

「バイオインフォマティクス・ ジャパン」の紹介

阿久津 達也

特定非営利活動法人バイオインフォマティクス・ジャパン
京都大学 化学研究所 バイオインフォマティクスセンター

講演内容

- NPOの概要
- 活動概要
- 京大バイオインフォマティクス・センターとの協力関係
- 研究紹介(自己紹介)
- 今後の展望

NPOの概要

NPO法人バイオインフォマティクス・ジャパンの概要

- **設立**: 2009年12月
- **初代理事長**: 金久實教授(現:京大名誉教授、特任教授)
- **現理事長**(2020年7月より): 阿久津達也
- **趣旨** (<https://www.bioinformatics.jp/ja/summary.html>)
 - ライフサイエンスに関する知識ベースの構築とゲノム解読に必要な知識処理技術の開発
 - 我が国独自のゲノム関連知識情報インフラを整備して、科学技術の振興及び経済の発展と社会の福祉に寄与
 - 以前はKEGGのアカデミック向けライセンス事業を行っていたが、現在はPathway Solutions社に移行
 - 現在は、京大化学研究所バイオインフォマティクスセンターの協力のもと、**バイオインフォマティクスの振興・普及**を中心に活動
- **現在の役員構成** (<https://www.bioinformatics.jp/ja/board.html>)
 - 理事長、副理事長、理事6名、監事1名

活動概要

活動概要(1)

- **理事会**: 毎年4月に開催
- **総会**: 毎年5月に開催
- **日本バイオインフォマティクス学会年会**(生命医薬情報学連合大会)での**ワークショップ主催**
 - 2020年(<https://www.jsbi.org/iibmp2020/program/workshop.html#ws2-3>)
「スーパーコンピュータとバイオインフォマティクス」
 - 2021年(<https://iibmp2021.hamadalab.com/workshop/>)
「人工知能と生命誌に基づく生命医科学のためのバイオインフォマティクス」
- **環境ウイルス研究集会**(微生物生態学会 環境ウイルス研究部会)の協賛
 - 2020年、2021年(<http://www.microbial-ecology.jp/?p=10204>)

活動概要(2)

- シンポジウム「**新たな創薬パラダイムの創出**」をNPO
法人**システム薬学研究機構**と共催
 - 2013年度(第3回)より毎年共催(<https://www.bioinformatics.jp/ja/events.html>)
 - 2020年度、2021年度はコロナ禍のため**中断**
- **バイオインフォマティクスセミナーシリーズ**
 - 京大バイオインフォマティクスセンターと一部共催
(<https://www.bic.kyoto-u.ac.jp/egis/seminar.html>)
- 講習会「**バイユーザのための化研スパコンシステム
利用法**」
 - 京大バイオインフォマティクスセンターと共催(https://www.scl.kyoto-u.ac.jp/Training/lecture/Hsem210521_Bio.html)

京大バイオインフォマティクス
センターとの協力関係

バイオインフォマティクスセンター

- 京都大学 化学研究所 附属 (所在地:京大宇治キャンパス)
- 2001年4月設立 (https://www.bic.kyoto-u.ac.jp/index_J.html)
- 初代センター長: 金久實教授(現特任教授・名誉教授)
- 化学研究所のスーパーコンピュータ運営に協力
 - HPE社製: Superdome Flex(576core) × 2 + Apollo 2000(142node)
- NPOとの協力関係: セミナー共催など

組織構成

- 化学生命科学研究領域(緒方研究室)
 - メタゲノム、巨大ウィルス
- 数理生物情報研究領域(阿久津研究室)
 - アルゴリズム、複雑ネットワーク
- 生命知識工学研究領域(馬見塚研究室)
 - 機械学習、データマイニング
- KEGGデータベース開発グループ(金久ラボラトリーズ)
 - KEGGデータベース開発

KEGGとゲノムネット

- KEGG (Kyoto Encyclopedia of Genes and Genomes)
 - 金久教授らが1995年より開発する生命化学情報データベース (https://www.genome.jp/kegg/kegg_ja.html)
 - 毎日、世界中から多数のアクセス
 - 多種類・大量の情報 + 検索ツール
 - 代謝パスウェイ、化合物、化学反応式、糖鎖
 - 医薬品・医薬情報、遺伝子分類情報
 - マッピングツール、類似化合物検索ツール
- ゲノムネット (GenomeNet)
 - ライフサイエンス統合データベース
 - 金久教授らが1991年より開発
 - バイオインフォマティクスセンターが提供
 - 独自データ + 海外データベースのミラーリング
 - DBGETシステムによる統合検索を提供

研究紹介 (自己紹介)

研究分野・目的

■ 研究分野

- **バイオインフォマティクス** (30年以上にわたり研究)
- 複雑ネットワーク、離散アルゴリズム、ニューラルネットワーク

■ 職歴

- 機械技術研究所(現産総研の一部)⇒群馬大学情報工学科⇒東大医科
研ヒトゲノム解析センター⇒京大化研バイオインフォマティクスセンター
(2001年10月より)

■ 目的

- **生命の数理的理解**
 - ヒトの設計図 \asymp 30億文字 \asymp CD-ROM一枚+ α
 - 数理的、情報科学的な原理があるはず!
- **数理的原理に基づく生命情報解析手法の開発**
 - 高校数学+ α で先端的な数理的研究が可能

このSection(展望の前まで)は阿久津個人の研究に関するものです

著書・研究テーマ

■ 主要著書

- バイオインフォマティクスの数理とアルゴリズム、共立出版、2008
- Algorithms for Analysis, Inference, and Control of Boolean Networks, World Scientific, 2018

■ 研究テーマ例

- 生体ネットワーク制御のための制御理論構築
- ニューラルネットワークの離散数理
- データ圧縮による生命の構築原理の理解
- 化学構造の情報解析・設計手法の開発
- RNA・タンパク質の配列・構造解析手法の開発
- グラフ(木構造)パターンマッチング・アルゴリズム

■ 研究の指針

- なるべく競争・競合は避ける
 - 重要と思われるがあまり研究されていないトピックを研究
 - 流行の課題でも、できるだけ別の視点から研究
 - 温故知新: 昔に流行したが、すたれてしまったトピックを研究

生体ネットワークの制御(1)

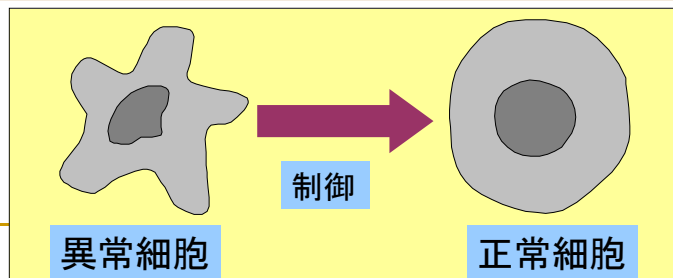
■ 研究目標: 生命システムの制御理論の構築

- 線形制御理論は確立しているが、生命システムは大規模かつ非線形
- 生命システムに対する制御理論を構築することが必要
- 創薬やがん治療などにつながる可能性がある
- 理論の構築に少しでも貢献したい

■ 現実が理論に先行: iPS細胞、ゲノム編集

■ 定式化(例)

- 入力: 細胞の初期状態と目標状態
- 出力: 制御すべき遺伝子と、その遺伝子へのシグナル

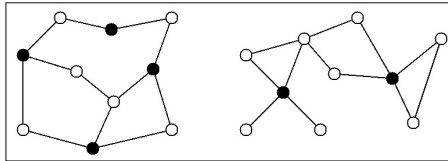


生体ネットワークの制御(2)

■ 最小支配集合を用いた制御

- 支配集合: どの頂点も、それに属するか、隣接
- 最小支配集合に対応する頂点(遺伝子)のみの制御で、細胞全体を制御可能
- 実際には制御すべき頂点が多すぎて、直接は応用できない
- しかし、重要な遺伝子の検出には有効である可能性

最小支配集合の例



Nacher & Akutsu: New J. Physics, 2012

■ ブーリアンネットワークの制御

- ブーリアンネットワーク(BN): 遺伝子ネットワークをデジタル回路で表現
- 目標状態が任意 ⇒ 多数の遺伝子の制御が必要
- 定常状態のみが目標状態 ⇒ 少数の遺伝子の制御で十分
- ただし、BNが単純化されすぎたモデルなので直接の応用は不可

Hou, Ruan, Ching, Akutsu: J. Theoret. Biol. 2019

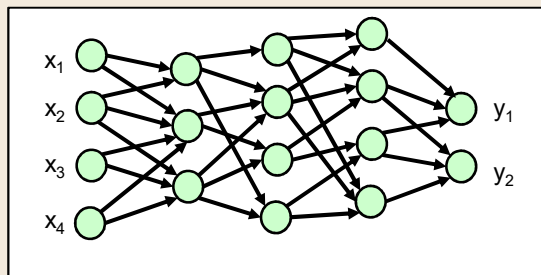
ニューラルネットワークからの知識抽出

■ 深層学習

- 予測精度は良いが、学習結果が人間にわかりづらい
⇒ 学習したニューラルネットワークからの知識抽出法(white box 化、見える化)に関する多数の研究

■ 我々の研究

- 学習したネットワークからのブール関数の抽出、確率的規則の抽出 [Liu, Melkman, Akutsu: Neural Networks 2020]



$$\rightarrow x_2 \vee (\overline{x_3} \wedge (x_1 \vee \overline{x_4}))$$

$$\rightarrow \Pr(y_1 | x_1=1, x_2=0) = 0.2$$

自己符号化器のデータ圧縮能力の理論解析

- **自己符号化器**: データ圧縮を行うニューラルネットワーク
- **頂点数、層数とデータ圧縮率**の関係を**離散モデル**(線形閾値関数モデル≡多層パーセプトロン)のもとで理論的に解析

d	architecture	type	constraint
$\log n$	$D/d/D$	Encoder/Decoder	BN
$\lceil 8\sqrt{2M \ln n} \rceil$	D/d	Encoder	BTN
$2\lceil \log n \rceil$	D/d	Encoder	BTN with parity
$2\lceil \log n \rceil$	$D/D^2/d$	Encoder	BTN
$\lceil \log n \rceil$	4 layers ($O(\sqrt{n} + D)$ nodes)	Encoder	BTN
$2\lceil \sqrt{n} \rceil$	$D/(\frac{d}{2} + D)/d/\frac{dD}{2}/D$	Encoder/Decoder	BTN
$\lceil \log n \rceil$	$D/n/d/n/D$	Encoder/Decoder	BTN
$2\lceil \log \sqrt{n} \rceil$	7 layers ($O(D\sqrt{n})$ nodes)	Encoder/Decoder	BTN

D : 入出力データの次元, d : 圧縮データの次元, n : サンプルデータの個数

BTN: Boolean Threshold Network(≡多層パーセプトロン)

BN: ブーリアンネットワーク

Melkman, Akutsu: IEEE Trans. Neural Networks and Learning Systems, in press

今後の展望

NPOの今後の展望

- 現在の活動内容の継続・発展
- 他NPO法人との連携強化
- 会員数の増加
- ソフトウェア開発
- シニア研究者・技術者の交流・発表の場の提供

阿久津の個人的展望です