Division of Statistics and Machine Learning
Department of Computer and Information Science
Linköping University

Master thesis project ”Biologically informed graph neural networks for perturbation effect prediction”

Background

More than 40% of patients do not respond to standard treatments for common diseases. One reason for this is complexity of the genomic response to drugs due to multitude of interactions between genes.

To predict perturbation effects of drugs, machine learning can be used. For example, DeepCOP [1] is based on deep learning to provide predictions. However, such methods do not consider information about the known molecular interactions between genes. On another hand, there are models that predict drug effects based on molecular interactions only, without using machine learning or gene expression data [3]. In this project, we aim to apply Graph Neural Networks and information about known molecular interactions such as Omnipath [2] to predict perturbation effects.

Data

- CMAP database contains perturbation effects for various drug and sample combinations [4].

Research questions

- Which graph neural network (GNN) models are most appropriate for predicting drug perturbation effects from gene expression data and information about drug targets?
- How accurately can GNN models predict the perturbation effects for new samples and for unseen drugs?
- How much do predictions obtained through the GNN models differ from predictions of some published ML models that only use gene expressions but not the network structure (for ex. DeepCOP)?

Prerequisites

- Good knowledge of Machine learning and Statistics
- Good programming skills

Contact and application

- Oleg Sysoev, oleg.sysoev@liu.se

References

[1] Woo, G., Fernandez, M., Hsing, M., Lack, N. A., Cavga, A. D., & Cherkasov, A. (2020). DeepCOP: deep learning-based approach to predict gene regulating effects of small molecules. *Bioinformatics*, *36*(3), 813-818.

[2] Türei, D., Korcsmáros, T., & Saez-Rodriguez, J. (2016). OmniPath: guidelines and gateway for literature-curated signaling pathway resources. *Nature methods*, *13*(12), 966-967.

[3] Roca, C. P., Sysoev, O., Eyre, E., Galan, S., Sinibaldi, D., Tedder, P., & Mangion, J. (2025). Regulation Flow Analysis discovers molecular mechanisms of action from large knowledge databases. *bioRxiv*, 2025-06.

[4] https://clue.io/data/CMap2020#LINCS2020