



生成AI時代におけるコミュニケーション科学研究の新展開 ——人間と情報の本質探求と人に寄り添う技術の協創に向けて

NTT コミュニケーション科学基礎研究所 (CS研) では、人と人および、人とコンピュータシステムとの間の「ここまで伝わるコミュニケーション」の実現に向けて、情報と人間の本質に迫る基礎理論の構築と、社会に変革をもたらす革新技术の創出に取り組んでいます。本稿では、昨今急激に進歩しつつある生成AI (人工知能) の動向を踏まえ、人とAIの共存に向けたCS研における最新の研究の取り組みについてその一部を紹介します。

キーワード：#コミュニケーション科学、#人工知能、#脳科学

なや ふとし
納谷 太

NTTコミュニケーション科学基礎研究所 所長

はじめに

2022年11月にOpenAI社が対話型生成AI (人工知能) であるChatGPTをリリースして以降、わずか2カ月で1億人ユーザに達するというニュースは世界中に大きなインパクトを与えました。「対話型」という文字どおり、人と自然に対話する機能を有し、自然な文章の作成や要約および翻訳に加え、図表を読解して質問に回答することや、人が指定したインストラクションに従って所望の画像や動画、音楽、さらにはプログラムさえ自動生成できる機能を有するものまで各社から発表され、高い注目を集めています。その背景には、深層学習の発展を支えたGPUの計算能力の大幅な向上に加えて、膨大かつ多様な言語資源を学習可能にしたTransformerと呼ばれる自然言語処理での大きなブレイクスルーとなる技術が生まれたことが挙げられます。膨大なデータを学習した結果である大規模言語モデル (LLM: Large Language Model) の1つであるGPT (Generative Pre-trained Transformer) と対話的にインタラクションするインタフェースを備えることで、研究者やエンジニアだけでなく、広く一般のユーザがスマートフォンなどで簡単に利用できる環境が整ったことが相乗効果を生み出しました。

こうした生成AIは、その精度や信頼性の改善に伴い、今後私たちの生活に確実に

浸透していくでしょう。最近発表されたGPT-4oでは、リアルタイムでの音声対話や、カメラで撮影した映像を見せながら質問するなど、マルチモーダルな入出力応答も扱えるようになり、さらなる注目を浴びています。しかしながら、現状の生成AIは、ユーザがどのような回答を求めたいかに応じて、プロンプトという指示により事細かに状況を言語的に与える必要があります。マルチモーダル化されたLLMとはいえ、LLM自体はあくまで言語情報をベースとして構築されており、映像や音声などのマルチモーダルな入力情報はLLMの入力形式に適合するようにトークン化されて処理しています。一方、人間は言葉を用いる以前からあらゆる感覚情報を脳内で処理して物事や概念を理解・記憶しており、これに対して言語という抽象化した記号体系を紐付けています。すなわち、現状のマルチモーダルLLMは、言語ではとらえきれないような五感情報を取り扱えないという限界が存在します。これは1990年に認知科学者のStevan Harnadが提唱した「シンボルグラウンディング問題⁽¹⁾」または「記号接地問題」と呼ばれ、AIに人間と同じように言語で表現される記号が指し示す物事に対して、実世界の概念や意味を結び付けて理解させることができるのかを問う根本的な問題です。加えて、現在のLLMは、人間と同じように、表情や態度や会話などから、相手の好みや感情、意図、知識、信念などを

察し、相手に対して気の利いた回答を返したり、あるいはAI自身がパーソナリティを持ち、信念に基づく回答を返すという段階には至っていません。そこには「心の理論⁽²⁾」という、人間のように他者の心の状態を推測し理解する能力をAIで実現できるのかを問う哲学的ともいえる大きな課題が残されています。

CS研は、1991年の設立当初から、「人と人との理解のメカニズムを学際的に解明」することを目的とし、情報と人間を結びコミュニケーション科学の分野を開拓すべく基礎研究に取り組んできました⁽³⁾。まさに、上記の「シンボルグラウンディング問題」および「心の理論」という本質的な問いを追究する研究テーマといえます。CS研設立以来33年を経た現在において、さまざまな情報の本質や人間の本質に関する深く理解に基づく新発見や革新技术を創出し続けていますが、人と人がお互いを理解するメカニズムの全貌の解明は容易には解決し得ない究極的な研究課題であり、研究が進展するほど、情報科学、工学、神経科学、心理学、社会科学、人文科学、哲学に限らず、医学、生物学、さらには数学など、さまざまな学際分野にまたがる幅広い視野に立ったアプローチの必要性が年々高まっていると実感します。

本特集では、学際的な研究分野を広げつつあるCS研の最近の取り組みにおいて、「情報の本質を極める」「人間の本質を極める」

「人と社会に寄り添う」「基礎理論の追究」という4つの切り口から、それぞれ代表的な研究事例の一部を紹介します。

情報の本質を極める

CS研では、コミュニケーションにおける情報を伝達するあらゆるメディアを対象にした情報処理技術に関する研究を進めています。近年、センシング技術の発展に伴い、これまで観測できなかった現象を見える化し情報化することができつつあります。中でも、「音」は私たちの日々の生活において非常に身近な存在ですが、通常はマイクロフォンを用いて計測することが一般的です。音は空気中を波として伝わりますが、マイクロフォンはそれが置かれた場所での音を計測することはできるものの、音がどのように発生しどのように空間を伝搬するかを詳しく知ることは困難です。CS研の最近の研究では、レーザー光を用いた光学技術と高速カメラを用い、音の波を動画像としてとらえ、「音を可視化」する光学的音響計測という技術を実現しました⁽⁴⁾。この技術では、レーザー光が音場を通過する際、音の粗密によって光の位相が変調される様子を、元の音による変調のないレーザー光と干渉させ、この干渉光を高速カメラで撮影することで音場をイメージングします。ただし、計測する際のノイズが非常に大きいいため、さまざまな音源数やノイズの条件を変化させシミュレーションで生成した音場データセットと深層学習技術を組み合わせることにより、雑音が多い状況下でもこれを除去し、クリアな音場の可視化を可能にしました。さまざまな音の可視化にとどまらず、心地良い音を再現する音響デバイスの設計や、騒音源の分析など、幅広い応用が期待できる技術です。これまで個体差が大きかったマイクロフォンで音の振動を

計測するのではなく、光周波数といった物理量を基準とした次世代標準をめざした超精密音圧計測技術への発展を見込んでいます。

上述の音や光に加えて、最近では生体信号の1つである心電図から、その元になっている心筋細胞の挙動に関する情報をAIで推定する技術についても研究を進めています⁽⁵⁾。医学的には心電図波形と病気との大まかな対応は解明されているものの、その根本となる心筋細胞レベルで何が起きているかを推定することは困難でした。この研究では、スーパーコンピュータ富岳を用いた心臓物理モデルによるシミュレーションにより、心筋細胞内におけるNa（ナトリウム）、Ca（カルシウム）イオンや伝導率、心臓形状などのパラメータを入力情報として、人工的に生成した心電図波形から、これを生成したパラメータを高精度に推定する逆問題を解く機械学習技術を考案しています。心電図を含む多様な生体データの計測とともに、その要因や関係性を解析する技術を構築することにより、個々人の遺伝的特性や生活習慣を反映させた健康状態の予測や、どのような薬や治療法が有効かを検証するテーラーメイド医療のシミュレーションを可能にするバイオデジタルツインの実現をめざした研究です。

このような生体内の心筋細胞や生化学反応などのミクロな挙動から、臓器としての心臓および、血液や他の臓器間との関係、さらには個としての人体全身におけるマクロな振る舞いまでをトータルで扱いモデル化することは極めて困難な課題です。これを解決するためには、体内で起きているさまざまな事象をとらえることに加え、どのような事象と事象とが相互に依存し、因果関係を有しているのかについて、膨大な組合せの中から正確な情報を高速に見出す技術が必要です。本特集記事『たくさん

タの中から素早く知識を発見——計算の枝刈りによる高速化手法を活用した厳密性を担保した大規模データ解析』⁽⁶⁾では、このような問題を解く足掛かりになる技術として、高速性と厳密性を備えた大規模データ解析について紹介しています。

人間の本質を極める

CS研は、人の情報処理メカニズムを科学的に解明することを目的として、視覚、聴覚、触覚などの感覚や、運動機能、情動に関する研究を推進しています。人間がなぜ今のような感覚や運動、情動のような機能を持つに至ったのかという普遍性を調べる研究や、個々人が持つ先天的な特性や、後天的に得た経験や学習により多様性が生まれる本質的な問いへの答えを追究する研究を含めて幅広く進めています。

例えば、私たちは歩く際に眼から入ってくる情報に基づき、無意識のうちに自分の歩行速度を推定し、最適な速度で歩くよう運動を調節していることが知られています。一方で、自分が動いておらず、TVなどで自転車に乗っている他者の一人称視点のカメラ映像を見るとき、その際に感じる速度感、自身が実際に自転車に乗っているときに感じる速度感と同じなのかを調べる研究を行いました。実験参加者がヘッドマウントディスプレイを装着して仮想的な廊下を歩く際に、縞模様のある左右の壁を前後に動かし、その際の歩行速度の変化を調べた結果、縞模様が粗いほど自身は早く歩いていると感じ、細かいほどゆっくり歩いていると感じます。しかし、歩行せずに単に縞模様が流れるさまを見る状況では、逆に縞模様が細かいほど早く動いていると感じ、粗いほど遅く動いていると感じる結果が得られました⁽⁷⁾。このことは、人間の脳内で運動と知覚における速度推定の過程が異なる

る（脳内に複数の速度計が存在する）という可能性を示唆しています。今後、メタバースなどのバーチャル空間において、高い臨場感を与えることや、VR酔いが生じにくい視覚映像を提示するインターフェースのデザインに貢献し得る成果です。

上記は、人の感覚や運動の普遍的能力に関する研究の例ですが、人の能力の多様性を表す身近な例として、利き手や利き足が挙げられます。多くの人は、成長するにつれて手や足の器用さに左右差が生じ、右利きの人は左手で文字を書くことは非常に困難ですし、逆に左利きを右利きに矯正した人は、運動の種類にもよりますがどちらの手もある程度器用に動かすことができます。CS研の最近の研究では、スマートフォンをぐるぐる回すだけで手足の器用さを簡便に計測し定量的に評価する新たな手法を考案し、利き手は非利き手に比べて繰り返し運動のばらつきが少ないことや、年齢によってそのばらつきがどのように変化するのか、さらには、このばらつきが生じる本質的なメカニズムを解明しました。本特集記事『ヒトの動きが“ばらつく”ことの本質』⁽⁸⁾で詳しく解説しています。

人間は感情的な生き物であり、そのときの自身の体調や心理状態によって、物事や他者に対する印象についての主観的な評価は曖昧性を伴います。主観的印象を問うアンケート調査などでは、その人個人の感情表現や感情認識に対する癖や、同じ問いに対する回答の揺らぎなどの曖昧性を伴います。このような不確実性は人間ならではの性質ですが、回答のパターンから本心を知ることが可能でしょうか。最近の研究⁽⁹⁾では、回答の癖や曖昧さを統計的に抽出・除去することにより、回答の信頼度を推定する手法について提案しています。

人と社会に寄り添う

これまで、情報および人間の本質を極める研究について述べてきましたが、ここではコミュニケーションの本質である情報と人をつなぐ、あるいは情報を介して人どうしや人と社会をつなぎ寄り添う研究について紹介します。

今や、情報通信ネットワークは日々の生活に深く浸透し、あらゆる社会活動を支えています。自然災害のみならず、人為的なミス等によるネットワークインフラの故障は甚大な影響を及ぼします。より堅牢で高信頼なネットワークを設計するには、構成する部品の故障や災害に対する耐久性を解析する信頼性解析および、どの部品を補強すべきかを発見する脆弱性解析などの課題があります。しかし、ネットワークを構成する部品がたった50個だけでもその組合せは1000兆パターンにもものぼり、従来技術では近似的な解析手法しか適用できませんでした。現実的な規模のネットワーク設計において非常に高い信頼性を評価するためには、近似ではなく厳密な解法が必要です。本特集記事『圧縮計算でめざす高信頼インフラ決定グラフを用いたネットワーク解析問題の高速解法』⁽¹⁰⁾では決定グラフというデータ構造を用いたネットワーク解析の厳密解法について詳しく紹介しています。

一方、人間科学の分野では、深層学習の技術を援用し、人の視覚情報処理メカニズムを数理的にモデル化しシミュレーションすることにより、人が映像に対して感じる不自然さや快適さなどの指標を定量的に評価できるようになってきました。本特集記事『人の知覚に寄り添った自然で快適な映像表示——人間の視覚情報処理モデルに基づく表示映像の最適化』⁽¹¹⁾では、人の視覚情報処理モデルに基づき、人にとって自然で快適と感じる映像を工学的に自動生成す

る技術について紹介しています。また、最近報道発表した記事⁽¹²⁾では、eスポーツ選手の格闘ゲーム対戦直前の脳波から、その後の勝敗に強くかかわるパターンが現れることを世界で初めて発見しました。人は機械とは異なり、身体的なスキルだけでなく、極度なプレッシャーや緊張などの「メンタル」的な要素が試合におけるパフォーマンスを左右します。この研究では、相手の出方に対してどのように試合運びをするかの戦略判断にかかわる脳波パターンと、逆境においても動じないよう感情制御することにかかわる脳波パターンが明らかになり、試合直前の脳波状態からその後の勝敗を機械学習モデルにより約80%の確率で予測することができることを示しました。これは、勝負事に臨む際の理想的な脳状態が存在することを示唆し、スポーツにとどまらず、手術に臨む医師や、的確な判断と操縦を要求されるパイロットなど、高度なスキルと熟練を要し、プレッシャーに対処する必要がある人々に対して、理想的な脳状態に近づけるような新たなメンタルトレーニングが可能になると考えられます。NTTが推進するIOWN (Innovative Optical and Wireless Network) 構想では、人のデジタルツインコンピューティングの実現を目標としていますが、熟練者の脳状態をデジタル化し、訓練者の脳状態から熟練者のそれに近づくプロセスをシミュレーションすることで能力拡張を実現することにつながります。

一方、育児・教育の場においては、幼児のこたばの発達に加え、他者との社会性の発達を調べるのが大きな課題です。幼児が他者に抱く親密さは友人関係の基盤であり、発達心理学分野での重要テーマですが、従来の大人による行動観察による手法では、非常に手間がかかるうえに、幼児に目的を悟られやすいという課題がありました。最

近のCS研での研究では、3～6歳の幼児に自分と他者の絵を描いてもらい、描かれた人物間の一番近い点どうしの水平距離が近いほど、その幼児が他者に抱く親密さが高く、それらが有意に相関することを明らかにしました⁽¹³⁾。大人の目が行き届かない場面も多い現代社会において、孤立やいじめなどの対人関係の問題をいち早く検知し、これを防ぐように気付きを与える支援技術へと発展させていく予定です。

基礎理論の追究

CS研では、長期的視野に立った研究開発を加速し、「知の泉」の源を一層強化するため、現代数学の基礎理論を研究する組織「基礎数学研究センタ」を2021年10月に設立しました⁽¹⁴⁾。2024年7月現在、8名の数学者が在籍し、数論、代数、幾何、表現論、解析、力学系などそれぞれ異なる専門領域を研究する研究者がそれらを横断・継続し、未知の数学的真理を探究と重要な未解決問題の解決や、物理学、生物学、医学など、他の学際分野での重要課題に対する現代数学によるアプローチの提案と新たな数学的対象の発見をミッションとして日々研究に勤しんでいます。NTT技術ジャーナル2024年7月号特集『未知に挑む数学研究と夢』⁽¹⁵⁾では、NTT基礎数学研究センタの所属メンバによる最近の主要な研究成果および展望について紹介されていますので、ご興味ある方は是非ご一読ください。

おわりに

CS研における最新の研究成果の代表例を紹介しました。生成AIをはじめとする技術が高度化するにつれ、私たち人間1人ひとりの多様性の本質の理解と、個々人の特性に応じて有益な情報をどのように処理

し伝えるのかといった、人に寄り添う技術の重要性はますます高まっていくと考えます。今後、人と人および、人とコンピュータが真に理解し合い、人とAIが協力してよりよい社会を創造できる未来に向けて、さらなる研究分野を切り拓いていきます。

参考文献

- (1) 土屋・中島・中川・橋田・松原・大澤・高間：“AI辞典 第2版,” 共立出版, 2003.
- (2) 子安：“心の理論：心を読む心の科学,” 岩波書店, 2000.
- (3) 研究所探訪：“情報通信を人間の思考の段階まで高めるコミュニケーション科学研究所,” NTT技術ジャーナル, Vol.4, No.9, pp.58-59, 1992.
- (4) 石川：“光を用いた音の可視化と精密計測,” 電信情報通信学会誌, Vol. 106, No. 9, pp. 849-854, 2023.
- (5) R. Nishikimi, M. Nakano, K. Kashino, and S. Tsukada: “Variational autoencoder-based neural electrocardiogram synthesis trained by FEM-based heart simulator,” Cardiovascular Digital Health Journal, Vol. 5, No. 1, pp. 19-28, 2024.
- (6) 藤原：“たくさんのデータの中から素早く知識を発見——計算の枝刈りによる高速化手法を活用した厳密性を担保した大規模データ解析,” NTT技術ジャーナル, Vol.36, No.9, pp.23-26, 2024.
- (7) S. Takamuku and H. Gomi: “Vision-based speedometer regulates human walking,” iScience, Vol. 24, No. 12, 2021.
- (8) 高木：“ヒトの動きが“ばらつく”ことの本質,” NTT技術ジャーナル, Vol.36, No.9, pp.27-30, 2024.
- (9) S. Kumano and K. Nomura: “Multitask item response models for responses bias removal from affective ratings,” Proc. of ACII 2019, Sept. 2019.
- (10) 中村：“圧縮計算でめざす高信頼インフラ——決定グラフを用いたネットワーク解析問題の高速解法,” NTT技術ジャーナル, Vol.36, No.9, pp.14-17, 2024.
- (11) 吹上：“人の知覚に寄り添った自然で快適な映像表示——人間の視覚情報処理モデルに基づく表示映像の最適化,” NTT技術ジャーナル, Vol.36, No.9, pp.18-22, 2024.
- (12) <https://group.ntt.jp/newsrelease/2024/07/18/240718a.html>
- (13) A. Shinohara, M. Narazaki, and T. Kobayashi: “Children’s affiliation

toward peers reflected in their picture drawings,” Behavior Research Methods, Vol. 55, pp. 2733-2742, 2023.

(14) <https://group.ntt.jp/newsrelease/2021/10/01/211001a.html>

(15) <https://journal.ntt.co.jp/article/27040>



納谷 太

CS研は、人間とAIが共存し、「ここまで伝わる」コミュニケーションと豊かな社会の実現に向けて、広く学際的な基礎研究に取り組むとともに、パートナーの皆様とのコラボレーションにより新たな価値創造に貢献していきます。

◆問い合わせ先

NTTコミュニケーション科学基礎研究所
企画担当
TEL 0774-93-5020
FAX 0774-93-5260
E-mail cs-jousen-ml@ntt.com