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)

Guest Scholar*
SONG, Jiangning
(Ph D)

*Monash University, Australia, 28 June 2022–25 February 2023

Students

YU, Coleman (D3)
TAKAGI, Motoshige (D3)
LI, Ruiming (D3)
OHTOMO, Masahiro (D3)
NAKASHIMA, Shogo (D3)
MU, Lixuan (D3)
SHIOTA, Koji (D2)
LIU, Chunting (D2)
MA, Yier (D1)
FUJITA, Satoki (D1)
GHAFOOR, Mamoona (D1)
CUI, Yan (M1)

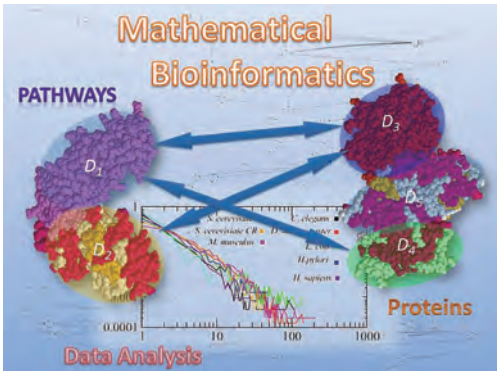
Guest Res Assoc

ZHANG, Bin — South China University of Technology, China, 23 May 2022–17 November 2022
SUN, Liangjie — The University of Hong Kong, China, 10 June 2022–25 August 2022

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

Guo, S.; Liu, P.; Ching, W-K.; Akutsu, T., On the Distribution of Successor States in Boolean Threshold Networks, *IEEE Transactions on Neural Networks and Learning Systems*, **33,** 4147-4159 (2022).
Kumano, S.; Akutsu, T., Comparison of the Representational Power of Random Forests, Binary Decision Diagrams, and Neural Networks, *Neural Computation*, **34,** 1019-1044 (2022).
Münzner, U.; Mori, T.; Krantz, M.; Klipp, E.; Akutsu, T., Identification of Periodic Attractors in Boolean Networks Using a Priori Information, *PLoS Computational Biology*, **18,** [e1009702-1]-[e1009702-27] (2022).
Sugihara, R.; Kato, Y.; Mori, T.; Kawahara, Y., Alignment of Single-Cell Trajectory Trees with CAPITAL, *Nature Communications*, **13,** [5972-1]-[5972-11] (2022).
Tamura, T., L1 Norm Minimal Mode-based Methods for Listing Reaction Network Designs for Metabolite Production, *IEICE Transactions on Information and Systems*, **104,** 679-687 (2021).

On the Compressive Power of Autoencoders Using Linear Threshold Activation Functions

Artificial neural networks have recently been extensively applied to bioinformatics. Among various models of artificial neural networks, autoencoders attract much attention because of their power to generate new objects such as protein sequences and chemical structures. An autoencoder is a layered neural network (Figure 1) consisting of an encoder which compresses an input vector to a lower dimensional vector, and a decoder which transforms the low-dimensional vector back to the original input vector (or one that is very similar). Although it is often mentioned that autoencoders perform dimensionality reduction, a kind of data compression, how data are compressed is not yet very clear. Therefore, we study the numbers of nodes and layers that are required to ensure that each vector in a given set of distinct input binary vectors is transformed back to its original using a autoencoder model with linear threshold activation functions. We show that for any set of distinct vectors there exists a seven-layer autoencoder with the optimal compression ratio, but that there is a set of vectors for which there is no three-layer autoencoder with a middle layer of the same size. We also study the numbers of nodes and layers required only for encoding, and the results suggest that decoding is more difficult than encoding.

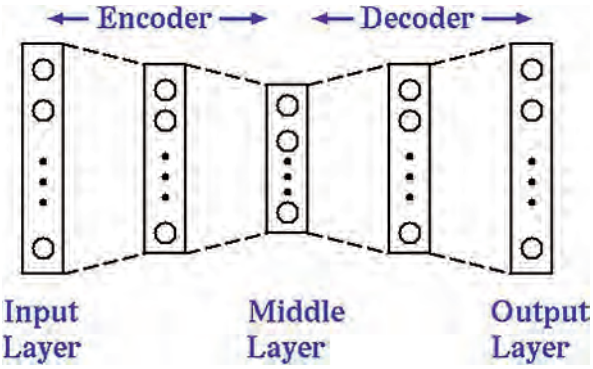


Figure 1. Architecture of an autoencoder.

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

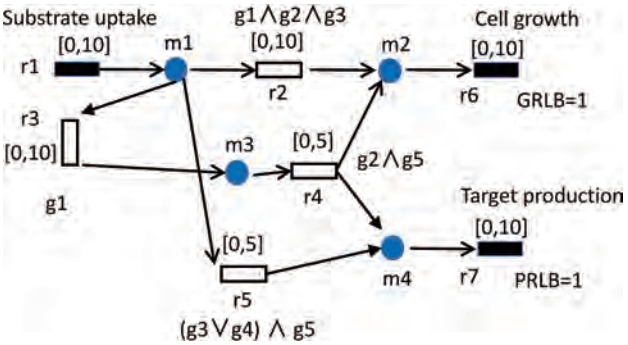


Figure 2. A toy example of the constraint-based model. Circles and rectangles represent metabolites and reactions, respectively. Black and white rectangles are external and internal reactions.