ANNOTATING ANTIBIOTIC RESISTANCE GENES USING GRAPH NEURAL NETWORK

goc Khoi Dang, Jun Chen
Introduction

- Antibiotic is the primary weapon to fight against the bacterial infection.
- Antibiotic resistance became a global threat due to the overuse and misuse of these medications.
- Antibiotic resistance protein (ARP) could deactivate or repel antibiotics.
Problem

- The protein labeling problem is a multilabel classification problem that can be processed quickly with the helping of computational protein function classifiers.
- More effectively labeling antibiotic resistance protein could help to improve the study of antibiotic resistance.
Approach

- Computational protein function classifiers.
- Building model to deal with antibiotic resistance protein function prediction
- We will use CARD dataset to train and evaluate the model.
Methods

- The model is Graph Neural network.
  - Each node represents an ARO, and the edge represents the relations between nodes.
  - The graph is trained with node features, edge features, and global features.
  - Each node has 12 node features which are the 10 top scores from PSI-BLAST [3], DIAMOND score [4], and Priority score [1].
  - The global feature is a vector of amino acid composition with a length of 20. The edge feature is just the relation of 2 nodes. The initial values would be assigned randomly from 0 to 1.

- We plan to use PyTorch and a graph network framework library.

- We plan to use Adam optimizer with a learning rate of 0.01.
Datasets

- 4,577 annotated ARP sequences from CARD database. The dataset will be splitted into training set and testing set with the ration 8:2.
Metrics for evaluation

- F-measure score (Fmax) is the maximum F-measure over confidence score thresholds t.
- Minimum semantic distance score (Smin) is the metric that considered the unbalanced information content (IC) of GO terms.
- Area under the precision-recall curve (AUPR)
Project milestones

- 3/14 - 4/04: Code implementation
  - Khoi: Implement graph neural network
  - Jun: Implement node feature extractions.
- 4/04 - 4/25: Evaluation and tuning
  - Khoi: Evaluation metrics implementation
  - Jun: Hyper parameters tuning
- 4/18 - 4/25: Presentation (Khoi + Jun)
- 4/25 - 5/04: Final report (Khoi + Jun)
Bibliography


