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

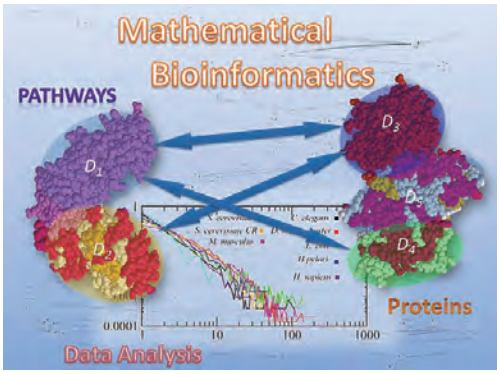
LIU, Pengyu (D3)
YU, Coleman (D3)
TAKAGI, Motoshige (D3)
LI, Ruiming (D3)
WANG, Feiqi (D3)
OHTOMO, Masahiro (D3)
NAKASHIMA, Shogo (D2)
MU, Lixuan (D1)
MA, Yier (M1)
YANG, Ziwei (RS)

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



Selected Publications

Liu, P.; Melkman, A. A.; Akutsu, T., Extracting Boolean and Probabilistic Rules from Trained Neural Networks, *Neural Netw.*, **126,** 300-311 (2020).

Akutsu, T.; Melkman, A. A.; Tamura, T., Improved Hardness of Maximum Common Subgraph Problems on Labeled Graphs of Bounded Treewidth and Bounded Degree, *International Journal of Foundations of Computer Science*, **31,** 253-273 (2020).

Wang, F.; Akutsu, T.; Mori, T., Comparison of Pseudoknotted RNA Secondary Structures by Topological Centroid Identification and Tree Edit Distance, *Journal of Computational Biology*, **27,** 1443-1451 (2020).

Zhu, J.; Wang, C.; Shurbevski, A.; Nagamochi, H.; Akutsu, T., A Novel Method for Inference of Chemical Compounds of Cycle Index Two with Desired Properties Based on Artificial Neural Networks and Integer Programming, *Algorithms*, **13,** [124-1]-[124-30] (2020).

Li, F.; Leier, A.; Liu, Q.; Wang, Y.; Xiang, D.; Akutsu, T.; Webb, G. I.; Smith, A. I.; Marquez-Lago, T.; Li, J.; Song, J., Procleave: Predicting Protease-Specific Substrate Cleavage Sites by Combining Sequence and Structural Information, *Genomics, Proteomics & Bioinformatics*, **18,** 52-64 (2020).

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.

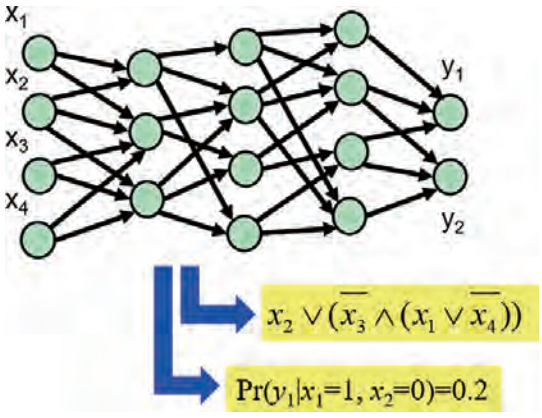


Figure 1. Our developed algorithms extract rules from trained neural networks in the forms of Boolean functions and conditional probabilities.

Grid-based Computational Methods for the Design of Constraint-based Parsimonious Chemical Reaction Networks to Simulate Metabolite Production: GridProd

Constraint-based metabolic flux analysis of knockout strategies is an efficient method to simulate the production of useful metabolites in microbes. Owing to the recent development of technologies for artificial DNA synthesis, it may become important in the near future to mathematically design minimum metabolic networks to simulate metabolite production.

We developed an efficient method for computing the design of minimum metabolic networks by using constraint-based flux balance analysis to simulate the production of useful metabolites. When the growth rate of this obtained parsimonious metabolic network is maximized, higher production rates are observed for many target metabolites when compared to existing methods. The set of reactions used in this parsimonious flux distribution consists of reactions included in the original genome scale model iAF1260. Under the conditions that the growth rate is maximized and the minimum cases of flux variability analysis are considered, the developed method produced more than 90% of metabolites, while the existing methods produced less than 50%.

The source code is freely available and is implemented in MATLAB and COBRA toolbox.

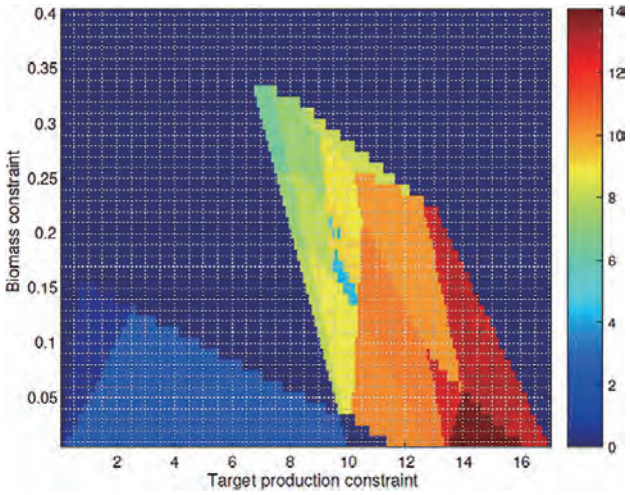


Figure 2. An example of the production rate heatmap for a target metabolite where each grid represents constraints for the ranges by the production rate and the growth rate.