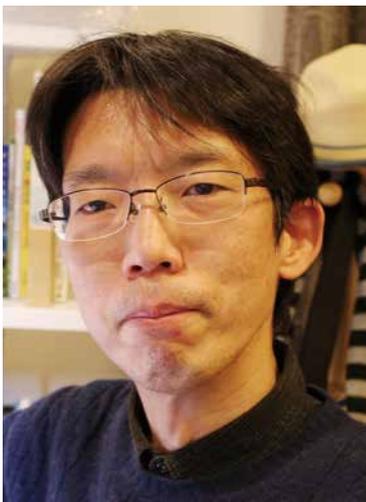


AFSA ニュースレター 研究紹介

今回登場する東京大学の渋谷哲朗さんと京都大学の瀧川一学さんは、ヒトゲノム計画後に誕生した情報科学分野で研究を重ね、今ではそれぞれが独自の研究を行っています。また、コラムではNTTの伝住周平さんが、伝住流作図法について語ります。伝住さんの写真の縦横のサイズの比率が黄金比($\phi \approx 1.618$)です。

interview 01



「究極の個人情報「ヒトゲノム」をどう守るか 「差分プライバシー」の実効性に期待

加速的に増え続ける「医療データ」。病気の診断や治療方針の決定など、私たちが医療データから受ける恩恵は計り知れません。その一方で、究極の個人情報である医療データの安全性をどう担保するかが大きな問題になっています。渋谷さんは、医療現場の医師や患者さんのために、より負担の少ない情報理論に基づく安全性を提供したいと、「差分プライバシー」を研究しています。

B04 班研究分担者

渋谷 哲朗 (しぶや てつお)

東京大学医科学研究所
附属ヒトゲノム解析センター 教授

「ムーアの法則」では、「集積回路に搭載される部品の数が2年ごとに2倍になる」と予測されています。一方、医療関係のデータは、現在、ムーアの法則を上回る10年で1000倍の勢いで増加しています。日々、医療データに接する渋谷さんは、「今こそ、集められた膨大なデータを解析したり、守ったりする“アルゴリズム”が必要です」と話し、AFSAで行われている研究に期待しています。

医療現場を支えるデータ

2003年、世界で初めてヒトの全ゲノム配列が解読されました。当時は、世界中の研究者が協力して、30億ドルの資金と13年の歳月をかけて、やっとヒト1人分の全ゲノムを読み解きました。私が所属する東大医科学研究所(医科研)のヒトゲノム解析センターは、スーパーコンピュータ(スパコン)を導入して、この「ヒトゲノム計画」に参加し成果の一翼を担いました。

しかし現在は、1台の装置で数十人分のゲノムを1日で解読できます。費用も1人分が数万円です。集められた

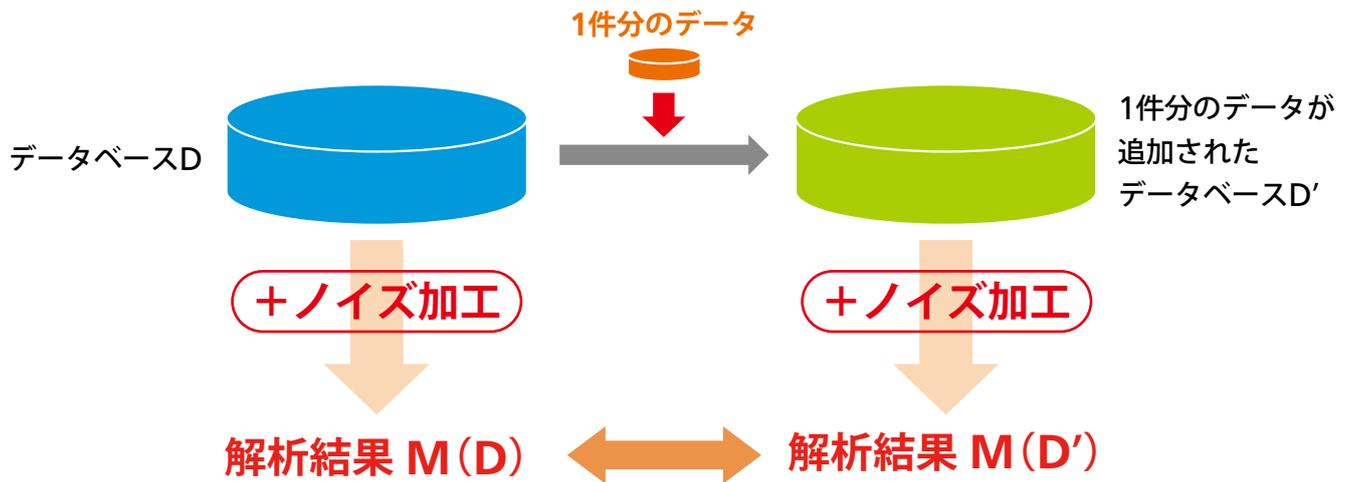
膨大なゲノムデータは、すでに先進的な医療機関で実際に活用されており、医科研附属病院でも、がんの患者さんのゲノムデータの比較をもとに薬や治療法を決めています。

ゲノムデータをはじめとした医療データの活用に興味をもち、データ解析用のアルゴリズムを提案したいと考えています。その一方で、医療データをどのように守るかが大きな課題だと認識しています。医療データは住所などとは違い、何があっても決して変えられない「究極の個人情報」だからです。医科研附属病院では、患者さんのデータを絶対に漏洩させないように、解析

用のスパコンを外界と遮断して、指紋認証が必要な部屋でしかデータを扱えません。ただ、このように厳重な管理は現場の大きな負担になっています。

「差分プライバシー」の鍵を握る ノイズ加工

情報理論によって、現場での情報セキュリティの負担を軽減できるのではと、期待が高まっています。2006年にハーバード大学の情報学者 シンシア・ドワークさんが提案した「差分プライバシー」は、情報理論では最高の賞と言われる「ゲーデル賞」を受賞するなど特に注目されています。



図：差分プライバシーの理論

データベースDとD'それぞれに対する解析結果によって、両者を区別することが(確率的に)困難ならば、「1件分のデータ」が区別できない。つまり「1件分のデータ」が追加されているかどうか分からないことになり、その結果としてデータは漏洩しない。

具体的には、漏洩しては困るデータベースDに1件分のデータを追加したデータベースD'を考えます。この1件分のデータが、DとD'の差分です。DとD'の両方を編集(ノイズ加工)して、最終的な解析結果のM(D)とM(D')が区別できなければ、DとD'のどちらに1件分のデータが入っているかわからないので、データは漏洩しません(図)。

ここで重要なのは、データとして解析に使える程度に小さく、そして元の

データが漏洩しない程度に大きな“ノイズ”を加えなくてはならないことです。現状では、それがどの程度のノイズなのか、またデータ解析のどの段階で加えるのが効果的かわかっていません。そこで私の研究室では、さまざまな解析手法で「差分プライバシー」が効力を発揮する「ノイズ加工」を探っています。うまくいくと、元のデータまでたどることができず、どんな攻撃を受けてもデータは守られた状態で十分な

解析ができるようになる可能性があります。

データが漏洩すると何が起るのか。具体的な脅威についてはわかっていないことも多いのですが、ゲノムデータだとして将来的には顔や指紋などが復元されないと限りません。技術の発展によって、そんな日を迎える前に「差分プライバシー」によって安全性を確立する必要を感じて研究しているのです。(取材・執筆/池田亜希子)

column



B01班 公募研究者

伝住 周平

Shuhei Denzumi

NTTコミュニケーション科学基礎研究所

言語知能研究グループ リサーチアシリエイト

良い図って どんな図だろう？

発表や執筆のための作図は多くの研究者にとって日常だと思いますが皆さんはどんな図を作っていますか？私は決定グラフというデータ構造の研究をしているのでよくグラフを描きます。必要な情報を過不足なく伝えられる図を作れるかはどんなテーマでも悩ましいですが、いろいろな制約をきっちり満たす図ができると嬉しいですね。私の場合だと「平面グラフ」「点線(実線) 辺は左(右)から出る」「節点が規定順に並ぶ」などの制約を満たす例が一意解のように得られたときは満足感がありました。

しかし、上のような論理的に記述できる条

件を満たすだけでは美しい図になってくれません。残念なことに、私は昔から美的センスが欠けていたため感覚で美しくする自信は皆無です。そこで、いつも外的な基準に頼っています。そう「黄金比」です。自然界や様々な芸術作品に現れるアレです。図全体の縦横比から節点や辺の大きさ、オブジェクト間の間隔に線や文字の太さまで黄金比を意識しながら作図しています。そのため私にとって黄金比は円周率や自然対数の底以上に普段使いの無理数です。実際には黄金比を取り入れたからといって美しくなる保証は何もないのですが、構成の指針として役立ちます。私の(特に評判は良くない)スライドテンプレートにもふんだんに黄金比を盛り込んでいます。そういえば、本コラムの文字数も黄金比で分かれていますね。皆さんの作図に対するこだわりも知りたいものです。

「科学研究に機械学習を活用し、革新を起こす」



科学研究において理解を促進し発見をもたらす機械学習の研究をしている瀧川さん。こうした研究は世界的に盛り上がりを見せており、「やっていて楽しい」と話します。さまざまな分野の研究者と積極的にコミュニケーションを図りながら、機械学習の新たな可能性を切り拓いています。

A01 班研究分担者
瀧川 一学 (たきがわいちがく)
 京都大学国際高等教育院附属
 データ科学イノベーション教育センター
 特定教授 (情報学研究科 兼務)

瀧川さんが博士号を取得した当時は、機械学習の需要はまだ少なく、就職に困ったそうです。しかし、今となつては、機械学習の“作り手”であり“使い手”としてのキャリアもある瀧川さんは、各所から求められる貴重な存在。「積極的に選んできた道ではないのですが、ようやく実を結び始めたみたいですね」と話します。

数字ではないデータを扱う

学生時代は情報科学の研究をしており、その後、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で職を得ました。ちょうど「ヒトゲノム計画」が終了し、遺伝子のデータを情報科学の技術で解析しようという動きが盛んになっていた時でした。遺伝子という数字ではないデータに情報科学の技術を転用する研究を始めたのをきっかけに、それ以来、生命科学や化学の研究に役立つ機械学習をつくる研究を続けています。

科学研究に機械学習を使う試みは、今、世界中で非常に盛り上がっています。私は主に分子や遺伝子のデータを

機械学習で扱っていますが、さまざまな分野に機械学習が奇与できる可能性が見えてきています。

また、科学研究で扱うデータには離散的な構造をしているものが数多くあります (図)。言語も離散的な構造をしており、ChatGPTのような大規模言語モデルは、実は分子や遺伝子のデータを処理するモデルと原理的によく似ています。分子や遺伝子についても大規模なデータを用いれば、ChatGPTが言語に起こしたような大きな革新を科学研究に起こせるかもしれません。

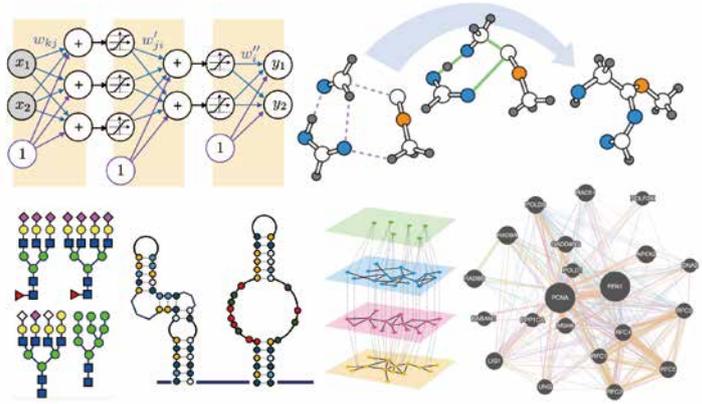
重要なヒューマンファクターの問題

科学研究に機械学習を活用するには、

まず機械学習が何をしているのかを理解していなくてはなりません。特に重要だと感じるのは、ヒューマンファクターの問題です。例えば、自分が持っている実験データを機械学習で利用すれば、何か良い結果が得られるのではと期待されることがあります。しかし、機械学習は学習したパターンにあったものしか予測できないので、研究者が持っている実験データに偏りがあると、そのデータからは本人が考えそうなことしか予測できないのです。自分が見逃していることや新しい考察を見たい場合には、まずは自分が気づいていない偏りに気づく必要があります。

また、機械学習を用いて、現象を理解したいと期待されることがあります。しかし、研究者が望む「理解」は非常に曖昧です。どんな情報を出せばその人が求める理解につながるのかについても議論を深める必要があります。

機械学習を正しく使うのは実は難しく、作り手が想定していない使い方をされるケースも多数あります。私たち作り手が積極的に現場にコミットするとともに、ユーザーにもある程度機械学習の中身を理解してもらえるように、コミュニケーションを図っています。



図：科学研究における離散構造の例
 科学研究が扱う問題は、集合や論理、系列、ツリー、グラフ、プログラム、言語など、離散構造をとともうものが多数ある。こうした数字ではないデータを扱う機械学習のアルゴリズムの研究開発をしている。

(取材・執筆／秦千里)

information

特集企画「ナラティブと人工知能」が人工知能学会誌に掲載

人工知能学会の学会誌『人工知能』Vol39, No.5 (2024年9月号)で、「ナラティブと人工知能」の特集が組まれました。AFSAのA01班のメンバーである須田永遠研究員、武富有香研究員、宇野毅明教授と北海道大学の坂地泰紀准教授が企画したものです。



現在さまざまな領域で注目されながらその全体像を捉えることが難しい「ナラティブ(物語)」をめぐる、多様な専門分野の研究者に寄稿していただきました。

執筆者らを集めて神田ラボで行った座談会の様子も収録されています。参加者が自らの「ナラティブ」を語りながら、他分野との共通点や相違点を見いだしていくダイナミックなディスカッションの様子を、ぜひ誌面でご覧ください。



座談会の様子

サマースクール「組合せ最適化セミナー」を実施

2024年8月6～8日にサマースクールを開催しました。京都大学数理解析研究所と共同で行った「組合せ最適化セミナー」では、組合せ最適化とその周辺分野で独自の研究を展開している3名の研究者を招いて、基礎から先端に至る体系的な講演をしていただきました(下記)。

参加者は89名。対面で行ったため、休憩時間にも参加者たちが時間を惜しんで議論する姿が見られました。



講演タイトル/講師

「作用素スケールリング入門」/ 相馬 輔 (統計数理解析所)
 「整数重みの組合せに着目したグラフの構造解析」/ 藤田慎也 (横浜国立大学)
 「主双対内点法: その始まりから今日まで」/ 吉瀬 章子 (筑波大学)
 ●講義資料と演習問題はこちらから
<https://www.kurims.kyoto-u.ac.jp/coss/coss2024/coss2024.html>

AFSA 共催「触媒コンペ」開催のお知らせ

化学の宝探しをしてみませんか?

「触媒コンペ(The Catalysts Design Arena)」という面白いコンペティションを企画いたしました。情報学的手法で、性能が良さそうだと予想される新規触媒(元素の組合せ)を投稿していただき、最も性能が高い触媒の提案者を表彰します。

通常の機械学習などのコンペティションとは異なり、提案していただいた新規触媒を化学者に実際に実験して評価してもらいます。自分が考えた物質を実験できる機会はなかなかないと思いますので、この珍しい企画にふるってご参加ください。投稿締め切りは、2024年11月20日

詳細は、下記URLより:

<https://afsa.jp/CDArena/>



「24年度第2回領域集会」を11月に東京で開催予定

2024年11月29～30日に2024年度第2回領域集会を開催します。対面とオンラインのハイブリッド形式で行い、対面の会場は東京神田付近のホールを予定しています。

最新情報はAFSAウェブサイト(お知らせ)でご確認ください。

「若手研究者の海外派遣支援プログラム」を今年度も実施

2023年度に、若手研究者の国際的研究活動を活性化することを目的にスタートしたAFSAプロジェクト関係者の海外派遣支援プログラムを、今年度も実施します。

2023年度には大学院生やポスドク研究員4名について、海外との研究交流を支援しました。

2024年度は、すでに3名の海外派遣が決まっているほか、支援プログラム参加者による報告会や意見交換会の開催を予定しています。



AFSA News Letter No.12

(2024年10月発行)

発行者 AFSA プロジェクト事務局
 所在地 〒606-8501 京都市左京区吉田本町
 京都大学大学院情報学研究所コンピュータアルゴリズム研究室内
 編集協力 サイテック・コミュニケーションズ
 デザイン 八十島博明、石川幸彦 (GRID)



<https://afsa.jp>

本領域に興味をお持ちの方は AFSA 事務局 (afsa-contact@algo.cce.i.kyoto-u.ac.jp) までお問い合わせください。