee. Integrative reconstruction of cancer genome karyotypes using InfoGenomeR. *Nature Communications*, 12:2467, 2021.