Bioinformatics Center
– Bio-knowledge Engineering –

https://www.bic.kyoto-u.ac.jp/pathway/index.html

Prof
MAMITSUKA, Hiroshi
(D Sc)

Senior Lect
NGUYEN, Hao Canh
(D Knowledge Science)

Students

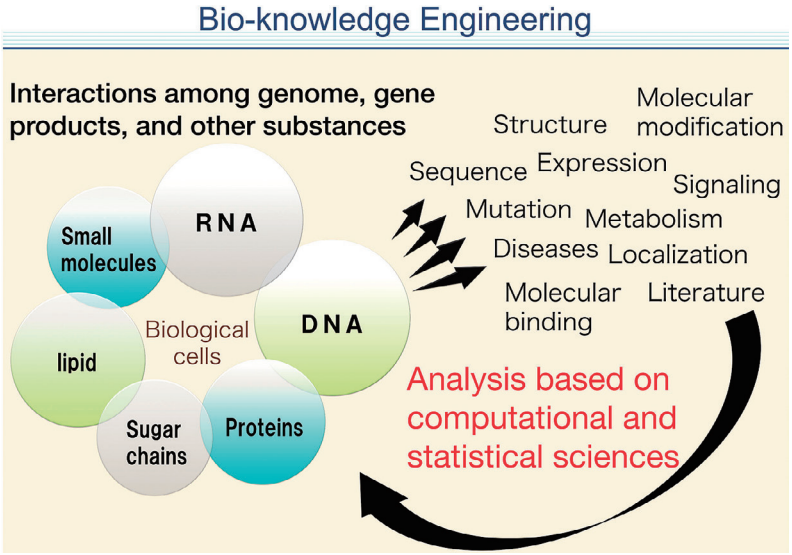
NGUYEN, Duc Anh (D3) NISHIKAWA, Emina (UG)

Guest Res Assoc

PETSCHNER, Peter (Ph D) Semmelweis University, Hungary, 28 November 2020– 27 November 2022

Scope of Research

We are interested in graphs and networks in biology, chemistry, and medical sciences, including metabolic networks, protein-protein interactions and chemical compounds. We have developed original techniques in machine learning and data mining for analyzing these graphs and networks, occasionally combining with table-format datasets, such as gene expression and chemical properties. We have applied the techniques developed to real data to demonstrate the performance of the methods and find new scientific insights.



KEYWORDS

Bioinformatics
Computational Genomics
Data Mining
Machine Learning
Systems Biology

Recent Selected Publications

Nguyen, C. H.; Mamitsuka, H., Learning on Hypergraphs with Sparsity, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **43,** 2710-2722 (2021).

Wimalawarne, K.; Mamitsuka, H., Reshaped Tensor Nuclear Norms for Higher Order Tensor Completion, *Machine Learning*, **110,** 507-531 (2021).

You, R.; Yao, S.; Mamitsuka, H.; Zhu, S., DeepGraphGO: Graph Neural Network for Large-Scale, Multispecies Protein Function Prediction, *Bioinformatics*, **37,** I262-I271 (2021).

Nakamura, A.; Takigawa, I.; Mamitsuka, H., Efficiently Enumerating Substrings with Statistically Significant Frequencies of Locally Optimal Occurrences in Gigantic String, *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI 2020)*, **34(4),** 5240-5247 (2020).

Nguyen, D. H.; Nguyen, C. H.; Mamitsuka, H., ADAPTIVE: Learning DAta-dePendenT, ConcIse Molecular VEctors for Fast, Accurate Metabolite Identification from Tandem Mass Spectra, *Bioinformatics*, **35 (14),** *(Proceedings of the 27th International Conference on Intelligent Systems for Molecular Biology (ISMB/ECCB 2019))*, i164-i172 (2019).

A Hypergraph Neural Network for Predicting Drug-drug Interactions

A drug-drug interaction (DDI) is a combination of two drugs causing side effects, which are unwanted reactions of human bodies. These side effects might be responsible for significant patient morbidity and mortality, and cost billions of dollars each year. Hence, predicting DDI is a very important task to guarantee drug safety in pharmacology. The traditional approach for predicting drug-drug interactions uses clinical trials on patients, which are time-consuming and costly. Recently, machine learning models have emerged as prominent tools for predicting DDI, which are fast and inexpensive.

Given drugs' information and known side effects of many pairs of drugs, one wishes to learn a model to predict side effects of all pairs of drugs. DDI is usually represented as a graph that nodes are drugs and edges are interacting drug pairs with side effects as labels. The task is to predict labels of all pairs of nodes in the DDI graph. Figure 1a shows an example DDI graph, where the dotted edge with question marks is the pair of drugs with labels to be predicted.

Existing work often uses graph neural networks to learn vector representations of drug nodes on the DDI graph and uses them to predict interactions. One drawback of this method is the lack of learning side effect representations. Side effects have complex relationships, for example, co-occurrences. Previous methods often represent each side effect as a one-hot vector indicating the presence of the side effect. This representation considers that side effects are independent, potentially under-utilizing the side effect relationships. Hence, it is necessary to learn representations of both side effects and drugs altogether.

To address the above drawback, we propose to encode DDI data with a hypergraph that a node in the hypergraph

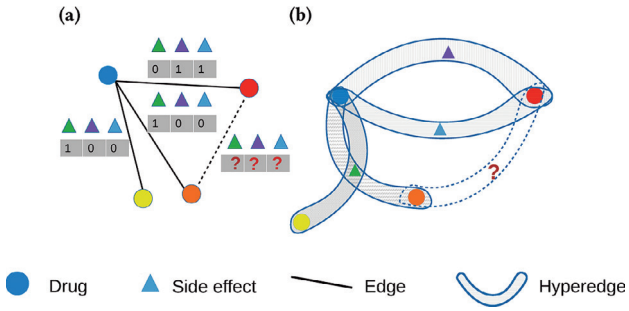


Figure 1. a) Traditional graph representation for DDI, b) Proposed hypergraph representation for DDI.

can be either a drug or a side effect and each hyperedge is a triple of two drugs and a side effect that they cause. Figure 1b illustrates an example of a hypergraph corresponding to the DDI graph in Figure 1a. We then propose CentSmoothie, a central-smoothing hypergraph neural network for predicting DDI, with a new assumption that in each hyperedge, the representation of the side effect node should be close to the midpoint of the corresponding subspace of the two drugs.

We conducted experiments on the benchmark DDI data to show the advantage of the prediction performance of our proposed method. We compared CentSmoothie with other state-of-the-art methods in terms of AUC (area under the ROC curve) and AUPR (area under the precision-recall curve). The results were shown in Figure 2. The results showed that our method (CentSmoothie) outperformed other methods in both AUC and AUPR, suggesting that our model was more suitable for DDI.

Reference

Duc Anh Nguyen, Canh Hao Nguyen, and Hiroshi Mamitsuka. "A survey on adverse drug reaction studies: data, tasks and machine learning methods." Briefings in bioinformatics22, no. 1 (2021): 164-177.

Method	MLNN	MRGNN	Decagon	SpecConv	HPNN	CentSmoothie
AUC	0.8372 ± 0.0050	0.8452 ± 0.0036	0.8639 ± 0.0029	0.8785 ± 0.0025	0.9044 ±0.0003	**0.9348** ± 0.0002
AUPR	0.7919 ± 0.0041	0.8029 ± 0.0039	0.8094 ± 0.0024	0.8256± 0.0022	0.8410 ± 0.0007	**0.8749** ± 0.0013

Figure 2. Performance comparison on the benchmark DDI data.