

開会挨拶

浅野哲夫 (北陸先端科学技術大学院大学学長)

北陸先端科学技術大学院大学は1990年創立の大学院だけの大学です。2年前に創立25周年を迎えました。この四半世紀に様々な変革がありました。大学院大学としての位置づけも大きく変わってきました。本学は25周年を契機に、それまでの3研究科(知識科学, 情報科学, マテリアルズ・サイエンス)を1研究科に統合しました。現代では、自分で目標を設定し、目標管理できる学生が求められています。研究科制は研究科の壁があって他研究科の講義科目が取れないという制約がありました。これでは多数の学生の要望に応えることができないということで、研究科の壁をなくすことにしました。新入生に聞いてみると、学部時代と専門を変えて勉強できそうな自由な雰囲気を感じられたからと動機を語ってくれました。私は学長に就任し、「知的たくましさ」と標榜しておりますが、これは「より幅広い経験と知を求めようとする挑戦力を持っていること」を意味しています。従来のように一つの専門だけを極めるのではなく、幅広い経験が求められる時代になっていると思います。

マテリアルズ・インフォマティクスも全く新たな学問分野で、ちょうどJAISTの挑戦に似ているのではないかと考えています。幅広い分野の研究者が集まらないとできないような、わくわくする分野ではないかと思えます。私もできれば研究面で貢献したいところですが、学長職の関係で表立って研究することができませんが、今のうちに勉強して、学長職を退いたときにマテリアルズ・インフォマティクスに何か関係する仕事ができないか、それを探すために今日やってきました。皆さんの活発な議論を期待しています。

略歴

大阪大学基礎工学部電気工学科卒業(1972)
大阪大学基礎工学研究科物理系修士課程修了(1974)
大阪大学基礎工学研究科物理系博士課程修了, 工学博士(1977)
大阪電気通信大学工学部講師 (1977)
同助教授 (1979)、同教授 (1988)
北陸先端科学技術大学院大学教授 (1997-2014)
同学長補佐 (併任,1999-2000), 評議員 (2002-2004)
学長補佐 (併任,2008-2010) , 情報科学研究科長(2012-2014)
大学院教育イニシアティブセンター長(併任 2010-2014)
学長(2014-)

趣旨説明

寺倉清之（物質・材料研究機構）

「科学における法則は簡潔な表現でなければならない」と Einstein が強調した。これに通じる主張として、「必要最小限以上の仮定をしてはいけない」という「Ockham の剃刀」が知られている（Ockham: 14 世紀の哲学者）。一方、統計学習の大家の Vapnik は「複雑な現実世界においては、よりよい予測をしようとするれば、理解を諦めなければならない」と主張した。

マテリアルズ・インフォマティクスなどのデータ科学の主要な目的は、現実世界を帰納的に扱おうとするものである。そのとき、上述の「理解と予測の相克」が問題になる。「分からなくても実際に役立つ予測ができればいい」という現実主義的な考えもあるが、「分からなければ知識、知恵にはならなくて応用が効かない」という言い分も理に適っている。「理解と予測の相克」の問題が内蔵する意味をより深く理解し、どうしたらこの相克を解くことができるかを考える、というのが本講演会の目的である。沢山の画像を入力すると、人工知能は猫らしい画像を作り出す。人工知能は大量の入力から情報縮約を行い、「猫」という概念を得たことに対応する。しかし、その概念が何なのかを説明してくれない。一体、人工知能は「猫」を分かってしまったのだろうか？煎じ詰めると、「分かるとは何か」に行きつくのであろうが。

略歴： 1971 年：理学博士（大阪大学）、1978 年：東大物性研助教授、1990 年：同教授、その後工業技術院産業技術融合領域研究所、産総研、北大、北陸先端大を経て、2015 年から 2 年間物質・材料研究機構フェローとして、情報統合型物質・材料開発イニシアティブ(MI²I)のプロジェクトリーダーを務めた。2018 年 4 月以後は同プロジェクトのエグゼクティブアドバイザー。専門は物性理論、計算科学、現在は物質科学への機械学習の活用に興味を持つ。

「分かる」を支援する AI とは？

溝口理一郎(北陸先端科学技術大学院大学・Laboratory for Applied Ontology (LOA),
ISTC-CNR, Trento, Italy)

将棋や囲碁のゲームにおいて AI システムが専門家を凌駕し、クイズの分野でもチャンピオンに勝つという華々しい成果を生んでいる。データから学ぶ方法論が実用レベルになり、Deep learning を代表とする様々な AI 技術が応用されている。それらは AI の工学的な側面での成功ということができる。一般に理系の学問は工学的側面と科学的側面があり、前者は「作る」ことが必須で、作ったものの有用性に重点が置かれ、後者は「分かる」ことに重点がある。その意味では昨今の AI ブームは前者に着目したものであると言える。しかし、有用性が重要な工学といえども、問題点や目的を「分かった」上で新しい技術が生み出され、ものが作られる事を考えることを考えれば明らかなように、「分かる」ことをないがしろにすることはできない。特に NIMS/MI²I で行われている AI 応用では「分かる」事の重要さは一層増す。残念ながら、現時点では機械学習に基づく AI システムは Black box になりがちであり、人間が納得する説明の生成能力に欠けるという問題が認識されつつある。AI システムが説明能力を持つためには概念の整備とその表現が必要となる。本講演では、良い結果を生み出すための AI 技術だけでなく、「分かる」ことを支援する AI について実例と共に述べる。具体的には、概念体系の抽出と組織化の方法論であるオントロジー工学の成果の一つとして機能オントロジーと機能分解木の応用について述べる。

略歴:1977 年大阪大学大学院基礎工学研究科博士課程修了。大阪大学産業科学研究所助手、助教授、教授を経て、現在北陸先端科学技術大学院大学特任教授、Associate researcher of Laboratory for Applied Ontology, Trento, Italy。パターン識別関数の学習、クラスタ解析、音声の認識・理解、エキスパートシステム、知的学習支援システム、オントロジー工学の研究に従事。人工知能学会会長、Intl. AI in Education(IAIED) Soc. President, J. of Web Semantics の Vice President, Editors-in-Chief 等を歴任。

情報・学習・理解：機械学習から機械理解の定式化に向けて

日高 昇平（北陸先端科学技術大学院大学）

近年の人工知能・情報科学における発展は目覚ましく、関連分野や一般社会においても、機械学習技術の応用が広がっている。こうした技術の多くは、大量のデータをうまく圧縮・要約し、まだ見ぬ未知のパタンに対する予測を行う手法の一種とみられる。こうした技術の基礎は少なくとも 1940 年代の Shannon らによる情報理論の確立に遡る。情報理論は、データの符号化・復号化による通信のモデルを提示し、通信容量の概念およびその定量的な限界を示す。Shannon & Weaver (1948)は、その著書の前文にて、こうした通信のモデルとしての情報理論は、情報の理解および因果性の探求へと続く入口であり、符号化・通信を数学的に扱う方法論を示すに過ぎないとしている。彼らはこれに続いて、情報理解の問題を扱うことの重要性を認識していた。

近年発展著しい機械学習技術も、この観点からは、特定クラスの復号器を符号から構築する方法論としてみなせる。囲碁における AlphaGo Zero のように、解の定まり得る良い構造をもつ一部のクラスの問題では人の知識をデータとして使わずとも自律的にデータの解釈を構築することが可能であることが示されつつある。しかし、一方で、科学的活動(知的探求)や社会一般など、ただ一つの解が存在するかも不明な問題では、それをいかに探索・学習すべきか究明すべき点が多く残されている。人間がこうした不定構造の問題に自律的に取り組む“機械学習”エージェントだとすれば、機械理解とも言うべき自律的に世界を解釈し、あるいはそれを修正する機能を備えていると言える。本講演では、人の理解の認知機構に学び、機械理解の定式化の可能性について論じたい。

略歴

平成 14 年九州大学理学部生物学科卒業。平成 19 年京都大学大学院情報学研究科博士後期課程修了，同大学博士（情報学）取得。平成 20 年 Indiana University にて博士研究員。平成 22 年北陸先端科学技術大学院大学知識科学研究科助教。平成 29 年同准教授。言語・認知発達、意味認知の計算論的メカニズムの解明を目的に、心理学実験・情報理論・機械学習・非線形時系列解析などを駆使した研究を行う。

機械学習からのフィードバックによる材料科学的理解の促進事例

田村 亮 (物質・材料研究機構 / 東京大学新領域)

現在、材料科学分野ではデータ科学を利用した材料研究であるマテリアルズ・インフォマティクスが注目を集めている。本講演会では、講演者が行なったマテリアルズ・インフォマティクス研究の中で、機械学習結果が材料科学的理解の促進に繋がった事例を2つ紹介する。1つは、材料測定結果を説明するための物理モデルを機械学習により推定する研究である[R. Tamura and K. Hukushima, *Phys. Rev. B* **95**, 064407 (2017)]。材料科学で導出された物理モデルの各パラメータはそれぞれ物理的意味を有しており、パラメータ値を知ること自体が対象材料の理解に繋がる。ここでは、磁性材料に対して、ベイズ推定を利用した研究を紹介する。2つ目は、嗅覚センサと機械学習を組み合わせることで、ニオイから特定情報を抽出することを目的とした研究である[K. Shiba, R. Tamura, G. Imamura, and G. Yoshikawa, *Sci. Rep.* **7**, 3661 (2017)]。この研究では、機械学習による予測結果を用いることで、有用材料の選択が行えることを示した。さらにその結果から、嗅覚センサでの使用に適した材料に対する理解が深まった。これらの具体的事例を通して、データ科学よる「材料科学的理解」について議論する。

略歴

2012年東京大学大学院理学系研究科物理学専攻博士課程修了。博士(理学)。2015年4月より、国立研究開発法人物質・材料研究機構 研究員。2018年4月より主任研究員。現在、同機構、国際ナノアーキテクニクス研究拠点(WPI-MANA)および統合型材料開発・情報基盤部門に所属。2017年4月より東京大学大学院新領域創成科学研究科メディカル情報生命専攻 講師を兼務。統計力学、材料科学、マテリアルズ・インフォマティクスの研究に従事。

データマイニングを活用した物質空間の理解に向けて

ダム ヒョウ・チー (北陸先端科学技術大学院大学・JST さきがけ)

計算機技術やアルゴリズム開発の発展により「人間のように学習するコンピューターシステムの構築」を目的とする機械学習の研究領域が急激に発展を遂げて、多くの汎用技術を作り、凄まじく応用範囲が広がった。少し遅れを取って提唱されたデータマイニングの研究領域は機械学習の研究領域とは大幅に重なる。データマイニングの研究では統計学、機械学習、および統計学習の技術を活用するが、機械学習とは少し異なり、「データから新しく有用な知識を見つける」を目的と設定される。データマイニングの研究は多くの機械学習技術を活用するが、機械学習の研究よりデータからの「知識」獲得および「発見」といった風味が強調され、要求される。そのため、機械学習の研究領域では教師あり学習によるデータ駆動型の予測が主流だが、データマイニングの研究領域では教師なし学習によるデータ構造の発見が注目される。また、データマイニングによるデータからの知識発見プロセスは、帰納法アプローチで行われているため分野知識の理解によるデータの記述や獲得した知識の評価が不可欠である。教師あり学習による「予測」問題は基本的には予測精度によって評価できるが、教師なし学習によるデータ構造の発見は客観的な統計基準などを活用する一方で、主観的な評価が導入される場合も少なくない。特に、計算機が自動的に発見された知識に対して、分野の知識の土台での説明性が評価され、いわば、「人間」が理解可能であることが求められる。

近年、データ科学を活用して自然科学の研究を促進化する試みが盛んになり、物理・化学の長年貯蓄された分野の経験・理論知識の土台で築かれたマテリアルズ・サイエンス領域においてもマテリアルズ・インフォマティクスの研究が注目を集めている。本講演会では、講演者が行なったデータマイニングを活用した材料設計研究の中で、マテリアルズ・サイエンスとデータ科学を融合させることを目指す際、両側の立場から見た期待と限界、またこれらに関して立場によって発生する相違について、研究事例を用いて紹介する。特に、材料のデータをマイニングして、新材料の物性を「予測」する知識の獲得や物性を発揮するメカニズムの「理解」を支援する情報の獲得する技術的課題を議論したい。

略歴: 2003年 北陸先端科学技術大学院大学 材料科学研究科 博士後期課程修了。同大学 知識科学研究科 特任助手 (COE), 特任講師(テニュアトラック)を経て, 同 知識科学研究科 准教授. JST さきがけ研究者. 結晶学・材料構造物性実験・第一原理計算・機械学習・データマイニング, 知識科学, 材料インフォマティクス等の研究に従事

機械学習は真の発見に寄与できるのか？

瀧川 一学 (北海道大学・JST さきがけ)

深層学習を含む機械学習は、データ駆動型の「予測」を与える汎用技術である。しかし、「予測」から「発見」や「理解」に至るためには、未だ方法論的に大きな壁が立ちはだかっている。2004年「予測と発見の科学」シリーズの第1巻として刊行された「統計的因果推論(宮川雅巳 著)」は、コンピュータが初めて卓近な道具になりつつあった1966年に、20世紀を代表する統計学者 Box が発表した論文“Use and Abuse of Regression”の話で始まる。回帰分析が実験研究よりも観察研究に使われる当時の状況を踏まえ、観察研究における回帰分析の目的を1)説明変数を観測したときの目的変数の予測、2)説明変数に外的操作を加えたときの目的変数への因果的効果の発見、の2つに大別した上で、Box は前者 1)を「use」、後者 2)を「abuse」と呼んだ。これは50年後の現在、「教師つき機械学習」にそのまま当てはまる。「相関関係は必ずしも因果関係を意味しない」は統計学のイロハである。統計学に精通した研究者ほど、機械学習はあくまでデータ中の相関関係の利活用のための道具である、という立場を強く支持するであろう。実際、検索も推薦も広告も制御も画像・音声・言語のパターン認識や変換も、AI ブームを牽引する成功例はこの用途 1)の例である。データにおいて観察された相関関係のみから因果関係を論じることは不可能であり、何か新たな枠組みが必要である。社会投入が進む現在、機械学習モデルのブラックボックス性は再認識され、例えば、所謂 FATE: Fairness(公平性), Accountability(説明責任), Transparency(透明性), Ethics(倫理)についての盛んな議論や、米 DARPA の Explainable AI の研究プログラムも始まっている。

さらに、興味が「科学法則の理解」である場合、相関関係の利活用では明らかに不満足である。本発表では、以上を十分認識した上で、機械学習をどのように活用できるか、現在何が研究されているか、について述べる。「探索と利用のトレードオフ」や「帰納推論の限界」の問題、Garbage in, garbage out とされるような「データ自体の質や量への根本的依存性」や「データのバイアス」の問題、そのままでは新規な発見はできない根拠となる「データが疎な外挿領域での低信頼性・予測不可能性」の問題、どういふ変数(記述子)が取り込まれば良いかという「交絡因子の過不足ない同定」や「疑似相関」の問題など、基本問題と事例を整理し、データ駆動な「予測」から「発見」や「理解」に至るための技術的課題を議論したい。

なお、「データ(経験)による法則性の理解」の本質には、「経験論か合理論か」「帰納と仮説形成」などの、太古より人類が世界を理解するために議論してきた科学哲学上の未解決の難問がある。この点は、不毛な議論や空論を避けるために理解しておくべきである。そもそも現象の真の法則が人間に理解可能な単純なモデルに還元できる保証はどこにもないし、データを説明できるモデルは複数存在しうる。現代の実用的シミュレーションモデルの多くは経験的パラメタやデータ同化を伴う複雑なものであるし、開いた系についての発見的表現に過ぎず、対象現象の本質的理解になっているかには議論の余地がある(*Science*, 263(5147): 641-646, 1994)。Von Neuman はパラメタの多い理論を見て「パラメタが4つあれば象を描ける」と揶揄したと言われるが、現代の予測モデルはそれ以上に多数のパラメタを持つ。こうした事実は、科学で現在も未解明の現象や、科学に対する社会ニーズが、陽なモデル化が著しく困難な程には、多様で多因子的で本質的に複雑であることを示唆している。中谷宇吉郎が「科学の方法」で述べたように、科学には本来限界があつて、広い意味での再現可能の現象を、自然界から抜き出して、それを統計的に究明して行く、そういう性質の学問である。この意味でも、かつて「科学の文法」と呼ばれた統計学が、機械学習や人工知能技術と融合した新たな情報技術として、複雑な現象を「統計的に」理解する

新たな手段となり得るのか再考する契機としたい。

略歴: 2004年 北海道大学大学院工学研究科 博士後期課程修了。同情報科学研究科 博士研究員(COE), 京都大学化学研究所 特任助手, 助手, 助教(薬学研究科 兼務), 北海道大学創成研究機構 特任助教(テニュアトラック)を経て, 北海道大学大学院情報科学研究科 准教授。JST さきがけ研究者。機械学習・データマイニング, データ駆動科学, バイオ/ケモインフォマティクス, 材料インフォマティクス等の研究に従事。人工知能学会 人工知能基本問題研究会(SIG-FPAI) 主査 (2018-2019)。

閉会挨拶

伊藤 聡 (物質・材料研究機構)

日本の産業界においては古くからインフォマティクス的な発想が重視されていた。80年代に定量的構造活性相関 (QSAR) が農薬の開発に使われていたし、90年代にかけて、日本の半導体業界が活況を呈すると、多くの技術者は J.C. Phillips の” Bonds and Bands in Semiconductors” を熟読した。90年代といえば、日本製のスパコンが世界を席卷していた時期であり、第一原理計算が身近となっていたにも関わらず、定性的な傾向、簡便な見積といったものが重視されていた。それは個々の数値の裏にある関係性を知ることが設計開発には必要だからであるが、それを知るには当時の計算機は能力が低すぎた。それから四半世紀が過ぎ、計算機の能力は当時に比べて6桁以上向上し、定量的な相関解析が手軽にできるようになってきたが、その意味するところを把握する重要性はますます高まっている。また、品質管理の観点から統計解析が用いられてきたが、そこでは品質保証や歩留まり率の向上を目的としている。そのためには原因説明が成されなければ管理はできない。

本講演会はこのような状況も踏まえて、私たちはインフォマティクス的なアプローチを通して何を知りたいのか、ということテーマとしたものである。「良い」マテリアルズ・インフォマティクスは予測性に優れているだけではなく、理解可能なものでなければならない。それは相関解析と因果推論の関係を意味しているように思われる。もちろん、相関解析と因果推論は全く別の概念であるが、マテリアルズインフォマティクスプロジェクト (MI²I) では今後、その間の橋渡しを行うことで、『予測する MI』から『作る MI』へ発展させていきたい。

本日はご参加いただきありがとうございました。

略歴

1985年筑波大学大学院修了。民間企業を経て、2011年理化学研究所計算科学研究機構、2017年より物質・材料研究機構情報統合型物質・材料研究拠点長。JST イノベーションハブ構築支援事業「情報統合型物質・材料開発イニシアティブ (MI²I)」プロジェクトリーダー兼任。専攻は計算材料科学。