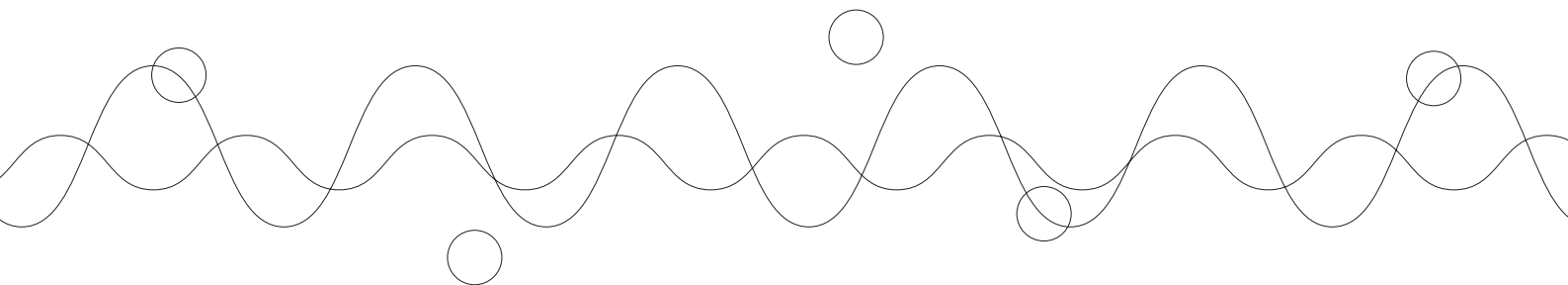


科学技術未来戦略ワークショップ報告書

「予測と発見」

大規模情報からの『知識』獲得技術



独立行政法人科学技術振興機構 研究開発戦略センター
Center for Research and Development Strategy Japan Science and Technology Agency

要 旨

独立行政法人科学技術振興機構（JST）研究開発戦略センター（CRDS）では、大規模な情報から、ユーザの目的にかなった『知識』を獲得するための情報技術に関して検討を行うため、「予測と発見－大規模情報からの『知識』獲得技術－」ワークショップ（平成 19 年 1 月 18 日、19 日）を開催した。

コンピュータを中心としたハードウェア・ソフトウェアの高度化により、数理工学的手法が実社会へ応用される可能性が従来よりも高まってきていることをふまえて、コンピュータサイエンスの分野と統計・数理科学の分野及びその応用分野の有識者にお集まりいただいた。コーディネーターを九州大学副学長の有川節夫教授にお引き受けいただき、有識者及び行政関係者 36 名、その他 JST 関係者 16 名の合計 52 名にて議論を行った。

初日の午前中に大規模情報から獲得されるべき『知識』がどのようなものであるかについて討論を行い、参加者全員の間で一定の認識を共有した後、3 つの分科会に分かれて『知識』獲得技術に関して以下の項目に関する集中討議を行った。

- ① 応用分野と目指すべき目標
- ② 研究開発課題
- ③ 研究推進方法
- ④ 獲得されるべき『知識』

研究開発課題の討議では、大規模情報から目的とする『知識』を獲得するための技術要素を、「取得」「データ構造化」「『知識』創出」「ヒューマンインタラクション」の技術要素に分類し、それぞれの技術要素の現状を俯瞰した上で、今後必要となる研究開発課題とその推進方法について議論を行った。概要は下記のとおりである。

今後研究開発すべき重要な研究開発課題として多くの研究課題が挙げられた。これらを、「取得」「データ構造化」「『知識』創出」「ヒューマンインタラクション」の技術要素にまとめると、下記のように概観できる。

- データ取得に関する課題
 - ・ モデルに基づいたアクティブセンシング
 - ・ ユーザのインテンションの記述方法
- データ構造化に関する課題
 - ・ 多分解能の時空間データを効率よく保持できる構造の設計技術
- 『知識』創出に関する課題
 - ・ ベイズモデル構築技術・ベイジアンネットワーク技術
 - ・ 超多次元・超ヘテロ・超大規模データマイニング技術（高速アルゴリズムを含む）
 - ・ 数理統計と機械学習の融合

■ ヒューマンインタラクションに関する課題

- ・ ジャストアシスト技術

■ その他

- ・ サイバースペース天文台

なお、『知識』獲得の技術要素の高度化だけでなく、統合可視化研究基盤システム、自律的常識知・身体知獲得技術、シナリオ・マイニング技術等、技術要素を統合して、実社会でのサービスを実現するための技術も非常に重要な課題であることが指摘された。

研究推進方法については、下記提案がなされた。

- 大規模データが生み出されるフィールドに密着した現場主義を重視すべき
- 統計解析に基づく方法論と機械学習に基づく方法論の相互協力により推進すべき
- 共通プラットフォームを構築して、グランドチャレンジ型の研究が推進されるべき

情報技術によって獲得されるべき『知識』についてまとめると、下記のようなになる。

『大規模情報から獲得される知識とは、個々のシステムや個人がもつビューを満たし目的達成のために使えるように、集積した情報を出発点にして一般化し、既知の知識や蓄積した経験・情報に照らし集約して得られるもので、この過程はスパイラルに繰返し適用され再生産される。』

大規模なデジタル情報への対応については、大規模な情報の中から求める情報を探し出す技術の高度化や、大規模な情報を支えるデータベースの高度化、情報の信憑性を検証するための技術開発等、総合科学技術会議もフォローアップしている経済産業省、文部科学省、総務省の3省連携戦略プロジェクト「情報集積活用基盤技術」が進行中である。これらは、インターネット上の情報検索、分析、管理を対象にした施策であり、大量情報からの価値の創出という面では共通点があるが、既存の技術を応用し、あるいは応用技術を実用化することを主眼にしたものがほとんどである。

したがって、このワークショップで検討しているように基礎研究からスタートして体系的かつ組織的な研究開発活動を展開し、現在の諸課題に答え、次世代のICT社会に求められる理論・技術を追究し、次世代の発展ための基礎を築いておこうとするものとは大きな違いがある。

なお、本報告書は、コーディネーターの有川節夫（九州大学副学長）、分科会リーダーの北川源四郎（統計数理研究所長）、有村博紀（北海道大学教授）、及び田中譲（北海道大学教授）の各位に執筆いただいたものを、CRDSでとりまとめたものである。

目次

1. ワークショップの目的	1
2. 大規模情報からの『知識』獲得技術について	5
2.1 『知識』の定義をめぐって	7
2.2 日本における『知識』獲得の研究の経緯	8
2.3 世界の動向	9
2.4 『知識』獲得の対象・手法	10
2.5 ワークショップの事前に整理した『知識』獲得関連の技術と研究開発の方向性	10
3. 大規模情報からの『知識』獲得技術に関する検討	23
3.1 個人のニーズに応じた『知識』獲得：パーソナライゼーション技術（分科会A）	25
3.1.1 応用分野と目指すべき目標	25
3.1.2 研究開発課題	26
3.1.3 研究推進方法	30
3.1.4 大規模情報から獲得する『知識』とは	30
3.2 ヘテロで大規模なデータからミクロな規則を発見する技術（分科会B）	35
3.2.1 応用分野と目指すべき目標	37
3.2.2 研究開発課題	38
3.2.3 研究推進方法	42
3.2.4 大規模情報から獲得する『知識』とは	42
3.3 サイエンス知・常識知・身体知・シナリオ知の獲得技術（分科会C）	50
3.3.1 応用分野と目指すべき目標	50
3.3.2 研究開発課題	52
3.3.3 研究推進方法	62
3.3.4 大規模情報から獲得する『知識』とは	62
4. 全体討議のまとめ	67
4.1 各分科会への要請と対応	69
4.2 対象とする『知識』の多様性	69
4.3 大規模情報から獲得する『知識』とは	70
4.4 今後研究開発すべき重要な研究開発課題	71
4.5 研究開発課題の推進方法	72
4.6 他のプロジェクトとの違い	73
付録1 ワークショッププログラム	77
付録2 参加者一覧	79
付録3 分科会メンバー構成	81

1

ワークショップの目的

1. ワークショップの目的

独立行政法人科学技術振興機構（JST）研究開発戦略センター（CRDS）は、科学技術の研究分野を俯瞰的に展望し、今後重要となる研究領域、課題を系統的に抽出し、社会ニーズの充足と社会ビジョンの実現に向けた研究開発のファンディング戦略を立案・提言している。その活動の一環として、重要研究テーマについて専門家によるワークショップを開催している。

現在の社会においては、センシング技術の高度化、インターネット等のインフラの高度化などにより、大規模な情報へのアクセスが容易になってきている。このような大規模情報の取得・蓄積は、サイバー世界だけでなく、実世界においても進展しており、これら大規模情報から有用な『知識』を発掘・獲得する技術の重要性は今後増大すると考えられる。

具体的な応用分野は、気象、重要インフラ、防災、医療、創薬、マーケティング、金融、日常生活など多岐にわたる。膨大な情報から有用な『知識』を獲得する技術体系を構築していくことにより、新たな製品・サービスを創出・高度化したり、大災害・大事故に至る前に有用な情報を重要インフラやプラントの事業者提供したり、ユーザ指向のマーケティングを実現できるようになる。こうした技術は今後、安全で信頼性が高く強靱な（ディペンダブルな）国、産業力を持った国の実現に貢献できる。

今回のワークショップでは、5年～10年先を見据えて『知識』獲得プロセスの高度化を目指した研究開発戦略について検討するため、この分野の研究の現状を俯瞰し、重要な研究課題を抽出し、その研究推進方法を時間軸まで含めて明確化することを目的とした。

2

大規模情報からの 『知識』獲得技術について

2. 大規模情報からの『知識』獲得技術について

有川 節夫（九州大学）

このワークショップの開催にあたり、大規模情報からの『知識』獲得技術に関するこれまでの研究開発の状況、現状及びこれからの方向や可能性について、網羅的ではないが、簡単に説明しておく。

まず、大規模情報からの『知識』の獲得技術とは、大量の情報から人間にとって有用な『知識』を見つけるための技術である。これは、1994年のAgrawalの結合規則の発見に関する研究を契機に発展したデータマイニングを中心に、機械学習や数理統計学、データベース技術にも深く関連した技術である(図2-2)。1990年代の高速ネットワーク、計算機の高性能化、大規模記憶装置、センシング技術の発展により、『情報爆発』ともいわれるような人間の能力を超えて増大する情報へのアクセス手段として不可欠なものになっている。また、こうした情報通信技術(ICT)に支えられた知識基盤型産業等に見られるように、社会システムとしても的確なデータや情報、知識に基づく判断が必要になっている。(図2-3)

2.1 『知識』の定義をめぐって

こうした『知識』について共通の理解を得るために、その定義を明確に定める必要がある。しかし、その定義については、データや情報のような概念についても、また数学や情報学のような学問分野についても同様であるが、

一定の明確で普遍的なものを示すのは困難であり、時には無意味でさえある。それにも拘らず、共通の認識とイメージを描き、課題を共有する必要があるので、ここで、現在考えられているいくつかの定義を簡単に紹介しておく。この二日間の議論において、『知識』という概念の定義を絶えず意識して議論していただき、ワークショップの終わりにはある程度の形が見えてくるようにしていきたいと考えている。

知識は、情報やデータと比較あるいは並んで議論されることが多い。広辞苑から引き出したものやその他の表現について、パワーポイント資料(以下、PPT資料と略す。)に示している(図2-4～図2-8)。こうして並べて見ると、『知識』の定義がある程度浮び上がってくる。Googleで、“知識とは”で引くと約14万3千件ヒットするが、『獲得』と共起するものが、約2万8千件、『発見』と共起するものは、4万7千件見つかる。なお、このワークショップを「予測と発見」としているが、英語の場合、“Knowledge Acquisition”と“Knowledge Discovery”については、ほぼ同数で約110万件ヒットするが、我々の関心事に近い“Data Mining”との共起ということで絞ると、『発見』の方が圧倒的に多くなる。

ウェブ上の百科事典「Wikipedia」には、かなりしっかりした知識の定義がある(図2-8)。この中でも暗黙知(tacit knowledge)、すなわち「知識のうち、

勘や直観、個人的洞察、経験に基づくノウハウのことで、言語・数式・図表で表現できない主観的・身体的な知」についての記述もあるが、これは、ある程度、機械学習やデータマイニングにより顕在化できると考えている。

2.2 日本における『知識』獲得の研究の経緯

次に、我が国における知識獲得に関する研究展開の歴史や経緯について、国内の多くの研究者が同じ目的意識でかかわった国のプロジェクトといえる文部（科学）省の科学研究費補助金による重点領域研究と特定領域研究に焦点を当てて振り返っておこう（図 2-9）。

知識獲得ということばは、1991 年度～1993 年度に展開された重点領域研究「知識科学における概念形成と知識獲得に関する研究（略称、知識科学）」において、知識科学という新しい視点を伴って知られるようになった。ここでは、大規模情報からの知識獲得だけでなく、推論や学習等による概念形成についても、人間によるものと機械によるものの双方から追究されている。この成果は、領域代表者の大須賀節雄先生と 5 つの研究班の班長の編集による単行本として出版されている（図 2-10）。ここでの「知識科学」ということばが、その後、北陸先端大の知識科学研究科へ繋がったともいわれている。

筆者（有川）は、大須賀先生の「知識科学」において「機械学習と機械発見の理論」の班を担当し、各種推論や帰納推論の理論と応用の研究を推進してきた（図 2-11・図 2-12）。いくつかの重要な理論的成果やバイオインフォマティクスのさきがけになる研究成果

が得られた。その中で、「巨大学術社会情報からの知識発見に関する基礎研究」が必要で、そのためには、「発見科学」という新しい永続性のある情報学の研究領域を確立することが喫緊の重要課題であると確信し、何回かのトライアルの後で、重点領域（スタート時点では、特定領域と改称）に採択され、1998 年度～2000 年度に亘って研究を推進してきた。科学哲学から論理、推論、学習、統計解析、巨大データベースからの知識発見技術、ネットワーク環境からの知識発見に至るまでの研究を、当時の日本の一線の研究者約 60 人の参加を得て、総合的かつ組織的に展開してきた。その成果は、Springer 社から“Progress in Discovery Science”として出版されている。

この特定領域研究の「巨大データベースからの知識発見」班の中心的研究者であった元田教授（当時、阪大）を領域代表とする「情報洪水時代におけるアクティブマイニングの実現」という実際の領域の課題に直接的に取り組む統制の利いた特定領域研究が 2001 年度から 4 年間に亘って展開された（図 2-13・図 2-14）。アクティブ情報収集やユーザ指向アクティブ・データマイニング、アクティブ・ユーザリアクション等の課題を明確に設定し、メンバーが分担協調して取り組み多くの具体的な成果を得て、特定領域研究の中でも高い評価を得て成功裏に終了した。

一方、2001 年度には、「国策の」特定領域研究（C）という枠組みで、科研費による情報学関係の研究では空前の規模で、「IT の深化の基盤を拓く情報学研究」が安西教授（慶応・塾長）を領域代表とし国立情報学研究所が強く支援した形で 5 年間に亘って展開された。これは、特定領域研究の模範とな

るような研究班の組織化と活動を展開したとして極めて高い評価を得ている。この特定領域研究の五つの柱のうちの一つが、国立情報学研究所の安達教授に率いられた大規模情報からの知識獲得に関するものであり、東大の喜連川教授や北大の有村教授を中心にウェブ上のテキストデータや半構造データ等を対象にして、高度な研究が展開されている。

この特定領域研究はスタート後に、通常の特定期域研究の枠組みに組み込まれた。そして、実質的にその後継といえる大規模の特定期域研究が 2005 年度から 5 年間の予定で喜連川教授を領域代表として「情報爆発時代に向けた新しい IT 基盤技術の研究」として大々的に展開されている。その中のひとつの班でこの知識獲得技術に関する研究が主としてネットワーク上の大規模情報を対象にして扱われている。

2.3 世界の動向

次に、世界の動向について、データマイニング (DM) とその基礎にもなる機械学習 (ML) 関係の国際会議の開催状況を PPT 資料に示している (図 2-15・2-16)。

DM に関しては、米国情報処理学会 (Association for Computing Machinery, ACM) の知識発見とデータマイニング国際会議 (Knowledge Discovery and Data Mining, KDD) が 1989 年から続いていて活況を呈している。また、ヨーロッパの欧州機械学習会議 (European Conference on Machine Learning, ECML) と欧州データマイニング国際会議 (European Conference on Principles and Practice of Knowledge Discovery, PKDD) が毎年合同で開

催され、環太平洋地域では環太平洋データマイニング国際会議 (Pacific-Asia Conference on Knowledge Discovery and Data Mining, PAKDD)、他に、日本発の先に紹介した発見科学プロジェクトの推進時にスタートした日本発の発見科学国際会議 (International Conference on Discovery Science, DS) がある。2000 年に入りスタートした、米国電気電子学会 (IEEE) の International Conference on Data Mining (IEEE ICDM) や米国応用数学会 (Society for industrial and Applied Mathematics, SIAM) の SIAM データマイニング国際会議 (SIAM Conference on Data Mining, SDM07) も定着している。

機械学習に関しては、機械学習国際会議 (International Conference on Machine Learning, ICML) が一番古く 1984 年から毎年開かれている。理論を中心にした国際会議も、アメリカでスタートした計算学習理論国際会議 (Computational Learning Theory, COLT) や日本発の計算論的学習理論国際会議 (International Conference on Algorithmic Learning Theory, ALT)、ヨーロッパの欧州計算学習理論会議 (EuroCOLT, 現在に COLT と合併)、帰納的論理プログラミングの機能的論理プログラミング国際会議 (International Conference on Inductive Logic Programming, ILP) 等、数多くの国際会議が毎年開催されている。また、データ管理国際会議 (ACM SIGMOD) や大規模データベース国際会議 (Very Large Data Bases, VLDB) 等のデータ工学に関する伝統ある国際会議でもセッションが設けられていて、活況が持続している。

2.4 『知識』獲得の対象・手法

獲得・発見の対象や手法について、最近の内外の資料を簡単に PPT 資料 2 枚にまとめている (図 2-18・2-19)。

まず、ACM の Special Interest Group on Knowledge Discovery and Data Mining (SIGKDD) からの DM についてのカリキュラムの提案書 (2005 年) には、発見する知識の例をいくつか挙げている。対話的なくり返し過程としての知識発見、異種類のデータを対象として、異種類の知識を発見すべきこと、知識の評価尺度、多様なアプリケーション等が示されている。

また、情報処理学会誌では、2005 年に特集を組んで、発見する知識や発見手法の例を組織的に解説している。ストリームマイニング、データスカシング、グラフマイニング、構造カーネル法、統計的異常検出、自然言語処理におけるマイニング技術応用、大規模ウェブアーカイブからの DM 等が挙げられている。

2.5 ワークショップの事前に整理した『知識』獲得関連の技術と研究開発の方向性

このワークショップを企画するに当たって JST 研究開発戦略センター (CRDS) の担当者がまとめた「知識獲得の観点からの関連技術の概略俯瞰」(図 2-20) と「知識獲得プロセスと求められる研究開発の方向性」に関する PPT 資料 (図 2-21) を示している。このワークショップでは、『知識』とは何かについての認識を深め、共有することに加えて、これらの技術俯瞰図や重要課題について見直しを行い、それを詳細化すること、研究の推進方法、投資する意義 (社会へのインパクト) 等についての具体的なアウトプットが求められている。なお、CRDS の担当者による研究者に対するインタビュー「大規模情報ハンドリング技術が求めるニーズの例」に関する PPT 資料も一枚つけている (図 2-23)。



オリエンテーション

大規模情報から獲得する 「知識」とは

有川 節夫(九州大学)

2007年1月18日
CRDS/JST



図 2-1

大規模情報からの知識獲得とは

- 大量の情報から、人間にとって有用な知識を見つけるための技術
- データマイニング研究
 - 結合規則発見の研究 [IBM, Agrawal '94] を契機に爆発的に進展
 - 機械学習・統計学・データベースの境界分野
 - 1990年代後半以降, 理論と応用の両面で活発に研究.
- 人間の能力の拡大
 - 人間の能力を超えた情報爆発への対応策の一つ
 - 大規模データへの新しいアクセス手段としての位置づけ



2007年1月18日

JST CRDS ワークショップ

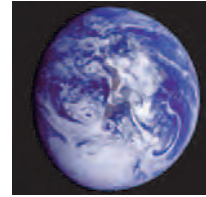


図 2-2

知識獲得支援の必要性

● 大規模データの出現

- 1990年代の高速ネットワークと、大規模記憶装置、センシング技術の発展による大規模データの出現
- 計算機の高性能化
- WWW (ワールド・ワイド・ウェブ)
 - 420万サイト・数十ペタバイト=10,000,000,000,000文字
 - vs. 米国国会図書館全収蔵印刷物=10テラバイト
- 人間の処理能力を超えた情報の増大. 「情報爆発」



U. S. National Archives

● 社会システムの変化

- 情報通信技術の重要性の増大(サイバー空間)
- 世界的なネットワーク接続と知識基盤型産業
- 社会と技術のオープン化(仕様とモデルの欠如)
- 不確実な環境・予測不能な危険の増大
(予測し得ない事態)
- 情報にもとづく判断の重要性



Calling center in Bangalore, India
(www.voanews.com/english/images/Call_Center.jpg)

2007年1月18日

JST CRDS ワークショップ



図 2-3

データ、情報、知識 (広辞苑)

● データ

- 立論・計算の基礎となる、既知のあるいは認容された事実・数値。
「実験ー」
- コンピュータで処理する情報。

● 情報

- あることがらについての知らせ。「極秘ー」
- 判断を下したり行動を起こしたりするために必要な、種々の媒体を介しての知識。「ーが不足している」

● 知識

- ある事項について知っていること。また、その内容。「豊富なー」
- 知られている内容。認識によって得られた成果。厳密な意味では、原理的・統一的に組織付けられ、客観的妥当性を要求し得る判断の体系

2007年1月18日

JST CRDS ワークショップ



図 2-4

データ、情報、知識（熊坂賢次）

● データ

- － 現実世界の事象に対する客観的な事実を表現し、潜在的な価値のみを持っている。

（例） 血圧:170-105

● 情報

- － データの潜在的価値が、受け手の意識や関心に影響を与え、受けての価値観を変化させると情報になる。

（例） 血圧が高いから木をつけなくては。

（例） アルコールを控えよう。

● 知識

- － 特定の問題解決に役立つ情報の集まり。
- － 情報を一般化あるいは抽象化することによって得られる。

（例） 脳梗塞が心配である。

2007年1月18日

JST CRDS ワークショップ



図 2-5

情報と知識(下嶋篤)

● 情報

- － ある状況と別の状況との間に性質の相関があるとき、一方の状況が他方の状況について知る手がかりになる。したがって、こうした相関が成立するとき、一方の状況が他方の状況についての「情報を担う」といい、これをもって情報の定義とすることができる。

（例）私の研究室の「在室」と「帰宅」の表示板； 表示板の針の示す「在室」、「帰宅」と研究室の内部の状況の間に相関がある。

● 知識

- － 人間やロボットなど、何らかの認知主体が、状況と状況の間の性質の相関に適応し、その結果として、ある状況の性質に基づいてた状況の情報を獲得し、これを自分の行動を統制するために使える方法で貯蔵したとき、この情報を、その主体の「知識」としてみよう。

（例） 表示板と内部の相関関係に適応し、表示板の「帰宅」という情報から、学生が私が不在と判断し、部屋のノックをするのをやめた。

（ナレッジサイエンス、紀伊国屋書店から）

2007年1月18日

JST CRDS ワークショップ



図 2-6

知識とは(1)

- 「正当化された真なる信念 (justified true belief)」(プラトン)
分析哲学での『知識』の古典的定義:ある認知者Aが「Xである」という知識を持つのは以下の場合、その場合にかぎる。
Aは「Xである」と信じており、かつ、
Aの「Xである」という信念は正当化されており、かつ
「Xである」は真である。
一言で言えば、「知識とは正当化された真なる信念である」。
- McCarthy, 佐藤雅彦の知識に関する研究やHalpern の“Reasoning about Knowledge”など、『信念』との絡みのものもある。
- 「個人の信念やスキルを『真理』に向かって正当化していく、ダイナミックで人間的／社会的なプロセス (A dynamic human/social process of justifying personal belief and skill towards the “truth”)」(泉田裕彦、「知識国家論序説 新たな政策過程のパラダイム」)
- 価値観あるいは価値によって重みづけられ、検索可能な形に体系化された情報群(生駒)
- 計算機科学での「知識:情報」は統計科学では「情報:データ」(北川)

2007年1月18日

JST CRDS ワークショップ



図 2-7

知識とは(2) wikipedia

- 心理学では、「**宣言的知識**」、「**手続き的知識**」と分類している。
 - － 宣言的知識を**knowing that**ということもある。
(例) 科学的法則についての知見(九九、地球上での重力定数など)
(例) 社会的規約についての知見(日本国の首都は東京である、赤信号が点灯したら停止するなど)
 - － 手続き的知識を**knowing how**ということもある。
(例) 箸の使い方、ピアノの弾き方
(例) 車の運転の仕方などが挙げられる。
- 形式化、伝達方法の観点から、ナレッジマネジメントの世界では知識は「**形式知**」、「**暗黙知**」に分類されることがある。
 - － 暗黙的知識
宣言的に記述することが不可能か、極めて難しい知見。手続き的知識や直観的認識内容は暗黙知とされる。例えば「美人」についての知識は誰でも持っているが、それを明確に定義することはできない。

2007年1月18日

JST CRDS ワークショップ



図 2-8

科研費「重点／特定領域研究」による展開

- 知識科学(大須賀 節雄) H3-5年度
「知識科学における概念形成と知識獲得に関する研究」
- 発見科学(有川 節夫) H10-12年度
「巨大学術社会情報からの知識発見に関する基礎研究」
- 情報洪水(元田 浩) H13-16年度
「情報洪水時代におけるアクティブマイニングの実現」
- 情報学(安西 祐一郎) H13-17年度(特定領域研究C;国策)
「ITの深化の基盤を拓く情報学研究」
- 情報爆発(喜連川 優) H17-21年度
「情報爆発時代に向けた新しいIT基盤技術の研究」

2007年1月18日

JST CRDS ワークショップ



図 2-9

知識科学の展開

大須賀節雄、有川節夫、波多野誼余夫、志村正道、西尾章治郎、大槻説平 編著(1996年、オーム社)

- 内容
 - 第1章 はじめに
 - 第2章 人間とコンピュータの知識獲得
 - 第3章 概念形成と知識獲得
 - 第4章 データベースにおける知識獲得
 - 第5章 CAIにおける知識獲得
 - 第6章 機械学習と機械発見の理論
- 【セールスポイント】
知識科学の最先端の成果を集大成
- 【発刊の目的と内容】
文部省重点領域研究「概念の形成と知識の獲得」の研究成果をもとに、知識科学の現在とその将来動向について、斯界の第一線の執筆陣がまとめたもの。概念形成・知識獲得のメカニズムを軸に、最先端の研究内容を重点領域研究の範囲にとらわれず、広い視野から解説している。

2007年1月18日

JST CRDS ワークショップ



図 2-10

発見科学(1)

- 知識発見の論理
 - － 知識発見の(ための)論理的基礎(科学哲学)
 - － 知識発見の計算モデル(computational model)
- 推論による知識発見
 - － 統計的アブダクション
 - － 帰納推論
 - － 帰納的論理プログラミング
- 計算学習理論に基づく知識発見
 - － 知識発見のための学習モデル
 - － 知識発見アルゴリズム
 - － 知識発見における知識形成

2007年1月18日

JST CRDS ワークショップ



図 2-11

発見科学(2)

- 巨大データベースからの知識発見
 - － 機械学習とデータマイニング技術による知識発見
 - － 大規模観測データ(から)の統計解析と信号抽出
 - － 数値データからの知識発見
- ネットワーク環境における知識発見
 - － 知識発見のためのメディアシステム
 - － CSCW
 - － 可視化
 - － 知識発見のためのネットワークエージェント
 - － ネットワークにおける検索と発見

2007年1月18日

JST CRDS ワークショップ



図 2-12

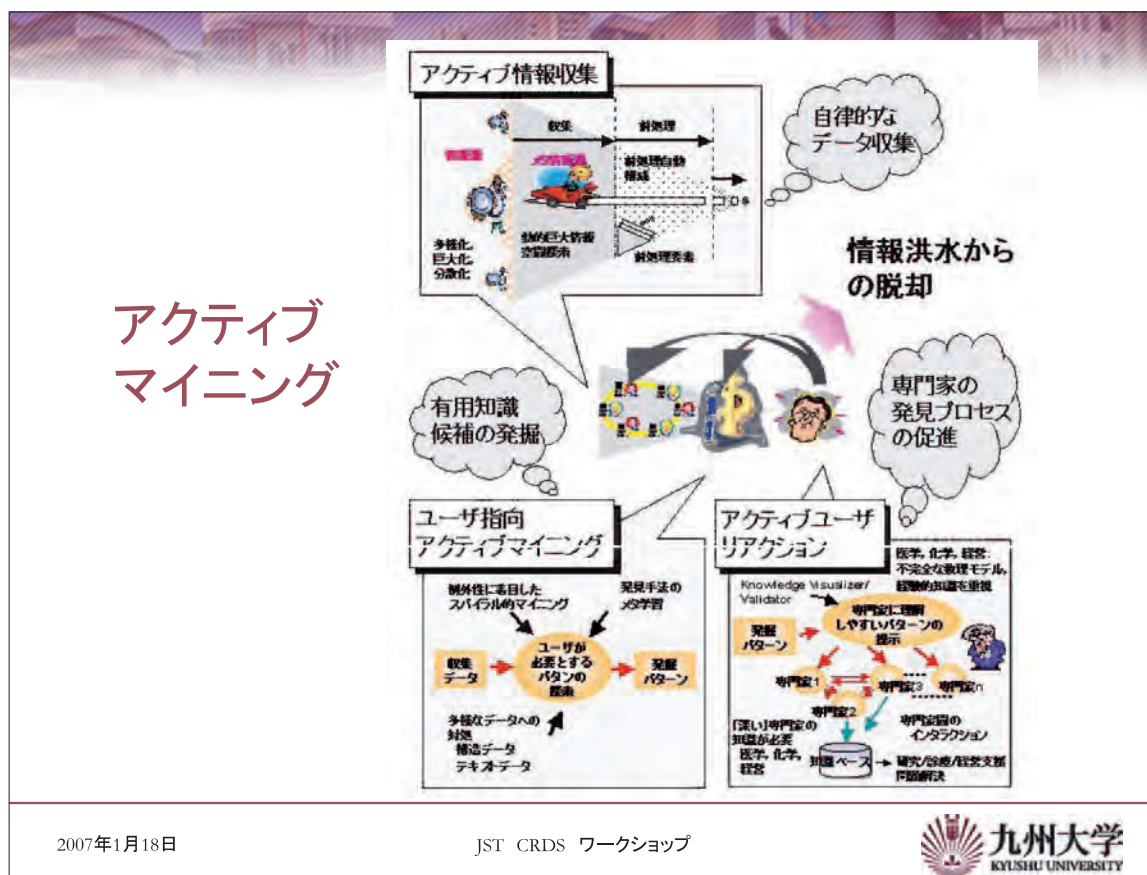


図 2-13

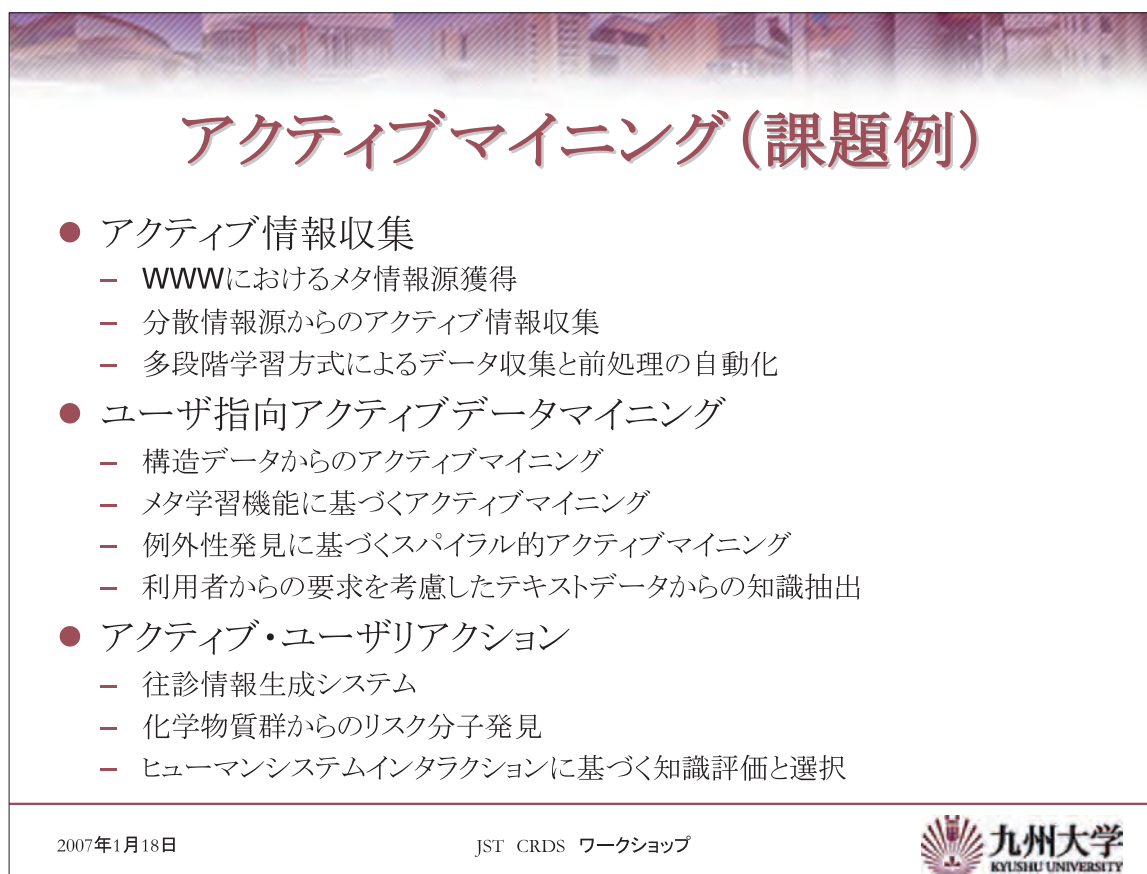
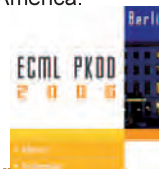


図 2-14

世界の動向(データマイニング)

- **ACM KDD (1989－)**
 - － **Int'l Conf. on Knowledge Discovery and Data Mining**, ACM SIG-KDD, North America. 12th in 2006. **Started as AAAI workshop in 1989.**
 - － Rapid growth of the area: 112 papers from 457 submissions in KDD2006 (50 oral, 55 posters, 7 industrial); 30 papers from 69 submissions in KDD1991 ("A tremendous international response" in KDD1991 preface).
- **ECML/PKDD (1997－)**
 - － **European Conference on Principles and Practice of Knowledge Discovery in Databases**, 10th in 2006. 46/36 papers from 564 paper submissions in PKDD2006.
- **PAKDD (1997－)**
 - － **Pacific-Asia Conference on Knowledge Discovery and Data Mining**, 11th in 2007 **since 1997**. 67 full and 33 short papers from 501 submissions in 2006. (PAKDD2008, Osaka, Japan, General co-chair: S. Tsumoto, PC co-chair: T. Washio)
- **DS (1998－)**
 - － **Int'l Conf. on Discovery Science**, Pacific-Asia and Europe, 9th in 2006 (DS2007, Sendai, Japan, Conf. chair: A. Shinohara, PC co-chair: M. Takeda)
- **IEEE ICDM (2000－)**
 - － **IEEE Int'l Conf. on Data Mining**, IEEE Computer Society, 6th in 2006, North America. 39 regular and 66 short from 451 submissions in 2004.
- **SIAM DM (2001－)**
 - － **SIAM Int'l Conf. on Data Mining**, SIAM, North America, 6th in 2006.



2007年1月18日

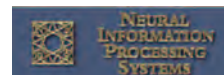
JST CRDS ワークショップ



図 2-15

世界の動向(機械学習関係)

- **ICML (1984－)**
 - － Int'l Conf. Machine Learning. 23rd in 2006 **since 1984**. 134 papers from 491 submissions in 2005.
- **NIPS (1987－)**
 - － Neural Information Processing Systems Conference. 20th in 2006.
- **UAI (1985－)**
 - － Uncertainty in Artificial Intelligence. 22nd in 2006 **since 1985**.
- **COLT (1988－)**
 - － Annual Conference on Learning Theory. 19th in 2006 **since 1988**. Successor of ACM COLT(Computational Learning Theory, 1988 -1998) and EuroCOLT (1991-2001)
- **ALT (1990－)**
 - － Int'l Conf. on Algorithmic Learning Theory. 17th in 2006 **since 1990**. (ALT 2007, Sendai, Japan)
- **ILP (1991－)**
 - － Int'l Conf. on Inductive Logic Programming. 16th in 2006.
- **IJCAI, AAAI, ...**
 - － Artificial Intelligence conferences
- **SIGMOD, VLDB, IEEE ICDE, ...**
 - － Databases and Data Engineering



2007年1月18日

JST CRDS ワークショップ



図 2-16

最近の資料から

比較的最近の内外の資料から、どのような知識が獲得・発見の対象になっているか、簡単に見ておく:

- 提案書: Data Mining Curriculum Proposal, ACM SIGKDD Curriculum Committee, Jan. 2005.
- 解説: 特集「最新! データマイニング手法」, 鈴木・鹿島(編), 情報処理学会誌, 2005年1月

2007年1月18日

JST CRDS ワークショップ



図 2-17

発見する知識の例(1)

提案書: ACM SIGKDD Curriculum Committee, Data Mining Curriculum, A proposal, ver1.0, April 30, 2006.

- Knowledge Discovery Process
 - － 対話的くり返し過程としての知識発見
- Different Kinds of Data
 - － 関係データ, 取引データ, 複合データ, 時間空間データ, テキストデータ, マルチメディアデータ, ウェブデータ, ストリームデータ, センサーデータ
- Different Kinds of Knowledge to be mined
 - － 分類, 予測, 回帰, 確率モデル推定, クラスタリング, 頻出パターン, 外れ値/例外検出, トレンド分析, 部分構造, ウェブリンク構造, 可視化
- Evaluation of Knowledge
 - － 規則/パターンの質と興味深さ: 精度, 再現性, 有用性, 関連性, その他の統計的尺度. サンプルング
- Applications
 - － 市場・経済調査, 意思決定支援, 自然科学・製造業のプロセス解析, 医療, バイオインフォマティクス, セキュリティ, 情報検索支援, 個別化, など. 産業応用とプライバシー. 標準化

2007年1月18日

JST CRDS ワークショップ

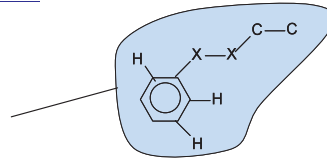


図 2-18

発見する知識の例(2)

解説:「情報処理」情報処理学会誌, 特集「最新! データマイニング手法」,
鈴木・鹿島(編), Vol.46, No.1, 2005年1月

- ○:発見する知識の例 □:知識発見手法
- ストリームマイニング(1章, 有村博紀・喜田拓也)
 - 売上げ上位商品リスト, 長距離電話の通信パターン, □オンラインマイニング.
- データスカッシング(2章, 鈴木英之進)
 - 年齢・資産・年収に関する銀行の顧客データのクラスタ
 - データ圧縮を用いたクラスタリングと予測(SVM)
- グラフマイニング(3章, 鷲尾隆)
 - 創薬における有機化合物の機能性(変異原性)の部分分子構造
 - グラフデータの部分構造パターン発見アルゴリズムAGM
- 構造カーネル法(4章, 鹿島久嗣)
 - ウェブ上のHTMLページの情報抽出と分類. 遺伝子制御ネットワーク
 - 構造カーネル法とサポートベクトル機械(SVM)
- 統計的な異常検出(5章, 山西健司・竹内純一・丸山祐子)
 - ネットワークとコンピュータの不正侵入の検出(外れ値・変化点・異常行動パターン).
 - 適応的学習に基づく外れ値検出. 情報理論的手法による動的モデル選択.
- 自然言語処理におけるマイニング技術応用(6章, 工藤拓・新保仁)
 - ウェブ上の製品レビューやアンケートの自動分類. 対訳例からの翻訳表現抽出.
 - 機械学習手法における部分構造パターン発見を用いた素性抽出
- 大規模Webアーカイブからのデータマイニング(7章, 豊田正史・喜連川優)
 - Web空間における新しい話題の発展過程の追跡. ある話題の履歴に対する質問応答.
 - 大規模Webアーカイブ技術. テキストマイニングとリンク情報マイニング.



```
<people>
<person age="25" id="608">
  <name>John</name>
  <email>john@abc.com</email>
</person>
<person id="609">
  <name>Mary</name>
  <tel>555-4567</tel>
</person>
</people>
```

「be staying at XX hotel」
⇔「XXホテルに滞在」

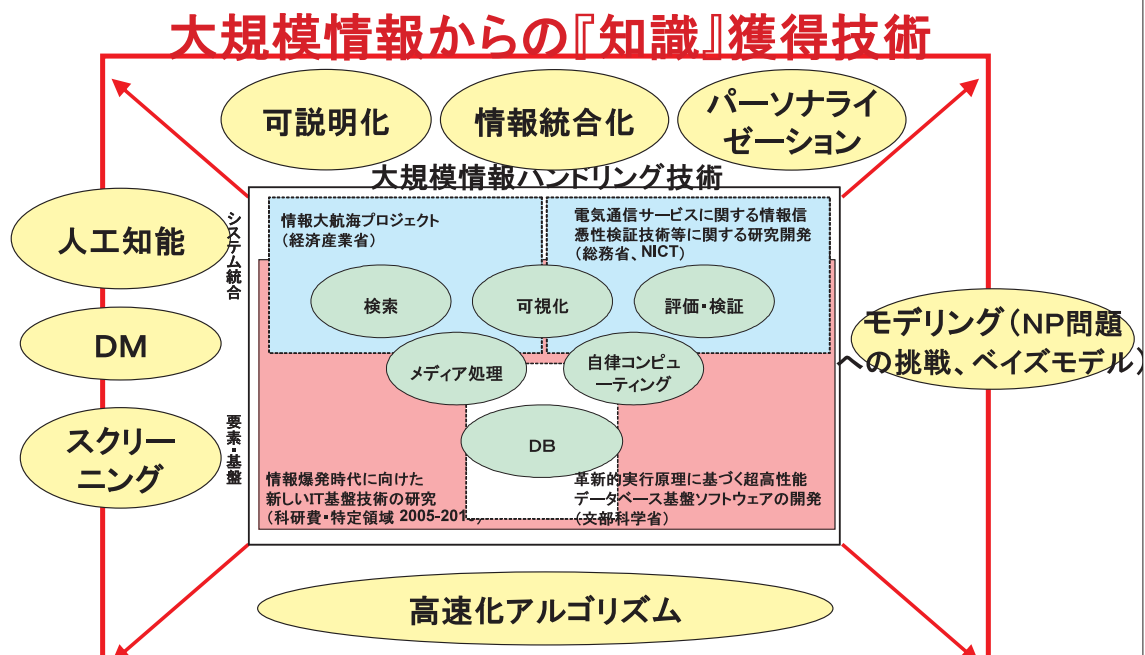
2007年1月18日

JST CRDS ワークショップ



図 2-19

『知識』獲得の観点からの技術の概略俯瞰 (再検討結果)



2007年1月18日

JST CRDS ワークショップ



図 2-20

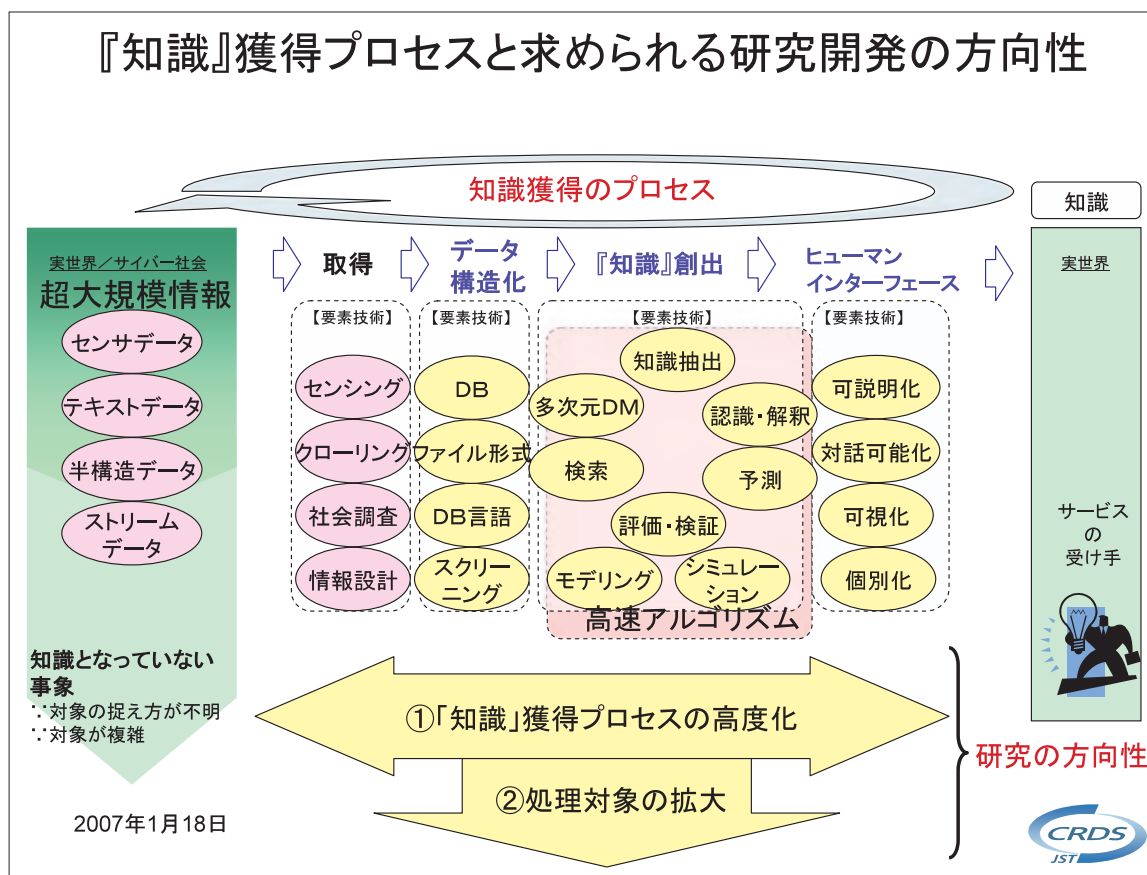


図 2-21

WSでのアウトプット

- 『知識』*とは（「定義」を与えるのは最後にする。一般的なある程度はあいまいさを許容した状態で、2, 3, 4の作業を行い、しかる後に、このグループでの「公式な」定義を与える。）
- 技術俯瞰図、重要研究課題（CRDS案の見直し、詳細化）
- 推進方法
- 投資する意義（社会へのインパクト、必然性）

* : CRDSの案としては以下を考えている段階で結論に至っていない。

対象とする分野の構造や原理原則など対象を理解している人が、その分野で抱えている問題点を解決するために、取得・検索・解析・発見する個別の具体的な新たな情報をいう。例えば、膨大な医療画像、今までの事例データから、発見が難しい癌などを発見するための情報をいう。他の例としてはヒヤリング結果を添付。

2007年1月18日

JST CRDS ワークショップ

九州大学
KYUSHU UNIVERSITY

図 2-22

大規模情報ハンドリング技術が求められるニーズの例 (インタビュー結果)

- ウェブアーカイブ (東大 喜連川優 先生)
ウェブサイトのリンクパターンの全データとその経時変化の履歴(5年分)
→ 社会の風評の判定、トピックスの抽出
- ライフサイエンス (東大 高木利久 先生)
GenotypeとPhenotype
→ 細胞中の遺伝子が関わる複数のプロセスの解明
日々生産される膨大な量の論文
→ 自分の求める遺伝子等に関する論文の抽出
- プラント制御 (阪大 鷲尾隆 先生)
設置されたセンサからのデータ(1ブロック2000以上)
→ 制御、故障や異常の予測
- 商品・サービス設計 (産総研 西田佳文 氏・本村陽一 氏)
WEB上に設置されたセンサからのデータ(アンケート、通販履歴等)
→ 最適な商品の提供方法や商品形状の最適化
- マーケティング (関西大 矢田勝俊 先生)
大規模店舗の購入履歴(どの顧客がいつ何を)
→ "優良顧客"の購買パターン、価格設定や仕入商品との関係

サイバー社会の情報
だけでなく、実社会の
情報も大きなニーズ

2007年1月18日

JST CRDS ワークショップ



図 2-23

おわり

2007年1月18日

JST CRDS ワークショップ



図 2-24

3

大規模情報からの 『知識』獲得技術に関する検討

3.1 個人のニーズに応じた『知識』獲得：パーソナライゼーション技術

分科会 A 北川 源四郎（統計数理研究所）

古典物理学に代表される機械的世界観のもとでは知識は「普遍の真理」を目指したものであった。しかしながら、生命科学や社会科学などの発展を経て、情報社会が実現した現代においては、知識のイメージも大きく変化しつつある。人類は大量の情報を蓄積し、広大なサイバー世界を構築しつつある。いまや人類は、実世界とサイバー世界の接点に存在しているともいえる。しかしながら、サイバー世界に普遍の真理は想定し難い。これからの知識は「普遍の真理」ではなく、それぞれの人間の目的に応じて、これまでに得られたデータ・情報・知識を総動員して、創造していくべきものとなるであろう。

確定的世界観においては、対象を記述する簡潔な方程式が知識のイメージに合致する。方程式と現在の状態がわかれば、未来永劫その動きを予言できるからである。確率的世界観のもとでは、対象を分布で捉えるようになったが、これまでは分布自体あるいはその集団の平均的な「代表」に関する推論を行ってきた。しかし、現実の世界は「個」の集まりである。平均的な人間で分布が構成されているわけではない。情報化以前の多くの情報が望めない状況では、平均的な対象を想定した推論は合理的である。しかし、大量の情報獲得が可能となった現在、「個」に特化した推論・判断・行動を行うことも可

能であり、またそれを実現してはじめて豊かな知識創造社会を構築できる。

3.1.1 応用分野と目指すべき目標

「パーソナライゼーション」

大規模情報を最大限活用するためには、個人のニーズに応じて知識獲得を行うための、パーソナライゼーション技術の研究開発が最も重要である。この技術を例示的に表現すると以下のようになる。

■ beyond MATRIX

実世界を個人レベルでキャプチャーし、データ同化技術を用いながら Virtualized Reality を構築する。集団レベルである MATRIX に比して、いわば次世代の MATRIX。

■ ジャストアシスト方法

■ 気の利いた万能エージェント

集団の平均的な人間を対象とした大量生産・大量消費時代のサービスに替わって、個々人のニーズを忖度して、過不足なくちょうどよいサービスを実施すること。

■ 個人のための予言者

一般的な知識と、相手に関する少ない情報を統合して、個人の判断や予測・行動にとって重要な情報を提供すること。テラーメイド医療・創薬・教育、マイクロマーケティング、

局所天気予報などが実現できる時代となっている。

3.1.2 研究開発課題

(1) 技術分野の現状

■ 取得技術

現代社会では、ランダムサンプリングの実施がほとんど不可能になっている、一方バイアスサンプリングされたデータの活用法が貧弱である。また、データデザインの方法論が未発達で、取得項目の選択などが個人のスキルに依存し、匠の技によってなされている。

データ構造化が不十分で、大規模データが巨大なゴミ箱の状態になっている。

■ 知識創出

知識創出の2つの方法、演繹的手法と帰納的手法、が断絶している。また、これまで利用されてきた知識獲得の方法の多くは、次元の呪いによって、大規模データへの適用が困難である。

■ ヒューマンインタラクション

これまでの技術は20世紀の大量生産・大量消費に対応して、平均的な人間に対応する「マスマーケティング」の状態に止まっていた。近年は、パーソナライゼーションの試みが始まっているが、層化のような擬似パーソナライゼーションの段階に止まっている。

(2) 研究開発課題

■ ベイズモデル構築技術

最も重要な課題は、知識創出のためのベイズモデルの構築技術である。長い歴史を持つベイズ推論は、知識に対する考え方の変化と計算手法お

よび計算機の高速化によって、ようやく実用的なものとなりつつある。ベイズモデル構築に関連する要素技術のさらなる発展は、今後ますます重要となる異種情報の統合や新しい推論手法の開発に不可欠である。

■ 演繹的手法と帰納的手法の統合技術

科学技術の進展に対応して、これまでの科学研究で用いられてきたトップダウンな演繹手法とボトムアップの帰納的手法の2つのアプローチを統合するための方法の研究開発が重要である。具体的には、従来のフォワード型のシミュレーションにデータの持つ情報を統合させるデータ同化や、高次元で非線形の複雑なシステムに対応できるフィルタリングの方法の研究開発が重要である。

■ ラーニング技術

機械学習に関しては、カーネル法の成功が著しい。今後も、カーネルの設計法のおよび構造学習の研究が重要である。

■ 「モデルに基づいた」アクティブセンシング技術

データ取得も知識発展スパイラルの一部として位置づけることが肝要である。モデルに基づくデータの取得法、すなわち取得すべきデータ項目や場所を選択する方法の研究が必要である。

■ データのアラインメント技術とクリーニング技術

大規模データの活用のためには、断片的なデータを統合するためのアラインメント技術や欠測値や異常値(外れ値)を適切に処理するためのクリーニング技術が不可欠である。言うまでもなく、これらの技術もモデルに基づくべきである。

■ 個別オントロジー構築技術

パーソナライゼーションの実現のためには、ヒューマンインタラクションに関連しては、個別のオントロジー構築技術が重要になる。

目
的

大規模情報からの
「知識」獲得技術について

分
科
会
A

分
科
会
B

分
科
会
C

ま
と
め

付
録

表1 技術俯瞰用マトリクス* (分科会A)

	大規模情報	取 得	データ構造化
研究開発課題 (これから何をすべきか)	■ 地球・宇宙科学データ ・地震予知関連観測データ ・地震波データ, 津波データ ・OBS (Ocean Bottom Seismograph) データ ・気象データ ・オーロラ・マルチモーダルデータ (画像、物理観測) ・全天 CCD 観測データ ■ マーケティングデータ ・POS データ ・消費者行動	■ 「モデルに基づいた」 アクティブセンシング技術 ・取得すべきデータ項目のデザイン ・取得すべき場所の選択	■ 「モデルに基づいた」 データのアラインメント技術とクリーニング技術
現状俯瞰 (今、どうなっているか)	・CRM (Customer Relationship Management) データ ・FSP (Frequent Shoppers Program) ■ 金融経済データ ・CRD (Credit Risk Database) ・ROI (Return on Investment) ■ 生態計測情報 ■ サイバー世界データ ・ニュースのマルチモーダルデータ (動画、音声、テキスト) ・事故・故障情報 ・WEB アクセスログ ・ユーザーの嗜好性の順序データ	・バイアスサンプリングされたデータの活用方法がない ・人のスキルに依存する項目・場所のデザイン	・巨大なゴミ箱

*大規模情報から目的とする『知識』を獲得するための技術要素を、「取得」「データ構造化」「『知識』創出」「ヒューマンインタラクション」の技術要素に分類し、それぞれの技術要素の現状を俯瞰した上で、これからどのような研究開発をするべきかについて議論を行った。A分科会の成果を表1にまとめた。

『知識』創出	ヒューマン インタラクション	目的とする『知識』
■ ベイズモデル構築技術 ■ 演繹的手法（トップダウン） と帰納的手法（ボトムアップ） の統合技術 ・ データ同化 ・ フィルタリング ■ ラーニング技術 ・ カーネルの設計	■ 個別オントロジー構築技術	■ 「パーソナライゼーション」 個別の人間生活に役立つ知識 （目的志向） ・ QOL の向上 ・ オーダーメイド医療 ・ オーダーメイド創薬 ・ テーラーメイド教育 ・ 局所気象予想 ・ マイクロマーケティング ・ カスタマイゼーション ・ 資源の有効利用
・ 演繹的手法と帰納的手法の 断絶 ・ 次元の呪い	・ マスマーケティング ・ 擬似パーソナライゼーション	・ 平均的知識

3.1.3 研究推進方法

(1) フィールドに密着した現場主義

データ解析や情報処理のような汎用性を持った方法の研究でも、成功例を見るとほとんどの場合、実際の現場の問題にチャレンジしながらその問題解決の中から汎用的な方法を確認したものである。したがって、パーソナライゼーションのための研究および技術開発においてもフィールドに密着した現場主義が不可欠である。また、日本発の独自の技術を開発し発信していくためには、輸入された課題ではなく、日本の現場に即して研究を推進していくことが必要である。

(2) 『知識』獲得の要素技術間の“相互干渉”

能動的な知識獲得は、それまでに獲得したすべての情報および知識を活用して実行される。モデルベースのデータの取得法はモデルの改善によって発展し、それがさらにモデルの改善につながる。知識獲得の方法自体も知識なのである。知識獲得の過程は一方向ではなく、知識発展のサイクルを構築することが肝要である。したがって、その要素技術の研究においても、お互いの研究領域を聖域とせず、お互いに積

極的に干渉しあう相互貫入の研究体制が必要である。

(3) ビジネスセクターを並行して開拓

大規模情報からの知識獲得技術を現実の問題解決の中で発展させていくためには、研究と同時にビジネスセクターを並行して開発していく必要がある。

3.1.4 大規模情報から獲得する『知識』とは

知識とは、個別の人間生活にとって役立つ“制約”である。

情報社会における知識は、従来のような普遍の真理とは異なって、個々人のニーズに応じて役に立つ“制約”である。広大な不確実性の空間に適切な“制約”を加えることによってはじめ、個人にとって有用な情報が獲得できるからである。ただし、この“制約”は設計概念、機能、モデル、プロセスを含む広い概念であり、また、知識発展のプロセスの進化にともなって変化する動的なものである。情報とはこの不確実性を減少させるものであり、それを実現するのは、適切なモデリングを用いる知識獲得技術である。

以上

大規模情報からの知識獲得支援技術

分科会A

大規模情報からの知識獲得支援技術
分科会 A

1

図 3-1-1

メンバー

- 北川源四郎 統計数理研究所【リーダー】
- 赤穂昭太郎 産業技術総合研究所
- 石井信 奈良先端科学技術大学院大学
- 田辺國士 早稲田大学理工学術院
- 照井伸彦 東北大学大学院経済学研究科
- 西田佳史 産業技術総合研究所
- 樋口知之 統計数理研究所
- 広瀬直毅 九州大学応用力学研究所
- 波多腰玄一 CRDS
- 楠本博之 CRDS
- 嶋田一義 CRDS

大規模情報からの知識獲得支援技術
分科会 A

2

図 3-1-2

目指すべき目標 「パーソナライゼーション」

- beyond MATRIX
- ジャストアシスト方法（個人のリクエストに対して、サービスをやりすぎず、やらなさすぎず）
- 気の利いた万能エージェント
- 個人の予言者

図 3-1-3

技術分野の現状

- 取得技術： バイアスサンプリングされたデータの活用法が貧弱
- 取得技術： 人のスキルに依存する項目のデザイン
- データ構造化： 巨大なゴミ箱？
- 知識創出： 演繹的手法と帰納的手法の断絶
- 知識創出： 次元の呪い
- ヒューマンインタラクション： マスマーケティング
- ヒューマンインタラクション： 擬似パーソナライゼーション

図 3-1-4

技術課題(1)

「モデルに基づいた」

- アクティブセンシング技術
 - 取得すべきデータ項目のデザイン
 - 取得すべき場所の選択
- データのアラインメント技術
- データのクリーニング技術

大規模情報からの知識獲得支援技術
分科会 A

5

図 3-1-5

技術課題(2)

- ベイズモデル構築技術
- 演繹的手法(トップダウン)と帰納的手法(ボトムアップ)の統合技術
 - データ同化
 - フィルタリング
- ラーニング技術
 - カーネルの設計
- 個別オントロジー構築技術

大規模情報からの知識獲得支援技術
分科会 A

6

図 3-1-6

『知識』の定義

- 個別の人間生活にとって役立つ“制約”
- “制約”とは、設計概念、機能、モデル、プロセスを含む広い概念
- 環境によって変化する動的なもの

図 3-1-7

研究開発の推進方法

- フィールドとの深い連携が必須（現場主義）
- 日本の現場の問題に即して推進
- ビジネスセクタを平行して開拓
- 『知識』獲得要素技術間の相互“干渉”

図 3-1-8

3.2 ヘテロで大規模なデータからミクロな規則を発見する技術

分科会 B 有村 博紀（北海道大学）

■ 背景

二十世紀は物質を対象とした科学・技術の時代として始まったといわれる。これに対して、二十世紀後半は情報を対象とした科学・技術が発達した時代であるといえる〔参考文献 1〕。二十世紀初頭の電子計算機の登場以後、情報をつかう科学・技術は著しく発展した。とくに、その最後の 10 年間である 1990 年代には廉価な高性能計算機と大容量記憶装置、汎地球的な高速ネットワークが一体となって、情報技術が社会に急速に普及した。これにより、実世界の大量の情報が情報世界に流れ込むことで、情報が世界中に瞬時に伝わり、膨大な情報がネットワーク上に蓄積される「サイバー世界」と呼ばれる巨大な情報の海が出現しつつある。

このような情報通信技術の急速な発展と普及を背景として、いまや大量の情報とデータをつかう情報技術が、安定した技術として社会を支える基盤技術となってきた。例えば、この 30 年間に急速に発達した「大規模情報処理技術」として、大量のデータを格納するデータベースや、ウェブサイトの集積から所望の情報をとりだす情報検索、無数のデータのやりとり（トランザクション）を交換する基幹情報システム、多数のセンサーを統合するセンサーネットワークは、その典型的な例である。しかし、知識社会を支える基盤技

術として、このような大規模情報技術だけでは十分でなく、膨大な情報からの『知識』の獲得をつかうための新しい科学・技術が必要になってきている。

■ 知識と知識獲得とは

これに対して、大量情報から『知識』を取り出すための新しい科学・技術が求められている。具体的には、上にあげたような大量情報を対象とした現在の情報技術は、「素材としての情報」をつかう情報技術であり、集められた情報の活用はすべて人間によって行われる。これに対して、この 10 年間に急速に増大した情報は、量質共に人間の能力を大きく超えており、大規模情報のもとで人間の知的活動を支える「知識としての情報」を対象とした情報技術が必要である。その核となるのが、大量情報から『知識』をとりだす「大規模情報からの知識獲得」技術である。

ここでいう『知識』とは、人間が実世界と情報世界についての認識・理解・判断を行うための高次の情報をいう。例えば、気象情報システムの例でいえば、日本全体で収集された「各地の気温・雨量・風向といった情報の集積」は低次の素材としての情報であり、そこから取り出された「台風が近づいている」という高次の情報は『知識』である。また、低次の情報から高次の情報を取

り出すための手順やそのパラメータに関する情報自体も『知識』といえる。人間や生物は、素材としての低次の情報（データ）をそのまま用いるのではなく、いったん、高次の情報に「加工」し、それらを「組合せ」てさまざまな目的に「適用」することで、自分を取りまく環境における活動をうまく行ってきたといえる。

ふりかえて、これまでの産業技術と情報技術とのアナロジーでいえば、これまでの情報技術が素材としての情報をあつかう第一次産業に対応していたのに対して、「知識獲得技術」は、情報から知識の生産とサービスをあつかう第二次産業と第三次産業に相当するといえる。すなわち、21世紀の情報技術は、第一次産業から、第二次産業・第三次産業へと進む変革の時期にあるといっても良い。「大規模情報からの知識獲得」技術は、新しい情報の科学・技術である。

■ 必然性・重要性・技術俯瞰：なぜ今「大規模データからの知識獲得」か？

それでは、なぜ今「大規模データからの知識獲得」の科学・技術の研究開発に取り組むことが必要なのだろうか。とくに、知識獲得のための情報技術は情報分野と応用数学分野で研究が行われて機械学習と数理統計が融合して生まれた。これらと、現在の知識獲得の違いは何だろうか。また、新しい科学・技術の実態はあるのだろうか。以下、説明する。

（1）ミクロなマクロ：伝統的なデータ解析や数理統計では、大量のデータをならして、平均や分散（さらには統計モデルの少数のパラメータ）といったマクロな特徴を見つけていた。これに対して、最近の知識獲得では、膨大な

細粒度のデータから、さまざまなタイプの個別データに対して成立する細粒度のパターンや規則性をみつける。例えば、最近の市場分析では、分単位の（up-to-the-minutes に）、個人ごとに購買される（individual transactions）、商品ごとに記録された（item-wise market basket data）、購買データが分析の対象であり、これを地域ごとや四半期ごとの大きなくくりで平均をとって分析するのではなく、さまざまなタイプの個人の購買行動に関する多数のミクロな規則性として解析する。また、解析のためのモデルも大きなパラメータをもつ。膨大なマクロデータに対して、このようなミクロな規則をみつけることは人間の能力を超え、従来の統計学に含まれない、新しいデータ解析の流れである。このような知識獲得技術の例として、結合規則発見（association rule discovery）に代表される頻出パターンマイニング技術がある。

（2）ヘテロなデータ：情報の海としてのサイバー世界においては、一つの事象に対しても、実世界のさまざまな側面からの多様な異種情報（これを以後「ヘテロな情報」という）が蓄積されており、これらを結合して知識獲得を行う必要がある。例えば、生命科学においては、生命活動に関して、ある遺伝子の働きを知るためには、DNA 配列の遺伝情報レベルから、3次元構造をもつタンパク質のレベル、それらの機能の相互作用からなる生体レベルまで、さまざまな階層のデータがあり、また、さまざまな生物種におけるデータを横断して研究する必要がある。したがって、生命データベースからの知識獲得を行うには、これらのヘテロな情報をあつかえる知識獲得技術が必要である。

また、従来の統計では正規分布のような少数の数値パラメータをもつモデルを用いることが多かったが、隠れ変数を含む多数のパラメータをもつモデルや、グラフや階層構造などの構造データを扱う技術が必要である。このような知識獲得技術の例として、因果関係を扱うためのベイジアンネットワークや、構造データを扱うためのグラフマイニング、機能を予測するためのサポートベクターマシンなどがある。

(3) 理論に裏付けられた新しい知識獲得技術の登場：機械学習分野においては、1980 年代から決定木学習アルゴリズム（Quinlan の ID3）やその枝刈最適化手法のような実用に耐える知識獲得手法が出てきたが、その数理については依然としてわからないことが多かった。これが、1990 年代に入って、計算学習理論の進展を背景にその原理が徐々に解明され、集約学習法であるブースティングや、カーネル法と最大マージン原理を組み合わせたサポートベクターマシン（SVM）等の理論的な裏づけをもち、実問題において安定して高い性能を発揮する新しい知識獲得技法が登場してきた。また、詳しくは分科会 A の報告を参照いただきたいが、数理統計分野においても、多数の隠れパラメータをもつ統計モデルの学習法や、モデル自体をデータから適切に選択するための情報量基準やベイズ推定の新しい研究成果が続々と登場し、広く実際の問題に適用されるようになってきた。これらが 1990 年代の大規模データと、センサーや自動計測などの実世界情報の収集技術、自然言語処理や HTML 等の構造化・分節技術と一緒に、新しい知識獲得の技術と応用を形成するにいたっている。

■ まとめと投資の意義

以上にのべたように、「大規模情報からの知識獲得」技術は、第一次の大規模情報処理技術を次の第二次の段階に進めて、21 世紀の知識社会を実現するための「大規模情報」と『知識』をあつかう新しい情報技術である。これは、長い歴史をもつ機械学習や数理統計、大規模データ処理技術をもとに誕生した新しい融合分野である。同時に、基礎研究の成熟を背景とし、この 20 年間のいくつかの技術的ブレイクスルーを契機として、現在理論と応用の両面で急速に発展しつつある情報に関する新しい科学・技術分野である。この「大規模情報からの知識獲得」技術は、今までさまざまな分野で独立して研究されてきたが、この機に統合的な視点からの研究・開発を行うことは、知識社会の基礎技術としての「知識獲得」技術を確立するために有効と考えられる。

■ 参考文献

- [1] 生駒俊明. 日本の進むべき道(7章), 科学技術と社会：20 世紀から 21 世紀への変容, JST 研究開発戦略センター編, 丸善プラネット, 2006.

3.2.1 応用分野と目指すべき目標

「大規模情報からの知識獲得技術」の研究は、次のような応用分野と研究目標を持つ。

■ 市民の生活品質向上のための知識獲得技術

- ・ 主役が「YOU」の時代
- ・ 個人差を考慮した医療・生活行動マイニング
- ・ 大量情報からの価値創出

- 安全安心社会のための知識獲得技術
 - ・ 巨大システムの障害検出
 - ・ 知識にもとづくシステムディペンダビリティ
 - ・ サイバー天文台
- 科学技術のための知識獲得技術
 - ・ data driven science/engineering・データ中心科学

3.2.2 研究開発課題

(1) 技術分野の現状

- 「データの取得」に関しては、「素材としての大規模データ」の取得・管理は可能になっているが、その膨大なデータを活用できていない。例えば、生命科学の自動計測データや、地球シミュレータの出力等、膨大なデータが生み出されているが、これらは限定された目的に利用されている状況である。
- 「データの構造化」に関しては、現在の知識獲得技術の主流は、単純な表形式データ・属性データを対象としている。後述するように、これをヘテロな高次元データへと拡張する動きが登場しつつある。
- 「『知識』創出の技術」に関しては、長い基礎研究を背景として、この20年ほどの間に急速に理論と技術が発展をとげ、理論的な裏づけをもつ各種知識発見手法が登場している状況である。具体的には、次のような知識獲得技術が登場している：
 - ・ サポートベクトルマシン (SVM) の学習
 - ・ ベイズモデル推論
 - ・ ベイジアンネットワーク

- ・ 協調フィルタリング
- ・ 離散データマイニング
- ・ グラフマイニング

限界として、非常に高次元のデータや、多数の隠れパラメータ、超大規模データへの適用は依然として困難である。

- 「ヒューマンインタラクション」に関しては、これを獲得した知識の利用法と考えると、現状は「知識を発見しっぱなし」の状況である。

すなわち、知識が利用者から容易に利用できるようになっていない。

(2) 研究開発課題

- 知識の入口である「データの取得」に関しては、次のような実世界データからのデータ取得技術が重要な課題の例である：
 - ・ 日常生活行動ライフログ
 - ・ 大規模センサーネットワーク
 - ・ 大量 WEB 情報
 その他に、大規模データ取得の技術基盤の確立も重要な課題である。また、実世界に近いデータとして、現実の有料コンテンツ・学術データを意識したアプローチも必要である。
- 「データの構造化」に関しては、単純な表形式データからヘテロな高次元データへと適用範囲を拡張することが必要である。そのためには、次のような研究開発が必要である：
 - ・ ヘテロデータの融合
 - ・ WEB 空間とリアル空間の融合
 - ・ 分散センサーデータの統合
 - ・ 不完全・欠損データの扱い
 - ・ 自然言語データの文節化・構造化
 - ・ といったデータの構造化

■ 「『知識』創出の技術」に関しては、「超高次元・超大規模・超ヘテロデータ」への挑戦が大事である。そのために、次のような超大規模データのための知識発見技術の開発が重要課題である：

- ・ 高精度学習アルゴリズム
- ・ データスカッシング技術
- ・ 因果と相関の発見
- ・ 背景知識とデータの融合
- ・ 不完全情報からの発見
- ・ 因果関係とダイナミクス発見
- ・ リアルタイム異常検出
- ・ シミュレーションによる発見
- ・ 構造パターン発見

■ 「ヒューマンインタラクション」は、情報の出口であり、獲得した『知識』を、人間が実世界での利用に適用可能にできなくてはならない。そのための、知識と実世界・人間系との界面をつなぐ技術や、知識獲得を核とする統合的なサービスを実現するための研究が必要である。そのために、次のような事項が研究課題として挙げられる：

- ・ パッケージ・ソリューションからサービスへ
- ・ 一気通貫なプロセス
- ・ 要素技術研究と適用研究

表2 技術俯瞰用マトリクス*（分科会B）

	大規模情報	取 得	データ構造化
研究開発課題（これから何をすべきか）	超高次元・超ヘテロ・超膨大なデータの出現 ～自動計測技術の大規模化・精緻化・簡便化 ■ サイバー世界データ ・WEB 情報 ・CGM（Consumer Generated Media）データ ・報道データ ■ マーケティングデータ ・消費者の購買履歴データ ・製品アンケートデータ ・コールセンター問合せデータ ■ 生活行動データ ・自動車運転者の認識行動データ ・事故データ（家庭内幼児事故）	～知識の入口 ■ 実世界データからのデータ取得技術 ・日常生活行動ライフログ ・大規模センサーネットワーク ・大量 WEB 情報 ■ 大規模データ取得の技術基盤 ・有料コンテンツ・学術データも大事	■ ヘテロデータの融合 ■ WEB空間とリアル空間の融合 ■ 分散センサーデータの統合 ■ 不完全・欠損データの扱い ■ 自然言語データの文節化・構造化
現状俯瞰（今、どうなっているか）	■ セキュリティデータ ・サイバー&フィジカルログ ・システムの運転データ ・機器の性能データ ■ 医学データ ・診断データ ・病院管理データ ・化学薬品活性構造データ ■ 科学データ ・地球データ ・大規模シミュレーションデータ ・エクソンアレイデータ ・分散センサーデータ ・文献データ	・膨大なデータを活用できていない - ー生命科学の自動計測データ - ー地球シミュレータ	・単純な表形式データ・属性データを対象としている

*大規模情報から目的とする『知識』を獲得するための技術要素を、「取得」「データ構造化」「『知識』創出」「ヒューマンインタラクション」の技術要素に分類し、それぞれの技術要素の現状を俯瞰した上で、これからどのような研究開発をするべきかについて議論を行った。B 分科会の成果を表 2 にまとめた。

『知識』創出	ヒューマン インタラクション	目的とする『知識』
「超高次元・超大規模・超ヘテロデータ」への挑戦 ■ 超大規模データのための知識発見手法の開発 ・高精度学習アルゴリズム ・データスカッシング技術 ・因果と相関の発見 ・背景知識とデータの融合 ・不完全情報からの発見 ・因果関係とダイナミクス発見 ・リアルタイム異常検出 ・シミュレーションによる発見	～知識の出口 ■ 人間が解釈・理解・応用可能な形の知識獲得 ■ 発掘した知識の適用 ・予測・制御・診断 ・戦略決定・意思決定の支援 ・組織内スキル移転 ■ 知識体系の構築へ ■ サービスとしての知識発見	■ 市民の生活品質向上のための知識 ・主役が「YOU」の時代 ・個人差を考慮した医療 ・生活行動マイニング ・大量情報からの価値創出 ■ 科学技術のための知識 ・データ駆動科学／技術 ■ 安全安心社会のための知識 ・巨大システムの障害検出 ・知識にもとづくシステムディペンダビリティ ・サイバー天文台
・理論的な裏づけをもつ各種知識発見手法の出現 - ーサポートベクトルマシン(SVM)の学習 - ーベイズモデル推論 - ーベイジアンネットワーク - ー協調フィルタリング - ー離散データマイニング - ー構造パターン発見 ・超大規模データへの適用は困難	・「知識を発見しっぱなし」利用者から容易に利用できるようになっていない - ーパッケージ・ソリューションからサービスへ - ー一気通貫なプロセスとしての必要性 - ー要素技術研究と適用研究	・限定された個別化 - ーデータ解析 - ーマーケティング

3.2.3 研究推進方法

(1) 対象領域と共通プラットフォーム

対象領域として、医療と e-Science、データ中心科学、コンテンツの分野が有望である。とくに、大規模データを中心とした情報技術は、データ中心科学として、学術研究方法に変革をもたらす可能性がある。知識獲得技術は、基礎研究が鍵となっている一方で、総合的な実世界の問題への適用研究も重要である。また、個々の要素技術の開発と共に、さまざまなタスクを一気通貫に組み立てるプロセスとしての知識獲得の視点が不可欠である。また、パッケージから、ソリューション、さらに、サービスとしての知識獲得も考慮する必要がある。

そのために、大きな目標をたて、グランドチャレンジを息が長くサポートするような方策が必要である。また、このような基礎的な研究開発をサポートするには、オープンな研究イノベーションサイトの設置による推進方法が有効であると考えられる。そこでは、競争原理だけでなく、むしろ一つの目的に向かって協調し、新しい技術を作り出すような機構を整備する必要がある。

見落としてはいけない点として、知識発見の推論技術としては大規模データに対する高速化ももちろんだが、例えば基礎技術の 1000 倍程度の単なるスピードアップは、多数の研究者が協力・競争すれば比較的容易に実現する。それにたいして、グランドチャレンジとしては、現在不可能なことを可能にするような技術開発目標を掲げ、上記のような協調によって研究推進することが重要と考えられる。

3.2.4 大規模情報から獲得する『知識』とは

知識とは、ある目的をもった未来の活動をうまく行うために再利用可能な、集積したデータや過去の経験を集約して得られる情報である。この知識は低レベルのデータや情報から得られる高レベルの情報である。そのため、思考や行動を行う人間や機械、生物などの主体にとって、環境や経験をコンパクトにまとめた「経済的な」情報であり、知識の適切な利用により、主体は限られた能力と変化する環境のもとで、より適切な活動を行うことが可能になる。この意味で、知識は、情報世界や実世界における人間や機械、生物にとって、その活動に有用な、集約された情報であるといえる。

知識は、単なるデータ集合や記録と比較して、次のような特性をもつ：

- ・ 集約性：集積したデータや過去の経験を集約して得られるコンパクトな表現である。
- ・ 目的性：ある目的をもった活動をうまく行うためのものである。
- ・ 即時適用性：容易にかつすぐに適用可能である。
- ・ 再利用性：メディアを通じて、記録し、伝達し、再利用することが可能である。
- ・ 多様性：書物に記述された宣言的な知識から、行動の最適化のための非言語的な知識まで表現に大きな幅がある。一般的で系統的な知識だけでなく、与えられた範囲内でデータを極大に一般化しただけの特殊な知識もありえる。



JST CRDSワークショップ

予測と発見：大規模情報からの『知識』獲得技術

分科会 Bグループ

H19.01.18, JST CRDS 予測と発見：大規模情報からの『知識』獲得技術ワークショップ

図 3-2-1



分科会Bグループ

推論技術班

◎有村博紀	北海道大学大学院情報科学研究科 教授
● 上田修功	NTT コミュニケーション科学基礎研究所 協創情報研究部長
● 喜連川優	東京大学生産技術研究所 教授
● 坂内正夫	国立情報学研究所 所長
● 津本周作	島根大学医学部 教授
● 宮野悟	東京大学医科学研究所 教授
● 本村陽一	産業技術総合研究所デジタルヒューマ ン研究センター 主任研究員
● 山西健司	NEC インターネットシステム研究所 データマイニング技術センター長 兼 主席研究員
● 鷲尾隆	大阪大学産業科学研究所 教授
● 佐々木和則	CRDSシニアフェロー
● 南谷 崇	CRDSシニアフェロー
● 伊藤義曜	CRDS主任調査員

H19.01.18, JST CRDS 予測と発見：大規模情報からの『知識』獲得技術ワークショップ

図 3-2-2



獲得する知識とは

■ 知識とは

- 集積したデータや過去の経験を集約して得られる * **一般的な * 情報**で、ある目的をもった**未来の活動をうまく行う**ために再利用可能なもの

■ 知識のもつ特性

- 集約性: 集積したデータや過去の経験を**集約**して得られる**コンパクト**な表現
- 目的性: ある目的をもった活動を**うまく行う**ために、
- 即時適用性: **容易**にかつ**すぐに**利用可能なもの
- 再利用性: 繰り返し**再利用**される。
- 多様性: 書物に記述された宣言的な知識から、行動の最適化のための非言語的な知識まで表現に**大きな幅**がある。

QA: *一般的: 系統的な、極大な。特殊な知識もありえる。

H19.01.18, JST CRDS 予測と発見: 大規模情報からの『知識』獲得技術ワークショップ



図 3-2-3



獲得する知識とは

- 知識とは、集積したデータや過去の経験を集約して得られる一般的な情報で、ある目的をもった未来の活動をうまく行うために再利用可能なもの

- 特性: 集約性、即時適用性、再利用性、多様性

- 獲得される知識の例

- 診療支援知識: 数万件の検査データから数百人のプロトタイプをまとめて得られた**新しいタイプの病気経過**
- 診断: 数百万件の診療記録から得られた診断予測。医師の診断決定やEBMIに用いる。
- 化学薬品データ: 活性構造を抽出
- 病院管理: 病院での医療活動データから得られた院内感染と医療事故の事例と一般的なパターン。感染や事故防止に用いる。
- 創薬における1000遺伝子のネットワーク。高脂血症薬関連ターゲット遺伝子の機能解明と創薬に用いる。
- 自動車走行シーンの運転者の認識のモデリング。交通事故の防止に用いる。
- 製品アンケートアンケート(1万人)や、コールセンター問い合わせ(数千万件)から得られた人間の消費行動のモデル。
- 購買履歴から得られたユーザの商品リコメンデーションのための規則
- 家庭環境における子供の事故がおきる原因の一般的な規則。子供の事故防止。
- 顧客分析における携帯銘柄推薦の規則
- 組織内スキル移転における文節化と言語化されたセールスノウハウ。上位オペレーターの暗黙知を、下位オペレーターに移管

H19.01.18, JST CRDS 予測と発見: 大規模情報からの『知識』獲得技術ワークショップ



図 3-2-4



図 3-2-5

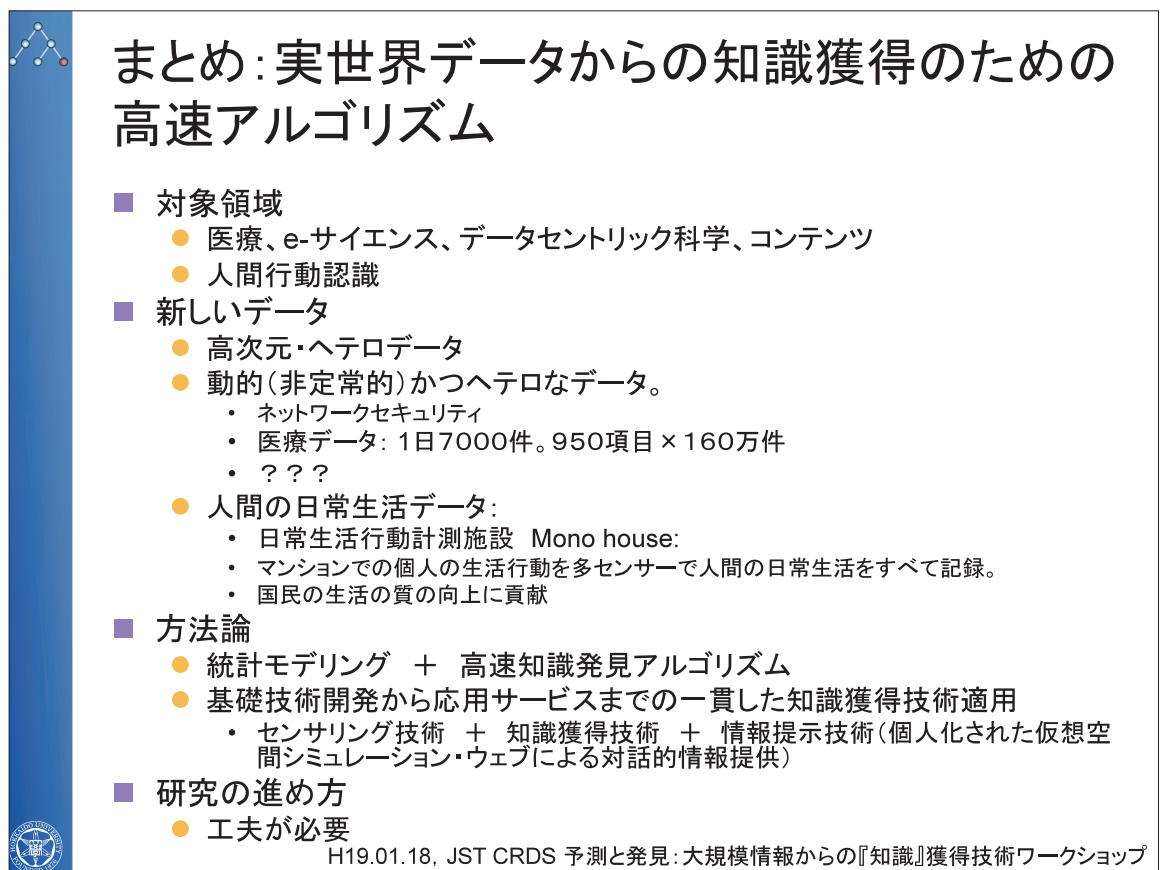


図 3-2-6



知識獲得

- 健康・医療マイニング
 - 人間の健康に関わる**ヘテロで多次元な情報**を統合し、質の良い医療を実現する。
- 生命マイニング
 - 生命を理解するための新しい情報科学が必要であり、それは超高次元・超大規模・超ヘテロデータからの予測と発見技術
- 生活行動マイニング
 - 日常生活行動に関するヘテロで多次元な情報を統合し、人間の行動様態をマイニングする。健康医療・個別化システムなど安心安全な社会を築くための基盤技術。例) 子供の事故予防。老人ケア。
- サイバー空間天文台
 - サイバー空間を観察するためのインフラとツールを研究開発する。(Cyberspace Observatory)
- 知識獲得技術に基づくシステムディペンダビリティ
 - 次世代安心安全技術のための知識獲得技術。大規模分散システムのマイニング。障害予兆検出技術。
- 実現のための要素技術
 - KAAS (Knowledge As a Service)。サービス指向知識獲得技術
 - 超高次元・超大規模・超ヘテロデータからの**高速知識獲得アルゴリズム**
 - 成長モデルを取り入れたマイニング



H19.01.18, JST CRDS 予測と発見: 大規模情報からの『知識』獲得技術ワークショップ

図 3-2-7



健康・医療マイニング

- 医療データ
 - 医療データ = **多次元ヘテロデータ。動的な成長データ。**
 - ・ 1日あたり7000件。950項目 × 160万件
 - ・ テキスト300GB/年。画像データ3TB/年
 - 病院の情報インフラ
 - ・ 数十個の異なるサブシステム(メーカーもことなる)からなる複合システム
- 知識獲得に基づく健康・医療支援
 - 診療支援
 - ・ 技術: 検査項目(2次元)の時系列をクラスタリングして、数百人のプロトタイプがひとつにまとめて、可視化。知見: **医師の知識と異なる新しいタイプの病気経過判明**
 - 診断予測
 - ・ 950項目 × 160万件の診療記録。医師の決定プロセスに近い肝硬変予測。EB MIに有効。
 - ・ 意外な結果: ネットワークからの異常発見と関係?
 - 化学薬品薬品データ
 - ・ 結果: 活性構造を抽出
 - 病院管理
 - ・ 問題: 院内感染防止。医療事故防止。膨大な広域データから知識発見の可能性
 - 進め方
 - ・ 専門化との長期の協力が可能。KAAS



H19.01.18, JST CRDS 予測と発見: 大規模情報からの『知識』獲得技術ワークショップ

図 3-2-8

生命マイニング

■ 超高次元・超大規模・超ヘテロ

- エクソンアレイデータ550万次元の表データ。
- 遺伝子多型データ(SNP, Single Nucleotide Polymorphism)
- 遺伝子の発現プロモータ
 - ・ 平均5個／遺伝子以上あるプロモータが遺伝子発現の主役
- シーケンス技術のロードマップ
 - ・ HWの進歩で100倍／年(1回・4時間のシーケンス数:2005年 $7 \times 10^4 \Rightarrow$ 2006年 7×10^4)

■ 遺伝子のイン・シリコ探索

- 高脂血症薬関連の1000遺伝子の270例のマイクロアレイ
- スパコンで2週間で遺伝子ベイジアンネットワークを推定。創薬ターゲット遺伝子を推定

H19.01.18, JST CRDS 予測と発見:大規模情報からの『知識』獲得技術ワークショップ

図 3-2-9

生活行動マイニング

■ 適用例:

- 自動車走行シーンの運転者の認識のモデリング
 - ・ 交通事故の防止
- 人間認識行動のモデル化:
 - ・ アンケート(1万人)、コールセンター問い合わせ(数千万件)、製品アンケート、購買履歴(通販の商品リコメンデーション)
 - ・ 子供の事故防止。
- 協調フィルタリング
 - ・ 顧客分析:携帯銘柄推薦
- 組織内スキル移転
 - ・ 上位オペレーターの暗黙知を、下位オペレータに移管

市民の生活品質向上のための知識獲得技術

■ 方法

- ベイジアンネット
 - ・ 多数の変数(隠れ変数を含む)間の統計的関係をモデル化。
- 協調フィルタリング
- データスカッシング・データ縮約
 - ・ 蓄積された大量データからの定数時間処理を可能。ビジネスでは時間が重要
- 超高次元・超大規模・超ヘテロデータからの高速知識獲得アルゴリズム

■ 共通プラットフォーム

- 未完成でもサービスを立ち上げて、市民にフィードバック。センシング。
- 家庭内モノデータベース & Mono house: 日常生活行動計測施設

H19.01.18, JST CRDS 予測と発見:大規模情報からの『知識』獲得技術ワークショップ

図 3-2-10



研究の推進方法2

■ サービスの重要性

- データマイニング技術：パッケージからソリューションから、サービスへ
- ユーザはDM技術を使う必要はない。ウェブページでデータ／要求を入れるだけ？
- プロセスとして一気通貫に考えることが重要だという認識ができたのはDMのおかげ
- ×ゲノム：TM／ウェブマイニングは×
 - ・ 専門家／サービスが欠けていた。

■ 価値の付け方

- データマイニングの垂直統合
- 要素技術と『適用技術』と市場開発
- 単にあてはめればよいというのではない。応用のセンスの問題

■ 新しい研究の方法論が必要



H19.01.18, JST CRDS 予測と発見：大規模情報からの『知識』獲得技術ワークショップ

図 3-2-11



研究の推進方法1

■ オープンな研究イノベーションサイト

- グランドチャレンジ的な基礎研究をオープンに、産官学そろえて、息が長くサポートすることが大事
- 同じプラットフォームで、共通形式で、共通目標を設定して、継続的に競い合う。(例：情報検索コンペ：ENTICIR)
- テキストマイニング、ブログマイニングは好適

■ 競争と協調

- (喜連川) 共通プラットフォームは、競い合うだけでなくナショナルチームでやる。(部品を持ち寄る。一緒にシステムを開発)。
- 多くの技術(例：検索エンジン)は、複合技術。

■ 基礎体力が大事

- (喜連川) 技術基盤がないとサステイナブルでない。
- (上田) Googleは基盤がしっかりしていて、体力があるから強い。
- オープンな研究イノベーションサイト
- (喜連川) 有料コンテンツ。(坂内) 学術分野もこれから大事。



H19.01.18, JST CRDS 予測と発見：大規模情報からの『知識』獲得技術ワークショップ

図 3-2-12

知識獲得

- 健康・医療マイニング
 - 人間の健康に関わる**ヘテロで多次元な情報**を統合し、質の良い医療を実現する。
- 生命マイニング
 - 生命を理解するための新しい情報科学が必要であり、それは超高次元・超大規模・超ヘテロデータからの予測と発見技術
- 生活行動マイニング
 - 日常生活行動に関するヘテロで多次元な情報を統合し、人間の行動様態をマイニングする。健康医療・個別化システムなど安心安全な社会を築くための基盤技術。例) 子供の事故予防。老人ケア。
- サイバー空間天文台
 - サイバー空間を観察するためのインフラとツールを研究開発する。(Cyberspace Observatory)
- 知識獲得技術に基づくシステムディペンダビリティ
 - 次世代安心安全技術のための知識獲得技術。大規模分散システムのマイニング。障害予兆検出技術。
- 実現のための要素技術
 - KAAS (Knowledge As a Service)。サービス指向知識獲得技術
 - **超高次元・超大規模・超ヘテロデータからの高速知識獲得アルゴリズム**
 - 成長モデルを取り入れたマイニング

H19.01.18, JST CRDS 予測と発見:大規模情報からの『知識』獲得技術ワークショップ

図 3-2-13

まとめ:実世界大規模データのための超高速知識獲得アルゴリズム

- 健康・医療マイニング
- 生命マイニング
- 生活行動マイニング
- サイバー空間天文台
- 知識獲得技術に基づくシステムディペンダビリティ
- 実現のための要素技術
 - KAAS (Knowledge As a Service)。サービス指向知識獲得技術
 - **超高次元・超大規模・超ヘテロデータからの高速知識獲得アルゴリズム**
 - 成長モデルを取り入れたマイニング

H19.01.18, JST CRDS 予測と発見:大規模情報からの『知識』獲得技術ワークショップ

図 3-2-14

3.3 サイエンス知・常識知・身体知・シナリオ知の獲得技術

分科会C 田中 譲（北海道大学）

3.3.1 応用分野と目指すべき目標

本分科会では、大規模情報からの知識獲得技術の応用分野として、（１）科学技術研究基盤システム、（２）ロボット、（３）ビジネス戦略における大規模情報からの知識獲得について討論を行った。プラントシステムへの応用に関しても討論したが、獲得すべき知識が異常予兆の検出に関する知識であるのに対し、実働システムでは、異常事態に至ることがないように運転が行われるため、異常事態に関する大規模なデータ集積が望めないという矛盾がある。シミュレーションによる異常事態データの収集も考えられるが、充分ではない。このような理由から、今回の調査対象からは除外した。

■ サイエンス知

英国が 90 年代後半にコンセプトを出し、2000 年代前半に大規模なプロジェクト推進を行った e-サイエンスは、EU の第 7 期フレームワークにおける重点課題であるリサーチ・インフラストラクチャの研究開発推進へと展開されており、高度な IT を駆使した研究開発支援基盤の研究開発が国内外で 21 世紀初頭の重要課題と認識されつつある。科学技術の研究開発現場では、実験計測装置の高度化、自動化、高速化により、

大規模な実験計測データがシステムティックに自動的に生み出されるようになってきており、その管理蓄積と分析、さらには、これらのデータからいかに有用な知識を抽出・獲得するかが大きな課題となっている。このような有用な知識を「サイエンス知」と呼ぶことにする。これらのデータは大規模文献情報とも関連付けられて仮説の生成と検証に用いられ、その結果、新たな仮説が生成され検証実験へと進展する。この仮説生成検証過程には 2 つの大きな特徴がある。第一の特徴は、知識獲得が終端でなく新たな過程の始端になり、全過程はスパイラルな繰り返し構造をなすという点である。第二の特徴は、データだけでなく、分析・発見・シミュレーションのツールが地理的にも組織的にも分散しており、ウェブ上の多くの公開資源を含んでおり、個々の研究者がこれらの多様なデータや情報を目的に応じて相互に関連付け、その結果に多様なツールを次々と自在に適用できなくてはならないという点である。分析・発見・シミュレーションの結果は研究者の新たな発想の刺激となるように可視化されることが必須である。可視化結果の一部を自在に選び取り、新たな仮説生成検証過程へとつなぐことが、研究者自身の簡便な操作で即座に実行できなくてはならない。このような「統合可視化研究基盤シス

テム」の研究開発は重要課題の一つである。

■ 常識知・身体知

将来のロボットにおいては、対話を介した人との相互作用と、センサーとアクチュエータを介した環境との相互作用の2種類の相互作用を通じて、大量の情報が入力され蓄積され、そこからロボットの推論と動作に関する有益な知識の獲得が自律的になされる必要がある。これらの有益な知識には、2種類が考えられる。1つ目は、人との対話を可能にし、環境の対象を把握し、言語による参照と対応づけ、それらに対する推論を可能にする「常識知」とよばれる知識である。2つ目は、ある目的のもとにロボットが身体を使って行う環境との相互作用において、ロボットの各種の微妙な動作と、各種センサーの時系列データ、さらにはその動作がおこす環境との相互作用の目的適合度を示す評価データ、これら3者の間の膨大な関連情報から知識獲得によって得られる知識であり、目的を上手に達成するためのロボット動作を実現するための身体制御に関する知識である。これを「身体知」と呼ぶ。環境および人との自律的相互作用を通じて行われる「自律的常識知獲得技術」と「自律的身体知獲得技術」の研究開発も今後考えるべき重要課題である。ロボットを人に置き換え、ロボットの動作制御信号を人の各種筋電位信号に対応付け、センサーを同様に用いると、ロボットの身体知獲得問題は、人の身体知の客観的抽出・獲得問題に変換される。達人の技を真似て技の習得を行う過程を、各種センサーによってデータ取得することにより、人の身体知に関する客観的知見を得ることができると期待でき

る。このような研究は「スキルサイエンス」と呼ばれる分野を開拓しつつある。

■ シナリオ知

ビジネス戦略、企業戦略においては、昨今、シナリオ・プランニングと呼ばれる手法が注目されている。想定される未来の事業環境を「シナリオ」と呼ばれる複数の一貫性あるストーリーに落とし込み、どのシナリオが実現しても対応でき得る戦略オプションを用意することで不確実な未来に備えようという考え方である。正確に唯一無二の未来を予測することを主眼とはせず、起こり得る可能性の高い未来を複眼的に捉えることで、合理的な経営上の意思決定をサポートすることを目的としている。シナリオ・プランニングを通じて将来の事業展開に対する仮想シミュレーションを繰り返すことによって、「未来経験」を重ねることができ、実際に危機がもし起こったときにも動揺せず、柔軟な対応をとることができる。シナリオという形式で表される知識を「シナリオ知」と呼ぶことにする。社会・経済状況に応じた企業戦略と企業活動の個々の事例は相互に関連する事象の系列として観察される。このような事象系列の中には、主体や客体、それらの間の関係を多様なレベルで抽象化することにより、多様な抽象度に応じて、共通の類似パターンが見出せることがある。これがシナリオ知である。事象ならびに事象系列の膨大な事例情報からシナリオ・プランニングに有用なシナリオ知を獲得する「シナリオ・マイニング技術」は重要かつ未開拓な研究開発課題の一つである。世界中の物語をデータベース化し、そこから物語のアーキタイプを自動抽出し、アーキタ

イブによる物語の分類や検索を可能にしたいという要求は古くからある。アーキタイプの自動抽出もシナリオ・マイニングに他ならない。シナリオ・マイニングの応用領域は非常に広いと考えられる。

3.3.2 研究開発課題

分科会Cでは、以下の4つの研究開発課題を提案する。

- (1) 統合可視化研究基盤システム
- (2) 自律的常識知獲得技術
- (3) 自律的身体知獲得技術
- (4) シナリオ・マイニング技術

各研究開発課題の内容を以下に示す。各課題の現状と今後取り組みが必要な課題については表3(1)~(4)にまとめて示す。

(1) 統合可視化研究基盤システム

大規模な計測・観測データベース、実時間計測・観測データ、あらゆるパラメータ値に対応した膨大な大規模シミュレーション結果のデータベース、解析ツールライブラリ、これらのデータとデータ間の関係に関する意味構造を記したメタ情報データベース、ローカル・データベースないしはウェブ上の公開情報として提供されるこれらの大規模情報源を用い、特定の問題意識に従って、これらの情報源から、インタラクティブ・ビュー技術によって、特定の条件を満たす情報群を特定の見方で自在に取得し、これらを自在に関連付け連携統合し、得られたデータ集合、情報集合から、既知の理論モデルを参考にしながら、数学モデルで記述できる規則性を発見する過程をトータルに支援する研究基盤システムが、一

つの重要な研究開発課題である。あらゆるデータ、情報、ツールを統一的に外在化し、編集、流通の対象とする知識メディア技術の確立と、知識メディア・オブジェクトの部品化と合成操作に基づいて、知識連携やツール適用が実現され、ビューの定義・操作や可視化定義・操作のフレームワークが確立されることが必要である。研究過程の全過程をスパイラルにシームレスに支援し、新知識の創生を促進する環境を構築しなくてはならない。

この課題の実現には、以下のような技術課題がある。大規模計測・観測データ、関連文献情報、さらには科学技術に関する既知の各種理論モデルを集めた大規模理論モデル・ライブラリを総合的に統合連携して用い、新たな理論モデルを自動ないしは半自動で構築する理論モデル構築技術は、未踏の研究開発課題である。文献に既に報告されているもののまだ注目されていない事実を、既知の事実と新しい観点でつなぎ合わせ、有益な新知識を創生する新関連性連鎖の発見技術も未踏の課題である。データ、情報、知識、ツールを含むあらゆる情報知識資源をビューをインタラクティブに定義・変更しながら可視化し、選択、ツール適用、結果の可視化も自在にインタラクティブに行えるインタラクティブ知識資源統合連携可視化技術と、可視化結果が新たなインタラクティブ情報アクセス空間を形成する技術に関しては、プロトタイプが研究されているに過ぎない。

(2) 自律的常識知獲得技術

第2の重要研究開発課題は、ロボットが目的に応じて人との対話、環境とのインタラクションを自律的に積極的に行うことにより、常識に関する自ら

の新しい知識を獲得し、自身の常識を自律的に拡張していく技術である。

この実現には、以下のような技術課題がある。自律的かつアクティブな常識知獲得技術と、必要な情報をえるための誘導対話技術は今後の課題である。現状では、自然言語文書を知識表現形式に自動変換し、これを用いて質問に答える QA システムがプロトタイプレベルで実現されているに過ぎない。

(3) 自律的身体知獲得技術

第3の重要研究課題は、ロボットが目的に応じて自らの動作を積極的に制御し、その結果生まれる環境とのインタラクションに関するデータや姿勢のセンシング・データとの関係を膨大に集め、そこから身体動作とその効果に関する一般規則を身体知として得る技術である。

この実現には、以下のような技術課題がある。一般化規則の抽出でなく、大規模データ関連性を制約している制約条件の抽出を行う制約知識の獲得創出技術と、目的に応じて、環境や自身の身体に積極的に働きかけながら、知識獲得に必要なデータを獲得する自律的知識創出過程の技術、身体動作の科学的評価理論の確立は、いずれも未踏の研究開発課題である。

(4) シナリオ・マイニング技術

第4の重要研究課題は、新聞記事やウェブ情報、内部議事録などに文章情報として記述される企業経営や政治経済情勢に関する事象情報や、株式市況データ、さらにはビジネス活動データベースなど大規模情報源とし、これらの事象系列や事象間関連グラフ、あるいは文章記述の膨大なデータベースから、種々の抽象レベルに対応して得られる類似のパターンをシナリオとして抽出する技術である。このようなシナリオの抽出は、シナリオ・プランニングと呼ばれる戦略プランニングに利用することができる。シナリオ・マイニングは、世界中の物語を集めた大規模物語データベースから、物語のアーキタイプを求め、アーキタイプの相違により物語を分類するのに用いることもできると期待できる。

シナリオ・マイニング技術は未踏の研究開発課題であり、現状では、人がシナリオ抽出を行うことを支援するためのシナリオ地図などの俯瞰地図を用いた思考法が提案されているに過ぎない。これらの図示を支援するツールさえも開発されていない。

表3 技術俯瞰用マトリクス* (分科会C)

(1) 統合可視化研究基盤システム

	大規模情報	取 得	データ構造化
研究開発課題 (これから何をすべきか)	■ 文献 大規模科学技術論文データベース ■ 大規模蓄積データ 大規模計測・観測データベース ■ 実時間データ 大規模実時間計測・観測データ ■ シミュレーション 大規模シミュレーションデータ (ファイル/データベース) ■ 解析ツール 大規模解析ツール・ライブラリ ■ メタ情報 所在地・利用法・関連性に関する 情報を含む大規模オントロジー これらの大規模情報が、ローカル・データベースとウェブ上のオープン・リソースの2種類の形態で存在	■ ウェブ上のリソースの自在な取得：インタラクティブ・ラッパー技術 ・ウェブ文書、ウェブ・アプリケーション、ウェブ・サービス、ブログなど、多様な形式と構造で公開されているウェブ上のリソースから、必要な形式に整理された大規模表形式データを取得抽出する技術 ・簡便かつ即座にユーザがインタラクティブに行えることが望ましい ■ アクティブ・データ取得 ・実時間でデータを生成する実験装置や観測システム、シミュレーションシステムに、取得すべきデータを指示 ・取得すべきデータは目的に応じて自動的に決定	■ インタラクティブ・ビュー ・目的に応じて、対象となる大規模データの表現構造、検索条件を自在に変えることで、見方や焦点を自在に変更しながら、知識抽出の対象データを取得する技術 ・ビューの定義とビューを介したデータ取得は試行錯誤的に行えることが望ましく、ビュー定義と、取得結果を可視化しながら直接操作で試行錯誤が可能なマンマシン・インタフェースが望まれる ■ 情報資源の統合連携 ・ウェブ上の公開情報から抽出された大規模表形式データを、各種データベース中の表形式データと共に、区別することなく、相互に関係データベース演算により連携統合して用いることができる技術 ■ スクリーニング技術 ・取得されたデータのスクリーニング
現状俯瞰 (今、どうなっているか)			・インタラクティブ・ビュー技術は未発達 ・ビューの重要性に対する認識は一部の研究者に限られるのみ ・ビュー定義は知識マイニング手法の一部となって分離されていない。両者を分離することにより、より自在な組み合わせが可能 ・情報資源の統合連携技術は未発達

*大規模情報から目的とする『知識』を獲得するための技術要素を、「取得」「データ構造化」「『知識』創出」「ヒューマンインタラクション」の技術要素に分類し、それぞれの技術要素の現状を俯瞰した上で、これからどのような研究開発をするべきかについて議論を行った。C分科会の成果を表3にまとめた。

『知識』創出	ヒューマン インタラクション	目的とする『知識』
■ 理論モデル構築技術 ・大規模計測・観測データ、関連文献情報、さらには科学技術に関する既知の各種理論モデルを集めた大規模理論モデル・ライブラリを総合的に統合連携して用い、新たな理論モデルを自動ないしは半自動で構築する技術 ・理論モデルの記述は、数式記述、論理記述、グラフ記述など、多種多様 ■ 新関連性連鎖の発見 ・文献に既に報告されているもののまだ注目されていない事実を、既知の事実と新しい観点でつなぎ合わせ、有益な新知識を創生する技術、 ■ 知識の統合連携技術 ・得られた知識を自在に連携統合して活用する技術 ■ 超高速アルゴリズム ・研究者の試行錯誤的、探索的研究過程を支援するには、レスポンス・タイムが充分短い必要がある	■ 統合可視化研究基盤システム以下の要素技術からなる。 ・データ、情報、知識、ツールを統一的に外在化表現し、相互に連携、適用可能な形式で外在化表現可能で、ウェブ技術を含むあらゆる情報知識資源をビューをインタラクティブに定義・変更しながら可視化し、選択、ツール適用、結果の可視化も自在にインタラクティブに行えるインタラクティブ知識資源統合連携可視化技術 ・可視化結果が、新たなインタラクティブ情報アクセス空間を形成 ・研究過程の全過程をスパイラルにシームレスに支援し、新知識の創生を促進	■ サイエンス知 ・数式記述、論理記述、グラフ記述など数学モデルで記述 ・新関連性連鎖に関する知識
・理論モデル構築に関しては、数式表現の族を、多項式や特定の級数展開に決めた場合には、既知の研究成果有 ・既知の各種理論モデルに関する大規模な知識ベースを持ち、大規模データから知識獲得する技術は未踏研究領域 ・新関連性連鎖の発見にかんしては、注目されていない関連性情報を、重要な関連性連鎖の発見に繋がる重要関連性として発見する技術が未発達	・知識メディア技術の基礎技術は現在研究されている ・インタラクティブ知識資源統合連携可視化技術はプロトタイプ・レベル ・統合可視化研究基盤システムは未踏研究課題	・理論モデルに関しては、数式表現の族を、多項式や特定の級数展開に決めた場合には、既知の研究成果有 ・関連性知識に関しては、頻繁に現れる単純な関連性の発見が対象

(2) 自律的常識知獲得技術

	大規模情報	取 得	データ構造化
研究開発課題 (これから何をすべきか)	■ ロボットと人の対話記録データベース ■ 実時間対話情報 ■ 常識知識ベース	■ インテンションの変化に応じたアクティブ情報取得 ・ インテンションの変化に応じ、対話を通じて必要な情報取得を行う自律的データ取得技術 ・ ロボットの現在の知識ベースを参照しつつ、必要な情報を積極的に取得	■ 対話記録情報の構造化 ・ 対話記録情報の中から、目的に応じて種々の情報を抽出するための、対話記録情報格納構造
現状俯瞰 (今、どうなっているか)		・ インテンションの表現法未発達 ・ インテンションに応じたアクティブ情報獲得技術は未踏課題	・ 対話記録を、推論に利用できるように構造化して格納する技術は充分には発達していない

『知識』創出	ヒューマン インタラクション	目的とする『知識』
■ 自律的アクティブ知識獲得 ・ 特定分野の専門知識の創出と、百科事典の知識の創出を共に行う必要有 ・ パーソナライズされた知識の創出・必要な情報をえるための誘導対話技術 ■ 常識の自動構築 ・ 対話と環境との相互作用の中から、既知の常識を参照して、新しい常識を自動構築	■ インタラクションと対話を通じたインクリメンタル知識獲得インタフェース ■ 常識エディター ・ 開発者が予めロボットに基本常識を格納するためのツール	■ 常識知
・ 自然言語文書を知識表現形式に自動変換し、これを用いて質問に答えるシステムはプロトタイプレベル	(未踏領域)	

(3) 自律的身体知獲得技術

	大規模情報	取 得	データ構造化
研究開発課題（これから何をすべきか）	■ 大規模マルチモーダル時系列データ (センサー・データ、内部状態、制御データ)	■ 非接触データ取得法 ■ 目的志向適応型多分解能取得法 ・ 目的に応じてアダプティブに分解能を切り替えながらデータを取得する技術 ■ アクティブ・データ取得技術 ・ インテンションの変化に応じて、積極的に環境との相互作用や、身体動作を制御し、必要なデータを取得する技術 ■ セマンティック信号処理法 ・ 信号の意味を理解しながら適応的に行われる信号処理技術	■ 多分解能時空間データ構造 ・ 途中で分解能が変化したり、複数の分解能のデータを同時に持つ時系列データのデータ構造 ■ インテンション記述
現状俯瞰（今、どうなっているか）	・ マルチモーダル時系列データ	・ 目的志向のアクティブなデータ取得技術は未発達	・ マルチモーダル時系列解析 ・ マルチモーダル多分解能時系列データのデータ構造、解析法の研究は不十分

『知識』創出	ヒューマン インタラクション	目的とする『知識』
■ 制約知識の獲得創出 ・一般化規則の抽出でなく、大規模データ関連性を制約している制約条件の抽出 ■ 自律的知識創出過程 ・目的に応じて、環境や自身の身体に積極的に働きかけながら、知識獲得に必要なデータを獲得する技術 ■ 身体知の有効性と評価の理論 ・身体動作の優劣は、目的適合性だけでなく、美的評価にもよる。身体動作の科学的評価の理論の確立が必要 ■ 実時間高速アルゴリズム ・自律的に身体に働きかけ、身体動作に関する実時間データ収集を行い、知識獲得技術によって身体知を実時間で修正し、再度身体に働きかけて結果を評価するには、実時間での高速な知識獲得技術の開発が必要	■ 解釈を伴う知識獲得 ・身体知の獲得は身体動作に対する解釈を伴う ■ スキル創造支援 ・得られた身体知を自在に連携統合して合成身体動作に関する身体知を創造し活用する技術	■ 身体知 ・環境（物理環境、自然環境、社会環境）における身体知 ■ 行動知 ■ 個人／集団のスキル
・一般化動作規則の抽出 ・自律的知識創出過程は未踏領域 ・身体動作の評価理論は未発達	・メタ認知可視化 ・単純動作の合成によるアニメーション作成支援技術	

(4) シナリオ・マイニング技術

	大規模情報	取 得	データ構造化
研究開発課題 (これから何をすべきか)	■ 企業経営、政治経済情勢に関する事象情報 (新聞記事、ウェブ情報、内部議事録) ■ 株式市況データ ■ ビジネス活動データベース	■ 目的に応じて定義されたビューを介した情報取得	■ 事象系列・事象関連グラフのデータベース化 ■ 事象と事象間の関係に関する解釈に関するオントロジー・データベース
現状俯瞰 (今、どうなっているか)		・ビュー定義とビューを介した取得はデータベースに対してのみ可能	・事象系列のデータベース化と検索技術に関してはテンポラル・データベースの研究のみ ・事象関連グラフのデータベース化と検索技術は未発達

『知識』創出	ヒューマン インタラクション	目的とする『知識』
■ シナリオ抽出 ・ 事象系列や事象間関連グラフ、あるいは文章記述の膨大なデータベースから、種々の抽象レベルに対応して得られる類似のパターンをシナリオとして抽出する技術 ■ シナリオ共創技術 ・ コミュニケーション・コラボレーション支援により、複数の利用者がシナリオ抽出を半自動で共創する技術 ■ シナリオの価値と評価の理論	■ 可視化・俯瞰地図自動／半自動作成技術 ■ シナリオマップ自動／半自動生成技術 ■ シミュレーションとの連携 ■ 情報デザイン支援・自動作成ツール	■ シナリオ知 ・ 物語の型 ・ ワークフローの型 ・ 手順の型
・ ナレッジ・マネージメント（計算機は内容を理解不能）	・ 情報デザイン ・ 知識マップ	

3.3.3 研究推進方法

研究推進は、応用分野を特定した Grand Challenge 型プロジェクトの形式で行うべきと考える。先駆的応用分野において、ミッション駆動型の研究推進を行い、運用レベルに近いシステムの研究開発を遂行することにより、個々の要素技術を独立に研究開発するだけではなく、革新的な統合連携基盤技術の研究開発と、ロバストな要素技術の研究開発が同時に遂行できると期待される。

社会的に意義のある研究推進対象分野としては、来るべき高齢化社会を念頭に置いた以下の2つの研究課題を提案する。

- ① 介護福祉における身体知、シナリオ知、常識知の統合的獲得・利用技術の研究開発
- ② 成人病臨床ゲノム治験と成人病予防におけるサイエンス知、身体知、シナリオ知、常識知の統合的獲得・利用技術の研究開発

この他にも

- ③ 生産・輸送システム支援における知識獲得利用
- ④ 教育・生涯学習支援における知識獲得利用

が考えられるが、以下では、特に①②について詳述する。

介護においては、介護者ならびに被介護者の身体知の獲得が重要である。介護者をロボットが勤める場合も同じである。また、介護福祉に関しては、将来起こりえる種々のシナリオを想定

したシナリオ・プランニングが重要であるが、このような種々のシナリオの抽出を膨大な実例事象データベースから求めることができれば大きな意義がある。介護ロボットにおいては自律的常識知の獲得も、被介護者とのコミュニケーションをスムーズに行うために極めて重要な課題である。介護福祉においては、これらの異なる知識は相互に関連しており、他の知識の獲得に利用される。これらの知識の統合的獲得・利用技術の研究開発が重要である。

成人病臨床ゲノム治験と成人病予防においては、臨床データや、ゲノムデータに基づき、治験の根拠となるサイエンス知の獲得技術が重要であるばかりでなく、この分野は統合可視化研究基盤システムの格好の応用分野となる。また、介護・福祉の場合と同様の理由から、身体知、シナリオ知、常識知をも含めた統合的獲得・利用技術の研究開発の格好の応用分野である。

3.3.4 大規模情報から獲得する『知識』とは

分科会Cでは、(1) 科学技術研究基盤システム、(2) ロボット、(3) ビジネス戦略における大規模情報からの知識獲得について討論を行った結果、獲得される知識を、「システムやユーザがそれを用いて社会と相互作用することにより拡大再生産されるもの」と定義するに至った。この定義には、知識獲得過程のスパイラルな永続性と自律性が特徴付けられている。知識獲得は決して一過性の処理ではなく、その結果が次の知識獲得過程の始点となる。また、知識獲得には、それに先立ち、目的に応じて特定の条件を満たす情報を特定の観点と焦点で取得するという

自律的かつアクティブな一面がある。
いわゆるビューの存在と、その定義・
操作の重要性を忘れてはならず、ビュー
の定義・操作がシステムによって自律

的に行われたり、ユーザによってアド
ホックに行われることに注目しなけれ
ばならない。

目
的

大規模情報からの
「知識」獲得技術について

分
科
会
A

分
科
会
B

分
科
会
C

ま
と
め

付
録

分科会C報告

図 3-3-1

知識

獲得・適用・解釈評価スパイラル過程の中で捉える

知識＝「システムやユーザがそれを用いて**社会と相互作用**することにより**拡大再生産**されるもの」

分科会Cが注目した「大規模情報から獲得される『知識』」

- **サイエンス知←統合可視化研究基盤システム**
 - － 科学技術知識（文献、計測・観測データ、シミュレーション・データ）
- **常識知←自律的常識知獲得技術**
 - － 特定分野、特定個人の常識知
- **身体知←自律的身体知獲得技術**
 - － 身体と物理環境、自然環境、社会環境との相互作用のスキル・技能
- **シナリオ知←シナリオ・マイニング技術**
 - － ワークフローや手順、物語の類型に関する知識

図 3-3-2

研究推進

- 応用分野を決めたGrand Challenge型プロジェクト
 - 先駆的応用分野において、ミッション駆動型の研究推進を行い、運用レベルのシステムの研究開発を遂行することにより、個々の要素技術の独立な研究開発ではなく、革新的な統合連携基盤技術とロバストな要素技術を共に研究開発することができる。
- 対象分野
 - 高齢化社会支援
 - 介護福祉における身体知、シナリオ知、常識知の統合獲得利用技術
 - 成人病臨床ゲノム治験と成人病予防におけるサイエンス知、身体知、シナリオ知、常識知の統合獲得・利用技術
 - 生産・輸送システム支援における知識獲得利用技術
 - 教育・生涯学習支援における知識獲得利用技術

図 3-3-3

4

全体討議のまとめ

4. 全体討議のまとめ

有川 節夫（九州大学副学長）

4.1 各分科会への要請と対応

初日の午後から3つの分科会に分かれて議論し、その中間状況の報告を初日の夕刻に行い、夕食後再び分科会に分かれて徹底した議論を行った。その結果に基づいて、二日目の午前中に十分な時間をとって各グループリーダーから報告し、全体での討論を行った。分科会Aは、応用分野として主にマーケティング・気象を、手法として統計的モデリングを中心に、分科会Bは、応用分野として医療・装薬・バイオを、手法としてはデータマイニング等を中心に、分科会Cは、応用分野として大規模プラント・ロボット・e-Scienceを、手法としてはヒューマンインタラクションを意識して議論をスタートした。しかし、同時に各分科会に『知識』獲得技術俯瞰用のマトリクスシートが配布され、それぞれの分科会で、自分達の分科会に割り当てられたアプリケーションから見た、大規模情報や情報の取得、データの構造化、『知識』の創出、ヒューマンインタラクション、求められる『知識』といった6つの項目を具体的に埋める作業を求め、それを二日目のリーダーによる分科会からの報告で使い、技術の現状を俯瞰し、研究課題を提示し、研究の推進方法を示し、同時に『知識』の定義もお願いした。

しかし、こうした予めお願いしてい

たフレームを超えた議論が各分科会でなされ、単にサイバー空間やリアル空間にある大規模情報からの知識獲得だけでなく、現在の高度情報社会、高齢化社会、急速に進展する先端学問分野において迅速かつ個別的に対応しなければならない『知識』獲得の技術が、非常に多くの分野や場面で必要になっていることが浮き彫りにされた。また、そうした中で、地道な学問としての深化と展開が必要であることも共通の認識となってきた。

4.2 対象とする『知識』の多様性

獲得すべき『知識』には多様性と多面性があることが明らかにされた。それは、センシング技術の進展・普及によって多様で夥しい量のデータや情報が容易に日常的に収集され、その用途・応用が多様であることに起因するものである。サイバー空間にある巨大な情報や科学的な観測・実験によって得られるデータはもとより、経済やマーケット、個人の日常の家庭内での生活にまで入り込んだパーソナルなデータに至るまで、情報・データは実に様々である。そして、こうした情報やデータからそれぞれの活動を支えるための『知識』の獲得が期待されているのである。『知識』も常識的な普遍性のある『知識』から使い捨てとさえ考えられる個別的でその場面だけでのみ有効なものまで

あり、それらが、例えば、ロボットが環境を認識して行動するのに不可欠であったりする。分科会 C では、常識知に加えて、サイエンス知、身体知、シナリオ知といった『知識』も考えなければならないと纏め上げている。こうした多様性に加えて、そうした知識を、誰が、どのような場面で、いつ、何のために使うのかという問題、つまり、知識を使う主体、場面、目的等も絡めて、その獲得技術を研究開発することも当然必要になる。

4.3 大規模情報から獲得する『知識』とは

初日の冒頭で紹介したように、これまでに考えられてきた『知識』の定義には多様性がある。我々は、それらに少なからずある種の違和感をもっていることを認識しながら、このワークショップを進め、最後に我々の『定義』を示そうとしてきた。各分科会で大規模情報から獲得する『知識』について議論していただき、その結果を全体会議で披露し、その後、このまとめに際して分科会のリーダーから最終的な考えを寄せていただいた。それは、各分科会のまとめの最後にあるように、以下のようなものになっている。

(1) 分科会 A では、『知識とは、個別の人間生活にとって役立つ“制約”である。』としている。

広大な不確実性の空間に適切な“制約”を加えることによってはじめて、個人にとって役に立つ情報が獲得できるからである。したがって、“制約”は、ポジティブなものであり、設計概念や機能、モデル、プロセス等を含む広い概念である。また、知識発展のプロセ

スにともなって変化する動的なものである。そして、『情報』との関係を、『情報』とはこの不確実性を減少させるものであるとして、それを実現するのは、適切なモデリングを用いる知識獲得技術であるとしている。

(2) 分科会 B では、『知識とは、ある目的をもった未来の活動をうまく行うために再利用可能な、集積したデータや過去の経験を集約して得られる情報である。』としている。

そして、こうした知識には低レベルのものから高レベルのものまであり、獲得する主体にとって、環境や経験をコンパクトに一般化し集約して、適切に利用することになる。限られた能力と変化する環境のもとで、より適切な活動を行うことが可能になる。また、『知識』は、データ集合や記録と違い、集約性、目的性、即時適用性、再利用性、多様性等をもっているとしている。

(3) 分科会 C では、『知識とは、システムやユーザがそれを用いて社会と相互作用することにより拡大再生産されるものである。』としている。

この定義には、知識獲得過程のスパイラルな永続性と自律性が包含されている。獲得された知識獲得は、その次の知識獲得過程の始点として使われる。また、知識獲得には、それに先立ち、目的にも関係したビューが存在する。

こうして、『知識』の定義を並べてみると、その内のどれが妥当でどれが不十分というようなことではなく、我々が意識していた知識に対する考え、感覚、側面が、重点の置き方を変えて、それぞれ明確に表現されているように思われる。

ここで、各リーダーには前日の打合せの時に話したことであるが、論理や一世を風靡した論理プログラム、帰納的論理プログラミングにおける知識の定義に触れておく。論理の世界では、あっさりいってしまえば、『知識』は、一般的な公理や（形式）理論における（数学的）公理であるといえる。それを出発点にして、推論の規則を適用して、次々に定理としての新しい『知識』が得られることになる。

論理プログラムは、いわゆる確定節の有限集合であるが、そこでは、

$$p(a, b, \dots, c) \leftarrow$$

の形の節をデータ、その集まりをデータベースと呼び、

$$A \leftarrow B, C, \dots, D$$

の形をした節も合わせて『知識』と呼んできた。この定義は、形式的なものであるが、データと知識の違いを端的に表現している。また、ここでいう、データベースから、上記のような『知識』を獲得するために「モデル推論」をはじめとする様々な機械学習の手法が研究されてきた。ここでいうモデルは、論理における解釈であり、上記の論理プログラムの場合には確定節の集合としての『知識』と考えていい。さらに、帰納的論理プログラミングというものも考えられてきた。そこでは、背景知識という一種の既知の『知識』を用いている。また、論理プログラミングの世界では、「制約論理プログラミング」という一歩進んだ実用的な問題解決のスキームも研究されてきた。ここでは、“制約”を満たす形で解を求めるわけである。これは分科会 A での“制約”と

通じるところがあり、それ自体『知識』であるともいえる。これらは、論理の枠組みでのことであり、他の枠組みでの知識の獲得にはそぐわないところもあるだろうが、ここでいう背景知識やモデル、“制約”としての知識という考えは、『知識』の定義を考える際に有効であろう。

さて、三つの分科会から提案された『知識』の定義を、それが大規模情報から獲得されるものであることや上記のことも考慮して、次のように纏めてみる。

『大規模情報から獲得される知識とは、個々のシステムや個人がもつビューを満たし目的達成のために使えるように、集積した情報を出発点にして一般化し、既知の知識や蓄積した経験・情報に照らし集約して得られるもので、この過程はスパイラルに繰返し適用され再生産される。』

この意味するところは、上記各分科会での定義であり、そこで行われているように若干の補足説明は必要である。上記の定義を共通の認識として持ち、各分科会で議論され、認識されたことでもって、補うのが妥当であろう。なお、暗黙知の獲得や表出については、別な考察が必要であろう。

4.4 今後研究開発すべき重要な研究開発課題

この大規模情報からの知識獲得に関して今後展開しなければならない重要な研究開発の課題は、各分科会のまとめ及び技術俯瞰マトリクスに示されている通りであるが、特に、重要と思われるものをいくつか挙げておく。

1. ベイズモデル構築技術・ベイジアンネットワーク技術
2. モデルに基づいたアクティブセンシング
3. 数理統計と機械学習の融合・発展
4. マーケティング・ビジネス戦略における知識獲得
5. パーソナリゼーション・大衆から個への時代への対応・ジャストアシスト方法
6. リスクマネジメントのための知識獲得・巨大システムの障害検出
7. サイバースペース天文台
8. テキスト・半構造データマイニング
9. 自然言語データからの知識獲得
10. 科学技術のための知識獲得・データ中心科学
11. ロボットのための知識獲得
12. 統合可視化研究基盤システム
13. 自律的常識知・身体知獲得技術
14. シナリオ・マイニング技術

この内、サイバースペース天文台は、巨大でヘテロなサイバースペースで何が起きているかを組織的・定常的に観測するもので、サイバーワールドとリアルワールド双方の危機管理のために特に重要であると考えられる。こうした、社会的要請の高い課題として、他に、高齢化社会を念頭においた、介護福祉における統合的獲得・利用技術、成人病臨床ゲノム治験と成人病予防における統合的獲得・利用技術がある。また、生産・輸送システム支援における知識獲得利用、教育・生涯学習支援における知識獲得利用の研究開発も重要である。

4.5 研究開発課題の推進方法

研究開発の方法に関しても各分科会からの報告の中で提案がなされているが、特に、重要なものを再掲しておく。

(1) 現場主義と一貫通貫性

フィールドに密着した現場主義を重視し、いくつかの応用分野を選びブランドチャレンジ型プロジェクトの形式で、ミッション駆動型の研究推進が必要である。また、ビジネスセクターを平行して開拓し、運用レベルの研究を展開すべきである。こうした推進方法は、大規模情報からの知識発見のように現場からの要請に明確に応えることが求められている課題については、極めて重要である。また、理論展開から現場に至る研究を貫く研究展開も重要である。

(2) 相互協力と相互干渉

知識獲得の方法論には、統計解析に基づくものと機械学習に基づくものも含めて、対立・競合する方法論もあるが、相互の協調を超えて、相互に干渉・貫入するといった姿勢が時として必要になる。また、データ取得、データ構造化、統計数理モデリング、データマイニング、推論、ヒューマンインタラクション等の各要素技術についても、個別独立ではなく、相互の干渉・貫入によって研究開発が進められる必要がある。そのような研究展開によって、実用に耐える革新的な成果が生まれるであろう。

(3) 共通プラットフォーム

このような研究開発をサポートするには、共通プラットフォームを構築し、オープンな研究イノベーションサイトを設置する必要がある。それを共有し利用することによって、相互協力と干渉が進み、新しい技術の創成が活性化される。

なお、現実の具体的な課題を出発点にして上記の推進方法により基礎研究を展開し、与えられた課題の解決を試みる。こうした基礎のしっかりした研究は、サポートベクターマシンに代表されるように、実際の課題を解決する強力な手段を提供するだけでなく、次世代の強固な基盤構築に寄与する。このような研究を展開するために、かなりの規模の優秀な研究者に対して重点的かつ組織的な財政支援が必要である。

4.6 他のプロジェクトとの違い

知識獲得・発見に関係する文部科学省関係（科学研究費補助金）による大型のプロジェクトについては、第2章で経緯を紹介してきたように、完了したものとしては、1998年度～2000年度に展開した特定領域「発見科学」があり、その後継として2001年度～2004年度の「情報洪水」がある。前者は、「発見科学」という新しい学問分野を確立しようとしたもので、科学哲学から実際の大規模データからの知識発見を含む体系化を目指したものであった。後者は、その中の実際的な問題に的を絞り、真にフィールドが必要とする問題に対してデータの収集に対してもアクティブに取り組む、効果的な知識発見をしようとするものであった。また、特定領

域研究（C）といういわば国策としてトップダウン的に計画された「情報学」は、文字通り情報学の基盤を強化するためのもので、情報学のほとんどすべての分野をカバーする包括的なものであった。その中に、ウェブを対象にした知識発見の研究も数件含まれていた。こうした研究プロジェクトによって、このワークショップでカバーすべき研究開発の基盤と基礎が準備され、研究グループが育成されてきたといえる。

「情報学」の後継とも位置づけられ、現在進行中の特定領域研究「情報爆発」は、「とりわけ21世紀に入ってから爆発的に増大しているデジタル情報がIT技術や社会にまで及ぼす様々な問題に対して総合的に挑戦しようとするもので、情報に関する喫緊の課題を「情報爆発」ということばに集約し、種々の情報分野から多面的に展開しようとするものである。」（「情報爆発」平成18年度研究概要の「まえがき」から。）主な研究項目は、（1）情報爆発時代における情報管理・融合・活用基盤、（2）情報爆発時代における安全・安心ITシステム基盤、（3）情報爆発時代におけるヒューマンコミュニケーション基盤、（4）情報爆発時代における知識社会ガバナンス、の四つであり本ワークショップで扱っている課題は、（1）の中で、ウェブ情報を対象にしたものも何件が含まれている。

二日目の冒頭で頂いた星野利彦文部科学省情報技術推進室長の話にもあったように、総合科学技術会議もフォローアップしている経済産業省、文部科学省、総務省の3省連携戦略プロジェクト「情報集積活用基盤技術」として、（1）情報大航海プロジェクト、（2）革新的実行原理に基づく超高性能データベース基盤ソフトウェアの開発、（3）電気

通信サービスに関する情報信憑性検証技術の研究開発、の三つがある。これらは、インターネット上の情報検索、分析、管理を対象にした施策であり、大量情報からの価値の創出という面では共通点があるが、こうした現在進行中の戦略科学技術イノベーションパイプラインでのプロジェクトは、既存の技術を応用し、あるいは応用技術を実用化することを主眼にしたものがほとんどである。

したがって、このワークショップで検討しているように基礎研究からスタートして体系的かつ組織的な研究開発活動を展開し、現在の諸課題に答え、次世代のICT社会に求められる理論・技術を追究

し、次世代の発展ための基礎を築いておこうとするものとは大きな違いがある。

なお、戦略科学技術イノベーションパイプライン網でも分かるように、すべてのパイプラインは、基礎から応用、応用から実用へというレイヤーに向いている。応用、実用レイヤーの研究には、それぞれ基礎、応用レイヤーの研究が先行している。このように、このワークショップで扱う研究開発は、現在進行中の情報大航海等のプロジェクトを基礎から支え、後続するこうした大型プロジェクトの基盤を構築し、新たな展開に備えるためにも必要不可欠なものである。

目
的

大規模情報からの
「知識」獲得技術について

分
科
会
A

分
科
会
B

分
科
会
C

ま
と
め

付
録

付 録

付録1 ワークショッププログラム

- ◆ 開催日：平成19年1月18日（木）9:00-21:00
1月19日（金）9:00-12:00
- ◆ 開催場所：研究開発戦略センター（CRDS）2F 大会議室

第1日 1月18日（木）

総合司会：丹羽邦彦 CRDS シニアフェロー

趣旨説明		9:00-9:15
1. 挨拶		
生駒俊明 CRDS センター長		9:00-9:05
2. 開催趣旨説明		
丹羽邦彦 CRDS シニアフェロー		9:05-9:15
合同セッション（大規模情報から獲得する『知識』とは）		9:15-12:00
座長：有川節夫 九州大学 副学長		
※ 今回のワークショップで議論すべき“大規模情報から獲得する『知識』”について、一定の共通認識を形成する。		
1. オリエンテーション 大規模情報から獲得する『知識』とは		9:15-9:45
有川節夫 九州大学 副学長		
2. 総合討論		9:45-12:00
参加者全員		
【昼食】		12:00-13:00
分科会1（技術分野の俯瞰と研究開発課題の抽出）		13:00～17:00
司会：各分科会リーダー		
※ 分科会に分かれ、特定のアプリケーションを設定し、技術分野の俯瞰と課題の抽出を行う。		
全体会議（中間報告）		17:00～18:00
座長：有川節夫 九州大学 副学長		
※ 技術分野の俯瞰、課題について分科会リーダーによる中間報告・意見交換		
【夕食】		18:00-19:00
分科会2（技術分野の俯瞰と研究開発課題の抽出）		19:00～21:00
司会：各分科会リーダー		
※ 分科会1の続きに加えて、研究推進方法を含めて討議を行い、翌日の報告内容をまとめる。		

目的

大規模情報からの『知識』獲得技術について

分科会A

分科会B

分科会C

まとめ

付録

第2日 1月19日（金）

総合司会：丹羽邦彦 CRDS シニアフェロー

全体会議（ワークショップ成果のとりまとめ）

9:00～12:00

座長：有川節夫 九州大学 副学長

1. 新規連携施策「情報集積活用基盤技術」について
星野利彦 文部科学省 研究振興局 情報課 情報技術推進室長
2. 分科会リーダーから報告
 - ① 技術俯瞰
 - ② 重要研究開発課題（アプリ依存の課題・共通の課題）
 - ③ 研究推進方法
3. 討議・まとめ

閉会挨拶

丹羽邦彦 CRDS シニアフェロー

以上

付録2 参加者一覧

有川 節夫	九州大学 副学長・・・・コーディネーター
有村 博紀	北海道大学大学院 情報科学研究科 教授・・・・分科会リーダー
北川源四郎	統計数理研究所 所長・・・・分科会リーダー
田中 譲	北海道大学大学院 情報科学研究科 教授・・・・分科会リーダー
赤穂昭太郎	産業技術総合研究所 脳神経情報研究部門 グループ リーダー
石井 信	奈良先端科学技術大学院大学 情報科学研究科 教授
上田 修功	NTT コミュニケーション科学基礎研究所 協創情報研究部長
大澤 幸生	東京大学 工学系研究科 助教授
小野田 崇	電力中央研究所 システム技術研究所 主任研究員
喜連川 優	東京大学 生産技術研究所 教授
児西 清義	株式会社ナレッジプロデュース 代表取締役
坂内 正夫	国立情報学研究所 所長
高野 明彦	国立情報学研究所 教授
竹林 洋一	静岡大学 創造科学技術大学院 創造科学技術研究部情報学部 情報科学科 兼任教授
田邊 國士	早稲田大学 理工学術院 客員教授
津本 周作	島根大学 医学部 教授
照井 伸彦	東北大学大学院 経済学研究科 教授
中島 秀之	はこだて未来大学 学長
西田 佳史	産業技術総合研究所 デジタルヒューマン研究センター 主任研究員
樋口 知之	統計数理研究所 副所長
広瀬 直毅	九州大学 応用力学研究所 助教授
古川 康一	慶應義塾大学政策・メディア研究科教授 兼 環境情報学部 教授
松本 裕治	奈良先端科学技術大学院大学 情報科学研究科 教授
宮野 悟	東京大学 医科学研究所 教授
本村 陽一	産業技術総合研究所 デジタルヒューマン研究センター 主任研究員
山西 健司	NEC インターネットシステム研究所 データマイニング技術センター長 兼 主席研究員
油田 信一	筑波大学 システム情報工学研究科 教授
横井 俊夫	東京工科大学 メディア学部 教授
鷲尾 隆	大阪大学 産業科学研究所 教授
星野 利彦	文部科学省 情報課 情報技術推進室長
木村 裕明	文部科学省 情報課 課長補佐
松尾 浩司	文部科学省 情報課 科学技術・学術行政調査員
大平 浩之	文部科学省 情報課 情報技術推進係長
小菅 一弘	東北大学大学院 教授・CRDS 特任フェロー
中島 啓幾	早稲田大学 理工学部 教授・CRDS 特任フェロー
今井 秀樹	中央大学 理工学部 教授・CRDS 特任フェロー

目
的

大規模情報からの
「知識」獲得技術について

分
科
会
A

分
科
会
B

分
科
会
C

ま
と
め

付
録

生駒 俊明	科学技術振興機構	研究開発戦略センター	センター長
浅見 康弘	科学技術振興機構	研究開発戦略センター	上席フェロー
永野 博	科学技術振興機構	研究開発戦略センター	上席フェロー
丹羽 邦彦	科学技術振興機構	研究開発戦略センター	シニアフェロー
佐々木和則	科学技術振興機構	研究開発戦略センター	シニアフェロー
南谷 崇	科学技術振興機構	研究開発戦略センター	シニアフェロー
波多腰玄一	科学技術振興機構	研究開発戦略センター	シニアフェロー
石正 茂	科学技術振興機構	研究開発戦略センター	フェロー
楠本 博之	科学技術振興機構	研究開発戦略センター	フェロー
雄山 泰直	科学技術振興機構	研究開発戦略センター	フェロー
伊東 義曜	科学技術振興機構	研究開発戦略センター	主任調査員
嶋田 一義	科学技術振興機構	研究開発戦略センター	アソシエイトフェロー
酒井 重樹	科学技術振興機構	研究領域総合運営室	副調査役
古川 雅士	科学技術振興機構	特別プロジェクト推進室	主査
永井 智哉	科学技術振興機構	研究領域総合運営室	主査
荒川 隆	高度情報科学技術研究機構	計算科学技術部	主任研究員

以上

付録3 分科会メンバー構成

◆ 分科会 A

北川源四郎	統計数理研究所 所長 【リーダー】
赤穂昭太郎	産業技術総合研究所 脳神経情報研究部門 グループ長
石井 信	奈良先端科学技術大学院大学 情報科学研究科 教授
田邊 國士	早稲田大学 理工学術院 客員教授
照井 伸彦	東北大学大学院 経済学研究科 教授
西田 佳史	産業技術総合研究所 デジタルヒューマン研究センター チーム長
樋口 知之	統計数理研究所 副所長
広瀬 直毅	九州大学 応用力学研究所 助教授
波多腰玄一	CRDS シニアフェロー
楠本 博之	CRDS フェロー
嶋田 一義	CRDS アソシエイトフェロー

◆ 分科会 B

有村 博紀	北海道大学大学院 情報科学研究科 教授 【リーダー】
上田 修功	NTT コミュニケーション科学基礎研究所 協創情報研究部長
喜連川 優	東京大学 生産技術研究所 教授
坂内 正夫	国立情報学研究所 所長
津本 周作	島根大学 医学部 教授
宮野 悟	東京大学 医科学研究所 教授
本村 陽一	産業技術総合研究所 デジタルヒューマン研究センター 主任研究員
山西 健司	NEC インターネットシステム研究所 データマイニング技術センター長
鷺尾 隆	大阪大学 産業科学研究所 教授
佐々木和則	CRDS シニアフェロー
南谷 崇	CRDS シニアフェロー
伊東 義曜	CRDS 主任調査員

◆ 分科会 C

田中 讓	北海道大学大学院 情報科学研究科 教授 【リーダー】
大澤 幸生	東京大学 工学系研究科 助教授
小野田 崇	電力中央研究所 システム技術研究所 上席研究員
児西 清義	株式会社ナレッジプロデュース 代表取締役
高野 明彦	国立情報学研究所 教授
竹林 洋一	静岡大学 創造科学技術大学院 教授
中島 秀之	はこだて未来大学 学長
古川 康一	慶應義塾大学政策・メディア研究科教授 兼 環境情報学部 教授
松本 裕治	奈良先端科学技術大学院大学 情報科学研究科 教授
油田 信一	筑波大学 システム情報工学研究科 教授
横井 俊夫	東京工科大学 メディア学部 教授
小菅 一弘	東北大学大学院 工学研究科 教授 (CRDS 特任フェロー)
丹羽 邦彦	CRDS シニアフェロー
石正 茂	CRDS フェロー

以上

目
的

大規模情報からの
「知識」獲得技術について

分
科
会
A

分
科
会
B

分
科
会
C

ま
と
め

付
録

**「予測と発見」
—大規模情報からの『知識』獲得技術—
ワークショップ報告書**

独立行政法人 科学技術振興機構 研究開発戦略センター

制作担当 生駒グループ

〒102-0084 東京都千代田区二番町3番地

電話 03-5214-7481

ファクス 03-5214-7385

<http://crds.jst.go.jp/>

2007年5月

©2007 CRDS/JST

許可なく複写・複製することを禁じます。
引用を行う際は、必ず出典を記述願います。
