

趣旨説明：データ科学と物理

寺倉清之（物質・材料研究機構）

データ科学における「理解と予測の相克」とは、統計学習の大家の Vapnik による次の言葉に表されている。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」と主張したり。この文言を真とすると、その対偶もまた真であり、それは「理解をしようとするれば、よりよい予測を諦めなければならない」ということになる。

マテリアルズ・インフォマティクスなどのデータ科学による研究では、現実世界を帰納的に扱い、例えば、目的とする量を存在するデータから予測することが行われる。多くの場合、「理解」よりも「予測」がまず狙われる。その上で、「相関関係」が得られ、この相関関係を使って、まだデータが得られていない系に対して高速に予測が行われる。Vapnik の主張は、理解を重視して相関関係を迫及した上で予測するよりは、予測性能を重視して予測する方がよりよい予測ができる、と云うことである。

一方、例えば物理では、観測データを生み出す基本法則（因果関係）を得て「理解」を得ることが重視される。その上で、予測は「因果関係」に則って行われる。正しい因果関係が得られれば、より正確な予測が可能になる。P. W. Anderson は“More is different”で²⁾、複雑な系になってもなお、その複雑系を支配する基本法則があるはずだと主張している。ただし、要素が多数集まった系を支配する法則は、要素の振舞を支配する元の法則とは異なっており、そのことを創発特性という。社会まで含めた自然界は全て、“要素”と“要素の集合体”、という層が順に重なってできた階層構造を成しており、要素の集合体の振舞は創発特性として理解される、というのが“More is different”の主張である。

データ科学で複雑な系の予測をするとき、Anderson が主張するような「階層構造」をより深く考慮して進める、と云うことは理解と予測の相克を克服するヒントになりはしないか？

1) <http://www.learningtheory.org/learning-has-just-started-an-interview-with-prof-vladimir-vapnik/>

2) Science 177, 393 (1972)

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2015年から二年間、その後2018年から同プロジェクトのエグゼクティブアドバイザーを務めた。現在、AIST 名誉リサーチャー・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。

第2回 MI²I・JAIST 合同シンポジウム
データ科学における予測と理解の両立を目指して

水素化物超伝導体研究における
人工知能技術と直観的理解

明石遼介（東京大学大学院 理学系研究科物理学専攻）

近年硫化水素が超高压下で高温超伝導体となることが発見された。この発見の経緯において特筆すべき点は、理論およびシミュレーションによる予測がまず先にあり、それに基づき実験が進められたことである。実験に先立つ予測を可能にした重要な要素として、人工知能技術を用いた結晶構造探索が挙げられる。様々な仮想的結晶構造について高精度の第一原理計算を行い、それらの結果からより安定な候補構造を生成する際に遺伝アルゴリズムや群知能の方法が利用された。こうして得られた候補構造のうちいくつかは、実験で実現していることが実際に示された。類似の事例はランタン水素化物研究においても見られる。本講演では水素化物における高温超伝導の発見において人工知能技術が果たした役割および、その後実験を通して提起された謎を解明するための理論的研究の進展について概観し、これを通して人工知能と人間の直観のスキームの違いおよび両者の協力のありかたについて考察する。

職歴：2014年東京大学大学院工学系研究科博士課程修了。理化学研究所創発物性科学研究センター特別研究員を経て、2014年9月より東京大学大学院理学系研究科助教。専門は第一原理計算に基づく固体物性のシミュレーション。主な研究対象は超伝導および密度汎関数理論基礎。

第2回 MI²I・JAIST 合同シンポジウム
データ科学における予測と理解の両立を目指して

放射光を活用したナノスケール可視化とデータ科学との融合

高橋幸生（東北大学多元物質科学研究所）

「百聞は一見に如かず」ということわざは、視覚情報が人間の判断にとっていかに決定的であるかを物語っている。現に様々な科学は「観る」ことを最大の解析手段として発達してきた。X線を用いたイメージング技術は、厚みをもった試料の非破壊で観ることを得意とし、光学顕微鏡、電子顕微鏡、プローブ顕微鏡とは相補的な関係にある。近年、放射光源・光学技術の進展とともに、X線イメージングの空間分解能は飛躍的に向上しており、シングルナノメートルに到達している。また、複雑化する材料の構造・機能情報を明らかにする必要性から、多次元化されてきており、量的に人間が直接扱える限界に達しつつある。この課題を解決するために、データ科学との連携によりイメージングデータから特徴的な情報を抽出する方法の確立が求められている。本講演では、第三代放射光施設 SPring-8 の放射光を活用した先端的ナノイメージング技術である三次元硬X線スペクトロタイコグラフィ(3D-HXSP)とデータ科学との連携によって実現した酸素吸蔵・放出材料粒子内の酸化反応の可視化について紹介する。また、本年度より建設が始まる次世代放射光施設での展開についても議論したい。

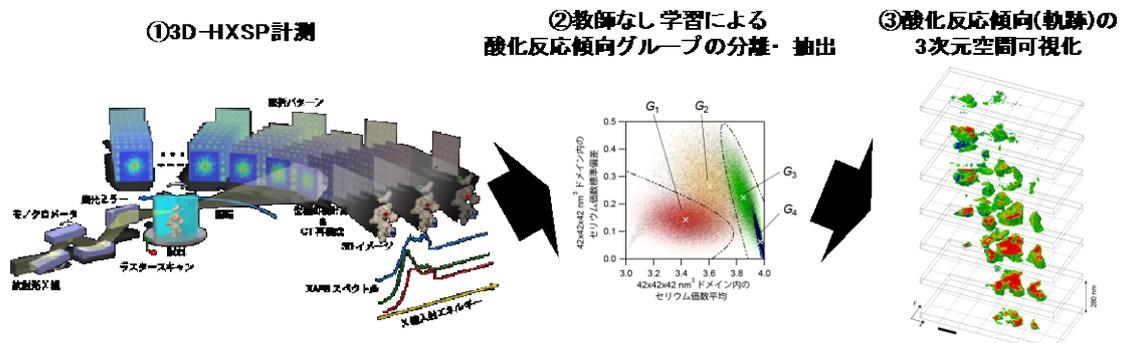


図 3D-HXSP 計測とデータマイニングによる酸素吸蔵・放出材料粒子内の酸化反応軌跡の可視化

職歴：2004年東北大学大学院工学研究科博士後期課程修了、博士(工学)、2004年日本学術振興会特別研究員 PD、2005年理化学研究所 基礎科学特別研究員、2007年大阪大学大学院工学研究科特任講師、2011年同准教授。2014年より理化学研究所放射光科学総合研究センター チームリーダーを兼務。2019年より東北大学多元物質科学研究所 教授。放射光を用いた新規イメージング技術の開発に従事。

第2回 MI²I・JAIST 合同シンポジウム
データ科学における予測と理解の両立を目指して

科学における理解と Tacit Knowing

橋本敬（北陸先端科学技術大学院大学）

人工知能技術は、広大な状態空間の高速な探索を可能にし新奇な戦略や物質の発見・予測に活用されるなど、科学技術の進展に貢献しているが、科学の目的である世界の理解への寄与については、批判的検討が始まった段階だろう。科学で目指す理解の一つは、種々の現象を成立させている、現象の裏にある、因果関係のメカニズムを明らかにすることである。そのようなメカニズムは直接経験できるわけではないので、アブダクションという推論が不可欠となる。科学者による大発見のような知識創造がいかにかき起こるかを考察した Michel Polanyi (“*The Tacit Dimension*”, 1966)は、そこには近位項（諸細目）から遠位項（包括的存在）を生み出す Tacit Knowing（暗黙的認識）という創発過程があると考えた。Tacit Knowing は科学におけるアブダクションのみならず、何かを認識したり技術に熟練したりする際に生じるだろう。だが、科学ではその包括的存在について「要するにこういうことだ」と言い切り、簡潔な命題や数式として表現できるところまで行きたい。それによって累積的知識創造という人類特有の文化進化が実現できる。講演では Tacit Knowing について紹介・吟味し、AI でもそのような創発が生じ得るのか、人間と AI との共創的文化進化はどんな意味で可能なのかを議論したい。

職歴：1996 年東京大学大学院総合文化研究科博士後期課程修了。博士（学術）。1997 年より理化学研究所脳科学総合研究センター基礎科学特別研究員，1999 年より北陸先端科学技術大学院大学知識科学研究科助教授，2009 年より教授，現在に至る。仮想制度研究所 VCASI フェロー，1998 年 SONY Computer Science Lab. - Paris 客員研究員，2001-2002 年 エディンバラ大学言語進化計算グループ客員研究員，2015-2016 年 Telecom ParisTech 情報ネットワーク学部客員研究員。複雑系の観点から言語の起源と進化・コミュニケーションのダイナミクス・制度設計の研究を行い，知識の創造・共有・活用を解明する知識科学の構築を目指す。

予測モデルを"理解"するための技術

原聡 (大阪大学)

概要：

本講演では、予測モデル（機械学習モデル）の判断根拠を理解するための技術について紹介する。高精度な認識・識別が可能な機械学習モデルは一般に非常に複雑な構造をしており、どのような基準で判断が下されているかを人間が窺い知ることは困難である。このようなモデルのブラックボックス性を解消するために、近年モデルを理解するための様々な技術が研究・提案されてきている。本講演ではこれらモデルの判断根拠を理解するための代表的な手法とそれらを取り巻く近年の展開について紹介する。

職歴：

2013.3 大阪大学大学院工学研究科博士課程終了 博士（工学）

2013.4～IBM 東京基礎研究所

2016.4～国立情報学研究所 JST ERATO 河原林巨大グラフプロジェクト 特任助教

2017.9～鷺尾研/大阪大学 助教

特徴選択、機械学習モデルの説明等を研究

第2回 MI²I・JAIST 合同シンポジウム
データ科学における予測と理解の両立を目指して

物理機構の解明を目的とした物質間の類似性の学習

ダム ヒョウ チ（北陸先端科学技術大学院大学 知識マネジメント領域）

自然科学に有効なデータ駆動型アプローチの適用はデータ記述とデータ駆動型アルゴリズムの適性に依存する。一般的なデータ駆動型アプローチでは対象のデータは属性・記述子によって記述されるが、記述子をどう決めるのかと、そして、それを用いて如何にデータ対象を比較するかはデータ駆動型アプローチの最も基本的な課題である。実務では記述子候補を想定し、データ駆動型アプローチの目的や達成度に応じて記述子を選択するのであるが、計算機とは言っても無限の可能性探索を行うことはできないので記述子候補を与える際に専門知識は多いに助けになる。

記述子が決まったとして、データ駆動的に結論を導くにいくつかの方式があるが、直感的に理解し易いのは類推による学習方法である。この場合データ間の比較方法を用いて相違・類似の評価が行われる。物質性質の背景にある物理機構の解明を目指す際に結論を導くために物質間どの側面での相違・類似かが本質的な話である。本講演では特定の物性値に注目した物質間の類似性をデータ駆動的に捉える手法の開発について紹介する。

職歴：2003年北陸先端科学技術大学院大学材料科学研究科博士課程修了。ハノイ大学物理学科講師を経て、2011年4月より北陸先端科学技術大学院大学知識科学研究科准教授。専門はマテリアルインフォマティクス。主な研究テーマは磁石材料およびデータマイニング手法の開発。