

# Knowledge Discovery

## 知識発見

---

### Toshinori Munakata ゲストエディター

我々は、オリジナルであり且つ有用な基本的性質や原理を自動的に見つけ出す技術に関心がある。

**概要：**コンピュータが有用な情報を自動的に見つけ出すという概念は、実用化が意図されているアプリケーションにおいては、魅力的で見込みのある側面である。人類の文明は、それ自身、過去の発見・洞察・考案・刷新を広大に積み重ねてきた結果といえる。その過去の蓄積を分類するといった、創造的人間活動をコンピュータに置き換えることは、壮大な人間の試みを表すものである。本特集号は、知識発見の全体像、もしくは、広い意味で、基本的性質や原理を自動的に見つけ出すためのコンピュータ技術を提供するものである。

本特集号がカバーしているトピックの選定基準は、発見された知識のクオリティと有用性に基づいている。例えば、オリジナルであって、些末的ではなく、基本的であり単純であるけれども、役に立つ知識に関心がある。単なる見込みの話よりは、信用できるある特定のアプリケーション事例を重視している。背景知識や各論文で採用されている技術の一般的説明は、省略するか極力抑えて、他の文献に委ねられている。大方の著者は、また、適用事例や参考文献についてもごく少数に留めている。

アプリケーション事例を通して、読者が知識発見の現状を把握できることを望んでいる。参考文献で説明されている基礎概念やアプリケーション事例を通して、より理解が深まるであろう。

本特集号は様々な問題を取り上げており、論文を分類するだけでも大変な仕事であるが、便宜上、本特集号を三部に分ける。第一部 - 一般問題 - では、知識発見の代表的なものを取り上げる。Tom Mitchell (カーネギーメロン大学) は、記号レベルの機械学習技術に力点を置きながら、データベースにおける知識発見とデータマイニングのアプリケーションや将来の見込みを調査している。彼が説明している、約一万人の妊婦の医学的記録が

らルールを帰納するというアプリケーションは、データベースにおける知識発見の試験的なものである。いくつかの技術が、マネジメント、科学、工学のような領域に適用されるけれども、大量データから影響を与える重要な特徴を抽出するという考え方は、首尾一貫している。

Raúl Valdés-Pérez (カーネギーメロン大学) は、科学の様々な分野を横断して共有される知識発見の側面を主張している。コンピュータは科学的所見を毎日作り出せはしないけれども、重要な科学的発見をサポートできる見込みはある。いくつかの基本概念を説明した後、どのように機械発見を科学に利用することが適切なのかを考察し、さらに、医療、数学、化学におけるアプリケーション事例にも言及している。

第二部 - 選ばれた技術 - では、知識発見のために特に重要な4つの技術を紹介する。Stephen Muggleton (英国のヨーク大学) は、一階述語論理による帰納を調査している。人工知能において、一階述語論理は重要になっているにも関わらず、今日の知識発見アプリケーションにおいては、共通の基盤技術になりえていない。本論文は、アプリケーションを通して、一階述語論理のもつパワーを示している。

混合データタイプのように、分かりやすく、過度には単純化されていない、実世界形式のデータを取り扱っていくことが、知識発見の将来の重要な一つの方向である。

LiMin Fu (フロリダ大学) は、ニューラルネットワークを使用することによって、根本的な知識を抽出するアプリケーションとそのための技術を論じている。人間の脳のアーキテクチャと機能のシミュレーションに基づいて、ニューラルネットワークは、実世界の現象を実際に予測すること、プロセスコントロール、パターン認識などを含む様々なタイプの問題に適用されている。例えば、入力から出力へのマッピング例が大量に与えられた時、ニューラルネットワークは、所与の訓練データだけでなく他の類似パターンのためのマッピングを繰り返して学ぶことができる。ニューラルネットワークは、ビジュアルな画像、オーディオセグメント、様々なタイプの信号を識別することも可能である。しかしながら、ニューラルネットワークをよくよく見てみれば、ネットワーク内の情報は、ニューロンの相互結合に関連づけられた数値の重みの集合にすぎないのである。これらの数値の重みを抽出、あるいは、意味づけを行って、よりハイレベルの知識に変換することが検討されてきているが、引き続き、挑戦のしがいがある課題となろう。

Kenneth De Jong (ジョージメイソン大学) は、遺伝的アルゴリズムとしてよく知られる「進化計算」による知識発見を調査している。惑星地球の五十億年の歴史は、次第に複雑な生命を生み出した。進化計算は、通常、遺伝学や進化論の原理に基づいて、次第に、精練された解を見つける。たぶんいつの日か、他の技術で見つけられなかった新しい知識が、進化計算を通して出現するであろう。De Jong は、進化計算が、よりよい戦略やヒューリスティクス、そして、通常の探索メソッドではまれにしか得られないよりよい解を同定することに役立ったアプリケーションを紹介している。

Wojtek Ziarko (カナダのレジャイナ大学) は、比較的新しくてあまり知られていない技術である、ラフ集合論を使った知識発見のアプリケーションを概説している。しかしながら、ラフ集合論は、不正確、あるいは不完全なデータに関するデータマイニングや推論の

ようなアプリケーションでは、読者の想像以上に広く採用されている。ラフ集合論は、すでに、データベースにおける知識発見の重要な選択肢の一つになっている。

第三部 - 特別なトピック - では、知識発見の特別な適用領域をいくつか紹介する。Kevin Knight (南カリフォルニア大学) は、様々な書式のテキストを処理しながら、自然言語処理における知識発見を論じている。本トピックは、ワードプロセッシング、インターネット・コミュニケーション、ウェブ・ブラウジング、機械翻訳といった今日的なコンピュータアプリケーションにおいて幅広く利用される。さらに、様々なタイプのデータがしばしば大容量になる状況で、知識発見の適用範囲を考えてみれば、本トピックは知識発見における実用上の関わり合いを含んでいるといえる。例えば、未来のワープロソフトは、スペルチェックや文法チェックのようなインテリジェント機能に加えて、知識発見機能あるいはそれから派生した機能を含むことが予想される。

Steffen Schulze-Kremer (ベルリンのマックスプランク研究所) は、知識発見が分子生物学、特に、ヒトゲノム計画においてどのように利用されるかについて論じている。近年、分子生物学の分野は進展しており、生物学データから知識を発見することを支援するためにコンピュータを使うことは、重要で新しいツールを使うことを意味している。逆に、自然に発見された知識は、コンピュータ内で情報がどのように処理されるべきかについてのヒントになるかもしれない。

最後に、Murray Campbell (IBM) は、公式試合でチェスの人間の世界チャンピオンを打ち負かした最初のコンピュータシステムであるディープ・ブルーについて述べている。ディープ・ブルーは、桁外れの計算パワーとチェスに関する知識を統合し、最良の人間とほとんど同じレベルでチェスのコマを動かすことができた。このアプローチは、計算パワーと高い精度で調整された知識に基づき、多くのコマを動かすルールの中で新しい探索戦略を開発するための大きな潜在力を秘めてい

る。本特集号でカバーされている領域や技術は、すべて網羅されている訳ではないことに注意しよう。例えば、統計的手法やデータベース技術といった、データベースからの知識発見に関するいくつかの共通の基盤技術は含まれていない(CACM, 1996年11月号, データマイニング特集を参照)。

## 将来の方向

各論文で示されている通り、知識発見は見込みのある、挑戦のしがいがある分野である。特に見込みのある事項は下記の通りである。

### 背景知識と連想知識の組み込み：

科学的発見からビジネス戦略まで、どのような技術が適用されようとも、単一の生データ源から知識を獲得するには、しばしば限界がある。ある連想知識を組み込むことが、プロセスの効率や発見される知識のクオリティを格段に向上させることがある。

### より分かりやすいタイプのデータ：

現在、大方の知識発見技術は、相対的に単純な形式のターゲットデータを取り扱っている。Mitchell が指摘している通り、ウルトラサウンドイメージや混合データタイプのような、より分かりやすい、過度に単純化されていない、実世界形式のデータを取り扱っていくことが、将来に向けて一つの重要な方向である。

### 人間とコンピュータの相互作用：

人とコンピュータは、それぞれ強みと弱みを持っている。例えば、コンピュータは、信じられない程速くて、バイアスのかからない

処理ができるが、常識や直感に欠ける。ディープ・ブルーは、ハイレベルの計算をやりとりするために行われる、人間とコンピュータの相互作用の好例である。

### ハイブリッドシステム：

本特集で論じられた各技術は、長所と短所を併せ持っている。ハイブリッドシステムは、ある特定のシステムの弱点を補うことが可能であり、知識発見への新しいアプローチを創出するであろう。

私がゲストエディターとして組んだ2つの人工知能の特集が出版された時(1994年3月と1995年11月),多くのインテリジェント技術が、実用的なアプリケーションへ広がり始めていた。今日,“インテリジェントコンピューティング”という言葉は、商品から産業的応用や科学的応用に至るまで、あらゆる所で聞くことができる。21世紀における知識発見の研究とアプリケーションは、近年の人工知能の研究とアプリケーションと同様の経過を辿るであろう。

将来、技術は進化し続け、新しい方法が考案されるであろう。我々は、それが何でいつ登場するかは分からないけれども。

---

Toshinori Munakata  
([munakata@cis.scuohio.edu](mailto:munakata@cis.scuohio.edu)) は、オハイオ州にあるクリーヴランド大学の計算機・情報科学科の教授である。

---

訳：山口高平(静岡大学・情報学部)