



# データの交わりに隠れた未知の知識を発見する ——無限の仮説を考慮して生体现象を解釈する AIモデルと高信頼医療ヘルスケアへの展望

近年さまざまな生体情報のデータ解析によって医療ヘルスケアの推進につながる数多くの知見が得られるようになってきました。一方で、人々の生命や健康と密接にかかわる分野でのAI（人工知能）の利用においては、特にその解析・予測・判断におけるリスク・不確かさの評価が重要な課題として注目されるようになってきています。本稿では、モデルが生体现象を説明する際に無限に存在し得る仮説を考慮することで、データ解析のリスク・不確かさを明示的にとらえさせようとする機械学習手法について解説します。

キーワード：#バイオ医療ヘルスケア、#データ解析、#組合せ論

## 未知の知識を獲得するための 機械学習

人工知能（AI）は私たちの生活に不可欠なものとなってきています。音声・音楽・画像・動画・テキスト・ゲームなど各種メディア制作や、動画配信サービス・商品推薦システムのバックエンドなど、さまざまな場面にAI技術が活用されています。特に近年の大規模言語モデルや生成AIの発展はめざましく、かつてAIの概念の黎明期にサイエンスフィクションとして描かれてきたような世界、例えば人類とAIが共存するような世界がすでに現実のものとなりつつあります。実際に、自然・社会に現れる多くの諸課題に対して、AIは私たち人類と遜色ない能力を発揮するどころか、場合によっては人類を超えるようなことも少なくありません。AIのさらなる発展は、おそらく私たち人類の生活をより豊かで便利なものへと導いてくれるでしょう。では具体的に、AIのさらなる発展にはどのような方向性があるのでしょうか。そして、その発展は今後の私たちの生活にどのような変革をもたらしてくれるのでしょうか。

NTTコミュニケーション科学基礎研究所では、AIのさらなる発展に関する重要な方向性の1つとして、「未知の知識の推論」能力に注目しています。この「未知の知識」という言い回しは一見すると逆説的な表現ですが、私たちは、実際にこのような能力

を養い・鍛え・発揮してきたことによって、現代の社会にまで発展してきたといっても過言ではありません。私たちは現実の世界で起こった現象から仮説を導き出し、さまざまな「未知」の現象に対しても想像を巡らせることで、あたかもそれらを「知識」の1つであるかのようにとらえることができます。そして、膨大な年月をかけて培ってきた知識や経験をもとに、「未知の知識」に対してもその確かさ・不確かさを「推論」することができます。例えば、かつて伝染病や未開の病原体は一度流行してしまうと、特定の地域に対して長期間甚大な被害を出し続けてしまいましたが、それに対して現代では新種の感染症や未知の病気に対しても、（ある程度のリスクを含んでいたとしても）有用な対処法や治療法が圧倒的な短期間で見出されるようになってきています。

人類は長い歴史の中でさまざまな経験を蓄積してきたことにより、このような統計的推論が可能となってきました。ただし、歴史が示すとおり、その経験の蓄積には膨大な年月を要します。そこで、私たちの研究チームでは、このような「未知の知識を推論」する能力をAIに実装し、人間を超えるような豊かな想像力とその統計的推論の高精度化を実現することで、私たちの生活に重大な変革をもたらされるかもしれないと期待しています。例えば、未知の新種の感染症の発生リスクを予測することに

なかの まさひろ  
中野 允裕

NTTコミュニケーション科学基礎研究所

よって世界的流行・パンデミックを未然に防ぐことができるかもしれません。あるいは近い将来、地球外生命体によって地球へ運ばれてくるかもしれない未知の病原体への対応手段をあらかじめAIが予測し準備することで、人類滅亡の危機を未然に回避することができるかもしれません。

本稿では、「未知の知識」の表現形式の中でも、特に構造的仮説に焦点を当てていきます。ここでいう構造的仮説とは、順列・分割・バイナリ木・カンブリア木・バイナリ系列・因子グラフ・長方形分割など、知識の効果的・効率的な表現を実現する離散構造のことを指しています。これらは、バイオ医療ヘルスケア応用を展望した生体情報処理の諸課題において頻出する未知の知識の表現法として特に有用です。例えば、ランキング・マッチング課題では「順列」が重要な役割を果たし、大規模データに対する検索システムに備わるランキング機能に活用されています。分類・クラスタリング課題では「分割」が活躍し、大規模データの中の隠れたクラスタ構造を発見するのに役立ちます。階層クラスタリング課題には「バイナリ木」が活用され、データの中に隠れたクラスタ構造の階層を発見するのに有用です。系統樹解析課題には「カンブリア木」が用いられ、例えば遺伝子の変異・融合・組み換えなど生体の系統的なイベント発生軌跡をとらえるために活用されています。このように、構造的仮説はデー

タの中に隠れた未知の知識を表出させるための有用な道具として注目されてきています。ここからは、構造的仮説を未知の知識の表現法とするAIについて、いくつかの応用事例とともに、それらを汎化する方法、そしてバイオメディカルヘルスケア応用への展望を紹介していきます。

### 関係データ解析

「不定愁訴」とは、患者が何らかの体調不良を自覚しているにもかかわらず、その医学的な原因を特定することができずに適切な治療処置を提供することができない事象を指しています。日本においては特に1960年代からこのような事象の発生が顕著になってきています。ここでは、不定愁訴をデータ駆動的に解釈する1つの有力な方法として関係データ解析と呼ばれる一手法を概説していきます。簡単のため不定愁訴にまつわる観測データを図1に示すような二値の行列で表すことにします。これは患者群を行（垂直方向）に、症状群を列（水平方向）に対応付けて、ある1つの要素はその行の患者がその列の症状を持っているかどうかを二値（図では白黒）で表現した行列データです。図1(a)のように、不定愁訴を表す観測データは一見すると無秩序の

ように見えるかもしれませんが、図1(b)のように、患者群の並び順と症状群の並び順を適切に入れ替えて、かつ赤い長方形分割の補助線を加えることで、ある種のクラスタ構造を表出させることができます。この赤い補助線として示している長方形分割は、各ブロック内になるべく似た要素を集まるように誘導しています。したがって、各ブロックは特定の患者群と特定の症状群の共通した類似性を暗示しているものともとらえることができます。これは広義の意味において、不定愁訴の中に表れる新しい病気を示唆し得るものともとえられるかもしれません。応用上は、非常に単純化した例として、各ブロックのある患者に対して有用であると分かった治療・処置が、同一ブロックに属する別の患者に対する有効な治療法に対する手掛かりとなり得るかもしれません。

関係データ解析の文脈における「未知の知識」とはこの場合、観測データを解釈するための長方形分割とらえることができます。そこで関係データから長方形分割仮説を推論するためのAIモデルの構築が研究されてきました。AIの力を借りてさまざまな長方形分割仮説に対してその確かさ・不確かさを推論することによって、不定愁訴を解消する手掛かりが得られることを期

待しています。では、このようなAIモデルの構築に対して求められるもっとも重要な要件とは一体何なのでしょう。その重要な要件の1つは、ありとあらゆる長方形分割仮説をAIが漏れなく想像できることにあります。不定愁訴のようなメディカルヘルスケア応用を見据えた場合、もしもAIが想像することすらできない仮説があり、その仮説の中に真実があったのだとすれば、そのAIによる推論は真実を見逃すことで結果的に私たちの生命・健康に重大な影響を及ぼしてしまうかもしれません。

このような動機から、関係データ解析の発展において、ありとあらゆる長方形分割仮説を漏れなく想像し、その確かさ・不確かさを推論することができるようなAIの構築が探求されてきました。歴史的には、2007年に正則分割と呼ばれるクラス、2009年に階層分割と呼ばれるクラスについて、それらクラスに属するすべての長方形分割を想像し得るAIモデルの構築に成功してきました。しかし、これら正則分割・階層分割は長方形分割全体の中のごく一部に制限されたクラスであり、依然としてあらゆる長方形分割を想像し得るAIモデルの構築にはいたっていませんでした。私たちの研究チームの貢献は、2020～2021年にかけて、2つの方式によってあらゆる長方形分

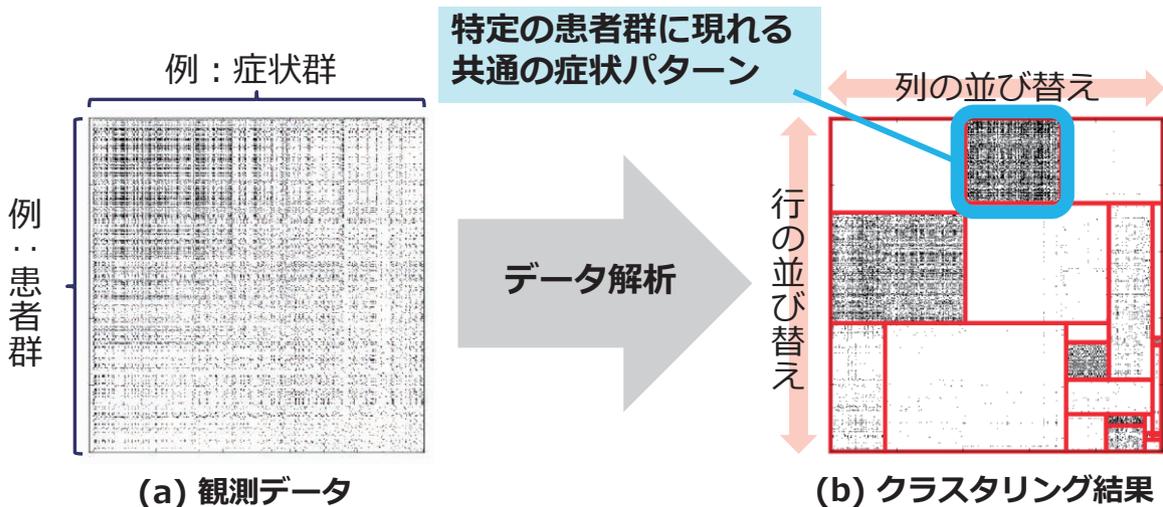


図1 不定愁訴応用を見据えた関係データ解析

割を想像し得るAIモデルの構築に成功したことにあります<sup>(1), (2)</sup>。

この成功は、組合せ論の知見に基づいて、あらゆる長方形分割の集合が、ある特定の順列のクラスと一対一対応関係を持つ事実を用いて、一見するとAIが直接想像することが容易でない長方形分割を、順列（つまり数字の並び順）というAIが取り扱いきやすい表現で間接的に想像する機構を導入したことによるものです。私たちは、実世界のいくつかのベンチマークデータをとおして、従来のように制限されたクラスにとどまらずにあらゆる長方形分割を想像し得る能力がAIモデルの高機能化・高信頼化に寄与することを確認してきました。この成果によって、AIにあらゆる長方形分割を漏らすことなく想像し得る能力を保証することができるようになりましたので、今後はその推論能力をさらに高精度化していくことにより、実社会の不定愁訴に対して関係データ解析の知見から、その解消につながる重要な洞察を見出ししていきたいと展望しています。

## 系統樹解析

細胞の分化や遺伝子の変異のように、時間経過とともに系統的な進化・発達構造を持つ対象からその系統樹を推論する技術は系統樹解析と呼ばれ、さまざまな応用事例において重要な役割を果たしてきています。

例えば近年、人工多能性幹細胞（iPSC：induced Pluripotent Stem Cells）の分化構造に対する系統樹解析は特に注目を集めている応用事例の1つといえます。iPS細胞はある種の分化万能性を持つことが知られており、体のさまざまな組織・器官・臓器へと分化誘導できる可能性が期待されています。そのため、その分化の原理を理解し適切に誘導することができるようになれば、患者から採取した体細胞をiPS細胞化し培養することで、移植用臓器・組織・器官を作成できるようになるかもしれません。また、従来採取が困難であった病変組織の細胞をつくり出すことによって、その発症原理の理解や有効な治療・薬剤の解明に役立てられる可能性も期待されています。別の観点からは、昨今の感染症の世界的流行・パンデミックにおいても、系統樹解析は発生源の推論や感染拡大を防ぐ手掛かりを見つげるための有用な道具として活用されています。一般に、感染症の持つ遺伝子は、その拡大の過程で変異・融合・組み換えなどの事象を経て系統的な変容を遂げていきます。そのため、その発展の系統樹を辿ることができれば、その発生源・感染経路への手掛かりとすることができるかもしれません。

系統樹解析の文脈における「未知の知識」は、図2に示すとおり、遺伝子の系統的な変異をカンブリア木として表すことができます。これは、親遺伝子が2つの子遺伝子

へ分岐するイベントと、2つの親遺伝子から1つの子遺伝子へ組み換えが起こるイベントを用いて、遺伝子の系統的な変異を系統樹として表しています。このような系統樹は一般に膨大な仮説を取り得ることから、AIの力を借りて系統樹を推論することのできる技術が期待されてきました。では、そのようなAIに求められるもっとも重要な要件とは何でしょうか。不定愁訴に対する長方形分割仮説と同様に、私たちはデータの中に隠れた系統樹、すなわちありとあらゆるカンブリア木構造を漏れなく想像し得る能力が重要であると考えています。それでは、ありとあらゆるカンブリア木仮説を漏れなく想像し、その確かさ・不確かさを推論することができるようなAIはどのようにして実現することができるでしょうか。興味深いことに、前述の関係データ解析における長方形分割と順列の関係とよく似た事実として、あらゆるカンブリア木の集合は特別な順列の集合と一対一対応関係を持っていることが組合せ論の分野において知られています。この事実を利用して私たちは、順列を介して間接的にあらゆるカンブリア木を想像し得る能力を持ったAIモデルの構築に成功しました<sup>(3), (4)</sup>。私たちは、このようなあらゆるカンブリア木を想像し得るAIモデルを用いてiPS細胞の分化構造を推論し、そのメカニズムを明らかにすることで、今後の再生医療の発展につながる重要な洞察を見出ししていきたいと展望して

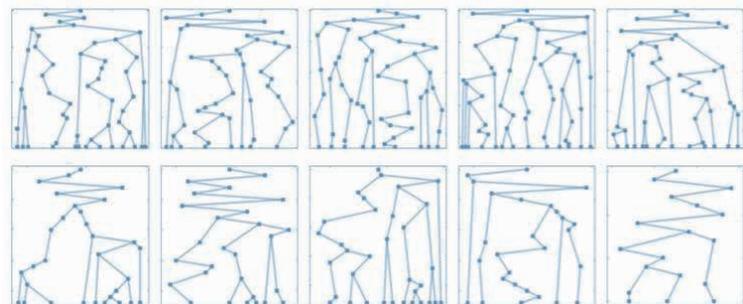
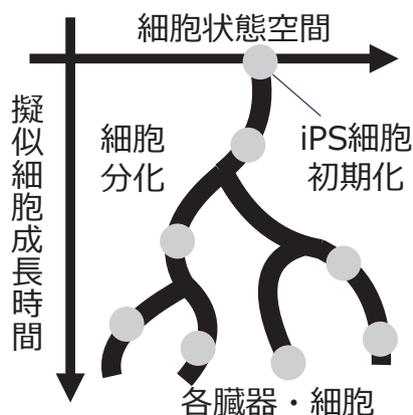


図2 細胞分化構造推論を見据えた系統樹解析

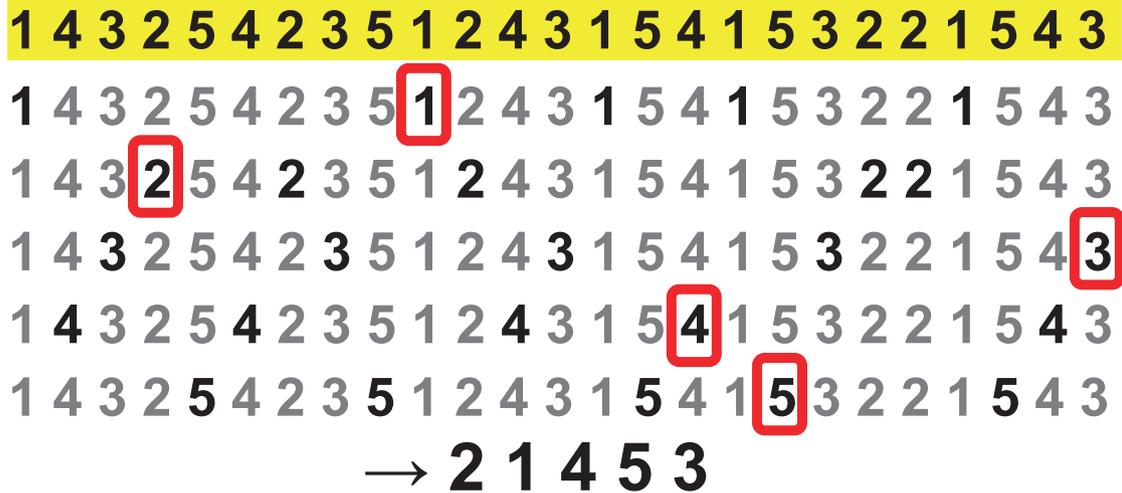


図3 スーパーペイズ法のご概念図

います。

今後の展望：スーパーペイズ構想

ここまで関係データ解析と系統樹解析の2つの応用事例に対して、それぞれ長方形分割とカンブリア木を「未知の知識」としてそれらを漏れなくAIに想像させ得る方法について紹介してきました。それでは、別の課題に対して他の構造的仮説をAIに想像させたい場合にはどうすればよいでしょうか。改めてそのような機構を1から構築していかないといけないのでしょうか。私たちはこのような対症療法なAIモデルの個別実装とは真逆のアプローチとして、対象とする課題に対して、AIが自律的に適切な構造的仮説を見出して想像できるような機構の実現をめざしています。このような構想を、私たちの研究チームでは「スーパーペイズ法」と呼んで2022年から提唱しています<sup>(5)</sup> (図3)。

これは、代数的組合せ論と極値組合せ論の知見をペイズ統計へ輸入したものととらえることができます。より具体的には、代数的組合せ論において、順列・分割・バイナリ木・カンブリア木・バイナリ系列・因子グラフ・長方形分割などのさまざまな構造的仮説は、順列からの全射または全単射

によって統一的に表現できることが分かってきています。そして、極値組合せ論において、「超順列」と呼ばれる冗長な数列はその部分系列にあらゆる順列を含んでいることが知られています。図3は長さ5の順列を含む超順列の例を表しています。あらゆる順列を含む「超順列 (superpermutation) 」になぞらえて、私たちはこのような構想をスーパーペイズと呼び、私たちユーザがその構造的仮説の表現形式をあらかじめ指定することなく、AIモデルが対象とするデータからデータ駆動的にその表現形式自体を推論することができる枠組みへ拡張をめざしています。

最近2025年になって私たちの研究チームではその理論基盤が整備できましたので、今後その本格的な実装に取り組みようとして始めています。私たちは「未知の知識を推論するAIをとおして高信頼メディカルヘルスケアの発展に貢献できるよう、努力を惜しまずに研究を進めていきます。

参考文献

- (1) M. Nakano, A. Kimura, T. Yamada, and N. Ueda: "Baxter Permutation Process," Proc. of NeurIPS 2020, pp. 8648-8659, 2020.
- (2) M. Nakano, Y. Fujiwara, A. Kimura, T. Yamada, and N. Ueda: "Permuton-induced Chinese Restaurant

Process," Proc. of NeurIPS 2021, pp. 27695-27708, 2021.

- (3) M. Nakano, D. Chijiwa, R. Shibue, Y. Fujiwara, R. Nishikimi, T. Iwata, A. Kimura, T. Yamada, and N. Ueda: "Permuton Process," OpenReview, 2024.
- (4) M. Nakano, H. Sakuma, R. Nishikimi, R. Shibue, T. Sato, T. Iwata, and K. Kashino: "Warped Diffusion for Latent Differentiation Inference," Proc. of AISTATS 2024, pp. 4789-4797, 2024.
- (5) M. Nakano, R. Nishikimi, Y. Fujiwara, A. Kimura, T. Yamada, and N. Ueda: "Nonparametric Relational Models with Superrectangulation," Proc. of AISTATS 2022, pp. 8921-8937, 2022.



中野 允裕

真実はいつも1つ。これは確かに真実かもしれませんが、しかし、解釈はいつも多重の有力な候補があり得るものです。メディカルヘルスケアにおけるAIの活用には、このような多重の仮説を陽にとらえられる能力が重要となってくるかもしれません。

◆問い合わせ先

NTTコミュニケーション科学基礎研究所  
 メディア情報研究部  
 生体情報処理研究グループ