Bioinformatics Center
– Biological Information Networks –

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

Prof
AKUTSU, Tatsuya
(D Eng)

Assist Prof
HAYASHIDA, Morihiro
(D Inf)

Assist Prof
TAMURA, Takeyuki
(D Inf)

Program-Specific Res
KATO, Yuki
(D Eng)

Students

POOLSAP, Unyanee (D3)
KAMADA, Mayumi (D1)
NAKAJIMA, Natsu (D1)
NARITA, Yuki (D1)
SAKUMA, Yusuke (M2)
ZHAO, Yang (M2)
LAI, Yeuntyn (M1)
MAKIGAKI, Shuichiro (M1)
MORI, Tomoya (M1)
RUAN, Peiying (M1)
UECHI, Risa (M1)
CONG, Xiao (RS)
SHI, Dongyue (RS)

Visitor

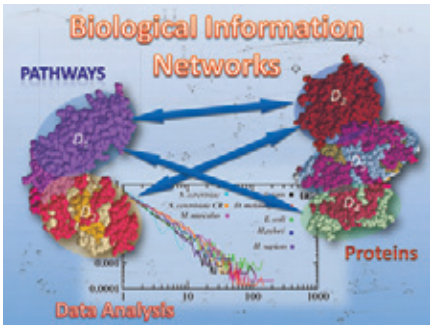
CHEN, Xi The University of Hong Kong, China, P. R., 2 June–13 August

Scope of Research

Due to rapid progress of the genome projects, 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 recently studying the following topics: systems biology, scale-free networks, protein structure prediction, inference of biological networks, chemo-informatics, discrete and stochastic methods for bioinformatics.

KEYWORDS

Scale-free Networks
Boolean Networks
Grammar-based Compression
RNA Secondary Structures
Chemical Graphs



Selected Publications

Akutsu T, Fukagawa D, Takasu A, Tamura T: Exact Algorithms for Computing Tree Edit Distance between Unordered Trees, *Theoretical Computer Science*, **421,** 352-364 (2011).
Hayashida M, Akutsu T: Comparing Biological Networks via Graph Compression, *BMC Systems Biology*, **4 (Suppl. 2),** S13 (2010).
Kato Y, Sato K, Hamada M, Watanabe Y, Asai K, Akutsu T: RactIP: Fast and Accurate Prediction of RNA-RNA Interaction Using Integer Programming, *Bioinformatics*, **26,** i460-i466 (2010).
Akutsu T: A Bisection Algorithm for Grammar-Based Compression of Ordered Trees, *Information Processing Letters*, **110,** 815-820 (2010).
Nacher JC, Hayashida M, Akutsu T: Emergence of Scale-Free Distribution in Protein-Protein Interaction Networks Based on Random Selection of Interacting Domain Pairs, *BioSystems*, **95,** 155-159 (2009).

Conditional Random Field Approach to Prediction of Protein-Protein Interactions Using Domain Information

Exploration of functions and interactions of proteins and domains is important for understanding cellular systems and biological networks. Many methods for predicting protein-protein interactions have been developed. It is known that mutual information between residues at interacting sites can be higher than that at non-interacting sites. It is based on the thought that amino acid residues at interacting sites have coevolved with those at the corresponding residues in the partner proteins. Several studies have shown that such mutual information is useful for identifying contact residues in interacting proteins.

We propose novel methods using conditional random fields for predicting protein-protein interactions. We focus on the mutual information (MI) between residues, and combine it with conditional random fields. In the methods, protein-protein interactions are modeled using domain-domain interactions. Therefore, MI is calculated for each domain pair as the maximum of MIs between residues.

We perform computational experiments using protein-protein interaction datasets for several organisms, and calculate AUC (Area Under the Curve) score. The results suggest that our proposed methods with and without mutual information outperform EM (Expectation Maximization) method proposed by Deng et al., which is one of the best predictors based on domain-domain interactions. Our methods based on domain-domain interactions are useful for predicting protein-protein interactions.

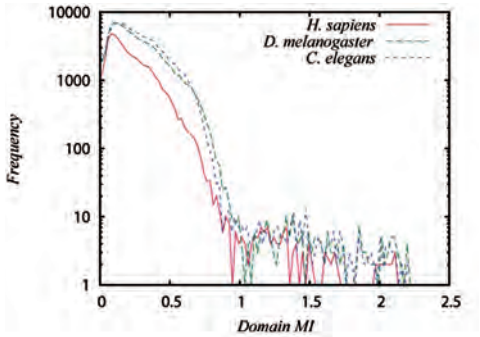


Figure 1. Distributions of domain MIs for *H. sapiens*, *D. melanogaster*, and *C. elegans*.

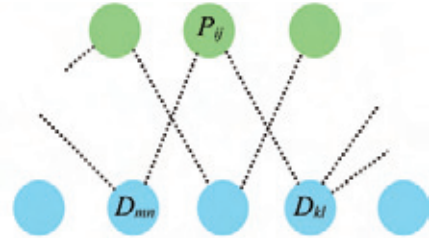


Figure 2. Markov random field model for protein-protein interactions.

Compound Analysis via Graph Kernels Incorporating Chirality

In chemoinformatics and bioinformatics, drug design is one of the main practical and industrial targets. In the development of new pharmaceuticals, in order to find lead compounds, researchers must often select a small subset of compounds from a vastly larger set that satisfies design requirements. After compound selection, high-throughput screening is a method for the synthesis and evaluation of compounds, although it requires considerable time and cost. Therefore, it is advantageous to reduce screening to only those candidates which have been filtered by computational prediction. A major approach to property prediction is to quantitatively analyze the structural features of a compound and find a connection between the target property and features analyzed. This methodology is known as a Quantitative Structure-Activity/Property Relationship (QSAR or QSPR).

In QSPRs, it is very important to achieve high accuracy. Though existing graph-theoretic kernel methods combined with machine learning techniques are efficient for QSPR model construction, they cannot distinguish topologically identical chiral compounds which often exhibit different biological characteristics. In this study, we propose a new method that extends the recently developed tree pattern graph kernel to accommodate stereoisomers. We show that Support Vector Regression (SVR) with a chiral graph kernel is useful for target property prediction by demonstrating its application to a set of human vitamin D receptor ligands currently under consideration for their potential anti-cancer effects.

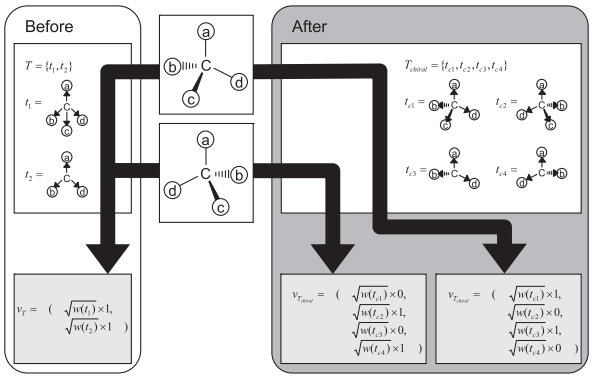


Figure 3. Difference in feature vectors yielded by chirality. Atoms are corresponding to labeled graph nodes, and a bond is corresponding to two oppositely directed graph edges with identical labels indicating the bond order.