Bioinformatics Center
– Mathematical Bioinformatics –

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

Prof
AKUTSU, Tatsuya
(D Eng)

Assoc Prof
TAMURA, Takeyuki
(D Inf)

Assist Prof
MORI, Tomoya
(D Inf)

Students

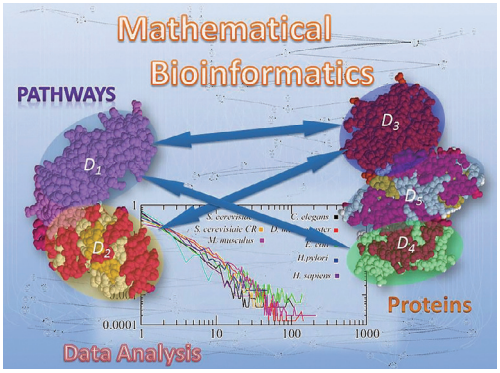
YU, Coleman (D3)
TAKAGI, Motoshige (D3)
LI, Ruiming (D3)
WANG, Feiqi (D3)
OHTOMO, Masahiro (D3)
NAKASHIMA, Shogo (D3)
MU, Lixuan (D2)
SHIOTA, Koji (D1)
LIU, Chunting (D1)
MA, Yier (M2)

Scope of Research

Due to the rapid progress of genome sequencing technology, whole genome sequences of organisms ranging from bacteria to human have become available. In order to understand the meaning behind the genetic code, we have been developing algorithms and software tools for analyzing biological data based on advanced information technologies such as theory of algorithms, artificial intelligence, and machine learning. We are currently studying the following topics: systems biology, scale-free networks, protein structure prediction, the inference of biological networks, chemo-informatics, and discrete and stochastic methods for bioinformatics.

KEYWORDS

Complex Networks
Boolean Networks
Neural Networks
Chemical Graphs
Protein Informatics



Recent Selected Publications

Akutsu T.; Jansson, J.; Li, R.; Takasu, A.; Tamura, T., New and Improved Algorithms for Unordered Tree Inclusion, *Theoretical Computer Science*, **883,** 83-98 (2021).

Cheng, C.; Ching, W-K.; Guo, S.; Akutsu, T., Discrimination of Attractors with Noisy Nodes in Boolean Networks, *Automatica*, **130,** 109630 (2021).

Nakajima, N.; Hayashi, T.; Fujiki, K.; Shirahige, K.; Akiyama, T.; Akutsu, T.; Nakato, R., Codependency and Mutual Exclusivity for Gene Community Detection from Sparse Single-Cell Transcriptome Data, *Nucleic Acids Research*, **49,** e104 (2021).

Kajiwara, M.; Nomura, R.; Goetze, F.; Kawabata, M.; Isomura, Y.; Akutsu, T.; Shimono, M., Inhibitory Neurons Exhibit High Controlling Ability in the Cortical Microconnectome, *PLoS Computational Biology*, **17,** e1008846 (2021).

Ma, Y.; Tamura, T., Dynamic Solution Space Division-based Methods for Calculating Reaction Deletion Strategies for Constraint-based Metabolic Networks for Substance Production, *Frontiers in Bioinformatics*, **1,** 716112 (2021).

Algorithms for Extracting Boolean and Probabilistic Rules from Trained Neural Networks

Recent progress of deep learning technologies has demonstrated the power of artificial neural networks in making predictions in various areas. Therefore, it is important to develop a methodology for interpreting how a trained neural network arrives at its predictions.

We develop two algorithms to extracting rules from a trained neural network consisting of linear threshold functions. The first one extracts rules in the form of Boolean functions, and outputs much more concise rules, compared with an existing one, if the threshold functions correspond to 1-decision lists, majority functions, or certain combinations of these. The second one is based on dynamic programming and extracts probabilistic relations between the input values and the output value in the form of conditional probabilities. Although this problem is NP-hard (theoretically difficult) in general, the proposed algorithm works in pseudo-polynomial time if each hidden layer consists of a constant number of neurons. The potential usefulness of these two algorithms is demonstrated by conducting several computational experiments.

We have also been applying deep learning technologies to various problems in bioinformatics, which include cancer subtype classification, protein cleavage site prediction, RNA post-transcriptional modification site prediction, and lysine post-translational modification site prediction.

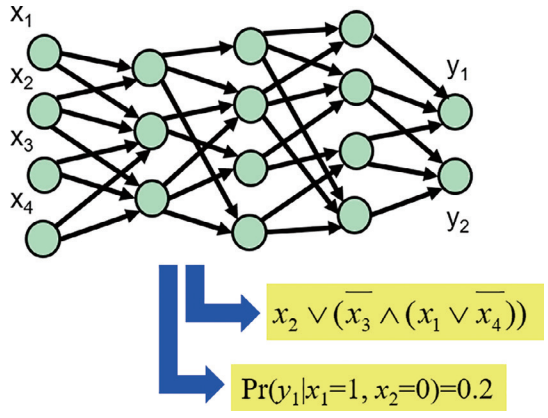


Figure 1. Our developed algorithms extract rules from trained neural networks in the forms of Boolean functions and conditional probabilities.

Gene Deletion Algorithms for Minimum Reaction Network Design by Mixed-integer Linear Programming for Metabolite Production in Constraint-based Models: gDel_minRN

Genome-scale constraint-based metabolic networks play an important role in the simulation of growth coupling, which means that cell growth and target metabolite production are simultaneously achieved. To achieve growth coupling, a minimal reaction-network-based design is known to be effective. However, the obtained reaction networks often fail to be realized by gene deletions due to conflicts with gene-protein-reaction relations.

Here, we developed gDel_minRN that determines gene deletion strategies using mixed-integer linear programming to achieve growth coupling by repressing the maximum number of reactions via gene-protein-reaction relations. Computational experiments were conducted in which gDel_minRN was applied to iML1515, a genome-scale model of *Escherichia coli*. The target metabolites were three vitamins that are highly valuable and require cost-effective bioprocesses for economics and the environment. gDel_minRN successfully calculated gene deletion strategies that achieve growth coupling for the production of biotin (vitamin B7), riboflavin (vitamin B2), and pantothenate (vitamin B5).

Since gDel_minRN calculates a constraint-based model of the minimum number of gene-associated reactions without conflict with gene-protein-reaction relations, it helps biological analysis of the core parts essential for growth coupling for each target metabolite. The source codes are implemented in MATLAB, CPLEX, and COBRA Toolbox. The obtained data and source codes are available on https://github.com/taketam/gDel-minRN

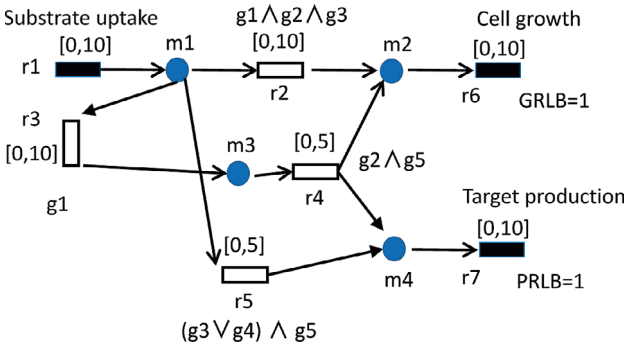


Figure 2. A toy example of the constraint-based model. Circles and rectangles represent metabolites and reactions, respectively. Black and white rectangles are external and internal reactions.