

1

学術創成としての知の構造化 —東京大学工学系研究科における試み—

松本 洋一郎

東京大学工学系研究科

巨大化した「情報・知識」の全体像の把握は容易ではない。同時に多くの解決されるべき課題も山積みされている。その解決には「情報・知識」を縦横に活用しなければならない。しかしながら、細分化・複雑化された知識を融合し、課題解決に役立てる合理的・体系的な手法はなく、膨大に蓄積された知識は活用されないままになっている。一方、細分化された学問専門領域内の活動は、飽和・成熟の段階にあり、個々の専門領域内での飛躍的發展よりは、むしろ領域融合型イノベーションへの期待が高い。個別領域における知識基盤の充実が重要であることはいうまでもないが、それ以上に領域間のインタフェーシングによる知の構造化、知識を活用できるネットワーク型知識基盤の構築が重要である。

知のさらなる活用に向けて

我々はあまりに多くの「情報・知識」に囲まれている。1990年代のIT革命以来、「情報ビッグバン」が叫ばれて久しい。あまりに巨大化した情報・知識の全体像の把握は容易ではない。同時に多くの解決されるべき課題も山積みされている。これらの解決には我々が手にしている情報・知識を縦横に活用しなければならない。しかしながら、細分化・複雑化された知識を融合し、課題解決に役立てる合理的・体系的な手法はなく、膨大に蓄積された知識は活用されないままになっているのではないだろうか。

人類は、さまざまな活動によって得た断片的な知見を領域化と領域内基本構造の発見によって知識化してきた。反面、知識の細分化と複雑・高度化は、専門家にとってすら専門領域外との連携が困難であるという状況を生んでいる。同時に、知識は幾何級数的に増大している。その結果、増え続ける膨大な知識の全体を把握することはきわめて困難となっている。一方で、学問は領域を細分化させることによって深化する。専門家は狭い領域を守備範囲としており、領域外の事柄については専門的判断を下すことができない。複雑化する問題を扱うためには、その全体像を把握することが必要であるが、それは個人の能力を超えてしまっていることも多い。一方、学問的専門領域について見てみると、きわめて細分化された専門領域内の活動は、ある意味で、飽和・成熟の段階にあり、今後は個々の専門領域内での飛躍的發展よりは、むしろ学際領域における領域融合型イノベーションへの期待が高いといえる。個別領域における知識基盤の充実が

重要であることはいうまでもないが、それ以上に領域間のインタフェーシングによって知識を構造化し、知識を活用できるネットワーク型知識基盤を構築することが必要であると考えられる。

知の構造化とは、要素と要素の関係性を明らかにすることであり、関係性には、階層性、因果性、関連性、類似性などさまざまな種類がある。知の構造化の目的に応じて相応しい関係性に着目し、分散する膨大な知識を関係付け、知識システムを構築することを知の構造化と呼ぶ。分散する膨大な知識を有効活用し、知的価値、経済的価値、社会的価値、文化的価値に結びつけるためには、知の全体像の把握が必要である。専門家は限られた領域の知識しか把握していないため、分野を越えて知識を活用するためには、知の構造化が不可欠となる。さらに、知を構造化し、可視化することで、知識を活用するためのさまざまな操作が可能となるであろう。もちろん、今までも多かれ少なかれ、こうした認識はすでに存在しており、暗黙のうちに実行されてきた。

このような背景のもと、東京大学工学系研究科では工学知の構造化プロジェクトを立ち上げ、社会技術のための知の構造化、学術創成のための知の構造化、ナノテクノロジーにおける知の構造化などの活動を展開してきた。学術創成プロジェクトでは、原子・分子の素反応から反応装置サイズの流動までのマルチスケールが複雑に相互干渉する「材料プロセスのマルチスケール解析」、さらに広範囲に物質レベルから社会的共通資本までが関連する「地球環境問題」を具体的な例題として、自然言語処理技術を発展的に援用するとともに、知的方法論の体系化手法を用いて、実際に知識の構造化とネットワーク型知識

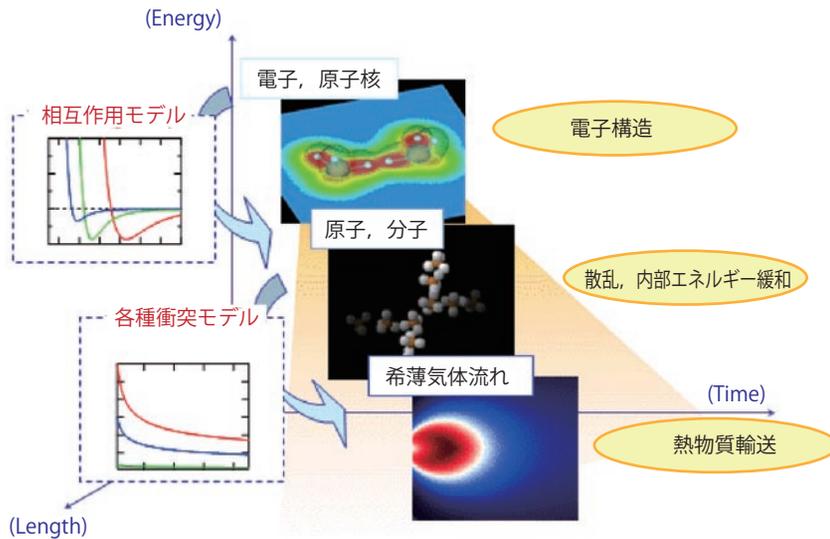


図-1 マルチスケール解析の概念図

基盤の構築と知識誘導システムなどのユーザインタフェースの実装を行ってきた。さらに上記の2つの例題に対する知識基盤構築過程での問題点や新たな手法を抽出し、普遍化した形で整理し学問分野を俯瞰する科学的研究手法として位置づけるとともに、さまざまな分野への適用を図り広く公開している。

さらに、それらの成果を受けて、東京大学では全学における文理協働を視野に入れた「知の構造化センター」を設置し、自律分散的に創造される知識を構造化することにより、知的価値、経済的価値、社会的価値、文化的価値に結びつけるための方法論を構築し、成果を広く社会に実装することを計画している。同時に、「情報の価値化・知識化技術協議会」を立ち上げ、企業、社会、学問の現場で生成される情報が急増する中、現場情報を収集、分析、可視化することで、経営の高度化、社会問題の解決、知識の構造化を促進することが重要との認識から、情報の価値化・知識化に関する技術課題について議論し、技術開発ロードマップ、標準ツールや標準フォーマットの作成・奨励を行い、必要に応じ、研究開発戦略の提言や関連プロジェクトの創出を目的とし活動を進めている。

学術創成プロジェクト

本プロジェクトでは、原子・分子の素反応から反応装置サイズの流動までの多重スケールが複雑に相互干渉する「材料プロセスの多重スケール解析」、さらに広範囲に物質レベルから社会的共通資本までが関連する「環境・社会問題」を具体的な例題として、自然言語処理技術を発展的に援用するとともに、知的方法論の体系化手法を用いて、実際に知識の構造化とネットワーク型知識基盤の構築と知識誘導システムなどのユーザインタフェースの実装までを行った。さらに上記の2つの例題に対する

知識基盤構築過程での問題点や新たな手法を抽出し、他の分野での知識基盤構築が容易に可能となるレベルまで普遍化した形で整理し学問分野を俯瞰する科学的研究手法として位置づけるとともに、さまざまな分野への適用を図り広く公開することを目的とした。

材料プロセスのマルチスケール解析

工学における複雑な事象を、さまざまなスケールの現象が影響しあう多重スケール構造として理解できる場合がある。従来は、重要と考えられるスケールのみを抽出して解析していたのであるが、急激に発達しているナノテクノロジー分野や複雑な現象が交錯する地球環境問題などでは、系の大域的振る舞いすら予想できない場合が多い。そこで、本研究では、材料プロセスにおけるさまざまな多重スケール問題に対して、具体的な問題設定による解析手法の構築と、個々の問題を介して得られた知見による多重スケール解析法に対する俯瞰的検討を行った。なお、材料に関する知識は、製造プロセス、構造・物性・機能、応用と、多様な側面を有する。多重スケール解析での表現が困難な問題については既存の文献情報も合わせ、知識の実装・有機的接続を行った。具体的には、最重要の半導体材料であるシリコンと、代表的なナノ材料である単層カーボンナノチューブを対象に広範な知識の構造化を試みた。

シリコンに関しては、代表的な製造プロセスである化学蒸着(CVD)法を対象に、量子力学から連続体力学に至る広範な物理現象を合理的に解析するマルチスケール解析手法を構築し、理論・実験両面での知識創成を行った。図-1にその概念を示す。まず、分子軌道計算によって多原子分子間の分子間相互作用のモデル化を行い、そのモデルを用いて、分子スケールの現象の解析が可能となる。さらに、分子衝突に関しては、古典軌道計算による

衝突断面積および散乱角モデルの構築を行い、さらにこれらの結果を直接シミュレーションモンテカルロ法に分子衝突モデルとして取り入れることによって、電子構造からマクロな熱流体構造を結びつけるマルチスケール解析の手法を確立した。多重スケール解析によって、希薄気体流れを3つのスケールで分離し、量子計算からボルツマン方程式レベルの現象までを合理的に結びつける手法を開発した。また、実際のCVD炉のシミュレーションまでを実行した。一方、実験的には、成膜初期の成膜遅れ現象、核発生・成長によるシリコンナノ粒子形成、多結晶連続膜での結晶配向発現等の機構解明を進め、文献情報と合わせ、製造プロセスとシリコンの形態・構造の関係を明らかにした。また、物性評価のための多重スケール解析を目指して、ワニ関数表現やクリノフ部分空間を用い、原子数の増加に伴う計算負荷を極力抑えた新しいオーダーN法を4種類開発して、タイトバインディング近似における大規模な系の計算を実証した。

単層カーボンナノチューブに関して、多重スケール解析による触媒からの成長のシミュレーションを実現し、そのメカニズムに基づき、実験的に高効率なCVD法の開発につなげた。特に、アルコール触媒CVD法の開発は、高純度の単層カーボンナノチューブを容易に合成できるようにした画期的な成果である。単層カーボンナノチューブ合成のもう一つの鍵は、触媒となる金属ナノ粒子である。種々の材料のスパッタ法での構造形成の実験的検討により、金属ナノ粒子の自己組織形成の機構解明も進めた。CVDの高温でナノ粒子を自己組織形成させるコンセプトのもと、触媒金属の担持量分布を1枚の基板上に網羅的に実現する簡易なコンビナトリアル手法を開発、さらに実際に1回の実験での触媒探索にも成功している。これはプロジェクト内の知識の深い共有・構造化により、発明・発見が加速できるという実例でもある。

本プロジェクト開始後、さまざまな研究分野で多重スケール（マルチスケール）のキーワードが使用されるようになってきている。論文検索サイトであるISI Web of Knowledge, Web of Scienceにおいて、多重スケールをキーワードに持つ発表論文数を調べると、5年前に比べ約3倍、10年前の約7倍、研究代表者が多重スケールという単語を使用し始めた15年前と比較すると、40倍にも膨れあがっており、多重スケール解析の分野で本研究が先導的役割を果たしてきたことを示唆している。また、多重スケールに関する研究は比較的新しいものとなるため、さまざまな分野にわたる多種多様な多重スケール問題に対して、これまでは、個別に多重スケール解析を行ってきた。これに対し、本研究では、一歩進んで、種々の多重スケール解析に対して、知識の構造化の概念に基づいて統一的な解釈を試みた。

環境問題における知識の構造化

環境・社会問題に関して、複雑な要因の交錯した問題を解決するために知識の構造化と俯瞰が緊急に求められている。本研究では、「地震防災を中心とした安全にかかわる社会問題」、および「水環境にかかわる問題」を対象とした。環境・社会問題の解決策設計のプロセスは、1) 問題の分析、2) 問題解決策の立案、3) 問題解決策の影響分析、4) 問題解決策の評価からなる。このプロセスのさまざまな局面において、さまざまな分野の膨大な知識を有機的に結合する知識の構造化の有効性が以下の方法によって実証された。

地震防災にかかわる社会問題では、既存の知識を1,000項目に整理して構造化した。それらをコーンツリーに基づく3次元情報可視化ツールを開発し、回転・拡大等の操作を通じて情報構造・問題の全体像を把握することを可能とし、安心・安全の立場から知識の構造化を試みている。この手法を用いることにより、大規模地震において多くの人命を救うためには、他のインフラ設備（橋梁、道路、その他）と比較して「既存不適格住宅の解消」が最重要な課題であることが導き出されることを示した。また、問題解決策の立案に必要な知識の構造化を検討した。これらは本特集の中で、堀井秀之氏により「社会問題解決のための知の構造化」として詳述されている。

水環境に関して、飲料水の微生物学的安全性をとりあげ、水を供給する立場から需要者へ安全性を説明するために必要な知識を構造化することを試みた。図-2に示すように、さまざまな要素がきわめて複雑に関連し合っていることが理解される。水道の安全については、そのリスク因子の特性に応じて、特徴的な知識の関係性が見出された。また、新規のリスク因子に対する最新の研究成果、水処理技術、水質基準の相互関係について整理した。安心・安全を脅かすリスクの特性と、その対策の特性に応じて相関関係を構造化し、「特性と対策の対応関係」を明らかにした。その対応関係に基づき、新たなリスクに有効な対策を見いだすことが可能となる。一方、今後の発展が期待される水処理技術に関連して、曝気槽内気泡流動の連続体スケール方程式におけるスケール分離を行い、分散相の挙動と連続相の挙動を合理的に接合する多重スケール解析手法を開発した。具体的には、微細気泡に対する表面活性剤の影響と気泡に働く力の解析、微細気泡発生システムの開発、縦型2次元チャンネルおよびテスト用曝気槽内における流動構造の計測に関する研究を通して、詳細なマルチスケール解析を行い、ミクロスケールの現象の制御によりマクロスケールでの性能を大幅に改善できることを示した。また、これらマルチスケール構造を持つ流れを高精度で解くための新しい計算手法の開発を行った。膨大な知識が複雑に関連する環

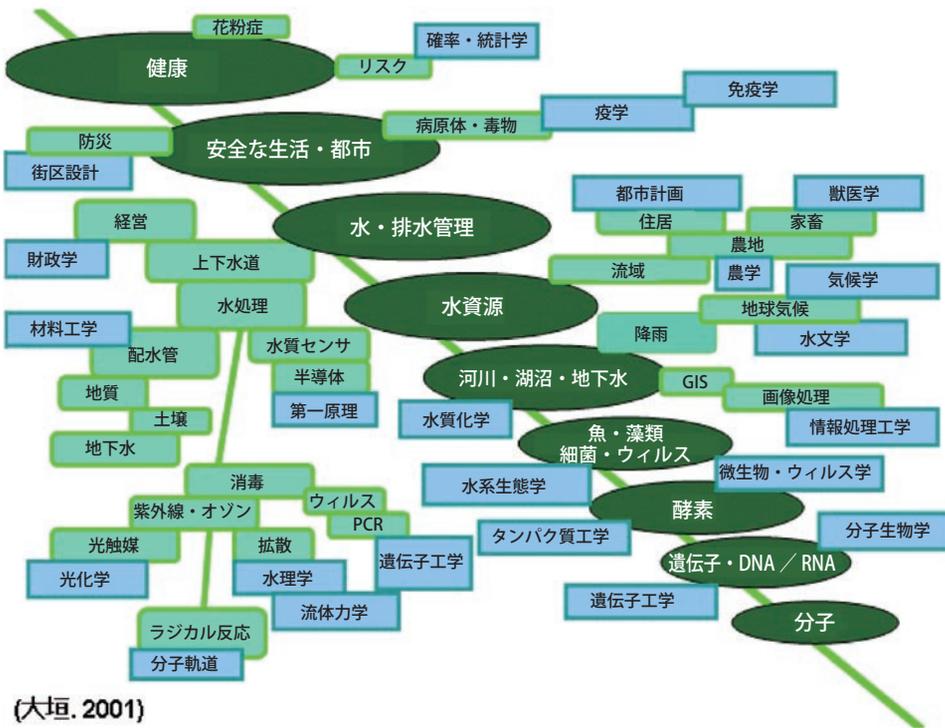


図-2 水処理に関連する知識群

(大垣, 2001)

境・社会問題においても、キーとなる要素については多重スケール解析が有効であるモデルを提案できた。

自然言語処理と知の構造化システム

知識構造化における情報処理技術の開発としては、言語テキストからの知識構築とそのアクセス手法に関する研究や知的方法論の工学における事例の抽象化と枠組みの設計などを進めるにあたり、ゲノム生物学分野での文献から情報抽出するシステムのための背景知識（オントロジー）を整備し、メドラインというこの分野のアプリケーションのデータ集合を例として、巨大な論文集合から有効な知識・情報を取り出し、それをほかの知識（たとえば、多種多様な外部データベース）と結びつけるテキストマイニングの手法を確立した。構造的な言語処理手法からの意味の処理と確率モデル・機械学習という、2つの手法を融合し、巨大テキストからの知識構造化手法を研究した。すなわち、膨大な知識源から有用な知識を効率的に収集し、ユーザが真に必要な知識を分かりやすい形で提示するシステムを構築するために、以下の要素技術を開発した。具体的には、言語処理と知識処理、視覚化の研究等、複数の研究を有機的に統合した基盤技術を確立することを目指した。また、これらの要素を統合、および実データを用いたシステム運用実験、ならびに知識獲得実験を行い、本技術が知識の効率化な活用を推進できることを示した。

HPSG による言語解析システム (Enju)：言語学的に妥当な文法に対する確率モデルの定義に成功し、妥当な解析結果をスコア付きで出力できる言語解析モデルを

確立した。また、コーパス指向の文法開発の方法論を提唱し、汎用な言語モデルを現実のコーパスに合わせて学習、精緻化する手法を確立した。これらの理論を実装した本システムは、深層意味構造の計算が可能で、かつ、現実のテキストを高速度で解析できる世界最初のものである。

自動用語抽出エンジン (TermEngine)：膨大なテキスト情報に対して、概念の包含性を基に用語としての妥当性を判断する「拡張 C/NC-value 手法」、および平均相互情報量を用いた用語の自動分類を行い、オントロジー情報を自動構築するエンジンの開発に成功した。これにより、用語を構成する形態素や構文のゆれ、また意味的なゆれを吸収し、テキストからの統合的な知識抽出が可能となった。本システムは日英両言語のテキストの解析に辞書の切り替えなしに対応できる。また、両言語共に同手法で高い精度の認識が行える言語非依存性を特徴とする。

知識類似度計算エンジン (SimEngine)：TermEngine により認識された用語とその分類の情報を用いて、知識間の意味的関連性を定量的に計算するためのエンジンの開発に成功した。本エンジンは、高度な拡張性を有し、計算手法を独自に定義、プラグインすることで容易に拡張が行える機構を備えている。

関連知識の視覚化エンジン (VizEngine)：SimEngine により得られた知識間の意味的関連性を基に、知識の関連性を視覚化し、全体のネットワーク構造を俯瞰可能とするエンジンの開発に成功した。本視覚化手法では、グラフ理論を基に、ドキュメント等の知識単位をノー

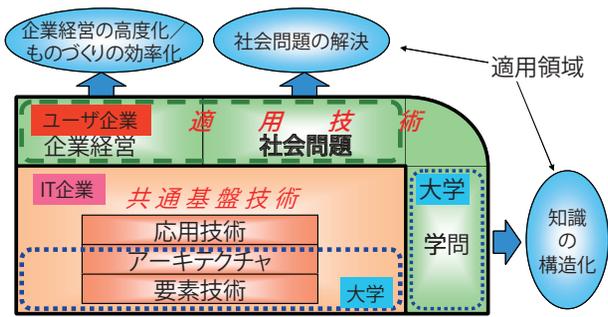


図-4 協議会において検討する共通基盤技術/適用技術と適用領域の関係

方法論を確立する。

- ・ **構造化手法研究開発**：自然言語理解，オントロジー構築，関連性認識などの技術の深化を図り，概念だけでなく，現実の時空間で変化していく情報を効果的に構造化・可視化する新たな手法を開発し，実用化を図る。事実情報だけでなく，感性的な評価を含めた情報を構造化・可視化するために，機械学習手法を活用し，自然言語処理技術の新たな展開を図るとともに，認知・コミュニケーション科学の知見を踏まえ，インタラクティブに知識を獲得，更新していく手法の確立を目指す。さらに，構造化知を目的に応じて提供するための共通知識基盤を開発する。

情報の価値化・知識化技術協議会

多様な環境問題への対応，安心・安全な社会の実現等，解決すべき多くの課題に直面している。科学技術は国家・社会・企業・個人等によるイノベーションの創出に大きく貢献し，複雑化する課題解決に資することが，従来にも増して，期待されている。特に高度情報通信ネットワーク社会を実現することを目標に e-Japan 戦略が進められ，世界有数のブロードバンド環境を得た我々は，社会基盤としての IT インフラをさらに活用し，新たな価値やイノベーションの創出に貢献する科学技術の発展を目指す必要がある。本協議会は，「情報の価値化・知識化技術」に関する産官学連携の場を提供することを目的に設立されたものである。実世界のあらゆる情報をモニタリングし，ダイナミックな情報分析・統合を行い，変化検知・予測・情報可視化・統合する「情報の価値化・知識化技術」は，上記課題を解決するための基盤技術の1つである。産業・学問・行政といったすべての社会活動の現場で急増する情報を収集・分析・統合・可視化することにより新たな知を連続的に生み出し，その協調サイクルの中で個々の活動の価値が高まることが期待される。すでに上記で述べたように，大学には「情報の価値化・知識化技術」を構成する要素技術が蓄積されてきて

いる。これらを真に活かすためには産学の密接な連携が必要である。先端技術の適用領域を設定し，産学が保有する要素技術をすり合わせることや，産業界のニーズを大学に伝え事業化を見据えた研究開発を進めることが重要である。本協議会で検討を行う，共通基盤技術/適用技術と適用領域の関係を図-4に示す。協議会においては，あらゆる情報の価値化・知識化を進め，それを活用するため，必要な共通基盤技術，適用技術，適用領域と取り組むべき課題の検討を行い，さまざまな具体的適用シーンを想定し，技術開発ロードマップを作成する。また，情報の価値化・知識化技術の普及を加速するため，情報や知識収集・活用を容易にする標準ツールや標準データフォーマットの策定やその奨励を行う。

「知の構造化」推進活動への参加を

領域化によって深化した知識は，非専門家はもとより専門家にとっても受容しがたい状況となっており，複雑化した知識の活用のためには，領域間の「インターフェシング」によって「知識を構造化」する必要があるといわれて久しい。また，学際領域における領域融合型イノベーションへの社会からの期待も高い。

東京大学工学系研究科では「知の構造化プロジェクト」を展開し，「知識の構造化」を実践し，「ネットワーク型知識基盤」の構築を試みてきた。その成果を受けて，東京大学では，文理協働を視野に入れた「知の構造化センター」を構想・設立し，自律分散的に創造される知識を構造化することにより，知的価値，経済的価値，社会的価値，文化的価値に結びつけるための方法論の構築に向けた活動を開始している。さらに，知の構造化によって得られる学術的成果を社会に実装するために，「情報の価値化・知識化技術協議会」を発足させ，産学連携の下に活動を行っている。多くの方々の参加を期待したい。

参考文献

- 1) 小宮山宏：知識の構造化，オープンナレッジ。
- 2) 松本洋一郎，小宮山宏監修：知識・構造化ミッション，日経 BP 社。
- 3) 学術創成のための知識の構造化とネットワーク型知識基盤の構築，<http://www.fel.t.u-tokyo.ac.jp/chisiki/>
- 4) 知の構造化センター，<http://www.cks.u-tokyo.ac.jp/>
- 5) 情報の価値化・知識化協議会，<http://chishikit.u-tokyo.ac.jp/>
(平成 19 年 6 月 10 日受付)

松本洋一郎 ymats@fel.t.u-tokyo.ac.jp

東京大学大学院工学系研究科教授。1972年東京大学工学部機械工学科卒業。1974年同大学院工学系研究科機械工学専門課程修士課程修了。1977年同博士課程修了。工学博士。同年同大工学部講師。1978年同助教授。1992年同教授。2006年同工学部長，工学系研究科長。