



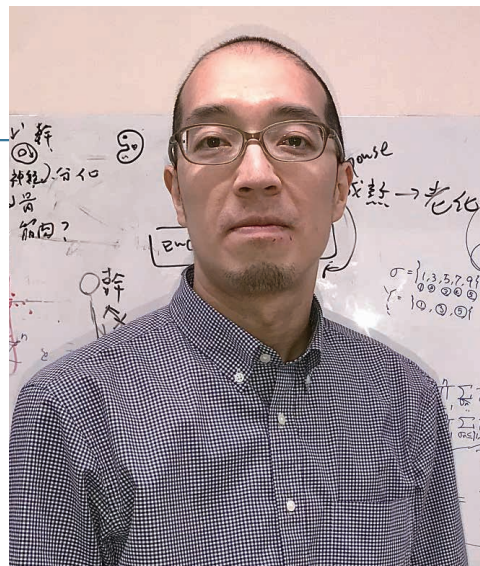
NTTコミュニケーション科学基礎研究所  
NTT物性科学基礎研究所  
特別研究員

中野 允裕 Masahiro Nakano

## 増え続ける無限のデータを解析するための「ノンパラメトリックベイズ法」

現在注目を集めている研究の1つが、解析データを実社会の応用課題へ活用する「統計的機械学習」です。従来の解析では、統計モデル・学習モデルの規模やパラメータを与えられたデータに応じて人手で設定するか、高コストなチューニングを行う方法が主に用いられていました。しかしこれらの方法では膨大な時間と労力がかかり、今後も肥大を続けるデータを分析することは困難です。そこでこうした課題を解決するため、モデルの規模やパラメータをデータの質や量に応じて自動的に調整する新たな手法が強く求められています。今回は中野允裕特別研究員に、現代データ社会が抱える課題解決に向けた「ノンパラメトリックベイズ法」の研究についてお話を伺いました。

◆PROFILE：2011年東京大学大学院情報理工学系研究科システム情報学専攻修士課程修了。同年日本電信電話株式会社入社。NTTコミュニケーション科学基礎研究所配属。2020年よりNTTバイオメディカル情報科学研究センタ兼務。確率過程（無限次元のパラメータ空間を持つ確率モデル）や極値組合せ論的普遍対象を用いた統計的機械学習法とその生体情報処理への応用に従事。



### 無限サイズの関係データを解析するための新たな手法を確立

■はじめに、「ノンパラメトリックベイズ関係データ解析」とはどのような技術でしょうか。

現在の社会におけるデータの容量・規模は年々増大し続けており、データの種類も多様多種となっています。こうした環境の中で、従来では人手によってモデルのパラメータを都度適切に設定して、各種の処理を行っていました。しかしこの手法では膨大な時間と労力を要してしまい、データ社会のさらなる発展に対応することは困難です。そこで、これらの諸問題を解決するのが、今回紹介する「ノンパラメトリックベイズ法」です。この技術は、統計モデル・学習モデルがモデルの規模やパラメータをデータ駆動的に（データの質や量に応じて自動的に）調整することで、従来課題であった人手による膨大な時間と労力を削減しています。

「ノンパラメトリックベイズ法」が活躍する1つの応用事例として、「関係データ解析」という技術を挙げることができます。ここでは、現代における不定愁訴の問題を例にその活用事例を考えていきます。不定愁訴とは、心身不調に対する自覚症状を訴えて検査しても原因となる病気が特定されず、有効な治療法が分からない状態のことです。その症状は多岐にわたり、さらに「体が

だるい」「疲れが取れない」「腕が重い」など複合的に現れることもあるため、患者個人単位で病気の特定は困難で社会課題になっています。しかし、もしこのような多くの患者と多くの症状の間関係データを集めてデータ分析を行えば、それぞれのデータの相関が明らかになり、データサイエンス的に何らかの共通の特徴を見て取ることができるのです。

しかしこの解析を行うにあたって大きな問題がありました。それは収集できるデータの数が無限（に近い）の場合に対処が難しいことです。例えば自覚症状の項目数は、人によってそれぞれ異なるものも含めるとかなりの数になります。また潜在的に不定愁訴となり得る患者の数も計りしれません。このように、患者と症状の関係データは、表形式の行方向と列方向に無限に大きくなっていく可能性があり、データを解析することは困難を極めます。

そこでこうしたデータを潜在的に解析するため、データ背後に潜む無限の可能性への「不確かさ」をデータ駆動的に行う技術が、2000年に創出された機械学習法の一分野である「ノンパラメトリックベイズ法」です。この技術の具体的な成果として、「遺伝子と病気の関係性の発見」というようなデータをクラスタリング（データ間の類似度に基づいてデータをグループ分け）解析する際に、データを長方形に分割し、その分割候補のあらゆる組合せパターンを生成できる新しい確率的生成モデルを構築しました（図1）。

## 例：遺伝子と病気の関係性の発見

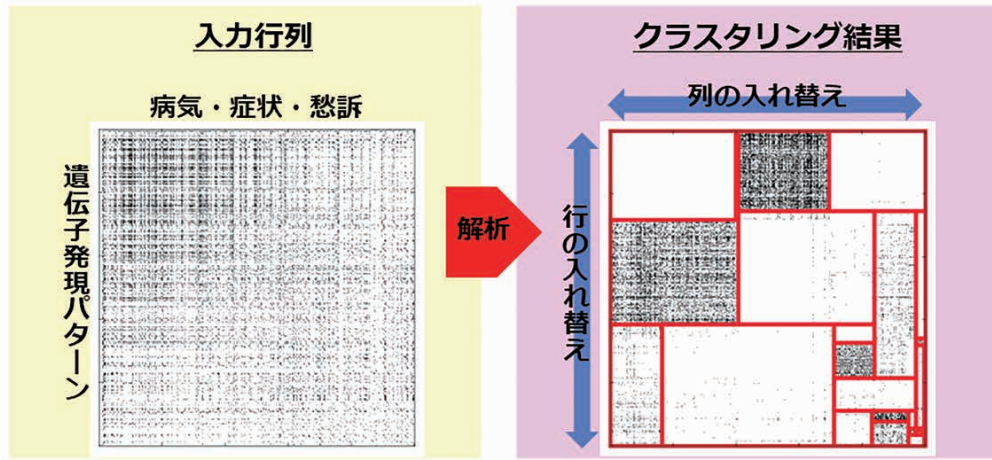


図1 ノンパラメトリックベイズ関係データ解析の概要

さらに最適な長方形分割をデータに合わせて調整しクラスタリングすることで、効率的なデータ解析方法を実現しました。これによって、従来では解析不可能だった「無限に近い」あらゆるデータを、データの質と量に合わせて解析可能にしています。

### ■そのほか、どのような技術をご研究に取り組まれているのでしょうか。

前述の技術と並行して、2022年から「スーパーベイズ関係データ解析」という研究を進めています。この技術を簡単に説明すると、ノンパラメトリックベイズ関係データ解析の「無限を扱う」という特徴を緩和したものです。

実はノンパラメトリックベイズ関係データ解析の実現に向けた最大の障壁は、推論アルゴリズムの構築でした。なぜなら「無限」を計算機上で積極的に活用することによって、データ解析に用いられる推論アルゴリズムの構築には「無限」が付きまとうことになり、それによりある種の無限ループの可能性を回避しなければならず、解析を行うモデルの設計難度が高くなってしまいます。これはノンパラメトリックベイズ関係データ解析における推論アルゴリズムで、確率的操作を用いることに起因する問題です。一般的に確率は、候補となる事象の発生確率を合計して常に100%とならなければなりません。しかし「無限の場合」の数では、合計が確率100%であったとしても、それぞれの場合の数そのものの確率は0%となります。つまり全体の「無限の場合の数」の中から、部分的な「無限の場合の数」を集めることで正の確率を持っているのです。そのため「確率0%の事象だから無視できる」と油断していると、知らず知らずのうちに無視した事象の確率が正となっていき、合計が100%とならずに破綻してしまうおそれが出てきてしまいます。

私は研究の道に進み始めてからすぐに、「無限を仮想的に扱う

ことができる」というノンパラメトリックベイズ関係データ解析の魔法のような売り文句に魅力を感じて惹かれていました。当時2000年代はちょうどノンパラメトリックベイズ法の第一次ブームのような時期でもあり、毎日のように新しい技術が出てきてとても刺激でした。しかし2012年ごろにAI（人工知能）が盛り上がりを見せ、第三次ニューラルネットワークブームが起これ、それとともにノンパラメトリックベイズの盛り上がりは相対的に落ち着いてきてしまったように思います。個人的な想像としてその原因は、「実用化に耐え得る技術ではなかった」の一言に尽きると 생각합니다。ノンパラメトリックベイズ法は、その最大の特徴である「無限を仮想的に扱う」能力のために、計算機上かなり取り扱いが難しいモデルになってしまったのです。このように実応用の観点で見たときに、ノンパラメトリックベイズ関係データ解析の構築はしばしば難度が高く、まずはすでに確立されている深層学習に基づいた方法を試してみるのが一番有望であるといっても過言ではありません。

しかし一方で、当時2012年ごろから第三次ニューラルネットワークで興隆を極めていた深層学習のように、何か構築度の容易さと実応用性が両立できる方法がないかを模索していました。そのブレークスルーのきっかけとなったのは、2021～2022年に出会った「極値組合せ論」です。

極値組合せ論とは、ある種の冗長さの中に生まれる秩序を扱う学問です。例えばとてつもなく長い数字の列（より正確に言えば、一様ランダムな順列）を考えると、その部分系列としてありとあらゆる短い順列（例：14523, 8245361）が高い確率で現れることが知られています。この極値組合せ論から得られた重要な洞察は、「十分に冗長で巨大なものを用意すれば、それは十分すぎるほどいろいろなものが表現できる」ということです。つまりノンパラメトリックベイズ関係データ解析においては「無限」が必要だと

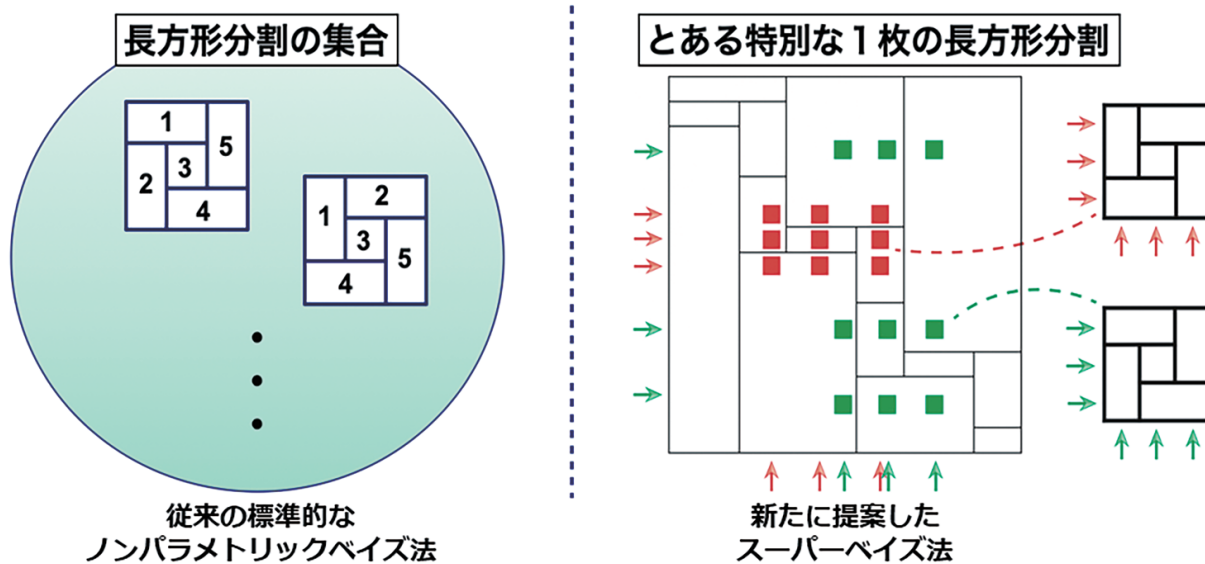


図2 ノンパラメトリックベイズ関係データ解析とスーパーベイズ法の比較

思われていたものを、もしかすると「十分に冗長で大きなもの」に置き換えることで代替できるのではないかと気が付きました（図2）。そしてこの「十分に冗長で大きなもの」というのは無限が絡むわけではないため、推論アルゴリズムの構築がずっと簡単なものになってくれる可能性を秘めています。このようにスーパーベイズ法は、「冗長性」という洞察を得たことによって、ノンパラメトリックベイズ関係データ解析の強みを持ちながら実応用に適した技術になり、これからの発展を導いてくれるのではないかと期待しています。

ノンパラメトリックベイズ関係データ解析の新しいブランチとして可能性を見出しました「スーパーベイズ法」の研究は、当該分野の重要国際会議 AISTATS2022にて Oral 発表 (Top 約 4%) の機会をいただいたことをきっかけに、それ以降も力を入れて取り組んでいます。このような極値組合せ論を積極的に活用する機械学習法は未開拓の領域であったため、査読者や聴衆から肯定的な反応がいただけたことで、研究継続の励みになりました。現在では機械学習・AI 分野全体の盛り上がりに対して、ノンパラメトリックベイズ法に力を入れている研究所は相対的にかなり少ないといえます。しかし研究はある種、人類全体の大海戦術によって発展していくものです。そのため今後もこのような取り組みを続けて当該分野で注目してもらうことにより、一緒に発展をめざす研究チームをさらに増やしていきたいと考えています。

**知見を活かしてNTTが掲げる医療健康ビジョンに貢献**

■これからの研究の展望を教えてください。

これからの展望として、NTTが掲げる「医療健康ビジョン」に

貢献したいと考えています。この目標は、人々が健康で希望を持ち続けられる未来のために掲げられたものです。NTTでは IOWN (Innovative Optical and Wireless Network) の構成要素の1つであるデジタルツインコンピューティングによって、人の身体および心理の精緻な写像 (バイオデジタルツイン) を実現します。そしてこれを通じて心身の状態の未来を予測することで、医療の未来ビジョンに貢献します。私自身としては、これまでノンパラメトリックベイズ関係データ解析やスーパーベイズ法で集めた知見を活かして、NTTのビジョン実現に貢献できるように考えています。

さらに具体的な取り組みとして、現在私が所属する生体情報処理研究グループでは「テレ聴診器」という研究も進めています（図3）。コンセプトを「着る・見える聴診器」としたこの技術は、聴診器をつけた患者のデータを遠隔地に送信することで、コロナ禍での遠隔感染リスクのない聴診や、遠隔の患者に対する緊急性の判断で異常の早期発見、患者自身による生体への理解促進・意識向上を可能にしました。

もちろん、バイオデジタルツインに向けた研究はこれだけにとどまりません。私が所属しているNTTコミュニケーション科学基礎研究所 (CS研) では、大阪大学との共創プロジェクト「PRIME」を通じて、生体の臓器をデジタル上に計算機モデルとして構築する長期的な研究目標に取り組んでいます。このようにテレ聴診器だけでなくさまざまな応用先を模索・実装していくことによって、NTTの医療健康ビジョンに貢献し、多くの人々がより豊かに生きられる世界を実現していきたいと考えて研究を進めています。

## コンセプト：着る聴診器・見える聴診器



図3 「テレ聴診器」の取り組み

### ■最後に研究者・学生・ビジネスパートナーの方々へ向けてメッセージをお願いします。

私はNTT入社前から多くのつながりに恵まれ、もともと大学院所属研究室の指導教員・嵯峨山茂樹教授がNTT出身、また大学院生時代にはNTTとの共同研究の中で亀岡弘和上席特別研究員にもご指導いただく機会がありました。当時からノンパラメトリックベイズ関係データ解析の分野では世界的に見てトップの研究機関の1つで、CS研（特に上田修功客員フェロー、山田武士前CS研所長）がその黎明期に重大な貢献をされていたこともあり、学生時代から魅力的な研究所でした。

そして入社してから感じたことは、NTT研究所では短期的な選択と集中だけに縛られることがなく、当人の熱意に応じた多様なテーマ選定・挑戦を歓迎してもらえる土壌があるということです。例えば私がNTTに入社した2011年ごろには、世界のコンピュータ科学が第三次ニューラルネットワークブームの影響を強く受けていました。しかしNTTはその潮流にとらわれることなく、私自身もノンパラメトリックベイズ関係データ解析の研究を続けることができました。もちろん世界の時流や研究のトレンドをとらえることは重要ですが、一方で研究はさまざまな分野の多様性と幅を持つことで思わぬ結果を得ることもあります。こうした中長期的な投資として多様なテーマを受け入れてもらえるNTTの環境は、研究者としてとても魅力を感じています。

そして現在私は、幸いにも後輩を指導する役割に立つことが多くなりました。さまざまな指導を行う中で実際に後輩が研究テーマを立ち上げる姿を間近で見た経験から、この記事を読んでいる若手の研究者の方にぜひアドバイスしたいことがあります。それは、研究において「無駄かもしれないと思うようなこと」をたくさん試してみることが大事だということです。お断りしておく、もちろん研究の進め方・モチベーションの高め方は人それぞれで

正解があるわけではありません。近年ではコストパフォーマンスやタイムパフォーマンスが強く叫ばれることも多く、これはそうした考え方とは逆の発想だと思います。

しかし例えば真っさらなキャンパスに新しく絵を描こうとしたとき、真っ白で汚れ1つない紙を前にして筆が全く進まなくなってしまうことがあります。そのとき何でも良いので一度思い切って紙を汚してしまえば、途端に心が軽くなりどんどん手が進む、と考えています。これは研究でも同じです。そしてもし最初に汚れてしまったとしても、絵が上書きされていくうちにその汚れは気にならなくなっていくものです。またよく言われる「サンクコスト効果」のように、人は何かにコストをかければ後付けで価値を見出そうという心理が働きます。もし何も研究が進まなくて困っているという状況に置かれた方がいらっしゃれば、「まず塊より始めよ」。身近に自分ができることから手を付けることで可能性を拡げ、これから新たな未来を切り拓いていってほしいと思っています。



(今回はリモートにてインタビューを実施しました)