戦略的創造研究推進事業
ERATO
追跡評価用資料

「湊離散構造処理系」プロジェクト
(2009.10～2016.3)

研究総括:湊　真一

2022 年 3 月

目次

要旨 1
プロジェクトの発足に至る経緯及び展開状況(まとめ図) 3
第 1 章 プロジェクトの概要 4
1.1 研究期間 4
1.2 プロジェクト発足に至る科学技術や社会の背景 4
1.2.1 科学技術の背景 4
1.2.2 社会の背景 5
1.3 プロジェクトのねらい 5
1.4 研究体制 6
1.5 プロジェクト終了時点での研究成果やその意義 6
1.5.1 離散構造処理の基盤研究 7
1.5.2 機械学習・制約充足応用研究 17
1.5.3 統計・マイニング応用研究 24
1.5.4 セミナー・ワークショップ開催とアウトリーチ活動 34
第 2 章 プロジェクト終了から現在に至る状況 36
2.1 プロジェクトの終了後の状況に関する基礎データ 36
2.1.1 調査方法 36
2.1.2 競争的研究資金の獲得状況 39
2.1.3 論文の発表状況 42
2.1.4 特許の出願・公開・登録状況 46
2.1.5 招待講演 47
2.1.6 受賞状況 47
2.1.7 ベンチャー企業の設立状況 47
2.1.8 報道関係 48
2.2 プロジェクトの進捗状況 51
2.2.1 離散構造処理の基盤研究 51
2.2.2 機械学習・制約充足応用の発展 62
2.2.3 統計・マイニング応用の発展 71
2.3 プロジェクト参加研究員の活動状況 75
2.4 第 2 章まとめ 76
第 3 章 プロジェクト成果の波及と展望 80
3.1 科学技術への波及と展望 80
3.1.1 アルゴリズムの研究 80
3.1.2 応用研究の展開 83
3.1.3 研究者の育成 86
3.2 社会への波及と発展 87

3.2.1 研究成果の社会への波及と展開..... 87
3.2.2 ベンチャー企業の創出..... 89
3.2.3 アウトリーチにおける成果と波及効果..... 91
【引用文献】 93

要旨

本資料は、戦略的創造研究推進事業のERATO「湊離散構造処理系」プロジェクト(2009年10月～2016年3月)において、研究終了後一定期間を経過した後、副次的効果を含めて研究成果の発展状況や活用状況を明らかにし、国立研究開発法人科学技術振興機構(JST)事業および事業運営等に資するために、追跡調査を実施した結果をまとめたものである。

科学技術の背景として、離散構造データを計算機上にコンパクトに表現し、モデルの解析、最適化などの処理を効率よく行う技法は、基盤技術として非常に重要である。一方、実問題を解くために、様々な工学的応用技術も加えて実用に耐えるアルゴリズムを実装することを重視した。

社会的な背景として、社会の中に蓄えられたデータを有効に扱い、価値ある情報を得ることが強く求められている。また、社会基盤としての通信、電力網等を代表とする複雑なネットワークの信頼性確保や故障からの迅速な回復等が現代社会の課題である。

これらの背景の下、本プロジェクトは(1)研究総括が1993年に発表したZero-suppressed Binary Decision Diagram(ZDD)を中心とした基盤ソフトウェアを整備し、公開・普及活動を行う、(2)データマイニングや設計問題等に対して、応用分野の研究者や企業と連携して、その効果を定量的に示す。(3)若い人材の育成に注力し、技術コミュニティを確立させる、という3つの目標を掲げた。

プロジェクトは離散構造処理グループ(湊真一：北海道大学(当時)、以下湊G)、機械学習・制約充足応用グループ(津田宏治：理化学研究所(当時)、以下津田G)、統計・マイニング応用グループ(鷲尾隆：大阪大学、以下鷲尾G)の3グループで行われた。

プロジェクト終了時点での研究成果とその意義について、①湊Gでは、ZDDを展開・発展させた様々な基盤アルゴリズムの開発を行った。また、ZDDを利用したスマートグリッド電力網への応用等の研究を行った。②津田Gでは、確率推論の融合、簡潔データ構造による大規模データの類似度検索と索引化技術、ビッグデータから新たな科学的発見をもたらす統計手法(LAMP法の開発)などの研究が行われた。③鷲尾Gでは、因果構造探索のためのデータ解析法としてのアルゴリズムDirectLiNGAMの開発などの研究を行った。応用ソフトウェアとして、任意の駅間の全経路を計算・描画するソフトウェアEkillionの開発・公開等を行った。

プロジェクト終了から現在に至る状況に関して、各種データベースによる業績(論文、特許出願状況、競争的資金の獲得状況、招待講演、受賞他)、文献調査(解説、原著論文など)、インターネットによる調査からなる基礎データ調査と、プロジェクト関係者へのインタビュー調査を行った。

論文について、成果論文が167報、発展論文が86報であった。特許について、期間中の国内出願5件、海外出願1件であった。終了後は、国内出願15件、海外出願7件であった。ベンチャー企業は3社が創立された。報道関係について、期間中3件、終了後44件であった。

プロジェクト終了後、各グループは獲得した助成金により、様々な共同研究者と共に研究を発展させた。

(1)湊Gの発展では、論理関数を表すデータ構造ZSDD等、また、ZDDの適用範囲の拡大として、ベイジアンネットワーク等への適用拡大を行った。さらに、ノード数の大きな通信ネットワークシステムの信頼性設計、配電ネットワーク設計等への応用を研究した。

(2)津田Gの発展では、LAMPの発展として、LAMPLINK、MP-LAMPを開発した。また、ベイジアン最適化等を改良したCOMBO、ChemTS等を使って、マテリアルズ・インフォマティクス(MI)、バイオ・インフォマティクス(BI)の分野への研究に発展した。

(3)鷲尾Gの発展では、大規模な非線形系についての因果関係推定の研究を行った。機械学習を使って、ナノポアセンサ等による抵抗・電流パルスを分析し、鷲尾が単一分子を特定する技術を大阪大学の谷口と一緒に開発した。この技術はベンチャー企業(アイポア株式会社)の創設に結び付いた。また、国立研究開発法人物質・材料研究機構(NIMS)と共同で、MEMSセンサへ機械学習を適用してガス種の識別を迅速に識別する技術を開発した。

(4)人材育成について、登録研究者総数16名に対して、10名がキャリアアップを実現し、また、登録されていない研究者の中で、3名がキャリアアップを実現している。国立研究開発法人理化学研究所革新知能統合研究センターチームリーダーに 3 名が就任(兼務)している。

プロジェクト成果の波及と展望について、科学技術と社会経済の観点から調べた。

科学技術の波及では、アルゴリズムの研究について、①ZDD原論文の被引用数調査等から、ゆっくりではあるが、着実に世界のZDD関連の研究に波及していることが分かった。②主なアルゴリズム論文(7報)の被引用数を調べた。類似の論文との比較指標FWCIによれば、COMBOが上位 1.8%に入る被引用数、それ以外も平均以上の被引用数であった。応用研究の展開について、①湊Gは、大規模なネットワークシステムの信頼性設計等の課題を解決する手法を提案、②津田Gは、ベイズ最適化等の手法が材料開発、薬品開発等に有効であることを示した。③鷲尾Gでは、簡単な構造のセンサと機械学習により高度の弁別機能を実現した。

研究者の育成について、多くの優秀な研究者が育ち、リーダーを輩出している。社会への波及と発展について、①大規模ネットワークについて、100程度のノードでの最適設計等を実現した。②MI、BI の研究を先導した。③ナノポアセンサによる検知、同定技術はベンチャー企業の創立に繋がった。さらに、臭気センサについて、NIMS で実用化のために企業との共同研究を推進している。④本プロジェクトの技術の利用によって創出されたベンチャー企業は、3社に上っている。⑤日本科学未来館のメディアラボでの、成果展示は20万人の来場者数を達成し、展示コンテンツは YouTube 上で公開され 295 万アクセスを超えた。海外からのコメントも数多く寄せられている。

プロジェクトの発足に至る経緯及び展開状況（まとめ図）

戦略目標、達成目標/プロジェクトの目標（ねらい）	インプット	アクティビティ/アウトプット	アウトカム（short/mid-term）		アウトカム（long-term）/インパクト
			～追跡調査時点	今後予想される展開	今後想定される波及効果
目的・目標	研究体制	研究成果のまとめ	プロジェクトから生み出された研究成果からトピックスを選び、研究成果とその展開について記入		想定される科学、社会、経済的な波及について記入

戦略目標

多様で大規模な情報から『知識』を生産・活用するための基盤技術の創出

プロジェクトの目標

1.基盤：ZDDを中心とした基盤ソフトウェアを整備し、効果的な例題とともに、公開・普及活動を行う。

2.応用：データマイニングや設計問題等において、国内企業や応用分野の研究者と連携し、訴求力のある大規模な実例題を基にして、その効果を定量的に示す。

3.普及：国内外の関連研究者と連携を密にするとともに、若い人材の育成にも注力し、技術コミュニティとして確立させる。

研究体制

研究総括
湊真一(北海道大学大学院情報科学研究科・教授)

研究体制

1.離散構造処理基盤グループ(湊真一)

2.機械学習・制約充足応用グループ(津田宏治)

3.統計・マイニング応用グループ(鷲尾隆)

研究成果のまとめ

論文

成果論文	167
発展論文	86

特許申請・登録

		期間中	終了後
出願	国内	5	15
	海外	1	7
登録	国内	5	1
	海外	6	0

国際会議の招待講演(研究者別)

	期間中	終了後	全体
湊	4	4	8
津田	0	3	3
鷲尾	1	3	4
合計	5	10	15

副次的成果物（ソフトウェア）

発表年	開発者	ソフトウェア名称
2011	湊	πDD
2011	鷲尾	DirectLiNGAM
2013	津田	LAMP
2016	湊	Graphillion
2016	津田	COMBO
2017	津田	MDTS
2017	津田	ChemTS
その他多数		

アウトリーチ活動

日本未来館へ出展→20万人の入場者
「フカシギの数え方」のYouTubeに現在295万アクセス

マネジメント

多くのセミナー・ワークショップ開催
→各年度平均25日のセミナーとワークショップ年2回開催
多くの研究者がキャリアアップした

～追跡調査時点

(1) 離散構造の基盤研究
・ZDD(についての基盤研究として、フロンティア法、πDD等多数の研究
・通信ネットワークの信頼性設計、配電ネットワーク設計への応用等

(2) 機械学習・制約充足応用研究
・統計手法LAMPの各種開発
・ベイズ法、モンテカルロ木探索のマテリアルズ・インフォマティクス(MI)、バイオ・インフォマティクス(BI)への応用研究

(3) 統計・マイニング応用研究
・線形、非線形系への因果構造探索の開発
・機械学習によるナノポアセンサー、臭気センサーの弁別機能実現

今後予想される展開

(1) 離散構造の基盤研究
・アルゴリズムを様々な分野で利用できるように体系化
・離散構造処理の「列挙問題」と「最適化問題」の統合により、確率モデリング、機械学習応用への発展
・大規模離散構造のための革新的アルゴリズム基盤技術の創出

(2) 機械学習・制約充足応用研究
・機械学習を応用したがん治療への展開
・機械学習を生命科学へ応用するための信頼性研究

(3) 統計・マイニング応用研究
・機械学習の教師無し分類器学習の研究
・ナノポアギャップ・センサの適用分野拡大(DNAシーケンサー、神経伝達物質の判別等)

科学技術的および社会・経済的な波及効果

科学技術の進歩への貢献

・ZDD、COMBO、ChemTS等のアルゴリズムの確実な世界への波及。
・MI、BI分野への先導的な成果による多大な貢献。
・新規のナノポアを使った分子センサ、MEMSセンサによる臭気センサの開発に貢献。
・多くの研究者がキャリアアップした。その結果、複数人が研究リーダーとして活躍している。

社会・経済への貢献

・社会的な重要なインフラの信頼性、経済性、安定性等の実現への基礎技術を築いた。電力のスマートグリッド研究は実証研究に展開している。
・ベンチャー企業3社を生み出した。すべてでアドバイザとして貢献が続いている。それに加えて、1社のアドバイザを務めている。1社から実際の製品が発売されている。
・日本の重要な産業に深く関係するMI、BI分野において先導的な研究を続けている。
・アウトリーチ活動によって、小学生を含めて社会的に大きな関心を集めた。

今後想定される波及効果

【基盤】
・ZDD、COMBO等のアルゴリズムの浸透・普及により、離散構造処理系基盤が拡充し、活用技術の創出がさらに進展する。その事により、大規模な情報を誰でも活用することが可能になる。

【応用】
・社会的な大規模ネットワークインフラの信頼性が向上し、経済性が改善され、故障からの復旧が早くなる等の波及効果が得られる。

・MI・BIが進展し、大規模なデータを活用した結果、以下が実現する。
　新規材料が次々に開発されて、日本の材料競争力が維持される。
　新規の薬の開発が進み、病気治療され、寿命が延びる。
　医薬品やバイオテクノロジーの開発が進む。

・超小型センサと機械学習の組み合わせが進展する事により、
　ガスの臭気検知が進み社会の安全性が向上する。
　果実の熟成検知等が普及して、安くて美味しい果実が出回る。
　ウェアラブル機器、IoTに適した高機能センサが開発される。

【普及】
・若い研究者が続々と生まれ、研究活動が活発化する。

・異業種との共同研究が増大し、成果が続々と生まれる。

第1章　プロジェクトの概要

本調査の対象である ERATO「湊離散構造処理系」プロジェクト(以後、本プロジェクトと記載)の概要を下記に示す。

1.1　研究期間

研究期間は 2009 年 10 月～2016 年 3 月。但し、最後の 1 年(2015 年 4 月～2016 年 3 月)は特別重点期間として、規模は縮小したが、すべてのグループを維持した。

1.2　プロジェクト発足に至る科学技術や社会の背景

1.2.1　科学技術の背景

20 世紀半ばのデジタルコンピュータの出現とともに、離散数学は様々な応用分野で飛躍的な発展を遂げている。本プロジェクトの中核となる離散構造は、離散数学および計算機科学の基礎をなすものであり、集合理論、記号論理、帰納的証明、グラフ理論、組合せ論、確率論などを含む数学的な構造の体系である。

離散構造の処理は、最終的には膨大な個数の場合分け処理を必要とすることが多く、計算時間や記憶容量の限界により所望の結果が得られないことがしばしばある。種々の離散構造データを計算機上にコンパクトに表現し、等価性・正当性の検証、モデルの解析、最適化などの処理を効率よく行う技法は、計算機科学、AI の様々な応用分野に共通する基盤技術として非常に重要である。

離散構造に関する最も基本的なモデルの 1 つとして論理関数があり、1986 年に R.E.Bryant が提案した Binary Decision Diagram(BDD; 二分決定グラフ)と呼ばれる技法が近年広く用いられている[1]。BDD は、元々は論理関数の処理技法であるが、同様な離散構造の一種である「組合せ集合データ」(膨大な組合せを列挙した集合データ)を扱うこともできる。このような集合データの処理に特化した BDD として、研究総括の湊が 1993 年に考案・命名したものが ZDD(Zero-suppressed BDD; ゼロサプレス型 BDD)である[2]。ZDD は、通常の BDD と異なる簡約化規則を持つ。ZDD は BDD より 100 倍コンパクトになる可能性がある。

一方、実問題を解くために様々な工学的応用に特化した技術領域が多数並列に存在している。本プロジェクトでは概念的・理論的な成果だけではなく、工学的な応用技術も加えて実用に耐えるアルゴリズムを実装することを重視した。

---

[1] https://ja.wikipedia.org/wiki/%E4%BA%8C%E5%88%86%E6%B1%BA%E5%AE%9A%E5%9B%B3(二分決定図)

[2] Minato S.-I., “Arithmetic Boolean expression manipulator using BDDs”, IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, E76-A(10), 1721-1729 (1993)

1.2.2　社会の背景

ネット社会の拡大や社会の巨大化・複雑化、コンピュータ化により、社会の中に蓄えられた巨大なデータを扱うための技術が強く求められている。離散構造数学や機械学習、データマイニング等はこれらのニーズに合致するものであり、その技術の進化・発展が求められている。さらに、通信、電力網、鉄道網等を代表とする複雑なネットワークの信頼性確保が現代社会の大きな課題である。

大規模ネットワークシステムの信頼性設計、故障解析、制約充足問題、データマイニングと知識発見、機械学習と自動分類、インフォマティクス、Web 情報解析など多種多様な課題に対して、大きな波及効果が期待される。

1.3　プロジェクトのねらい

本プロジェクトでは、BDD/ZDD 技術を新しい視点でとらえ直し、汎用的な離散構造処理系として再構築を行うことを前提に、次の 3 点を全体構想とした。

(1)社会的重要性が高い複数の応用に適用し、従来技法に比べて、10 倍～100 倍以上の性能向上を達成する。

(2)長期に渡って活用される基盤ソフトウェアを構築・提供する。さらに、それを使いこなせる人材を育成し、産業界に還元する。

(3)長期に渡って国内外から参照されるような成果ソフトウェアや学術専門書などにまとめる。

その上で具体的な目標と計画を以下のように定めた。

(1)基盤：離散構造処理の体系化と基盤ソフトウェアの整備。様々なデータ構造や目的に応じた処理系を検討する。また、ZDD を中心とした基盤ソフトウェアを整備し、効果的な例題とともに、公開・普及活動を行う。

(2)応用：実例題への適用と効果の提示。データマイニングや設計問題等において、国内企業や応用分野の研究者と連携し、訴求力のある大規模な実例題を基にして、その効果を定量的に示す。

(3)普及：人材育成とコミュニティの確立。シンポジウムや各種イベント、テレビ会議などを介して、国内外の関連研究者と連携を密にするとともに、若い人材の育成にも注力し、技術コミュニティとして確立させる。

本プロジェクトの研究領域を図1-1に示す。計算機科学の基礎部分には分野横断的な計算理論の領域がある。一方、実問題を解くために様々な工学的応用に特化した技術領域が多数並列に存在している。その中間層として、本プロジェクトが研究対象とする離散構造処理系の技術領域が存在する。この層は、上下の層との境界は明確ではないが、研究の指向性が異なる。ここでは、概念的・理論的成果だけではなく、実用に耐えるアルゴリズムを実装することを重視するとともに、技術基盤としての簡潔さや汎用性を重視する。この指向性こそが、

科学と工学の中間に位置する、いわゆる「Art」であるというのが本プロジェクトの理念であった。

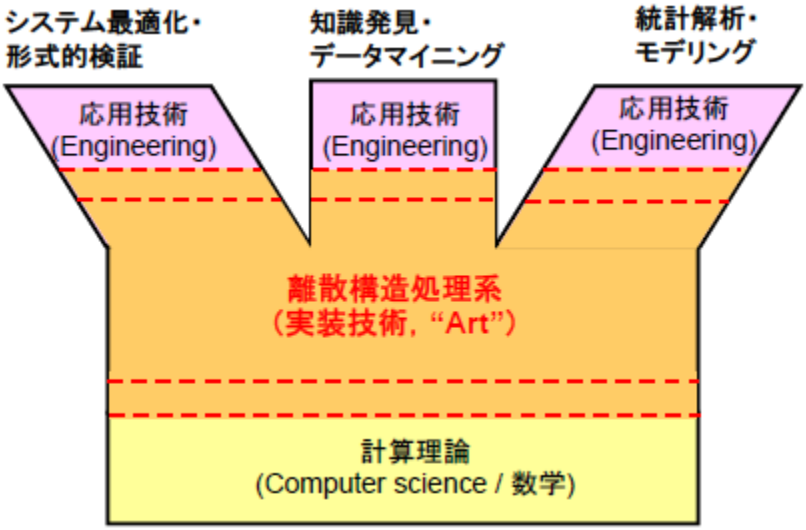


図 1-1　本プロジェクトの研究領域[3]

1.4　研究体制

研究体制、組織は年々変化していった。2015 年度は特別重点期間として本プロジェクトを継続した。

表 1-1　研究グループと人員および実施場所(2015 年 11 月 1 日時点)

グループ名	離散構造処理基盤グループ	機械学習・制約充足応用グループ	統計・マイニング応用グループ	事務所
実施場所	北海道札幌市	東京都目黒区	大阪府大阪市	
リーダー	湊 真一	津田 宏治	鷲尾 隆	技術参事 事務参事 事務員
研究員	5 名	1 名	1 名	
技術員	3 名(常勤)	0 名	1 名(常勤)	
研究補助員	4 名(時給制)	1 名(時給制)	5 名(時給制)	
計	13 名	3 名	8 名	3 名
総計	27 名			

(*)　研究推進委員:宇野毅明(国立情報学研究所情報学プリンシプル研究系)、富田悦次(電気通信大学　名誉教授)、羽室行信(関西学院大学経営戦略科)、森田裕之(大阪府立大学経営経済学科)

1.5　プロジェクト終了時点での研究成果やその意義

本プロジェクトは BDD、ZDD というグラフを基盤とした、離散構造処理の基盤研究を行うグループが展開している[4]だけでなく、機械学習、データマイニング等の AI に関係する研究グループも加えて研究を行った。

[3] 終了報告書より
[4] 湊真一、「BDD/ZDD を用いたグラフ列挙索引化技法(特集　BDD/ZDD を用いた新しい列挙索引化技法(フロンティア法)とその応用)」、 オペレーションズ・リサーチ、57(11)、 597-603 (2012)

1.5.1　離散構造処理の基盤研究

本研究では、BDD/ZDD 技術のさらなる展開のための基盤研究として、ZDD を効率的に構築するフロンティア法の開発と応用、そして例えばスーパーマーケットの商品とその中から買い物籠に入れた商品のように、項目は多いが少ない項目で 1 セットになるような事例についての BDD/ZDD の高速アクセス技術、系列集合の ZDD 処理のアルゴリズム、順列集合の処理アルゴリズム(πDD)の開発といった基盤アルゴリズムの開発を行うとともに、応用として ZDD を利用したスマートグリッド電力網への応用等の研究を行った。また、開発した ZDD 処理系 Graphillion を公開し、専門書も発刊した。

(1) フロンティア法の開発とグラフパス列挙への適用

①　フロンティア法に関する基盤技術の開発

従来は ZDD を構築する方法として、Apply 法[5]が使われてきた。この方法は最終の ZDD を求める過程で、中間の ZDD が多数生成され、中間の ZDD サイズが急増することが多い。これが計算のボトルネックになる。一方、トップダウンによるアルゴリズムとして、Knuth の Simpath 法が知られている[6]。これをさらに拡張、一般化したフロンティア法の開発を行った。Apply 法と Simpath 法をパス列挙の ZDD を生成するアルゴリズムで比較すると、Simpath 法の方が圧倒的に(例題にもよるが数十倍以上)速いことが分かった[2]。グラフをたどりながら ZDD をトップダウン、幅優先で生成する技法は、パスを列挙する Simpath 法だけでなく、さまざまな列挙問題に応用することができる。そのようなアプローチを総称して以後フロンティア法と呼ぶ。簡単な原理を図 1-2 に示した。

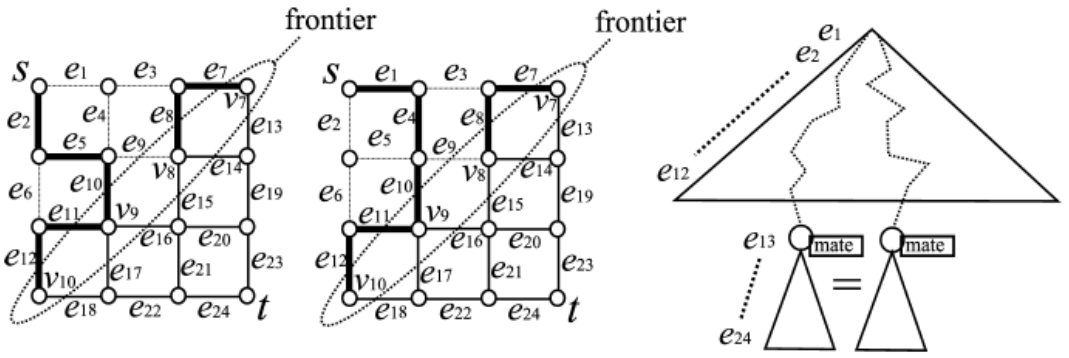


図 1-2　フロンティア法の簡単な説明[7]

[5] 湊真一、「BDD/ZDD の技法と離散構造処理系」、人工知能学会誌、27(3)、232-238 (2012)
[6] 川原純、斎藤寿樹、鈴木拡、湊真一、吉仲亮、「ZDD によるパスの列挙」、数理解析研究所講究録、1744、35-41 (2011)
[7] 終了報告書より

図 1-2 に格子グラフのパスを列挙する場合の途中状態の例を示す。辺 e_1 から e_{12} までの場合分けが終わっているとして、($e_2, e_5, e_7, e_8, e_{10}, e_{11}, e_{12}$)を選択した場合(左側の格子グラフ)と($e_1, e_4, e_7, e_8, e_{10}, e_{11}, e_{12}$)を選択した場合(右側の格子グラフ)とで途中状態を比較すると、残りの辺で節点 v_8 と v_{10} を連結し、さらに v_7 と終点 t が連結できれば良いということで両者は一致しており、以後の辺の取捨選択の制約条件は全く同じであることがわかる。これを右側の ZDD の模式図に当てはめると、e_{13} より下位のサブグラフが完全に一致するということであり、このように途中状態が等価であると判れば、e_{13} のレベルで節点を共有してしまうことができる。s-t パス列挙問題では、このような等価な途中状態が非常に多く発生するため、この共有の効果は大きい。

さらに、上記の破線で囲まれている節点、すなわち、処理済みの辺と未処理の辺の両方に接続している節点の集合を Knuth はフロンティア(frontier)と呼び、このフロンティアを始点から終点まで徐々に移動させながら、必要最小限の途中状態を記憶して処理を行う。フロンティアが小さくてすむかどうかは、与えられたグラフの形と、それに応じた辺の変数順序付けに依存する。平面グラフに近くて細長い形のグラフであるほうが都合がよい。

②フロンティア法の応用事例

(ｉ)グラフのパス列挙

グリッドグラフでのパス列挙は図 1-3 のように、辺を重複することなくスタートの s 点から、ゴールの t 点までのルートが幾つあるかという問題である。図 1-3 は n=2 の例で、全部で 12 のルートがある。n が増えると急激にルートの数が増えることが知られている(図 1-4)。ZDD を用いた数え上げ手法の最適化を行い、n=21(21×21 のグリッドグラフでの列挙)で世界記録を達成した(2012 年 9 月)。

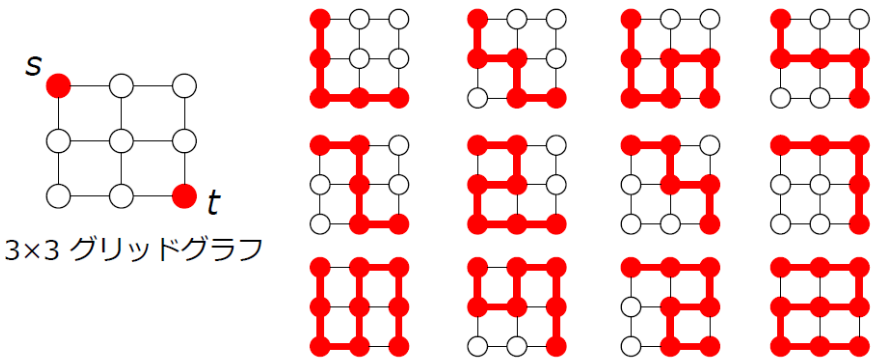


図 1-3　n=2 のグリッドグラフでパスルート[8]

[8] http://www.lab2.kuis.kyoto-u.ac.jp/jkawahara/frontier/frontier_ver0_3.pdf

その後、2013 年 2 月にノルウェーの研究者によって n=24 を達成されたが(方式は不明)、2013 年 4 月には n=25、12 月には n=26 に成功し、世界記録を更新した。この成果は日本科学未来館に展示され、好評を博した(3.2.3 参照)。図 1-4 は n=25 のグリッドのパスを計算した結果である。約 8.4×10^{149} の数を正確に計算できた。

n	#path
1	2
2	12
3	184
4	8512
5	1262816
6	575780564
7	789360053252
8	3266598486981642
9	41044208702632496804
10	156875803046475001321410 0
11	1824132915142480492414708852 36
12	64528039343270018963357185158482118
13	694506647615213616642747015489073589964 88
14	22744971467681273963182645932798986338761332 3440
15	2266745568862672746374567396713098934866324885408319028
16	687454456091499315876315631324892328245879459680994572854193 06
17	63448146112379639713102975407955244004494439868664806936463693878 55336
18	17821128408420651298933849466523252751678380657047676559314524746058266927825 32
19	1523344971704879993080742810319229690899454255323294555776029866737355060592877569255844
20	396289219982303756020729951713336250210633970573946377151523711337701068236403570670447206494039 8
21	31374751050137102720420538137382214513103312193698723653061351991346433379389385793965576992246021316463868
22	7559702866673453396615191233152226193531037320724094811673914104795179257927436312349870388833176349872711714404439792
23	554354293552374770099143184890614379306903799709643313325569586464840084073348855445663869240208757112420600854085134829339457 20
24	123717122312070647583387448626735708323730419890129435396787270804849516955159304856413945507921530371918580282125122809266003045813867910 94
25	84029748578811334710070837454368091272960542937753835498247426239370284978982152569291785770839709601216256025060273165497184021064940499783756042474 08

8.4×10^{149}

図 1-4　n × n グリッドグラフの n=25 の数え上げ [8]

(ⅱ)パズル(ナンバーリンク、スリザーリンク)の解法

その他、フロンティア法の応用事例の 1 つとして、グリッドグラフ上に与えられた制約を満たすようなパスやサイクルを見つけるパズル(リンクパズル)があり、その中でナンバーリンク、スリザーリンクなどを対象とした実験を行った[9]。図 1-5 はナンバーリンクの問題例とその解答である。このナンバーリンクの問題は、6×6 の正方形のテーブルであり、その各セルは空白であるか、1、2、3 の自然数のいずれかが 1 つ記入されていて、1、2、3 の各数字はちょうど 2 回ずつ出現しているものである。解答者はちょうど 2 回出現するすべての同じ数字同士の組を線で結ぶが、それらの線は互いに交差したり、接してはならない。

サイズを変えた複数の問題について、既存の Sugar という手法[10]による計算時間とフロンティア法を使って ZDD を生成させて計算した計算時間を比較した。結果を表 1-2 に示した。その結果、10×10 以下の少ないグラフに対しては圧倒的に ZDD が速い。20×15 のグラフについては問題によって、両者の計算時間が大きく異なる結果が得られた。例題が少ないため

[9] 吉仲亮、 岩下洋哲、 川原純、 斎藤寿樹、 鶴間浩二、 湊真一、 「種々のリンクパズルへの応用(特集 BDD/ZDD を用いた新しい列挙索引化技法(フロンティア法)とその応用)」、 オペレーションズ・リサーチ、57(11)、 616-622 (2012)

[10] http://bach.istc.kobe-u.ac.jp/sugar/puzzles/

どちらが速いと結論できない。この結果を見る限り、ZDD は既存の専用アルゴリズムに対して同等の計算時間を実現している。

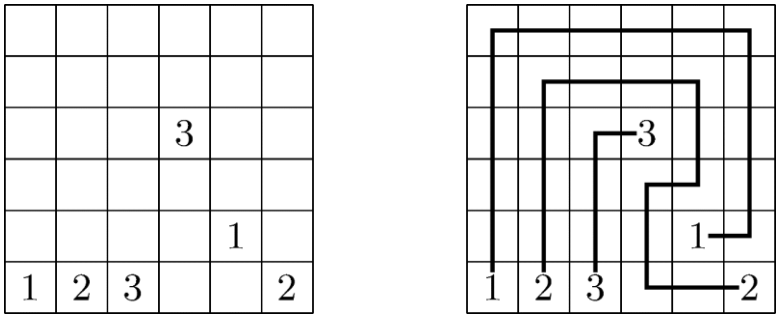


図 1-5　ナンバーリンクの問題例とその解答[11]

表 1-2　ナンバーリンクにおける既存の手法 Sugar と ZDD を用いた手法による計算時間(秒)[12]

問題番号 (size)		Sugar	提案手法
1	(8 × 8)	1.179	0.008
15	(8 × 8)	1.336	0.004
30	(10 × 10)	1.896	0.016
43	(10 × 10)	1.924	0.072
52	(10 × 10)	1.520	0.012
64	(10 × 10)	1.704	0.136
72	(10 × 10)	1.388	0.012
79	(10 × 10)	1.496	0.220
85	(20 × 15)	7213.079	862.518
99	(20 × 15)	14.097	1658.772

③超大規模 BDD/ZDD の高速アクセス技術

数千の項目中の少ない項目を選んで 1 セットになるような事例が数多くある。例えば、スーパーマーケットで買い物籠に入れた商品、自然言語中で相関がある単語のセット、文書データベース中のキーワードのリスト、関係するインターネット Web ページ集まり、人の遺伝子データ中の複数の突然変異点などである。このような状態では、有効な組み合わせ(セットと呼ぶ)に含まれない項目の数が多く、それらの項目は 0 辺になり、1 辺へ結ばれるセットに含まれる項目は非常に少ないため、図 1-6 のようなアンバランスな ZDD が得られる。この図では、0 辺は 1000 以上のステップが必要であるが、1 辺は 1～10 ステップしか必要がないことを表している。このような応用事例は多い。

[11] http://www.orsj.or.jp/archive2/or57-11/or57_11_616.pdf
[12] http://www.lab2.kuis.kyoto-u.ac.jp/jkawahara/frontier/frontier_ver0_3.pdf

一般的に、ZDD の操作では ZDD の高さに線形の計算時間が必要である。ZDD がアンバランスな場合、項目探索は幾つかのアイテムとの組み合わせしかない場合でも、k 番目の節点の 1 端接続の存在を確認するには、数千の節点を探索するステップがあるため、節点探索操作は特に非効率的である。これは、ZDD による操作が、配列とリンクリストに基づくデータ構造を持った古典的な手法よりも数百倍以上遅くなる可能性があることを意味する。この問題に対処するための実用的な方法を提案した[13]。

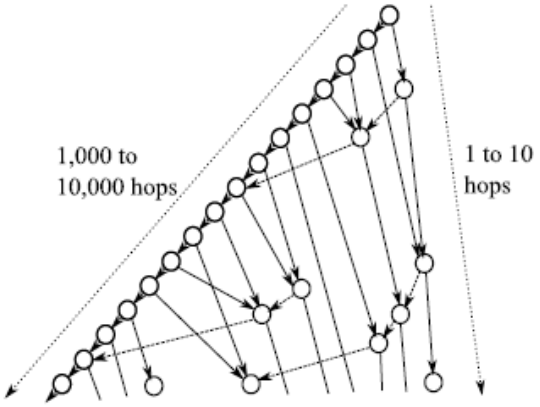


図 1-6　アンバランスな ZDD。下からのノードの高さをレベルと呼ぶ[14]

その方法とは、図 1-7 に示すように ZDD の全てのノードに Z-Skip-Links と呼ぶポインタを付加することである。ポインタの指し示す先は事前に決めた 0 辺の長さとする。ZDD の t レベルの節点を探索するとき、最初に根の Z-Skip-Links を参照し、指定された節点 s のレベルがターゲット t のレベルより高いときは、その Z-Skip-Links へジャンプしそこから検索を続ける。節点 s のレベルがターゲット t のレベルより低いときは、ジャンプを行わず 0 辺を 1 ステップ下がり、探索を続ける。

この方法は非常に単純だが、到達可能性は通常の線形探索として明確に保証されている。Z-Skip-Links の合計ジャンプ長と同じくらい検索時間を短縮できる。Z-Skip-Links には副作用がないため、ZDD 操作は、基本的なデータ構造と操作コードを大幅に変更することなく、従来の BDD/ZDD パッケージに簡単に実装できる。この方法の重要な問題は、Z-Skip-Links のスキップ長の設計である。スキップ長が長いほど節約時間が長くなるが、スキップ長が長すぎると、ターゲットを超える可能性が高くなり、スピードアップの可能性が低くなる。ここでは、時間計算量の平均的なケース分析と、特定の n に対する最適なスキップ長について説明する。特定の n に対する最適なスキップ長を計算すると最適スキップ数を αopt、平均

[13] Minato S.-I., “Z-skip-links for fast traversal of ZDDs representing large-scale sparse datasets”, LNCS 8125, 731-742 (2013)
[14] https://www-alg.ist.hokudai.ac.jp/~thomas/TCSTR/tcstr_13_63/tcstr_13_63.pdf

検索コストを Copt とすると、αopt=(√n+1)/2、Copt=√n+1/2 で表され、n=10000 とすると、αopt=50.5、Copt=100.5 となり、リニア探索より約 50 倍速くなると見積もられた。

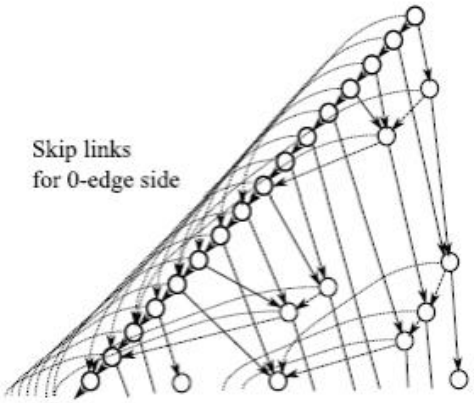


図 1-7　Z-Skip-Links を持つ ZDD[15]

(2)BDD の計算量に関する理論的成果

Bryant らが 1986 年に発表した BDD の初期論文において、示されていた BDD の計算量に関する予想は、おそらく正しいと信じられていたが、初めて反例を発見し、この問題に終止符を打った。本予想は、BDD 演算の計算時間が入出力サイズに線形であるとしたものであったが、本成果は、ある条件を満たす場合に、計算量が線形で抑えることができないこと(反例が存在すること)を証明したものである。反例は最悪ケースに近い状況で発生するため、実用的な意味で、BDD 演算の有効性が否定されるものではないが、計算機科学分野で極めて著名な論文において提示された予想を覆すものとして理論的に大きなインパクトを与えることができた。

(3)ZDD を利用した系列集合処理、順列集合処理と集合間類似アルゴリズムの研究

ZDD を利用した様々なアルゴリズムを開発した。

①系列集合の処理アルゴリズムの開発

系列データとは情報を時系列などのある順序で並べたデータである。例えば、テキスト文書、遺伝子情報、時系列イベント等が挙げられる。計算理論の分野では言語と称される。使用する文字の集合(アルファベット)Σ= {a, b, c} のとき、{aaa, aba, bbc, bc} は系列集合の一例である。

通常の ZDD は組み合わせ集合を表すので、文字の出現順序の違い(例えば {ab, ba})や、文字の重複(例えば {aa, aba, ab, aabb})は区別できない。これを改善する方法として、文字

[15] https://www-alg.ist.hokudai.ac.jp/~thomas/TCSTR/tcstr_13_63/tcstr_13_63.pdf

の種類と出現位置の 2 つを組みにして符号化したものを ZDD のアイテム変数に割り当てれば、系列集合を ZDD で表現できる[16]。例えば、系列集合｛aaa, aba, bbc, bc｝は｛$a_1a_2a_3$, $a_1b_2a_3$, $b_1b_2c_3$, b_1c_2｝のように符号化でき、その ZDD は図 1-8 で示される。

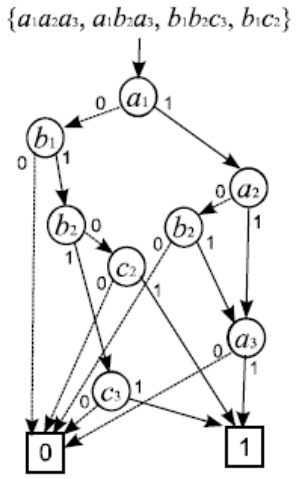


図 1-8　アイテム符号化により ZDD で系列集合を表した例[17]

その後 Loekito[18]らによって開発された系列を扱う処理系 SequenceBDD(SeqBDD)が開発された。これは ZDD とほとんど同じデータ構造であるが、ただ 1 点だけ、変数順序付けの規則が異なる。通常の ZDD では、すべての親節点と子節点の間で、あらかじめ固定された変数順序が守られているのに対し、SeqBDD では、0 枝側のみ変数順序に従うが、1 枝側は変数順序の制約を外しており、記号の重複や逆順が許される[19](図 1-9)。この SeqBDD に従うと図 1-8 の文字列のセットは図 1-9 のように表される。

[16] Kurai R., Minato S.-I., Zeugmann T., “N-gram analysis based on zero-suppressed BDDs”, JSAI 2006, 289-300 (2006)
[17] https://pdfs.semanticscholar.org/aebd/7b9392a2533ba30609bd410def851e0eb5e6.pdf
[18] Loekito E., Bailey J., Pei J., “A binary decision diagram based approach for mining frequent subsequences,” Knowledge and Information Systems, 24, 235-268 (2010)
[19] 湊真一、「BDD/ZDD を基盤とする離散構造と演算処理系の最近の展開」、IEICE Fundamentaals Review 4(3)、224-230 (2010)

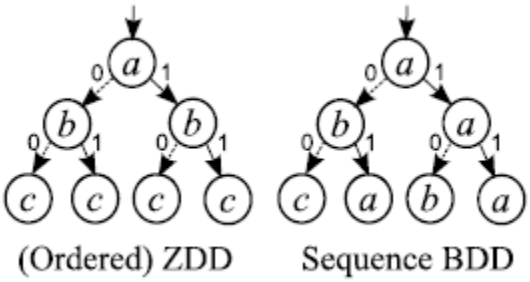


図 1-9　ZDD、SeqBDD の変数順序規則の違い[20]

さらに、この SeqBDD を基に、文字列解析において良く利用されている SuffixTree[21]の技術を融合した SuffixDD を開発し、系列集合に対する BDD/ZDD の拡張を行った。

SuffixDD はテキスト T の、全ての部分文字列のセットを表現するよう作成された SeqBDD である。図 1-10 に、テキスト aabac の SuffixDD の例を示す。テキスト aabac の全ての部分文字列は a, b, c, aa, ab, ac, ba, aab, aba, bac, aaba, abac, aabac である。この図で、SuffixDD のルールで部分文字列を得ることができる。すなわち、先頭からノードを辿る時、そのノードの 0 枝を辿るときにはそのノードの文字は使わず、1 枝を辿るときにはその文字を使うという規則である。例えば、先頭の a では実線の 1 枝と点線の 0 枝に別れている。先頭から 1 枝の実線を全て辿ると aabac を得る。先頭の a で 0 枝の点線へ分かれ、b に来ると 1 枝の実線は 3 番目の a に繋がっており、文字 b は有効になり、次の a が 0 枝点線であれば、結果として b の文字列を得る。a が 1 枝実線であれば、ba が有効になり、次の c の枝が 0 枝点線であれば、ba 文字列が得られる。このような操作を全てのノードに対して行うと全ての文字列が得られるグラフになっていることが分かる。

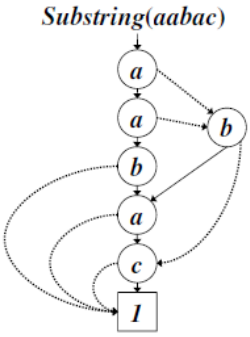


図 1-10　テキスト aabac に対する SuffixDD の例[22]

[20] https://pdfs.semanticscholar.org/aebd/7b9392a2533ba30609bd410def851e0eb5e6.pdf
[21] SuffixTree とは、与えられた文字列の接尾部(Suffix)(文字列のある位置から末尾までの文字列)を木構造で表すデータ構造であり、多くの文字列操作の高速な実装に利用されている。https://ja.wikipedia.org/wiki/%E6%8E%A5%E5%B0%BE%E8%BE%9E%E6%9C%A8
[22] https://www-alg.ist.hokudai.ac.jp/~thomas/TCSTR/tcstr_10_42/tcstr_10_42.pdf

②順列集合の処理アルゴリズム(πDD)の開発[1]

有限個の対象から幾つかを取り出しそれを順に並べたものを順列と呼ぶ。すなわち、取り出したものには1,2,…,nのような順番を表す番号が付与できる。順列集合を処理するとは順番の並べ替えを表現する離散構造を構築し、解析することである。この応用にはルービックキューブ、15パズル、トランプ、あみだくじ等がある。

順列を次のように定式化する。順列$\pi=(a_1, a_2, a_3, \cdots, a_n)$をk番目のデータをakと定義する。例えば、$\pi=(4,2,1,3)$と表したとき、1番目の値は$1\pi=4$「$\pi$の順列の1番目の値4」という定義になり、$2\pi=2$、$3\pi=1$、$4\pi=3$と表現する。このように定義すると$\pi_1=(3,1,2)$、$\pi_2=(3,2,1)$とした時、$\pi_1\pi_2$はどのようになるかを考えてみると、$1\pi_1\pi_2=3\pi_2=1$、$2\pi_1\pi_2=1\pi_2=3$、$3\pi_1\pi_2=2\pi_2=2$となるので$\pi_1\pi_2=(1,3,2)$となる。一般に$\pi_1\pi_2\neq\pi_2\pi_1$である。更に、基礎となる順列の2つの項目の交換に着目した。$\tau(x,y)$は項目xと項目yの交換を表す。πDDの鍵となるアイディアは、順列πの最大個数を$\dim(\pi)$とすると如何なる順列πも交換を最大$(\dim(\pi)-1)$回繰り返せば交換の組み合わせに分解できるということによる。例えば、(3,5,2,1,4)を実現するには基本順列$\pi_e=(1,2,3,4,5)$に対して、$\tau(2,1)\tau(3,2)\tau(4,1)\tau(5,4)$の順で交換を実行すれば良い。逆に(3,5,2,1,4)から$\pi_e=(1,2,3,4,5)$を得るには、交換の順番を逆にして、$\tau(5,4)\cdots\tau(2,1)$の順で交換すれば良い。

これらを前提にして数字の交換$\tau(x,y)$をZDDのノードにする構造を使い、1枝へ行く場合は数字の交換を行い、0枝は交換しないとすると、ZDDのグラフにより順列集合を表すことができる。πDDの例を図1-11に示した。開始は$\pi_e=(1234)$の数字列から開始する。各ノードで1枝が数字の交換有り、0枝が交換無しでパスを1終端から辿ると最終的にトップに辿り着く組み合わせによって{2143,2431,4321}が得られる。πDDは、多数の順列を要素として含む順列集合をコンパクトかつ一意に表現し、和集合、積集合、差集合、直積に関する新しい演算処理が実装されており演算処理を効率よく行える。ルービックキューブの解を実行可能な時間とメモリー量で求めることができた。あみだくじへの応用では、世界ではじめてオーダー13の数え上げ(本数13本(1,2,3,…,13)のあみだくじで逆順(13,12,11,…,1)を生成する方法の列挙)に成功した。フロアプラン(配置問題)への応用[23]や暗号システムの設計や解析にも活用できるのではないかと期待されている[24]。

[23] 終了報告書より

[24] https://cy2sec.comm.eng.osaka-u.ac.jp/miyaji-lab/CFG/2011/minato.html

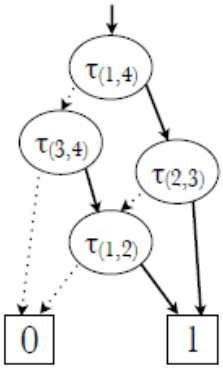


図 1-11　πDD の例。{2143, 2431, 4321} = { τ (1, 2) ・ τ (3, 4)、 τ (1, 2) ・ τ (1, 4)、 τ (2, 3) ・ τ (1, 4) } [1]

(4) ZDD を利用したスマートグリッド電力網への応用

早稲田大学と共同で、ZDD を用いたスマートグリッド上の効率的な計算手順の開発を行った[2],[25](図 1-12)。具体的には、ZDD を利用して、変電所設備 72 ヶ所、スイッチ数 468 個からなる典型的なスマートグリッド構成において、天文学的に膨大なスイッチ構成(ON/OFF)の組み合わせを調べ、与えられた電力品質条件を満たしつつ、かつ送電損失を最小化する最適構成を得ることに成功した。配電網の組み合わせ的な総数は 10 の 140 乗に及ぶが、制約を満たす構成は約 10 の 63 乗(割合ではわずか 10 の 78 乗分の 1)であることを示し、これらの構成数を正確に算出することに成功した。構成数の算出は市販 PC のスペックでも数十分で計算が可能で、ZDD の節点数(メモリ量)は約 110 万個(779MB)であった。

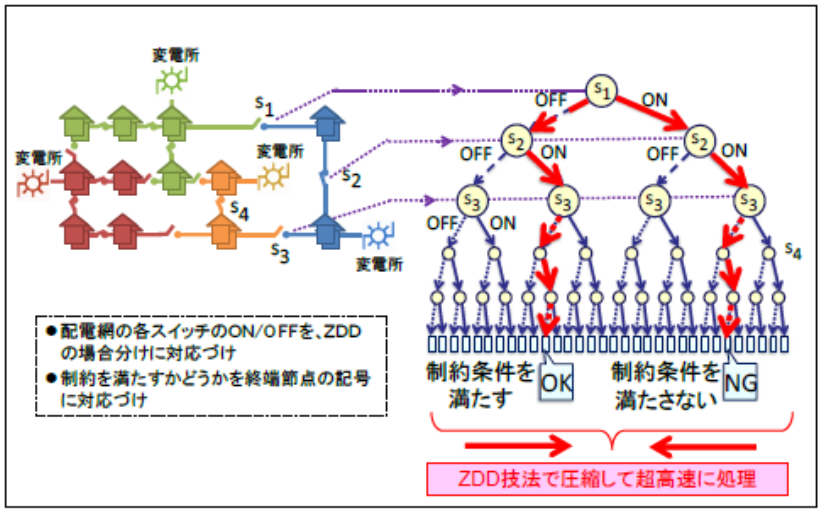


図 1-12　ZDD によるスマートグリッド配電網のモデル化[26]

[25] 井上武、高野圭司、渡辺喬之、川原純、吉仲亮、岸本章宏、津田宏治、湊真一、林泰弘、「フロンティア法による電力網構成制御」、 オペレーションズ・リサーチ、 57(11)、 610-615　(2012)
[26] 終了報告書より

(5)ZDD 処理系 Graphillion の開発と公開

大規模グラフデータに対して、検索や最適化、列挙を行うパッケージ Graphillion を開発、公開した。Graphillion は、C または C++で開発されたコアな ZDD エンジンに対して、Python によるインタフェースを兼ね備えたもので、Open source MIT license により 2013 年 6 月より公開されている[27]。さらに、大阪サブグループが開発し公開しているツール群「Nysol」[28]でも本成果を公開している。

Graphillion で用いているアルゴリズムの解説と利用法をまとめた専門書を ERATO メンバーで分担執筆し、2015 年 4 月に出版した[29]。また、電子情報通信学会誌にも本成果に関する特集記事を執筆し、研究者コミュニティへのアウトリーチ活動を行った。

1.5.2 機械学習・制約充足応用研究

本研究では大規模データの高速探索手法の開発、論理プログラミングと確率推論の融合、簡潔データ構造による大規模データの類似度検索と索引化技術、ビッグデータから新たな科学的発見をもたらす統計手法(LAMP 法の開発)、モンテカルロ木探索によるコンピュータ囲碁・将棋への応用などの研究が行われた。

(1)大規模データの高速探索手法の開発

大規模データの高速探索手法として類似しているセットを検索する手法の開発とタンパク質データに対して応用し、全ペア類似度検索を実施した。

①類似しているセット検索のための複合ソート法の開発

複合ソート法とは、長さ 1 文字列が n 個与えられた時、ハミング距離[30]d 以内のペアを全列挙する方法である(図 1-13)。

---

[27] http://graphillion.org

[28] http://www.nysol.jp/

[29] ERATO 湊離散構造処理系プロジェクト著、湊真一編、「超高速グラフ列挙アルゴリズム:〈フカシギの数え方〉が拓く、組合せ問題への新アプローチ」森北出版(2015 年 4 月 8 日)

[30] ハミング距離：等しい文字数を持つ 2 つの文字列の中で、対応する位置にある異なった文字の個数である。

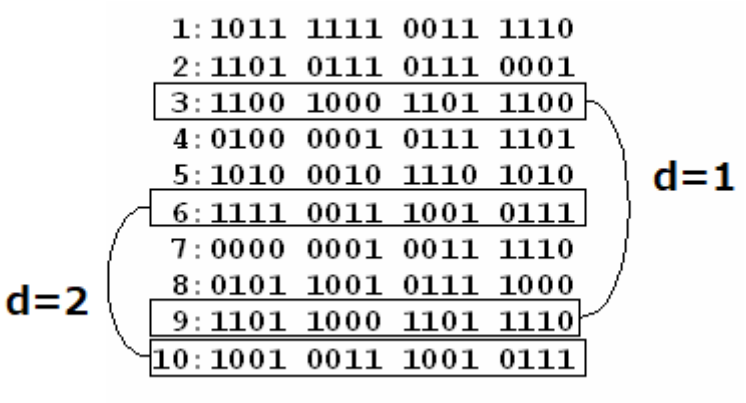


図 1-13　ハミング距離の例[31]

具体的には d 個の文字を全通り選んでマスクする。図 1-14 は d=2 のマスクの例である。マスクした列以外の文字列が等しければ、類似していると判断する。この検索を高速に行う手法として、ブロックマスクによる方法を開発した。これは文字列を k 個のブロックに分割し、d 個のブロックを全通りマスクする。このような工夫によって、ソートの回数を劇的に減らすことができる(図 1-15)。

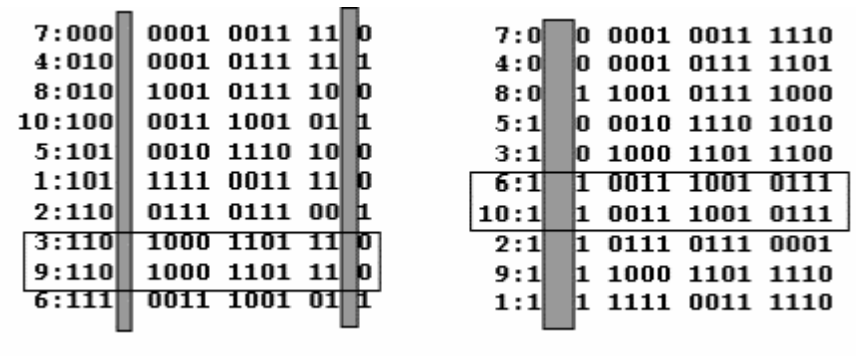


図 1-14　d=2 以下のハミング距離の探索結果[31]

[31] http://www-erato.ist.hokudai.ac.jp/html/php/symposium3_docs/MSM_tsuda.pdf

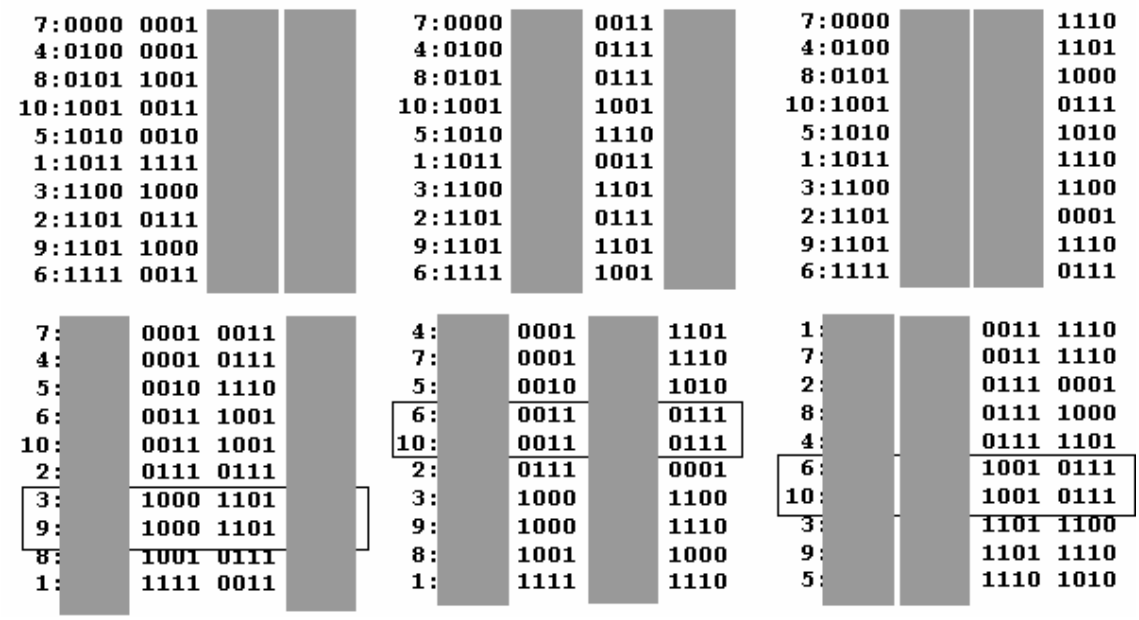


図 1-15　文字列を k=4 個のブロックに分割し、d=2 個のブロックを全通りマスクした例[31]

これは文字列の例であるが、画像、音声など実数値を持つデータについては、実数値を持つ複数の数値、すなわち多次元の特徴ベクトルとして表現し、ベクトルを二値(0 と 1)の文字列に変換することを考える。この場合、2 つの似たベクトルは 2 つの似た文字列へ変換することが重要である。ベクトル空間のデータが似ているというのは 2 つの考え方で表現されることが多い。1 つは距離が近い、もう 1 つはベクトルの角度が近いの 2 つであり、距離が近いのはユークリッド距離、角度が近いのはコサイン類似度と呼ばれている[32]。どちらが適しているかは目的によって異なる。この開発においてはコサイン類似度によって、ベクトルを 0 と 1 の文字列へと変換した。その結果を前述の文字列のハミング距離による検索を適用した複合ソートを提案した(SketchSort)[3]。

② 複合ソート法による全ペア類似度検索の応用

蛋白質の配位子(リガンド)と結合可能部位(ポケット)の幾何学的な 3 次元構造が分かるとその類似性から機能推定や薬物設計分野におけるスクリーニングに活用できる。図 1-16 は幾何学的特徴の抽出を行って、コサイン類似度を求める手続きの模式図である。このような作業により、蛋白質構造データバンク(PDB)[4]に登録されている全蛋白質に含まれる 100 万個以上のポケットに対する全ペア類似度検索を行い、発見された類似ペアをデータベース PoSSUM として、Web 上に公開した[33,5]。

[32] ユークリッド距離とコサイン類似度：https://enjoyworks.jp/tech-blog/2242
[33] https://integbio.jp/dbcatalog/record/nbdc01144

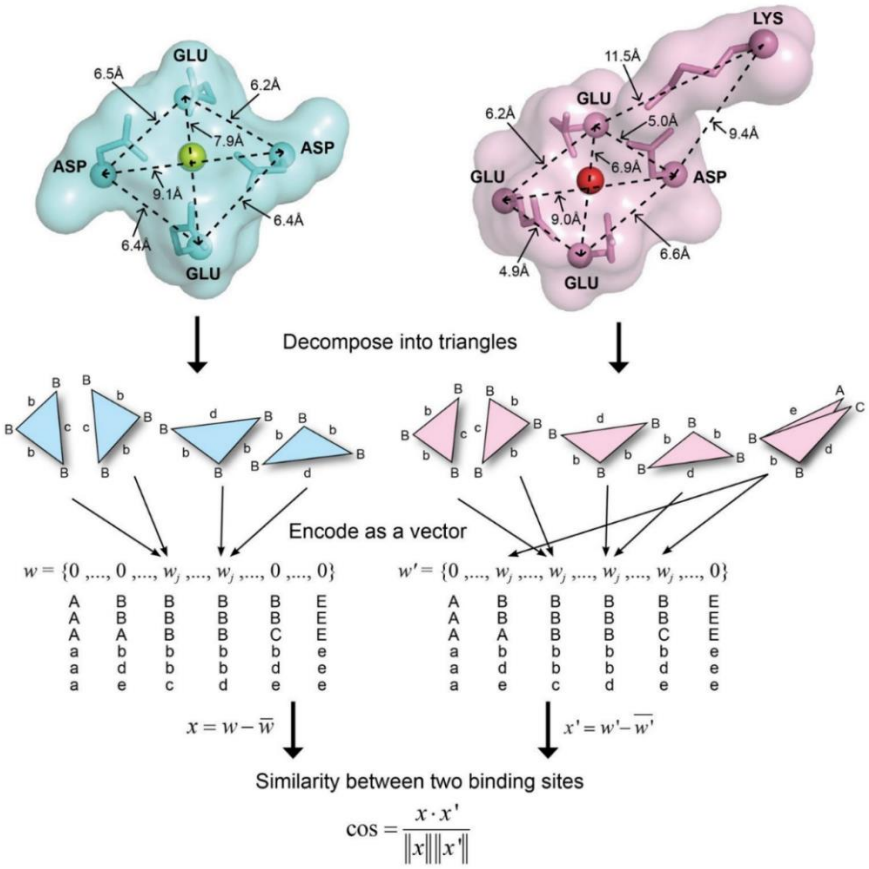


図 1-16　蛋白質からの幾何学的特徴の抽出[34]

(2) 論理プログラミングと確率推論の融合

論理的な知識表現と確率推論を融合した枠組みを検討した[6]。特に、ベイジアンネットワーク(BN)において組み合わせ的に増大する確率計算を、BDD や ZDD を用いて効率的に処理する方法を考案し、それらの適用実験を行った。ベイジアンネットワークの例を図 1-17 に示した。各 BN ノードは上流側の BN ノードの確率変数に依存する条件付き確率テーブル(CPT)を持つ。

[34] http://www-erato.ist.hokudai.ac.jp/html/php/symposium3_docs/MSM_tsuda.pdf

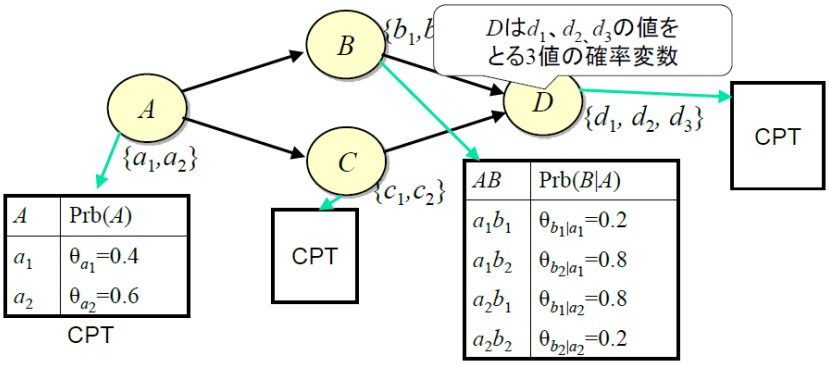


図 1-17　ベイジアンネットワークの例[35]

さらにすべての BN の確率分布を列挙し、展開して 1 つの式にする(Multi-Linear Function:MLF 式)。その際、1 つの確率変数に対して 2 種類の論理変数を用いる。

－λ変数：BN の観測値と矛盾すれば 0、しなければ 1

－θ変数：確率分布の数値(シンボルとして扱う)

図 1-17 の例について、全てを上記によって展開した式(MLF 式)は以下の通りである。

$$
\begin{aligned}
& \lambda_{a_1}\lambda_{b_1}\lambda_{c_1}\lambda_{d_1}\theta_{a_1}\theta_{b_1|a_1}\theta_{c_1|a_1}\theta_{d_1|b_1c_1} \\
+ \quad & \lambda_{a_1}\lambda_{b_1}\lambda_{c_1}\lambda_{d_2}\theta_{a_1}\theta_{b_1|a_1}\theta_{c_1|a_1}\theta_{d_2|b_1c_1} \\
+ \quad & \lambda_{a_1}\lambda_{b_1}\lambda_{c_1}\lambda_{d_3}\theta_{a_1}\theta_{b_1|a_1}\theta_{c_1|a_1}\theta_{d_3|b_1c_1} \\
+ \quad & \lambda_{a_1}\lambda_{b_1}\lambda_{c_2}\lambda_{d_1}\theta_{a_1}\theta_{b_1|a_1}\theta_{c_2|a_1}\theta_{d_3|b_1c_2} \\
+ \quad & \ldots \\
+ \quad & \lambda_{a_2}\lambda_{b_2}\lambda_{c_2}\lambda_{d_3}\theta_{a_2}\theta_{b_2|a_2}\theta_{c_2|a_2}\theta_{d_3|b_2c_2}
\end{aligned}
$$

BN 表現を ZDD に表した例を図 1-18 に示した。この ZDD の 1 終端に繋がるパスによって、図 1-17 の B に関しての MLF 式をすべて表現できる。

$$
\begin{aligned}
\mathrm{MLF}(B) = {} & \lambda_{a1}\lambda_{b1}\theta_{a(0.4)}\theta_{b(0.2)} + \lambda_{a1}\lambda_{b2}\theta_{a(0.4)}\theta_{b(0.8)} \\
& + \lambda_{a2}\lambda_{b1}\theta_{a(0.6)}\theta_{b(0.8)} + \lambda_{a2}\lambda_{b2}\theta_{a(0.6)}\theta_{b(0.2)}
\end{aligned}
$$

例えば、第 1 項は図 1-18 の右側の実線パスによって表現できている。ZDD で表すことにより、MLF 式が指数関数的に大きくなっても類似した部分を共有することで、場合によっては何十倍もの縮約効果を得ることができる。

[35] https://www.ieice.org/publications/conference-FIT-DVDs/FIT2010/pdf/F/F_024.pdf

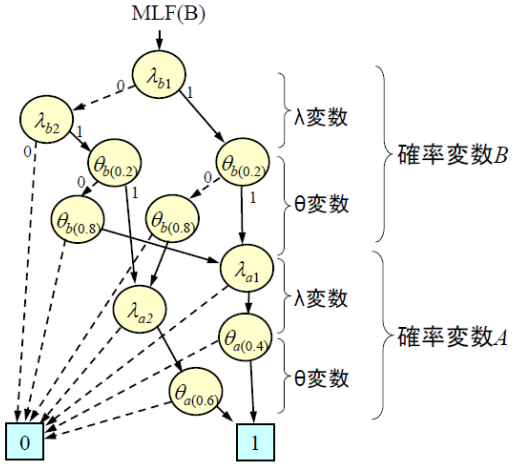


図 1-18　ZDD で表した BN 表現[36]

確率推論においては、従来、組み合わせ的に増大する確率計算が大きなボトルネックとなっていたが、本プロジェクトにおける BDD、ZDD 技術と連携することにより、実用的なスケールでの確率推論が可能となると考えられる。

(3) 簡潔データ構造による大規模データの類似度検索と索引化技術

簡潔データ構造は、ビット配列や木構造などのデータに対して、基本的なデータ検索を効率的に処理することが可能なコンパクトなデータ構造で、rank や select といった索引辞書を定義することで実現されていることが多い。本プロジェクトでは、簡潔データ構造を用いた大規模データの類似度検索のほか、ビット配列や木構造だけでなく、ZDD 構造における簡潔データ構造として、DenseZDD を開発し[4]、データ検索における高速処理を実現するなど、現実的なデータ利用局面において重要な技術の開発を行った。DenseZDD は静的 ZDD の索引データを圧縮した形で保持し、さらに、そのデータメンバーを高速に操作できるようにした。この成果は JST　CREST「ビッグデータ統合利活用のための次世代基盤技術の創出・体系化」(研究総括：喜連川優)「ビッグデータ時代に向けた革新的アルゴリズム基盤」(研究代表者：兵庫県立大学教授　加藤直樹、研究期間：2014 年 10 月 1 日～2121 年 3 月 31 日(予定))にも繋がった。

また、具体的には、創薬関係の Drug-Target ネットワークを索引化するための簡潔データ構造 Interval-Splitting Tree を開発した[7]。小さな化合物とタンパク質間の機能的相互作用を分析することは、ゲノム創薬に不可欠である。様々な化合物-タンパク質相互作用に関する豊富な情報が最近の分子データベースを利用するため、そのようなデータベースを最大限に活用するための強いニーズがあり、大規模で新しい機能的な化合物-タンパク質ペアを見つける方法が求められている。化合物-タンパク質ペアのデータベースで類似性検索

[36] https://www.ieice.org/publications/conference-FIT-DVDs/FIT2010/pdf/F/F_024.pdf

を効率的に実行する簡潔な間隔分割探索木アルゴリズム(Succinct Interval-Splitting Tree：SITA)を開発した。SITA は、間隔分割ツリーと呼ばれるデータ構造を開発することで時間とスペースの両方の効率を実現した。これにより、検索空間の不要な部分を効率的に枝刈りでき、簡潔なデータ構造で探索木をコンパクトに表現できる。2 億を超える化合物-タンパク質ペアや、基質-製品ペアを含む大規模なデータベースを使って、類似した化合物-タンパク質ペアや基質-製品ペアを取得する機能について、SITA を実験的にテストし、SITA が他の可能なアプローチよりも優れていることを示した。

(4)ビッグデータから新たな科学的発見をもたらす統計手法(LAMP 法の開発)

自然科学では新しい現象を見つけたとき、系のゆらぎや観測のあいまいさを考慮した上で、その結果の信頼性を担保する必要がある。科学データの解析において、この信頼性担保には、統計検定が欠かせない。統計検定では、誤発見の確率を示す検定値(P 値)が計算され、あるしきい値(一般には、0.05)以下の場合にのみ、信頼しうる科学的発見として認められ、論文に記すことができる。

観測できる対象(例：DNA の変異)が増えると、誤発見の確率も高くなる。誤発見を避けるには、対象数が増えれば増えるほど、発見の基準を厳しくしなくてはならない。一般的な多重検定法では、P 値に大きな補正係数を掛けて(補正 P 値)、それでも 0.05 以下の場合のみ発見とみなす。最もシンプルでよく用いられるボンフェローニ法では、n 個の対象があれば、P 値に n を掛けて補正し、それでも 0.05 以内であれば、発見として認める。その結果、観測対象が増えたのに、科学的発見が減るという奇妙な現象「ビッグデータのパラドックス」が起きる場合がある。

特に、複合的な組み合わせ因子を考えると、対象数 n が爆発的に大きくなるため、ほぼ発見は不可能となってしまう。このため、細胞の IPS 化を引き起こす 4 つの転写因子などに見られる組み合わせ因子を、データから見つけ出すことは困難であった。

本研究では、従来よりも格段に正確な補正 P 値を計算できるアルゴリズム LAMP(Limitless-Arity Multiple testing Procedure、無限次数多重検定法)を開発した。LAMP では、出現頻度の低い組み合わせは誤発見率を変化させないという数理的性質に注目し、超高速アルゴリズムを用いて無為な出現頻度の低い組み合わせを特定し取り除くことによって、補正係数を大幅に削減している。また LAMP では通常のボンフェローニ法と比べて、統計的な検定の精度を保ったままで、補正係数を十分に低くすることができる。この手法を用いて、ヒトの乳がん細胞株の遺伝子発現データを再解析したところ、これまで見過ごされてきた、最大 8 個の転写因子の組み合わせが乳がん細胞の増殖に関与していることを発見できた(図 1-19)。

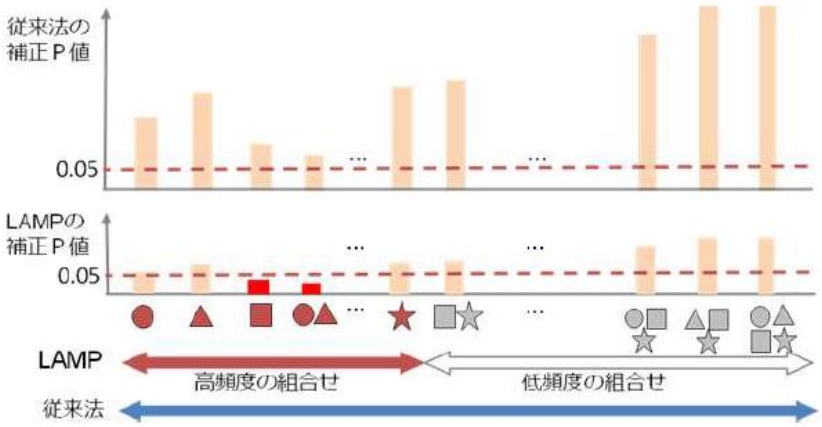


図 1-19　LAMP による組み合わせ因子発見[37]

(5) モンテカルロ木探索によるコンピュータ囲碁・将棋への応用

モンテカルロ木探索[38]を応用したコンピュータ囲碁やコンピュータ将棋プログラムの開発に取り組んだ。囲碁、将棋は人間の思考過程をモデル化する上でも有益な題材であるだけでなく、近年は、コンピュータ将棋プログラムとプロ棋士の対戦イベントなどもあり、注目度の高い応用分野となっている。

ゲームでのモンテカルロ木検索は、最後までプレイしたシミュレーション結果に基づいて構築する。ゲームの勝敗の結果に基づいてノードの値を更新して、最終的に勝率が高いことが見込まれる手を選択する。

当プロジェクトでは、膨大な組み合わせから生成されるゲーム木の効率的な表現ならびに探索手法を研究することにより、他分野への応用が可能な基盤技術の開発に取り組んだ。

1.5.3　統計・マイニング応用研究

グラフ系列からの知識を発見する手法の開発、因果構造探索のためのデータ解析法としてのアルゴリズム DirectLiNGAM の開発、機械学習における劣モジュラの最大化の理論解析、超高次元データに関する一般的な非圧縮直接推定の研究を行った。また、ビジネスデータを用いたマイニング検証実験や任意の駅間の全経路を計算、描画するソフトウェア Ekillion の開発・公開、簡易スクリプト言語の Ruby 上で利用できる ZDD ライブラリの開発、大規模表構造テキストデータを高速に処理するコマンド群 MCMD(M コマンド)の開発・公開等を行った。

[37] 終了報告書より
[38] モンテカルロ木探索：最も良い手を選択するために使われ、ランダムサンプリングの結果に基づいて探索木を構築する。

(1) グラフ系列からの知識発見

現在、情報技術の発展により、膨大なデータを蓄えることが可能になった。しかし、蓄積された大規模なデータから人手で有用な情報を発見することは非常に困難である。このような背景から、大規模なデータから有用なデータを抽出するデータマイニングの研究が盛んに行われている。例えば、人間をグラフの頂点、人間関係をグラフの辺とすると、ある時点の人間関係はグラフで表現することができる。常に変化する人間関係は時間とともに変化するグラフによって表される。人間関係に限らず、遺伝子が頂点、相互関係が辺である遺伝子ネットワークは進化の過程で新規遺伝子の獲得・既存遺伝子の欠落・突然変異等、変化するグラフによって表現可能である。

本研究では、グラフで表されたデータ集合から頻出する部分グラフを発見する手法の研究を行った。2008 年に鷲尾らが開発した Graph Transformation Sequnce Mining(GTRACE)の高速化の手法を提案し、その後 GTRACE の改良を続け高速化を行った。

図 1-20(a)はグラフ系列の集合の例である。この例では g(1)から g(4)へと変化する 4 つの状態、各状態には 1 から 5 の 5 つの頂点 ID からなるグラフ系列を表している。図 1-20(b)は(a)のグラフ系列の部分系列で、頻出する系列になっている。複数のグラフ系列からこのような頻出する部分系列を見出し、列挙するアルゴリズムを開発するのが本研究の目的である。

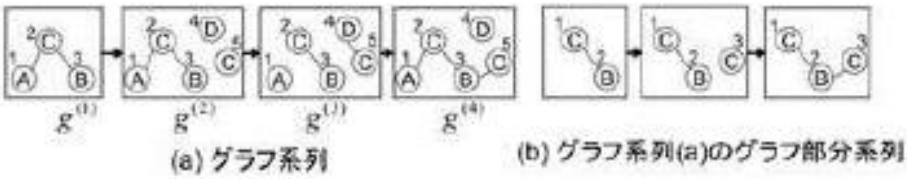


図 1-20　観測グラフ系列とそのグラフ部分系列の例[39]

開発した GTRACE は 2 つのステップから成る。初めに、複数のグラフ系列について、各グラフ系列を表 1-3 による返還手続きにより、vi、vd、vr、ei、ed、er の並びに変換し、変換系列にする(図 1-21)。変換系列は 2 つのグラフ間の差異のみが表されているので、グラフ表現を直接用いた系列よりも簡潔である。次のステップで、グラフ系列ごとに得られた変換系列群から、指定回数以上頻出する変換部分系列を求めれば、頻出変換部分系列(FTS)を列挙することができる。

[39] https://www.jstage.jst.go.jp/article/pjsai/JSAI2011/0/JSAI2011_1D12in/_pdf/-char/ja

表 1-3　グラフ系列データのための変換規則[39]

頂点追加 $vi^{(j,k)}_{[u,l]}$	ラベルが l，頂点 ID が u である頂点を $g^{(j,k)}$ へ追加し，$g^{(j,k+1)}$ へ変換
頂点削除 $vd^{(j,k)}_{[u,\bullet]}$	頂点 ID が u である頂点を $g^{(j,k)}$ から削除し $g^{(j,k+1)}$ へ変換
頂点ラベル変更 $vr^{(j,k)}_{[u,l]}$	頂点 ID が u である頂点のラベルを l に変更し，$g^{(j,k)}$ を $g^{(j,k+1)}$ へ変換
辺追加 $ei^{(j,k)}_{[(u_1,u_2),l]}$	頂点 ID が u_1 と u_2 である頂点間にラベル l の辺を追加し，$g^{(j,k)}$ を $g^{(j,k+1)}$ へ変換
辺削除 $ed^{(j,k)}_{[(u_1,u_2),\bullet]}$	頂点 ID が u_1 と u_2 である頂点間から辺を削除し，$g^{(j,k)}$ を $g^{(j,k+1)}$ へ変換
辺ラベル変更 $er^{(j,k)}_{[(u_1,u_2),l]}$	頂点 ID が u_1 と u_2 である頂点間の辺のラベルを l へ変更し，$g^{(j,k)}$ を $g^{(j,k+1)}$ へ変換

頂点削除と辺削除の変換規則，vd と ed は頂点 ID の指定のみで変換可能なので，引数 l はダミー変数であり，'•' で表す．$u_1 \leq u_2$ を必ず満たす．

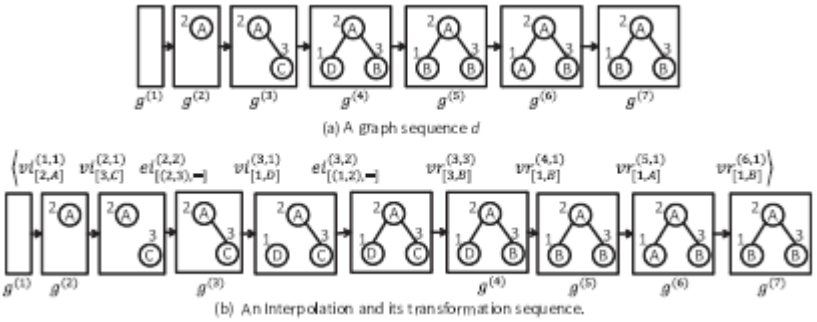


図 1-21　グラフ系列とその変換結果[39]

GTRACE ではグラフ系列中の連続する 2 つのグラフで、その構造は大きく変化しないことを仮定している。しかし、グラフデータを収集する際に、時間分解能が低い場合、収集されたグラフ系列の連続する 2 つのグラフ間で、グラフの大部分が変化する可能性がある。このようなデータに関する解析に GTRACE は適さない。このような課題を克服するために、FRISSMiner(Frequent, Relevant and Induced Subgraph Subsequence Miner)を開発した[8]。図 1-22 はそのアルゴリズムの概要を示す。

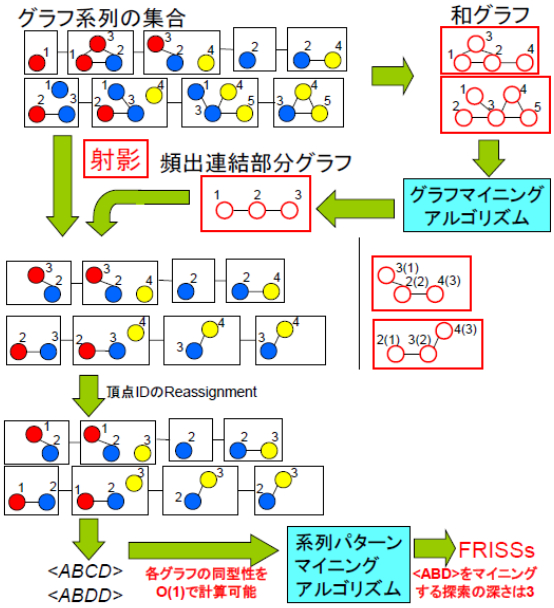


図 1-22　FRISSMiner のアルゴリズムの概要[40]

(2) 因果構造探索のためのデータ解析法

因果関係とは、単なる統計的な相関関係とは異なり、「変数 x の値を大きくした時に、対応する変数 y の値も大きくなる傾向にある」という必要がある。2 つの変数の間に相関関係が見られたからと言って、そこに因果関係があるとは限らないということである。

相関関係を前提に、因果関係を求める研究が様々に行われてきた[41]。統計的因果推論の分野では、(線形)構造方程式モデル(Structural Equation Model：SEM)と呼ばれる統計モデルが主に用いられる。構造方程式モデルは、ある変数が別の変数に影響を与えその変数がさらに別の変数に影響を与えるという関係を、それら変数から構成される方程式の組として表し、データがその方程式に従って生成されていると仮定するモデルである。そしてそのモデルの下で、データから方程式がどのようなものであるかを推定する。本プロジェクトでは LiNGAM モデル(Linear Non-Gaussian Acyclic Model)[42]についての改良研究を行った。

LiNGAM モデルは構造方程式モデルの特別な場合である線形非巡回モデルのもとで、データの生成過程に非正規分布を仮定するモデルである(図 1-23)。係数行列 B を独立成分分析[43]の手法を使って推定する。

---

[40] http://www-erato.ist.hokudai.ac.jp/html/php/seminar5_docs/inokuchi.pdf
[41] 小川雄太郎著、「Python による因果分析－因果推論・因果探索の実践入門」、マイナビ出版(2020 年 6 月 30 日)
[42] Shimizu, S., Hoyer, P. O., Hyvaerinen, A., Kerminen, A., "A linear non-gaussian acyclic model for causal discovery" , J. Machine Learning Research, 7, 2003-2030 (2006)
[43] 多変量の信号を複数の加法的な成分に分離するための計算手法である。各成分は、ガウス的でない信号で相互に統計的独立なものを想定する。
https://ja.wikipedia.org/wiki/%E7%8B%AC%E7%AB%8B%E6%88%90%E5%88%86%E5%88%86%E6%9E%90

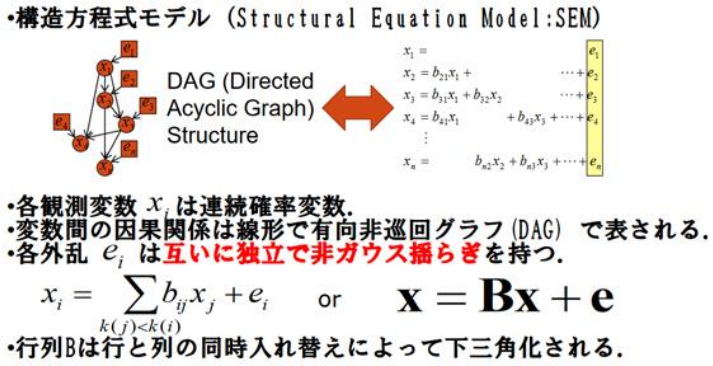


図 1-23　LiNGAM モデル[44]

本研究では、独立成分分析は使わず、係数行列を回帰分析と独立性の評価を繰り返すことにより、観測変数 x_i の因果的順序を 1 つずつ推定する手法を開発した(DirectLiNGAM モデル)[9]。基本原理を図 1-24 に示した。

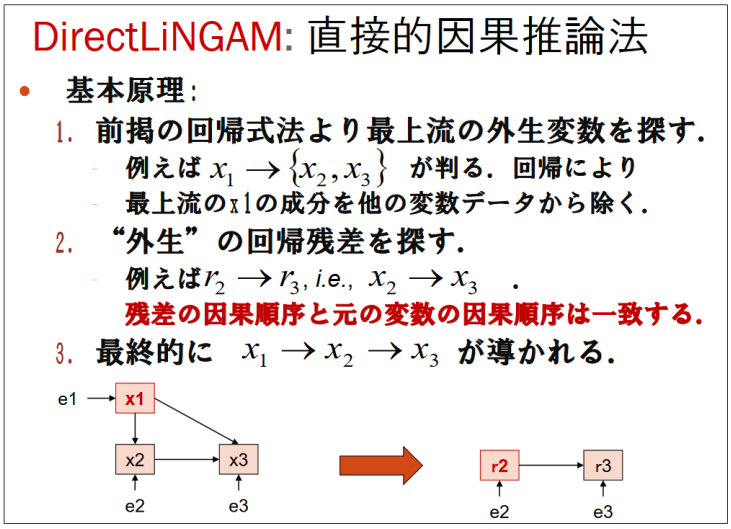


図 1-24　DirectLiNGAM モデルの基本原理[44]

(3) 組み合わせ論的計算による高次元データからの知識発見

高次元データからの解析は従来から行われており、分類・判別・回帰式等の手段で解析が行われてきた。これは基本的な機械学習・統計モデリングの問題である。すなわち、高次元データ集合から評価関数を求め、評価関数の計算から評価となる数値を求めることである。表 1-4 はその例である。

[44] https://www.sas.com/content/dam/SAS/ja_jp/doc/event/sas-user-groups/usergroups2015-g-01.pdf

表 1-4　機械学習における集合関数の最適化の例[45]

問題	有限集合	評価関数の例
特徴選択	説明変数の索引	2 乗誤差
能動学習	サンプル	Fisher 行列の比
クラスタリング	サンプル	カット関数，MDL
画像セグメンテーション	画素	エネルギー（カット）
グラフ構造学習	ノード	相互情報量

機械学習において、劣モジュラ関数とその最適化理論の重要性が認識されるようになったのは、ここ 10 年以内のこととされている[46]。本研究は劣モジュラ関数とその展開・応用を目指した研究である。

劣モジュラ関数とは劣モジュラ性を持つ集合関数[47]であり、図 1-25 に示すように、一般の連続関数の凸性を集合関数で表したのが劣モジュラ性であり、劣モジュラ性は集合関数を用いて以下のように定義する。

集合関数　$g(X_s)$　s.t.　$X_s \subset X$

劣モジュラ性　$g(X_{s1} \cup \{i\}) - g(X_{s1}) \geqq g(X_{s2} \cup \{i\}) - g(X_{s2})$　但し　$X_{s1} \subseteq X_{s2} \subseteq X$

具体的には、「限界効用逓減効果」[48]や「規模の経済性」などがその例である。集合関数は対象とするモデルの目的によって選ばれる(例えば、対数尤度関数等)。

[45] 河原吉伸、「機械学習における劣モジュラ性の利用と組合せ論的アルゴリズム」、 オペレーションズ・リサーチ、 58(5)、 267-274 (2013)
[46] 河原吉伸、永野清仁、「劣モジュラ最適化と機械学習」、講談社、2015 年
[47] 集合関数とは集合を引数とするスカラー関数である。
[48] 一般的に、財の消費量が増えるにつれて、財の追加消費分(限界消費分)から得られる効用は次第に小さくなる、とする考え方。

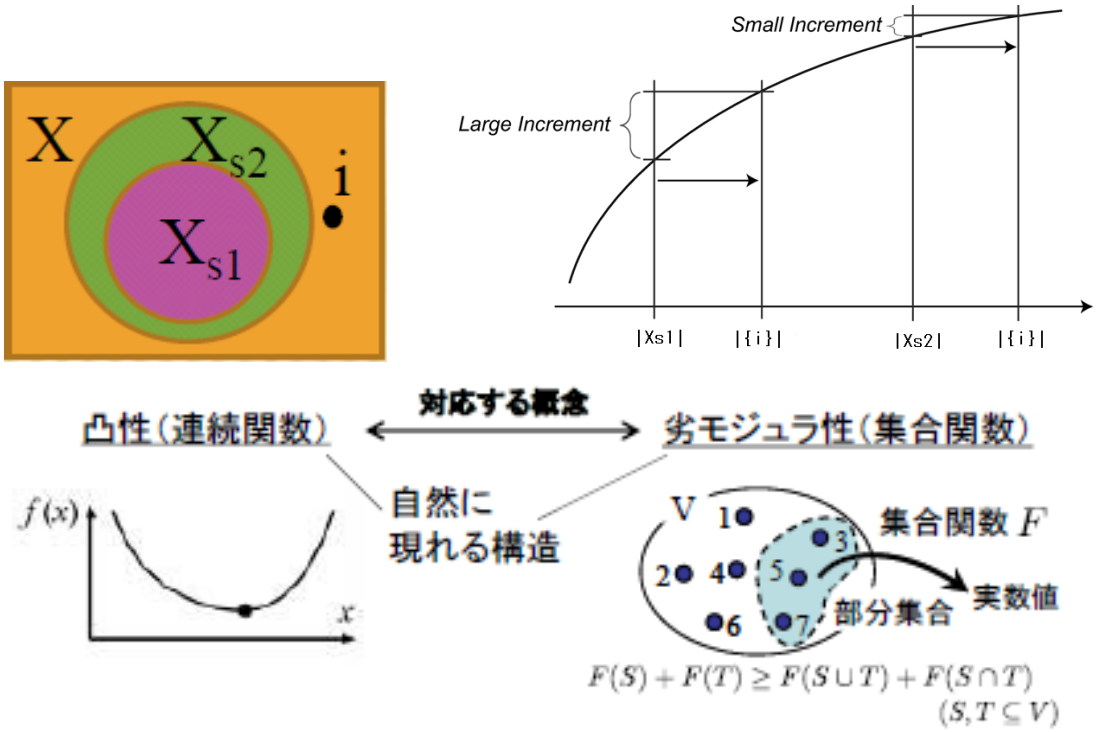


図 1-25　劣モジュラ関数の概念[49]

本研究は、劣モジュラ関数の最大化問題の方法論や理論解析の研究である。特に、限られた説明変数の下、結果を最も良く説明する最良の組み合わせを求める組み合わせ問題について、劣モジュラ関数の最大化問題を研究した[10]。

(4) 超高次元データからの情報推定・知識発見

近年のネットワークセンシングや生物情報科学、環境科学の実験測定技術、リモートセンシングの進歩を通して、超高次元ベクトルで表現されるデータが大量蓄積される時代となった。一方、データマイニングでは、解析目的に応じた諸条件でデータを場合分けした沢山のデータセグメントを作り、セグメント毎に傾向分析を行うのが一般的である。しかし、数十次元以上では、推定精度やロバスト性が大きく損なわれる。この問題に対して、次元圧縮を用いて問題の軽減を図る手法が研究されてきた。これは情報削減による近似推定であり、結果の解釈が難しい。本プロジェクトでは、超高次元データに関する一般的な非圧縮直接推定の研究を行った。

---

49 https://www.mext.go.jp/b_menu/shingi/gijyutu/gijyutu17/002/shiryo/__icsFiles/afieldfile/2012/09/12/1325657_02_1.pdf
http://ibisml.org/ibis2010/session/ibis2010kawahara.pdf

初めに、様々な超高次元データ分布について解析を行い、「球面集中効果」、「確率密度集中効果」、「スパース効果」と呼ばれるいくつかのケースに分類されることを見出した[50]。本プロジェクト以前に「球面集中効果」については、効果を打ち消す人工的な歪データを与える方法を提案した[51]。本プロジェクトでは「確率密度集中効果」について、確率密度集中効果を打ち消す人工的状態分布を与える高精度、ロバストな推定法を提案した[52]。「スパース効果」についてはランダム部分標本推定を多数重ねる高精度、ロバストなアンサンブル推定法を提案した[11]。

しかし、これらの成果では数千を超える本質次元で十分な推定が得られない場合がある。これらの問題のブレークスルーを実現する一般原理を様々な方向から探求した。

まず、ランダムサンプリングによって次元の呪いの影響を軽減しつつ、高本質次元空間上の事例の類似性やデータ密度を機械学習手法に反映する基礎原理を探求した。その結果、種々の類似性尺度や密度評価方法を確立することができた[12]。また、密度推定を行う上で外れ値などの外乱の影響を効果的に排除する基礎原理として Density Power Divergence[53]を導入する方法を確立した[54,13]。

(5)ビジネスデータを用いたマイニング検証実験

ニュース記事からの市場動向予測や、Web 上のクチコミデータの分析等の実データを用いた機械学習と統計技術の適用実験を行った。

市場動向の予測については、強化学習を利用した市場予測のシミュレーションを実施した[55]。具体的には、ブルームバーグ社が日々配信する日本語の経済関係のニュース記事(12年間)のデータを解析の対象とした。ニュース記事の含まれる語句を機械的に処理し、その中のキーワードと株価の関連性を解析した。株価評価の表現として、「回復する」「株価が上昇する」(好評表現)、「下落する」「上値が重い」(不評表現)などの事象を表す表現を選び、評価表現という極性を付与した評価表現辞書を作成した。これを種表現として、例えば、「上値が重い」の前後の文章は「しかし」等の表現が出現しない限り不評表現で構成されていると仮定して、新たな語彙表現を追加することを行い語彙の充実を図った。これらの語彙を使

---

[50] 鷲尾隆、「情報時代の高次元データマイニング」、電子情報通信学会誌、94(8)、679-683 (2011)

[51] Kido K., Kuwajima H. Washio T., "A range query approach for high dimensional euclidean space based on EDM estimation", Proc. of SIAM Data Mining Conf.(SDM08), 387-398 (2008)

[52] Phuong N.V., Washio T. Higuchi T., "A new particle filter for high-dimensional state-space models based on intensive and extensive proposal distribution", Int. J. Knowledge Eng. and Soft Data Paradigms, 2(4), 284-311 (2010)

[53] Density Power Divergence：確率密度分布間の差異を測る手法の 1 つ。他に、KL-ダイバージェンスが知られている。

[54] 十河泰弘、 植野剛、 河原吉伸、 鷲尾隆、 「Density Power Divergence を用いたロバスト能動回帰学習」、 人工知能学会論文誌、 28(1)、13-21 (2013)

[55] 前川浩基、 中原孝信、 岡田克彦、 羽室行信 、「大規模ニュースデータと株価収益率の予測可能性について」、 第 10 回 人工知能学会 ファイナンスにおける人工知能応用研究会(SIG-FIN)、2013 年 3 月 19 日

ってモデルを構築し、将来の予測を行う。種々のモデルがあるが、本研究ではナイーブ・ベイズモデル[56]を採用し、このモデルによってどのような評価表現が上昇(下落)に確率的に結びつきやすいかを認識させた。

投資モデルは、直近のニュース記事を学習したモデルで評価し、売買指示を出す。直近どの範囲までの記事を対象とし、どこまでの予測をするのかについて全60通りの組み合わせを用意し、日々将来の日経225先物の上下動を予測しながら11年間運用した。実験の結果、いくつかのモデルの成績が偶然では得られない結果を残すことが明らかとなった。

(6)応用ソフトウェアの開発

任意の駅間の全経路を計算するEkillion、ZDDを利用できるRuby、大規模表構造テキストデータを高速に処理するコマンド群MCMD等の応用ソフトウェアを開発した。

① Ekillionの公開

大規模グラフデータからの高速パス列挙を可能とするパッケージGraphillionを応用し、JR 大都市近郊区間において任意の駅間の全経路を計算し地図上に描画するソフトウェアEkillion(図 1-26)を開発/公開した[57]。ZDD アルゴリズムの威力を一般の人々に分かりやすく示す応用例となっている。

[56] https://avinton.com/academy/naive-bayes/
[57] http://www2.nysol.jp/apps/ekillion

図 1-26　Ekillion による経路列挙の結果画面[58]

② ZDD on Ruby

本プロジェクトのコア技術である ZDD を、簡易スクリプト言語である Ruby 上で利用できるようにした拡張ライブラリを開発した。ZDD が提供する全ての演算と関数を Ruby の演算子/関数として利用することが可能となり、応用ソフトウェアの開発が容易となる。

③ MCMD の公開

大規模表構造テキストデータを高速に処理するコマンド群 MCMD(M コマンド)を開発/公開した[59]。このソフトウェアは、知識発見プロセスにおいて最も時間を要するデータスクリーニングや前処理を得意とし、一般の PC でも数億レコードのデータ処理が実現できる。また並列処理にも対応しており、Hadoop[60]との比較実験において、平均的に 2 倍程度の処理効率を達成した。

---

[58] 終了報告書より
[59] https://www.nysol.jp/
[60] Hadoop とはデータを複数のサーバに分散し、並列して処理するミドルウェア(ソフトウェア基盤)のこと。

1.5.4　セミナー・ワークショップ開催とアウトリーチ活動

多くのセミナー(年度平均 25 日)、ワークショップ(年 2 回、講演件数平均 37 件/回)を開催し、積極的に研究の交流や情報の共有を図った。さらに、アウトリーチ活動において、日本科学未来館へグリッドグラフを対象としたパスの数え上げに関する展示を行い、大きな反響を得た。展示のコンテンツとしてアップした YouTube の動画「フカシギの数え方」が現在 295 万アクセスを超えている。

(1)多くのセミナー・ワークショップ開催

本プロジェクトでは、各グループリーダーの研究拠点が離れていることもあり、多くのセミナー・ワークショップの開催により研究の交流や情報の共有を図った。具体的には、各年度平均 25 日セミナーを開催し(表 1-5)、年 2 回(合計 12 回)のワークショップを開催し、平均 37 件/回の講演(表 1-6)が行われた。

表 1-5　ERATO セミナーの開催日数

年度	開催日数(日)
2010	34
2011	32
2012	34
2013	32
2014	20
2015	17
平均	28

表 1-6　ワークショップの講演件数

ワークショップ	講演件数(件)
2010 年度夏	30
2010 年度冬	25
2011 年度夏	43
2011 年度秋	38
2012 年度夏	43
2012 年度秋	35
2013 年度夏	36
2013 年度秋	38

ワークショップ	講演件数(件)
2014 年度春	45
2014 年度秋	46
2015 年度夏	26
2015 年度秋	36
平均	37

(2)アウトリーチの成果

本研究の成果として、アウトリーチの成果が挙げられる。日本科学未来館のメディアラボで、第 11 期展示として，湊プロジェクトの成果展示(グリッドグラフを対象としたパスの数え上げに関する展示)を行った。「フカシギの数え方」と題した本展示は予想外の反響があり、2012 年 8 月 1 日から 2013 年 4 月 15 日までの期間中、20 万人を超える来場者数となった。小中高校生のほか，大人にも十分楽しんでもらえる展示内容であった。また、展示のコンテンツとして作成した「フカシギの数え方」の動画(格子グラフを対象として組みあわせ爆発の様子をわかりやすく示したもの)は YouTube 上で現在 295 万アクセスを超えている。「アルゴリズム技術」に関しては，従来なかなか一般人に理解してもらえるようなコンテンツはなかったので、本展示で開発したコンテンツはその意味で画期的なものであった。

第2章　プロジェクト終了から現在に至る状況

2.1　プロジェクトの終了後の状況に関する基礎データ

2.1.1　調査方法

調査は、文献調査(プロジェクト報告書、解説、原著論文など)、インターネットによる調査、各種データベースによる業績(論文・特許・受賞他)の調査からなる基礎データ調査と、プロジェクト関係者や外部有識者へのインタビュー調査を行った。これに基づき、本プロジェクト期間中の成果の調査時点での発展状況及び波及効果等についてまとめた。

(1)基礎データ調査の方法

基礎データ調査については、基本的にプロジェクトメンバーを対象として、プロジェクトの研究に関連した成果の発展状況について、文献による成果の把握と、論文や研究助成金の獲得状況等のデータ調査を行った。各項目について利用したデータベースと調査範囲等を下記に記す。

①論文[61]

本プロジェクト期間中の論文は、2009 年～2016 年に発表された論文及びプロジェクトの終了報告書に成果論文としてリストアップされている論文とした。成果論文リストの中で in press、to be submitted 等と表記があり、その後発表されたものについても基本的には期間中の論文とした。以下、これらの論文を「成果論文」と定義する。

本プロジェクト終了後の論文は、2017 年 1 月以降に発表され、かつ本プロジェクトメンバーが著者になっている論文を収集した(ただし、本プロジェクト期間中の成果論文に含むものは除く)。収集した論文の中で、本プロジェクトとの関連を Abstract、引用関係等で確認し、関連のあるものを関連論文としてリストアップした。以下、これらの論文を「発展論文」と定義する。

データベースは、主としてエルゼビア社の Scopus を利用し、補完的にクラリベイトアナリティクス社の Web of Science を利用した。

各論文についての評価の 1 つである FWCI(Field Weighted Citation Impact)[62]、及び Journal の指標となる CiteScore2018[63]についても収集した。

---

[61] ここで示す論文の定義は、文献データベース Scopus で収集を行い、Scopus の文献タイプ“Erratum”を除く成果の記述の含まれるすべてのものを対象とした。

[62] FWCI(Field Weighted Citation Impact)：1 文献あたりの被引用数を世界平均(年別・分野別・文献タイプ別に算出)で割った数値。

[63] ジャーナル評価指標(Web of Science の Impact Factor と同様の指標)。各論文の出版年に相当する CiteScore を取得。

②競争的研究資金の獲得状況

プロジェクトメンバー全員を対象として、本プロジェクトの研究内容に関連している研究課題について調べた。表 2-1 はその中で、競争的研究資金の総額が 1 千万円以上のものを抽出して示した。

データベースとしては、調査対象者の所属する研究室や本人の WEB サイト及び KAKEN 科学研究費助成事業データベース等の競争的研究資金に関する検索サイトと、補助的に Google 等の検索サイトを利用した。

③特許の出願・登録状況

本プロジェクト期間中の特許は、プロジェクト終了報告書の成果リスト記載の特許とした。本プロジェクト終了後の特許は 2016 年 4 月以降に出願されかつ、プロジェクトメンバーが発明者の特許とし、データベースは、主に PatentSQUARE(パナソニック社)を利用し、補助的に特許情報プラットフォームと espacenet を利用した。

④招待講演

プロジェクト関係者の本プロジェクト終了後の招待講演を調査対象者の所属する研究室や本人の WEB サイトの調査、Google 等の検索サイト、①で記述した文献データベースの会議録等を併せて収集し、かつインタビューの際に主な招待講演について確認した。

⑤受賞

プロジェクトメンバーの本プロジェクト終了後の受賞を調査対象者の所属する研究室や本人の WEB サイトの調査、Google 等の検索サイトで収集後、研究総括及びグループリーダーに確認した。

⑥ベンチャー

インターネット検索やベンチャー情報の記載のある DB を用いて検索した。

⑦報道

日経テレコン及び、プロジェクト関係者の WEB サイトの情報を収集した。

⑧参加研究者の動静

終了報告書を元にプロジェクト参加研究者を特定し、プロジェクト参加時の職位及び、終了時の職位、現在の職位を検索した。

(2)インタビュー調査の方法

インタビュー調査は本プロジェクトの主なメンバー、数名について、実施した。プロジェ

クト関係者には、基礎調査で知り得た情報の本プロジェクトとの関連や、その後の展開等についての情報を収集した。

2.1.2　競争的研究資金の獲得状況

本プロジェクト期間中から現在までのプロジェクトメンバーも含めた競争的資金(1 千万円以上)の獲得状況を表 2-1 に示す。本プロジェクト以外の研究助成金の獲得件数は湊が 7 件、津田が 9 件、鷲尾が 5 件、合計 21 件であった。研究代表者としては、湊が 2 件、津田が 4 件、鷲尾が 2 件であった。

表 2-1　競争的資金の獲得状況(金額 1 千万円以上)

JST　科研費　AMED　JSPS

研究期間(年度)	研究種目	研究課題	研究代表者	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	金額(億円)
2009～2015	ERATO	湊離散構造処理系	湊 真一	JST	JST	JST	JST	JST	JST	JST										9.59
2012～2016	科研費 新学術領域研究 (研究領域提案型)	多面的アプローチの統合による計算限界の解明	渡辺 治 (連携研究者: 浅野 孝夫、 茨木 俊秀、 今井 浩、 戸田 誠之助、 丸岡 章、 湊 真一、 牧野 和久、 河原林 健一、 浅野 哲夫、 加藤 直樹、 エイビス デビッド、 徳山 豪、 山下 茂、 瀧本 英二、 堀山 貴史、 森 立平)				科研費	科研費	科研費	科研費	科研費									1.34
2013～2016	科研費 基盤研究(A)	超高次元データ空間における統計的推定・シミュレーション原理の開発と応用展開	鷲尾 隆					科研費	科研費	科研費	科研費									0.46

研究期間（年度）	研究種目	研究課題	研究代表者	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	金額（億円）
2013～2016	科研費 基盤研究(A)	建築・都市分野における離散数理基盤の構築と大規模最適化への展開	加藤 直樹 (連携研究者:湊 真一、宇野 毅明)																	0.47
2013～2017	科研費 新学術領域研究 (研究領域提案型)	ナノ構造情報に基づいた機能探索	田中 功 (研究分担者:吉矢 真人、大場 史康、津田 宏治)																	1.46
2013～2017	科研費 新学術領域研究 (研究領域提案型)	ナノ構造情報のフロンティア開拓－材料科学の新展開	田中 功 (連携研究者:松永 克志、柴田 直哉、武藤 俊介、谷口 尚、太田 裕道、北岡 諭、菅野 了次、高草木 達、津田 宏治、足立 裕彦、佐久間 健人、髙野 幹夫、森 博太郎、樋口 知之)																	0.28
2013～2018	CREST[ビッグデータ統合利活用のための次世代基盤技術の創出・体系化]研究領域	自己情報コントロール機構を持つプライバシ保護データ収集・解析基盤の構築と個別化医療・ゲノム疫学への展開	佐久間 淳 (主なる共同研究者:津田 宏治)																	2.73
2014～2019	文部科学省 ポスト「京」で重点的に取り組むべき社会的・科学的課題に関するアプリケーション開発・研究開発 重点課題アプリケーション開発	重点課題(1)生体分子システムの機能制御による革新的創薬基盤の構築	奥野 恭史 (津田 宏治)																	12.55

研究期間（年度）	研究種目	研究課題	研究代表者	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	金額（億円）
2015～2017	科研費 基盤研究(B)	超高次元データ解析アルゴリズムに基づく呼気診断センサのハード・ソフト双方向開発	吉川 元起 (連携研究者:鷲尾 隆、柴弘太)																	0.15
2015～2017	科研費 新学術領域研究 (研究領域提案型)	ナノ構造情報のフロンティア開拓－材料科学の新展開(国際活動支援)	田中 功 (連携研究員:松永 克志、武藤 俊介、谷口 尚、北岡 諭、菅野 了次、高草木 達、津田 宏治)																	0.49
2015～2019	科研費 基盤研究(S)	離散構造処理系の基盤アルゴリズムの研究	湊 真一 (研究分担者:有村 博紀、瀧川 一学、宇野 毅明、堀山 貴史、津田 宏治、鷲尾 隆)																	1.34
2015～2019	研究成果展開事業 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	伊藤 聡 (津田 宏治)																	22.50(最大)
2015～2020	CREST [ビッグデータ統合利活用のための次世代基盤技術の創出・体系化]研究領域	離散構造統計学の創出と癌科学への展開	津田 宏治																	－
2015～2020	CREST[新たな光機能や光物性の発現・利活用を基軸とする次世代フォトニクスの基盤技術]研究領域	超解像「生理機能」イメージング法の開発と細胞状態解析への応用	永井 健治 (主なる共同研究者:鷲尾 隆)																	－
2016～2017	科研費 新学術領域研究 (研究領域提案型)	材料インフォマティクスに適した機械学習法の開拓	津田 宏治																	0.12

研究期間（年度）	研究種目	研究課題	研究代表者	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	金額（億円）
2016～2018	JSPS 人材育成事業 頭脳循環を加速する戦略的国際研究ネットワーク推進プログラム 理工系	グローバル分子技術実装ネットワークの構築	竹田 精治（鷲尾 隆）																	1.05（最大）
2016～2019	科研費 基盤研究(A)	実世界知識基盤形成のための次世代半構造マイニング技術の研究	有村 博紀（研究分担者:宇野 毅明、湊 真一、平田 耕一、伊藤 公人、下薗 真一、喜田 拓也）																	0.44
2016～2021	CREST[計測技術と高度情報処理の融合によるインテリジェント計測・解析手法の開発と応用]研究領域	機械学習と最先端計測技術の融合深化による新たな計測・解析手法の展開	鷲尾 隆																	-
2018	AIP 加速 PRISM 研究(AIP-PRISM)	ターゲット検証用ツール抗体の開発	津田 宏治																	-
2018～2023	CREST[Society5.0を支える革新的コンピューティング技術]研究領域	学習/数理モデルに基づく時空間展開型アーキテクチャの創出と応用	本村 真人（主なる共同研究者:有村博紀、湊 真一、坂井 哲）																	-
2020～2024	科研費 基盤研究(A)	離散構造処理系に基づく列挙と最適化の統合的技法の研究	湊 真一																	0.34
2020～2024	科研費 学術変革領域研究(A)	社会変革の源泉となる革新的アルゴリズム基盤の創出と体系化	湊 真一（領域代表）																	8.56

2020 年 4 月 11 日調査
2020 年 12 月 1 日更新
2021 年 2 月 28 日確定

2.1.3 論文の発表状況

(1)本プロジェクト成果に直接関わる論文

期間中、終了後を通して、投稿された論文数とその被引用数から求められた FWCI 値を表

2-2 に示した。成果論文が 167 報、発展論文が約半分の 86 報であった。FWCI から求めた TOP%の 10%以内が成果論文では全体の約 8%、発展論文では約 13%であった。報告書に挙げられている論文を中心に、本プロジェクトと直接かかわると判断できる論文数をその被引用数の推移と共に図 2-1 に示す。また、被引用数上位 5 報の概要を表 2-3 に示す。プロジェクト期間中の論文が、プロジェクト終了後も安定して引用されている。被引用数 1 位、2 位の研究は材料に関する最適化の研究である。5 位は遺伝子座に関する研究である。3 位は、DirectLiNGAM、4 位は COMBO アルゴリズムの研究であり、本プロジェクトの中心研究である。

表 2-2　プロジェクトの論文投稿状況一覧

	論文数	FWCI 平均値	FWCI Percentiles　TOP%				
			0.01%以内	0.1%以内	1%以内	10%以内	10%圏外
成果論文	167	0.88	0	0	0	14	153
発展論文	86	1.11	0	0	0	11	75

2020 年 10 月 8 日更新
2020 年 12 月 1 日更新
2021 年 2 月 28 日確認
(FWCI 値及び、FWCI Percentiles の値は 2020 年 10 月 14 日の値)

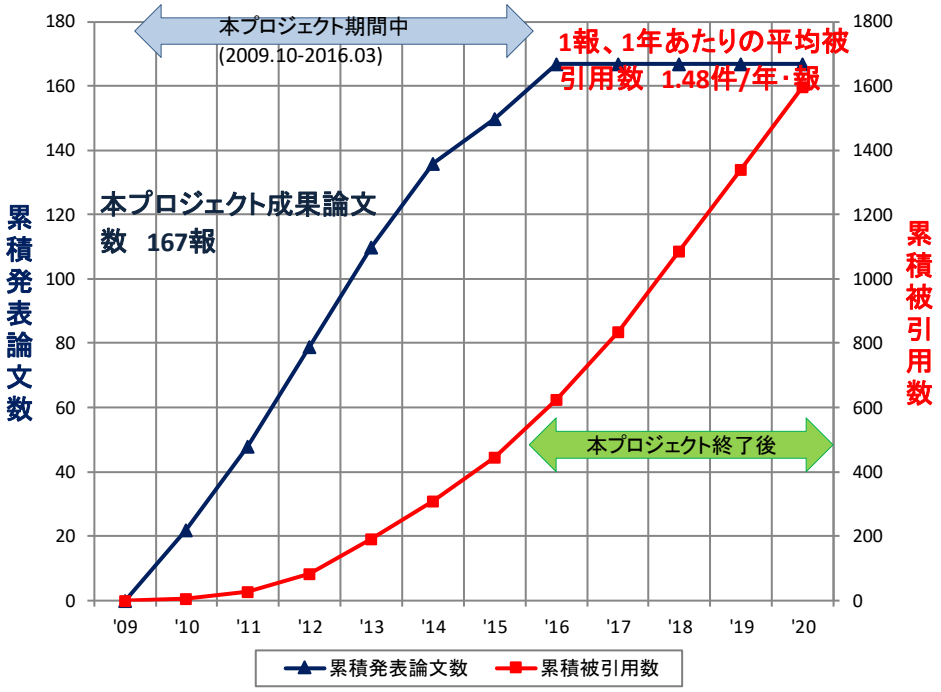


図 2-1　プロジェクトの成果論文の発表論文数と被引用数の推移
(検索 DB：Scopus　検索日　2021 年 1 月 5 日)

表 2-3　プロジェクトの成果論文の内、被引用数上位 5 報の論文概要

No	著者名	タイトル	出版年	出版物名	巻(号)	論文番号/ページ	被引用数	DOI	文献タイプ	FWCI	CiteScore
1	Seko A., Togo A., Hayashi H., Tsuda K., Chaput L., Tanaka I.	Prediction of Low-Thermal-Conductivity Compounds with First-Principles Anharmonic Lattice-Dynamics Calculations and Bayesian Optimization	2015	Physical Review Letters	115 (20)	205901	194	10.1103/PhysRevLett.115.205901	Article	7.93	15.1
2	Seko A., Maekawa T., Tsuda K., Tanaka I.	Machine learning with systematic density-functional theory calculations: Application to melting temperatures of single- and binary-component solids	2014	Physical Review B - Condensed Matter and Materials Physics	89 (5)	54303	127	10.1103/PhysRevB.89.054303	Article	5.72	7.0
3	Shimizu S., Inazumi T., Sogawa Y., Hyvärinen A., Kawahara Y., Washio T., Hoyer P.O., Bollen K.	DirectLiNGAM: A direct method for learning a linear non-gaussian structural equation model	2011	Journal of Machine Learning Research	12	1225-1248	118		Article	3.99	7.7
4	Ueno T., Rhone T.D., Hou Z., Mizoguchi T., Tsuda K.	COMBO: An efficient Bayesian optimization library for materials science	2016	Materials Discovery	4	18-21	98	10.1016/j.md.2016.04.001	Article	6.11	0.3
5	Kam-Thong T., Czamara D., Tsuda K., Borgwardt K., Lewis C.M., Erhardt-Lehmann A., Hemmer B., Rieckmann P., Daake M., Weber F., Wolf C., Ziegler A., Pütz B., Holsboer F., Schölkopf B., Müller-Myhsok B.	EPIBLASTER-fast exhaustive two-locus epistasis detection strategy using graphical processing units	2011	European Journal of Human Genetics	19 (4)	465-471	66	10.1038/ejhg.2010.196	Article	2.03	7.1

(検索 DB：Scopus　検索日：2021 年 1 月 5 日　FWCI、CiteScore は 2020 年 10 月 14 日時点)
CiteScore は出版年の値。

(2) 本プロジェクトの成果の発展、または本プロジェクトから波及した研究内容の文献

本プロジェクトの成果の発展、波及の論文数の推移を図 2-2 に示した。発表論文数 86 報であり、被引用数は順調に増加している。被引用数上位 5 報の概要を表 2-4 に示した。上位 1 位、3 位、4 位は無機材料最適化、有機分子の最適検索、2 位は分子探索アルゴリズム ChemTS の論文である。この 1 位～4 位の論文はすべて津田の論文である。5 位は鷲尾によるナノポアによるバクテリア検出の論文である。

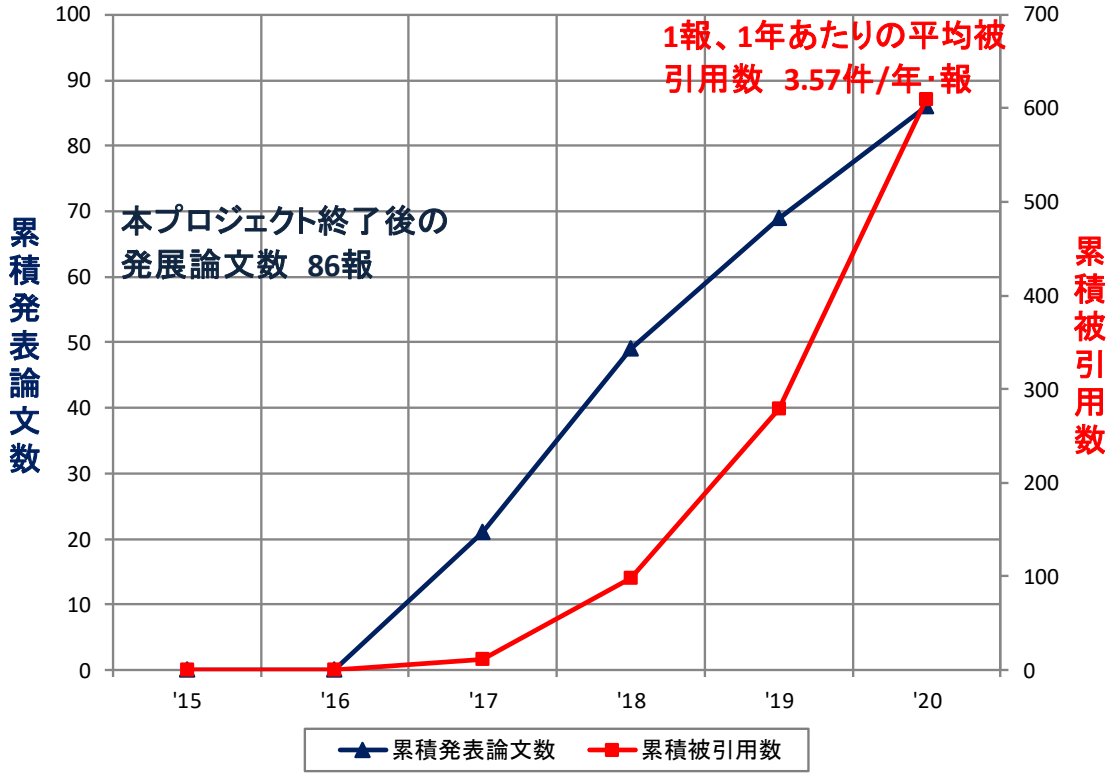


図 2-2　プロジェクト終了以降の発表論文の発表数と被引用数の推移
（検索 DB：Scopus　検索日　2021 年 1 月 5 日）

表 2-4　プロジェクトの発展論文の内、被引用数上位 5 報の論文概要

No	著者名	タイトル	出版年	出版物名	巻(号)	論文番号／ページ	被引用数	DOI	文献タイプ	FWCI	CiteScore
1	Ju S., Shiga T., Feng L., Hou Z., Tsuda K., Shiomi J.	Designing nanostructures for phonon transport via Bayesian optimization	2017	Physical Review X	7 (2)	21024	94	10.1103/PhysRevX.7.021024	Article	7.30	22.1
2	Yang X., Zhang J., Yoshizoe K., Terayama K., Tsuda K.	ChemTS: an efficient python library for de novo molecular generation	2017	Science and Technology of Advanced Materials	18 (1)	972-976	53	10.1080/14686996.2017.1401424	Article	2.91	7.1

No	著者名	タイトル	出版年	出版物名	巻(号)	論文番号/ページ	被引用数	DOI	文献タイプ	FWCI	CiteScore
3	Yamashita T., Sato N., Kino H., Miyake T., Tsuda K., Oguchi T.	Crystal structure prediction accelerated by Bayesian optimization	2018	Physical Review Materials	2 (1)	13803	39	10.1103/PhysRevMaterials.2.013803	Article	3.48	1.6
4	Sumita M., Yang X., Ishihara S., Tamura R., Tsuda K.	Hunting for Organic Molecules with Artificial Intelligence: Molecules Optimized for Desired Excitation Energies	2018	ACS Central Science	4 (9)	1126-1133	29	10.1021/acscentsci.8b00213	Article	2.78	12.2
5	Tsutsui M., Yoshida T., Yokota K., Yasaki H., Yasui T., Arima A., Tonomura W., Nagashima K., Yanagida T., Kaji N., Taniguchi M., Washio T., Baba Y., Kawai T.	Discriminating single-bacterial shape using low-aspect-ratio pores	2017	Scientific Reports	7 (1)	17371	25	10.1038/s41598-017-17443-6	Article	1.38	4.8

(検索DB：Scopus　検索日：2021年1月5日　FWCI、CiteScoreは2020年10月14日時点)
CiteScoreは出版年の値。

2.1.4　特許の出願・公開・登録状況

本プロジェクトの期間中と終了後の調査時点に至るまでの特許出願状況を表2-5に示す。期間中の国内出願が5件、すべてが登録されている。出願人は、2件が日本電信電話株式会社(NTT)と北海道大学の共同出願、残り3件が国立研究開発法人物質・材料研究機構(NIMS)、国立研究開発法人産業技術総合研究所(産総研)から出願されている。プロジェクト終了後は15件出願されている(内、6件国内登録)。内訳はNTT単独の出願が5件、企業と大学の共同出願が3件、NIMS単独の出願が2件、大阪大学単独の出願が4件、NIMSと大学の共同出願が1件となっている。

国際出願について、期間中に1件が国際出願されており、産総研から出願された。終了後に7件海外出願が行われた。出願人については、大阪大学単独が3件、大阪大学と発明者の共同出願2件、NIMS単独が2件であった。

本プロジェクトにおける特許出願の特徴は、企業あるいは国立研究開発法人からの出願や共同出願が多いことである。このことは企業や国立研究開発法人との共同研究が多いことを示している。期間中より終了後の出願が多いのは大きな特徴である。このことは、企業等との共同研究が増えたこと、応用研究における成果が大きかったことを示している。国際出願が多いことも大きな特徴である。これは発明の内容が重要であり、世界的に権利を取得

したいという考えの表れとみなせる。また、大阪大学の出願特許は大阪大学発のベンチャー企業、アイポア株式会社の基盤となっている。

表 2-5　プロジェクトの特許出願状況一覧

	出願件数		登録件数	
	国内	海外	国内	海外
プロジェクト期間中	5	1	5	1
プロジェクト終了後	15	7	6	0
合計	20	8	11	1

2020 年 4 月 1 日検索
2020 年 11 月 16 日更新
2021 年 2 月 28 日確定

2.1.5　招待講演

期間中の招待講演は国内学会 19 件(湊 16 件、湊、津田、鷲尾の連名 1 件、その他 2 件)であった。国際学会が 5 件(日本開催を含む)(湊 4 件、鷲尾 1 件)であった。終了後は、国内 5 件(湊 4 件、津田 1 件)、国際 10 件(湊 4 件、津田 3 件、鷲尾 3 件)であった。

2.1.6　受賞状況

期間中に 5 件の受賞があった。2010 年に湊が 2010 年電子情報学会の情報・システムソサイエティ論文賞(先見論文)、2012 年に井上、湊が 2012 年電子情報通信学会の IN 研究会情報ネットワーク研究会研究賞、2012 年に羽村ほかが人工知能学会の金融情報研究会 2012 年優秀論文賞、湊が 2013 年に電子情報通信学会の基礎・境界ソサイエティ　Fundamentals Review ベストオーサー賞、白井、森田、中元、Chung、高嶋が平成 26 年度データ解析コンペティション最優秀賞を受賞した。

終了後に全部で 4 件の受賞があった。井上と湊が 2017 年度人工知能学会の研究会優秀賞、石岡、川原、水田、湊、栗原が 2018 年度日本計算機統計学会の論文賞、鷲尾と原が 2020 年度日本顕微鏡学会の論文賞、鷲尾が 2017 年に IBM(米国)から 2016 IBM Faculty Award を受賞した。

2.1.7　ベンチャー企業の設立状況

本プロジェクトに関連して、設立されたベンチャー企業は以下の 3 社であった。

株式会社 Magne-Max Capital Management は 2011 年に設立された投資助言会社。パートナーとして契約した金融機関や資産運用会社に対してのみ投資助言を行う。湊がアドバイザである。

MI-6 株式会社は 2017 年に設立された IT を活用した材料開発の研究、開発、コンサルティング、プロジェクトマネジメント等を行う会社である。津田がアドバイザである。

アイポア株式会社は2018年に設立された人工知能を活用したIoTナノ粒子センサの開発および販売を行う大阪大学発のベンチャー企業である。鷲尾はアドバイザ、共同研究者の大阪大学谷口正輝が創立し、現在は非常勤の取締役である。

2.1.8 報道関係

期間中、終了後を通して報道されたタイトル、報道日、報道機関を表2-6にまとめた。報道件数は期間中3件、終了後44件、合計47件であった。応用研究の成果の発表が26件、ベンチャー関係が9件、受賞関係が3件、連携・共同関係の発表が2件、その他が7件であった。報道した報道機関は35機関に及んだ。多く報道した報道機関は、日経速報ニュースアーカイブが11件、日本経済新聞電子版(ニュースとセクションを含む)が13件、日経プレスリリースが8件、科学新聞が7件等となっている。

表2.6 報道されたタイトル、報道日、報道機関の一覧

報道タイトル	報道日	報道機関
スマートグリッド、CO_2削減効果は15.5トン－北大と早大	2012/02/24 ～ 2012/03/16	朝日新聞朝刊、日刊工業新聞、電経新聞、電気新聞、科学新聞、
最先端の開拓者たち－世界的権威が認めた　超高速アルゴリズム 電力危機に挑む－湊真一氏 北海道大学大学院 情報科学研究科 教授	2012/09/27	日経コンピュータ
<コラム:素領域>	2014/11/14	科学新聞
シミュレーション連携で人工知能開発を加速-NECと産総研が共同研究室を設立	2016/04/05 ～ 2016/04/15	マイナビニュース、Biz/Zine、BCN Bizline、クラウドWatch、日本経済新聞電子版セクション、化学工業日報、日刊工業新聞、電気新聞、日刊自動車新聞、日本情報産業新聞、科学新聞
東京大学教授津田宏治氏――機械学習でビッグデータ解析、医学から材料開発貢献(次世代の先導者)	2016/04/14	日経産業新聞
Emerging Tech－新産業－白熱する日本のAI研究、狙いは深層学習の凌駕	2016/06/20	日経エレクトロニクス
都道府県の隣接 1098 億通りと算出　病気などの地理的分析へ 北大・岡山大	2016/09/02	共同通信ニュース、秋田魁新報朝刊、東京新聞朝刊、静岡新聞朝刊、四国新聞朝刊
「データサイエンスアワード2016」最優秀賞を九州大学大学院内田研究室が受賞［一般社団法人データサイエンティスト協会］	2016/10/17	@PRESS

報道タイトル	報道日	報道機関
慢性心不全の予後を退院時BNP値から高確率で予測－国循	2016/11/18	QLifePro 医療ニュース
JSTなど、複数の疾患遺伝子を解析特定する手法開発－がん発症を早期に見分け	2017/01/18	日刊工業新聞
経営ひと言/北海道大学･湊真一教授「説得力が物言う」	2017/03/15	日刊工業新聞
NIMS×東大、機械学習により熱流を制御するナノ構造物質の最適設計に成功	2017/04/18 ～ 2017/04/28	日本経済新聞電子版ニュース、日経速報ニュースアーカイブ、日経プレスリリース、マイナビニュース、科学新聞
理化学研究所と共催！科学技術系オープンソースと、深層学習を学ぶ［クリーク・アンド・リバー社］	2017/05/10	PR TIMES
データ分析・活用でビジネスや産業の発展に貢献したプロジェクトや企業・団体を表彰する「データサイエンスアワード2017」のファイナリストが決定［一般社団法人データサイエンティスト協会］	2017/09/28	@PRESS
データ分析・活用による産業への貢献を表彰する「データサイエンスアワード」2017年の最優秀賞は産業能率大学総合研究所／東京地下鉄が受賞！［一般社団法人データサイエンティスト協会］	2017/10/24	@PRESS
人工知能が膨大な計算を3600分の1に削減 － 東大が転移学習を組み込んで成功	2017/11/15 ～ 2017/11/17	日本経済新聞電子版ニュース、日経速報ニュースアーカイブ、日経プレスリリース、マイナビニュース、科学新聞
賢くなるAIで物質内部の界面探査	2017/11/24	科学新聞
心不全患者個々の心血管イベント発生を予測可能な数学的モデルを作成－国循	2018/03/09	QLifePro 医療ニュース
富士通など、材料設計におけるAIの有用性を実証―高イオン伝導率の全固体リチウムイオン電池用固体電解質開発を効率化	2018/03/16	日本経済新聞電子版ニュース、日経速報ニュースアーカイブ、日経プレスリリース
AIで有機化合物探索　理研、深層学習で高速評価	2018/04/18	日刊工業新聞
心不全患者における心血管イベント予測因子をデータマイニングで特定－国循	2018/06/29	QLifePro 医療ニュース
東大と理研など、AIによる有機分子の設計とその実験的検証に成功－機能性分子の設計に道筋	2018/08/24	日本経済新聞電子版、日経速報ニュースアーカイブ、日経プレスリリース

報道タイトル	報道日	報道機関
東北大・産総研・理研、人工知能でタンパク質を自動設計	2018/08/31 ～ 2018/09/04	日本経済新聞電子版、日経速報ニュースアーカイブ、日経プレスリリース、マイナビニュース、QLifePro 医療ニュース
東大、人工知能でスペクトルを従来の約 2 万倍の速さで「解釈」する新手法を開発	2018/09/06 ～ 2018/09/14	日本経済新聞電子版、日経速報ニュースアーカイブ、日経プレスリリース、科学新聞
ベンチャー探訪、分子動力学計算による化合物デザイン技術が強みのモジュラス、自社パイプラインを構築してまずは IND までを目指す	2018/09/24	日経バイオテク
材料開発、AI で脱・職人技、東工大、ロボと実験反復、理研、分子構造から探索。	2018/10/13 ～ 2018/10/15	日本経済新聞電子版、日経速報ニュースアーカイブ、日本経済新聞朝刊
ナノポアセンサと AI 技術でインフルエンザウイルス型を高精度に識別－阪大	2018/11/28 ～ 2018/11/30	QLifePro 医療ニュース、科学新聞
インフルエンザウイルスを 1 個レベルで認識するナノバイオデバイスを開発	2019/02/04	MONOist
マイクロ RNA、高感度で検出、尿でがん早期発見、名大発イカリア、21 年実用化、術後の再発チェック薬に。	2019/03/26	日経産業新聞
AI 素材探索の MI-6、数千万円を資金調達	2019/04/24	日経速報ニュースアーカイブ 、日本経済新聞電子版
【資本参加】千葉功太郎氏(コロプラ元副社長)ら	2019/04/24	レコフ M&A 情報
AI 素材探索の MI-6、開発強化へ数千万円調達。	2019/05/09	日経産業新聞
有用物質、AI の手で、理研や物材機構、合成に成功(NextTech 2030)	2019/05/10	日本経済新聞電子版、日経速報ニュースアーカイブ、日経産業新聞
【資本参加】リバネス	2019/06/03	レコフ M&A 情報
産総研－NEC、AI で化学プラント早期復旧、対処手順を的確導出	2019/07/26	化学工業日報
「2030 年のまち」提案―新興企業、見守り・農業…課題解決(CEATEC2019)	2019/10/11	日経産業新聞
MI 専業の MI-6、素材開発の時流捉え成長、現場目線のソフトも	2019/10/31	化学工業日報
投資 AI と人間(2)株にも「モテ期」あり―国産 AI、ヤフーで測る。	2019/11/06 ～ 2019/11/07	日本経済新聞電子版、日本経済新聞朝刊
特集：AI 人工知能から人工知性へ	2020/01/01	日経サイエンス
阪大、複雑な分子をフロー合成、機械学習で最適化	2020/03/16	化学工業日報
モジュラス株式会社への出資を決定［東大 IPC］	2020/05/13	PR TIMES

報道タイトル	報道日	報道機関
「例外」を発見するAI「BLOX」の開発－AIを用いた革新材料の開発に新たな道筋－［横浜市立大学］	2020/05/28	Digital PR Platform
理研、特殊な物性探すAI、研究者の予想外物質発見も。	2020/06/11	日経産業新聞
JSTと阪大、神経伝達物質を単一分子レベルで検出・識別する手法を開発	2020/07/10	日本経済新聞電子版、日経速報ニュースアーカイブ、日経プレスリリース
【新刊案内】材料およびプロセス開発のためのインフォマティクスの基礎と研究開発最前線 監修：船津 公人(東京大学大学院教授)発行：(株)シーエムシー・リサーチ［CMCリサーチ］	2020/08/03	PR TIMES
NIMSと東大、機械学習により薄膜作製プロセスの高速化を実現	2020/10/21	日本経済新聞電子版、日経速報ニュースアーカイブ、日経プレスリリース
AI 技術とナノポアセンサでウイルスの複数種識別に成功！ ――一回の検査で複数のウイルス、感染症の原因特定に期待［大阪大学］	2020/11/10 ～ 2020/11/12	Digital PR Platform、QLifePro医療ニュース

2020年3月30日検索
2020年11月16日更新
2021年2月28日確定

2.2 プロジェクトの進捗状況

終了後のグループの発展状況を以下にまとめた。

2.2.1 離散構造処理の基盤研究

ZDD技術の改良・発展として、論理関数を表すデータ構造ZSDDとZd-DNNFを開発、厳密被覆問題に使われるKnuthのDLXのZDDによる改良等を研究した。また、ZDDの適用範囲の拡大として、ベイジアンネットワーク、バンディット問題、ナップザック問題等への適用拡大を行った。さらに、通信ネットワークシステムの信頼性設計、配電ネットワーク設計、ホットスポットクラスター検出等の応用を研究した。

(1)ZDD技術の改良・発展

ZDD技術の終了後の改良・発展として、論理関数を表すデータ構造としてZSDDとZd-DNNFの開発、KnuthのアルゴリズムDLXのZDDによる改良について以下にまとめた。

①ZSDDとZd-DNNF(NTTと共同)

BDDと同じく論理関数を表すデータ構造としてSentential Decision Diagram(SDD)が2011年に発表された[64]。SDDはBDDの一般化であり、BDDより小さく論理関数を表現できる

[64] Darwiche A., “SDD: A new canonical representation of propositional knowledge bases”, In IJCAI, 22, 819-826 (2011).

ことや、BDD で行える様々な操作が SDD でも行えることが分かっている。BDD と SDD の違いは次の通りである。

BDD では変数を 1 つ選び、それが 0 か 1 で場合分けする操作(Shannon 分解)に基づいて論理関数を分解したが、SDD では Shannon 分解を一般化した(X、Y)分割[14]によって論理関数を再帰的に分割する。(X、Y)分割は、変数の集合 X を選び、X に属する変数がどのような値を取るかで場合分けする操作である。左側の変数を prime、右側の変数を sub と呼び、prime による場合分けを行う。prime が「空の場合を含まず、互いに排反かつすべてを満たしている」条件を満たしている場合を(X、Y)分割と言う。(X、Y)分割において、X がちょうど 1 つの変数からなる場合が Shannon 分解に対応している。

また、BDD では変数を 1 つずつ選んで場合分けしていくので、変数順序は線形順序であった。一方、SDD の変数順序は vtree という木構造で表される。vtree は根付き二分木で、葉が論理変数と 1 対 1 に対応する。但し、vtree は順序木なので、左の子と右の子を区別する。SDD を作るときは、vtree を根からたどって、今いる vtree の頂点の左の子の部分木の葉にある変数たちを X、右の子の部分木の葉にある変数たちを Y として(X、Y)分割を構成し、それを再帰的に繰り返す。例として、論理関数 f=(A∧B)∨(B∧C)∨(C∧D)を表す SDD と、その変数順序を定める vtree を図 2-3 に示した。

図 2-3(a)は vtree を表していて、葉には対応する変数、葉以外の頂点には頂点番号を記している。図 2-3(b)は丸ノードが 1 つの(X、Y)分割を表しており、四角形が(X、Y)分割の要素を示している。左が prime、右が sub に対応している。また、丸ノードの整数は対応する vtree の頂点番号である。

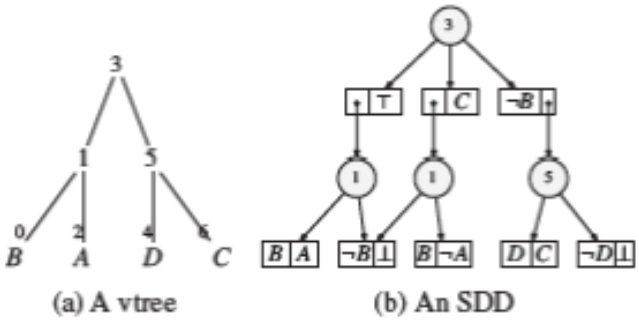


図 2-3　(a)変数順序を定める vtree、(b)論理関数
f=(A∧B)∨(B∧C)∨(C∧D)を表す SDD[14]

図 2-3(b)の SDD が f=(A∧B)∨(B∧C)∨(C∧D)を表していることを確かめてみる。まず、下の段の丸ノードから読むと、一番左の丸ノードはラベルが 1 なので、vtree の頂点 1 に対応している。vtree の頂点 1 は左の子が B、右の子が A なので、({B}、{A})分割を作れと言っている。結局、一番左の丸ノードは(B∧A)∨(¬B∧⊥)=A∧B を表している。同様に、中央

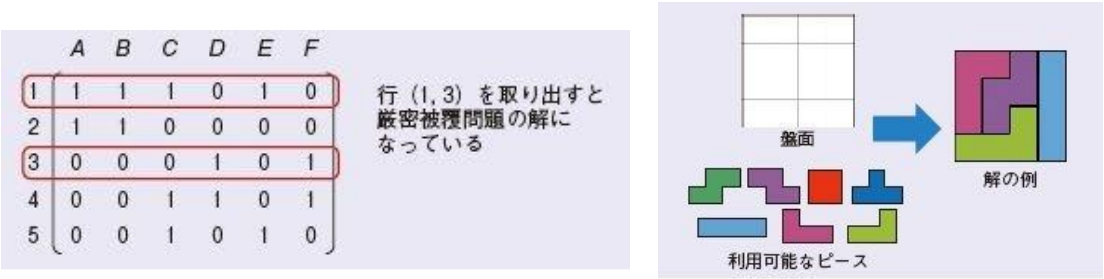
の丸ノード(ラベル 1)は(¬B∧⊥)∨(B∧¬A)=¬A∧B を表している。次に、一番右の丸ノードはラベルが 5 なので、vtree の頂点 5 に対応している。vtree の頂点 5 は左の子が D、右の子が C なので、({D}、{C})分割を作れと言っている。結局、一番右の丸ノードは(D∧C)∨(¬D∧⊥)=C∧D を表している。最後に、一番上の丸ノードはラベルが 3 なので vtree の頂点 3 に対応している。vtree の頂点 3 は左の子の部分木に変数 A、B があり、右の子の部分木に変数 C、D があるので、({A、B}、{C、D})分割を作れと言っている。結局、一番上の丸ノードは((A∧B)∧⊤)∨((¬A∧B)∧C)∨(¬B∧(C∧D))を表している。これは確かに f=(A∧B)∨(B∧C)∨(C∧D)の({A、B}、{C、D})分割になっている。結局、一番上のノードは変数の集合 X={A、B}に対し、X が A∧B であるか、¬A∧b であるか、¬B であるかで場合分けしていた訳である。

ZSDD は以上の SDD の主要な部分、(X、Y)分割、vtree による変数順序は同じにして、終端の表現や部分グラフ枝刈りの方法を変更した。その結果 ZSDD は、厳密なサブセットとして ZDD を包含しており、少ないブール関数を扱うときには ZSDD は SDD より簡潔である。ZSDD は SDD や ZDD よりコンパクトな代替手段として使用できることを示した。その後、特定のグラフのサブ構造を表す ZSDD について、トップダウンのコンパイルアルゴリズムを開発した(一種のフロンティア法)[14]。結果の ZSDD は同じグラフのサブ構造を表す ZDD より小さいことを実験的に確認した。

さらに、ZSDD のスーパーセットである Z-d-DNNF を開発した。Zd-DNNF は ZSDD の制限を緩和し、トップダウンコンパイルに必要な複雑な手順を改善した[15]。制限を緩和したが、モデルのカウントやモデルの列挙などの重要な探索については支障なく行えることを確認した。ZSDD に代わりにこの表現を使用して、s-t 単純なパスに対して、より効率的なコンパイルを実現できた。

②Knuth のアルゴリズム DLX の ZDD による改良

各種パズルの中で、厳密被覆問題という分野が知られている(図 2-4(a)(b))。その他、現実問題では、マンションの適切な間取りを考える問題、電子回路基板上で部品の配置を考える問題等が挙げられる。



(a) 厳密被覆問題の例　　(b) ポリオミノの例(テトロミノ)

図 2-4　厳密被覆問題とポリオミノの例[15]

この問題のすべての解を求める方法として、DLX 法が知られていた[65]。DLX 法は、すべての答えの候補を順番に並べていく総当りの探索によって厳密被覆問題を解く方法である。この方法は答えの数が多くなると膨大な時間が掛かる欠点がある。

本研究では ZDD の仕組みを使って、発見した解をすべて覚えておく。このようにすると、それを新しい解のヒントとして使うことができ、探索を大幅に高速化することができる[16],[17]。図 2-5 に ZDD による解の集合の表現を示す。この図では、ZDD 上では根ノードから終端に達する経路が 2 種類あり、それぞれがテトロミノの 1 つの解を表している。ZDD にはそれぞれの解が対応する経路の共通部分を圧縮できる特徴がある。この特徴によって膨大な数の解の集合を小さなグラフで表現することができる。その結果、例えば、盤面の大きさが 10×10 枡のテトロミノ敷き詰め問題には 7 京 2713 兆 5606 億 4890 万 6621 個の解が存在する。これを ZDD で解の集合を表現すると、1647 万 6396 頂点からなるグラフとして表現することができ、コンピュータ上では数百 MB 程度の大きさになる。この方法で DLX 法に対して最大 1 万倍高速に厳密被覆問題を解くことができた。

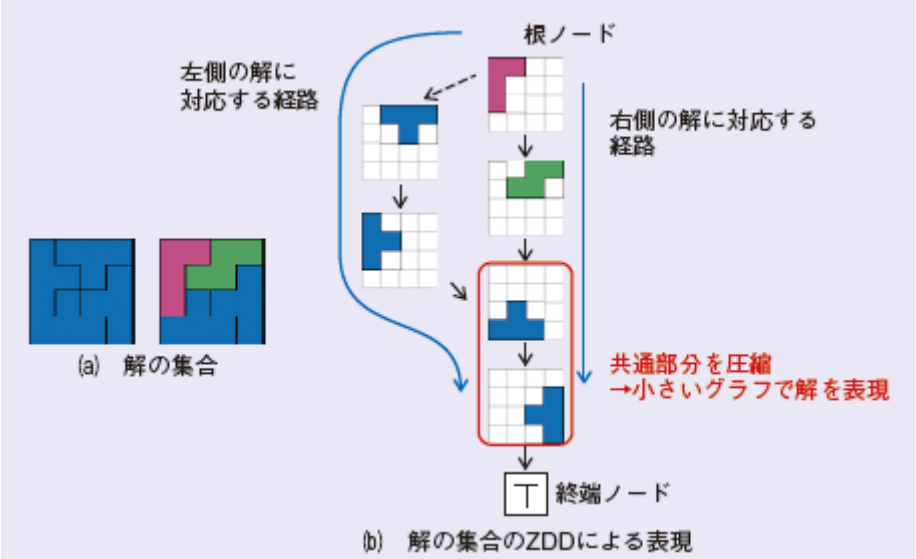


図 2-5　ZDD による解の集合の表現[17]

(2) ZDD の適用範囲拡大

ZDD の適用範囲について、ベイジアンネットワーク、バンディット問題、ナップザック問題等への適用拡大を行った。

[65] Knuth D.E.,"Dancing Links", Millenial Perspectives in Computer Science, 187- 214 (2000)

①ベイジアンネットワーク

ベイジアンネットワーク(BN)は、因果関係を確率により記述するグラフィカルモデルの1つで、複雑な因果関係の推論を有向非巡回グラフ構造により表すとともに、個々の変数の関係を条件つき確率で表す確率推論のモデルである。ベイジアンネットワークによる推論はNP困難問題であることが知られている。BNをZDDにコンパイルして効率的で正確な推論を実行する試みを行った[66]。ZDD を使用した正確な推論の計算時間は、ZDD のサイズによる線形時間に短縮される。また、キャッシュメモリ技術により、推論をさらに時間短縮できる。ただし、BNのサイズが大きくなるにつれて、BN の計算時間が増大するとともに、ZDDのサイズが大きくなり、ZDDのコンパイルができなくなる問題があった。

本研究では、従来のZDDを使った方法の改善のために、BNを分割して個別にZDDにコンパイルする方法を提案した[18]。BNの分割には従来から知られていたd-分離[67]を使う。図2-6にd-分離の規則を示した。

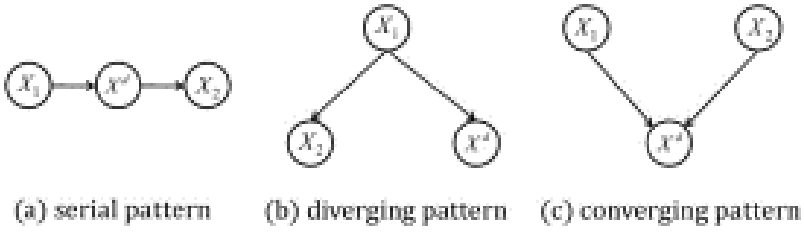


図2-6　d-分離の規則。(a)(b)がd-分離、(c)d-結合[18]

図2-6(a)のserial　patternのd-分離によって分離して、条件付きではあるが、BNを独立した部分に分割する。ネットワーク全体の巨大なZDDを生成するより効率的である。しかし、BN の分割が非常に多くなると、そのためにかなりの時間を消費する可能性がある。本研究では、BNの分割に使用するd-分離を1頂点に制限し、Tarjanの頂点カットアルゴリズム[68]を使用するとBN頂点の数を線形時間で実行できることが分かった。実験では、1頂点のd-分離が殆どのBNに存在することを示しており、BNを1頂点のd-分離で分離すると、従来のZDDの方法よりコンパイルと推論の両方の速度が大幅に向上することが分かった(表2-7)。

---

[66] Minato, S.-L., Ken, S., Taisuke, S., "Compiling Baysian networks by symbolic probability calculation based on Zero-suppressed BDDs", IJCAI 2007,2550-2555 (2007)

[67] Geiger, D., Verma, T., Pearl, J., "Identifying independence in Baysian networks", Networks, 20(5), 507-534 (1990)

[68] 多数のサイトがある。例えば、
https://inzkyk.xyz/algorithms/depth_first_search/strong_components_in_linear_time/

表 2-7　従来の ZDD と本研究の手法の比較[18]

Dataset	Conventional ZDD-based Method [Minato 07]				Proposed Method					
	ZDD Size	Compil-ing(ms	Mar. (ms)	Arb. (ms)	ZDD Size	Compil-ing(ms)	Mar. (ms)	Arb. (ms)	Time (*d*-sep) (ms)	No. of *d*-sep
ALARM	34,299	57	9	651	10,291	26	2	106	15	5(27,5,3,2)
WIN95PTS	26,477	104	15	231	27,397	107	7	79	6	4(71,3,2)
HEPAR2	51,000	126	12	885	40,308	77	6	281	15	7(48,9,5,4,2)
HAILFINER	294,608	467	39	3,330	188,088	317	33	1,753	28	6(45,4,3,2)
PATHFINDER	31,549	2,633	14	483	31,553	2,621	20	461	3	1(108,2)
MILDEW	15,310,511	664,097	17,545	522,670	15,310,711	688,067	8,786	225,250	1	1(34,2)
WATER	25,629	874	12	175	25,629	874	12	175	1	0
PIGS	73,715	517	28	97	48,588	509	23	69	198	41 (324,13,8,6,4,3)
BARLEY	time out	time out	time out	time out	time out	time out	time out	time out	0.52	2(46,3)
DIABETES	time out	time out	time out	time out	time out	time out	time out	time out	23	2(409,3)
WIN95PTS (Problem1)	24,460	84	7	248	5,299	62	4	46	1	3(26,8,3,2)
BARLEY (ngtilg)	1,091,779	4,360	354	461,975	1,089,670	4,049	171	222,881	55	2(15,4,2)

②バンディット問題

バンディット問題とは、複数のアームと呼ばれる候補から最も良いものを逐次的に探す問題である。アームとはスロットマシンの比喩から来ている。本研究では敵対的バンディット組み合わせ最適化(BCO)の問題を検討した。敵対的バンディット問題とは、報酬が何らかの確率分布に従うことを仮定しない。そのため時間変化があるバンディットのような、「報酬が確率的ではあるが非定常的」な設定も扱える。しかし、高い報酬を達成できるとは限らない。このような性質があるため、刻々と変化するオンライン最適化問題の一種ということになる。

組み合わせバンディット問題を簡単に説明する。K 本のアーム $i\in[K]$が存在し、さらにその組み合わせであるスーパーアーム $X\in 2^{[K]}$を考える。X は $2^{[K]}$の組み合わせの中の 1 つであるとする。各時刻 t=1、…、T でスーパーアーム X を選択する。その時に得られる利得を考える。時刻 t にアーム $i\in K$ から得られる利得 $r_t(i)$とすると、時刻 t にスーパーアーム $X\in 2^{[K]}$から得られる利得 $r_t(X)$は各 i から得られる利得の和$\Sigma_{i\in X} r_t(i)$となる。個々の利得 $r_t(i)$は観測できず、その組み合わせである時刻 t で選んだスーパーアーム X_t による利得 $r_t(X_t)$のみしか観測できないとする。このように定式化すると、組み合わせバンディット問題は次式の R_T を最小にするスーパーアームを選ぶ問題となる。

$$R_T=\max_{X\in[T]} \Sigma_{t\in[T]} r_t(X) - \Sigma_{t\in[T]} r_t(X_t)$$

この式の第 1 項はすべて最善のスーパーアームを選んだときの総利得であり、理論的な利得最大値になる。第 2 項は実際の総利得である。

本研究では ZDD でスーパーアーム集合を表現し、その ZDD でサンプリングと共起確率を計算する手法を開発した[19]。ZDD によりスーパーアーム集合を効率的に表現することがで

き、ZDD 上で最適化を実行することによって、既存の方法より数百倍高速であることを実験的に確認した(図 2-7)。

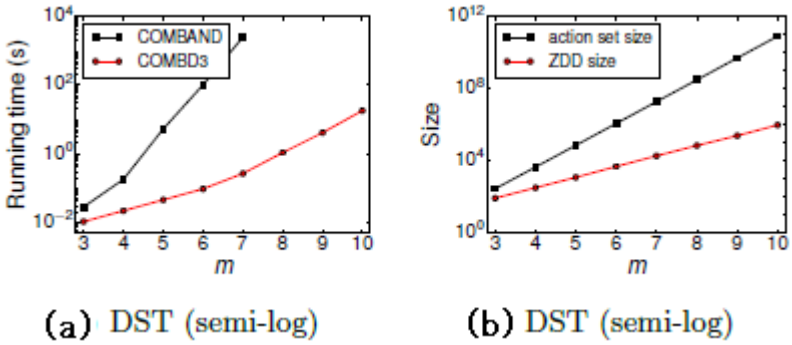


図 2-7　敵対バンディット問題でのソフトウェアによる性能比較。DST は人工的に作られた例題(3m ノード)。(a)COMBAND は既存、COMBD3 は開発したソフトウェアにおける計算時間。(b)データサイズの比較。[19]

③ナップザック問題

ナップザック問題は様々なバリエーションがあり、最も単純なナップザック問題を ZDD の手法で解くことはすでに実現されている[69]。0-1 多目的ナップザック問題は、複数の目的関数を同時に良くするような解を求めるもので、近年の研究目標の 1 つである。

0-1 多目的ナップザック問題について、そのすべてのパレート最適解[70]を列挙するアルゴリズムとして ZDD データ構造を使用した動的計画法[71]を開発した[20]。このアルゴリズムでは、Bazgan らによって提案された方法[72]により ZDD を使用して、実行可能なすべての解をコンパクトに格納し、本質的でない部分的な解を可能な限り迅速に刈り込む方法である。アルゴリズムの出力として、すべてのパレート最適解にインデックスを付ける便利な ZDD を取得できる。実験結果を表 2-8 に示した。問題の種類 3 タイプについて、要素 n の数を変えて計算した。目的関数 m が 2 個では Bazgan の方法と大差ないが、m が 3 個と 4 個に対して、計算時間が短く高速であった。

---

[69] Minato S.-I., "Zero-suppressed bdds and their applications. International Journal on Software Tools for Technology Transfer", 3(2), 156-170 (2001)

[70] 資源が無駄なく配分された状態のこと。「資源配分する際、誰かの効用(満足)を犠牲にしなければ、他の誰かの効用を高めることができない状態」のことをいう。

[71] 動的計画法(Dynamic Programming、 DP)は、対象となる問題を複数の部分問題に分割し、部分問題の計算結果を記録しながら解いていく手法を総称してこう呼ぶ。

[72] Bazgan C., Hugot H., Vanderpooten D., "Solving efficiently the 0-1 multi-objective knapsack problem", Computers & OR, 36(1), 260-279 (2009)

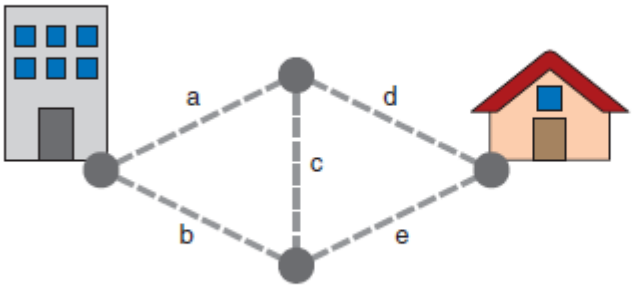
表 2-8　平均計算時間[20]

			Bazgan's				Proposal			
Type	n	m	random	$\mathcal{O}^{max}$	$\mathcal{O}^{pot}$	$\tilde{\mathcal{O}}^{pot}$	random	$\mathcal{O}^{max}$	$\mathcal{O}^{pot}$	$\tilde{\mathcal{O}}^{pot}$
	200	2	5.31	**3.05**	4.32	9.79	10.78	5.81	**4.35**	12.96
1	100	3	**18.42**	19.64	19.17	59.57	15.98	11.85	**7.59**	28.92
	70	4	**19.20**	29.03	21.89	81.88	9.84	7.34	**5.59**	30.77
	150	2	7.16	6.15	**6.02**	8.42	11.86	10.37	**6.63**	9.99
2	60	3	**10.81**	16.06	12.94	20.47	5.84	7.78	**3.69**	10.12
	50	4	**70.17**	105.25	**70.17**	126.93	17.12	25.08	**9.71**	40.93
	120	2	18.37	**9.24**	11.60	9.89	35.31	19.55	18.82	**18.22**
3	50	3	52.91	40.81	41.47	**36.29**	31.30	23.38	**20.95**	22.83
	35	4	33.39	29.91	29.26	**26.27**	7.50	6.43	**5.96**	6.71

(3) 各種応用の実現

①　通信ネットワークシステムの信頼性設計[21], [22]

現代においては、通信ネットワークは重要な社会基盤であり、故障なく動作しなければならない。しかし、通信ネットワークのリンクやノードにおける部品が故障するのは不可避である。そのような場合でも機能が存続するように通信ネットワークは設計されなければならない。図 2-8 は通信ネットワークの簡単な例である。記号 a, b, …, e は設置するリンクの候補を表している。リンクには故障の確率とリンクのコストが分かっているものとする。掛けられるコストを決めたとき、最も信頼できる通信ネットワークをどのようにすべきかという問題を考える。



Link	Fail probability	Cost
a	5%	2
b	8%	3
c	9%	3
d	7%	3
e	6%	1

図 2-8　コストを含んだ通信ネットワークの簡単なモデル[22]

この問題を考える時、最も直接的な考え方は、一定コストを決め、それを満足するリンクの組み合わせをすべて数え上げ、それぞれの組み合わせの故障の確率を計算するという方法である。この方法には 2 つ問題がある。1 つは通信ネットワークのサイズが大きくなると、コストを満足する組み合わせが指数関数的に増えることである。もう 1 つは、候補の通信ネ

ットワークの故障の確率を計算する時間も指数関数的に増加することである。そのため、この方法は 20 以下のノードにしか使えない。

本研究では、通信ネットワークにおける故障のパターンを圧縮した BDD によって表現する方法を開発した。図 2-9(a)に示した簡単な通信ネットワークにおいて、同図(b)にリンクが繋がっていながら故障があるパターンのすべてを列挙した(16 パターン)。このパターンについて BDD を使って圧縮したのが同図(c)である。BDD の図においては、リンクを BDD のノードとして表現し、リンクの有無を 1 と 0 に割り当て、1 の場合は実線、0 の場合は点線で次のノードに進む。同図(b)の緑の円で表された故障パターンは、同図(c)の BDD においては緑の道筋で表されている。16 の故障パターンは BDD においては 10 のノードに減少している。BDD 表現の特性によって、故障パターンの共通部分が省略できたためである。

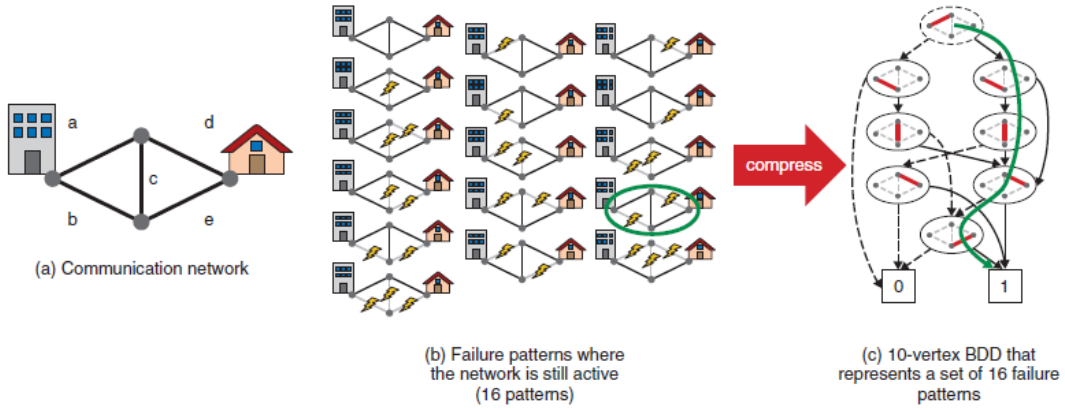


図 2-9　故障パターンを表示するために BDD を使う[22]

まず、BDD を使用することにより、BDD 頂点の数に比例した時間でネットワークの信頼性を正確に評価できる。BDD の圧縮率が 99%の場合、ネットワークの信頼性を評価するには、元の計算コストの 1%に減らせる。次に、BDD を使用して、ネットワークにリンクを追加することで達成できる改善の量を見積もることができる。このような見積もりにより、高い信頼性が達成できそうもない候補ネットワークを効率的に捨てることができる。これらにより、最適なソリューションを見つけるために必要な計算コストが削減できる。

②配電ネットワーク設計[23]

期間中に研究したスマートグリッドにより、ZDD を使って、変電所設備 72 ヶ所、スイッチ数 468 個からなるスマートグリッド構成において、与えられた電力品質条件を満たしつつ、かつ送電損失を最小化する最適構成を得ることができた。太陽光発電や風力発電を含んだ分散型発電(DG)の導入においては、できる限り装置や送電ラインの変更・追加を行わずにできることが望ましい。配電系統の容量を増やすには、DG の配置とネットワークの再構成、

すなわちスイッチの再構成を適切に行う必要がある。DG 配置の決定には電力の流れの非線形性が含まれ、スイッチの数 n が増加するにつれて、再構成の解の候補数が $O(2^n)$ の次数で指数関数的に増加するため、計算上難しい問題である。

本研究は、前述の期間中の研究を発展させたものであり、ZDD 構造を使って巨大なスイッチの組み合わせを圧縮した方法で表現し、それを使って最適な解を見つける研究である。本研究では、ルートセクション(つまり、給電点に隣接するセクション)と非ルートセクションの電力損失を個別に計算できることを前提とする。ただし、電力系統の容量はルートセクションと非ルートセクションに依存するため、このような独立性を想定することはできない。たとえば、ルート以外のセクションのみを考慮した DG 配置では、ルートセクションでの過負荷や、ルートセクションでの未知の電圧上昇による電圧障害が発生する可能性がある。ルートセクションと非ルートセクション間の相互作用を処理するために、ZDD ベクトル(ZDDV)と呼ばれる ZDD の拡張データ構造を利用する[73]。ZDDV は、構造と値をコンパクトに保持でき、それらの間で効率的な数値算術演算を実行できる。ZDDV を使用して、ルートセクションと非ルートセクションの統合構造とその電力系統の容量の両方を表す。

図 2-10 に実験に使った 432bus ネットワークを示す。ネットワークの放射状の機能を使用して、ネットワークごとの問題をツリーごとのサブ問題のセットに分解し、セットごとにすべてのサブ問題を解決し、次に、ZDDV を使用して解決されたサブ問題の組み合わせを集約し、ZDDV の数値算術演算を使用して最適なネットワークを見つける。本研究は早稲田大学教授林泰弘と共同で研究が行われた。

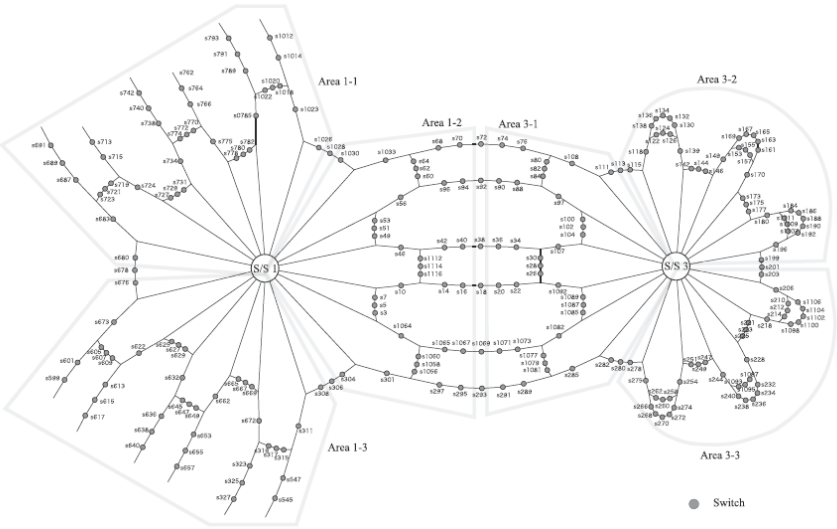


図 2-10　実験に使った 432bus ネットワーク[23]

[73] Minato S.-I., Somenzi F., "Arithmetic boolean expression manipulator using BDDs.", Formal Methods in Syst. Des., 10, 221-242 (1997)

この結果、列挙型計算により最適なDG配置を同時に考慮する構成の最適化手法を開発した。実験では、235個のスイッチを持つネットワークのサイズを計算できた。9.90×10^{29}の構成可能性数が存在する場合でも、解を処理できた。また、DG 接続要求率が 50%の場合、17.1時間で1.49×10^{18}の最適構成が得られた。

③ホットスポットクラスター検出[24]

ホットスポットとは、感染症の発生が著しい場所や災害がひどい場所を表す言葉である。このような場所を検出することは基本的で重要なことである。都道府県や都市などの地理的領域によって分割された統計データを使って、集中的に発生するクラスター領域を検出するホットスポットクラスター問題は一般的で古典的な問題の 1 つである。従来からホットスポット領域を評価するために地域統計を使った研究が多く行われている。例えば、表2-9に一例を示す。

表2-9　地域統計データの一例[24]

Region-ID	Population	Occurrence
1 (Hokkaido)	5,401,210	1,004
2 (Aomori)	1,338,465	279
3 (Iwate)	1,289,470	322
4 (Miyagi)	2,324,466	443
...	...	...
47 (Okinawa)	1,461,231	258

ホットスポット抽出の既存研究では、集中度の指標として「スキャン統計量」と呼ばれる尤度比[74]が広く用いられている。尤度比は次の式で表される。

$$\lambda(Z) = \begin{cases} \dfrac{\left(\frac{c(Z)}{n(Z)}\right)^{c(Z)} \left(\frac{c(Z^c)}{n(Z^c)}\right)^{c(Z^c)}}{\left(\frac{c(G)}{n(G)}\right)^{c(G)}}, & \text{if } \dfrac{c(Z)}{n(Z)} > \dfrac{c(Z^c)}{n(Z^c)} \\ 1, & \text{otherwise.} \end{cases}$$

Z：対象領域　Zc：非対象領域　G：全体領域

n(Z)：対象領域の総人数　c(Z)：対象領域の現象数

尤度比を最大化する領域の組み合わせを求めることによって、最も可能性の高いホットスポットの組み合わせを求めることができる。すなわち、λ(Z)を最大化する領域の組み合

[74] 尤度比(ゆうどひ)とは、尤度(検査における感度や特異度など)の比であり、比率として実数で表す。なお、尤度(なりやすさ、起こりやすさ)は確率であり、通常は比率として0～1で表すが、%として0%～100%で表す場合もある。

わせが目標である。従来は、λ(Z)ではなくlogλ(Z)を計算して最大化していた。計算コストはそれほど大きくないが、式には累乗や対数などの非線形演算が含まれているため、機能的動作の分析は簡単ではない。

実際の計算では、すべての領域のデータを「使う/使わない」に場合分けし、網羅的にスキャン統計量を計算して、指定した閾値を上回る領域の組み合わせをすべて出力する。その際、まず、十分大きな閾値で探索して統計量の最大値を求める。その後、それより少し小さい閾値で探索すると現実的な個数の解が生成される。しかし、大きな閾値で解なしの場合でもそのまま探索すると指数関数的な時間がかかる。

この課題に対して、本研究において、スキャン統計量には部分的な単調性が存在することを見つけた。すなわち、n(Z)>n(Z')かつc(Z)<c(Z')ならばλ(Z)＜λ(Z')が成立する。その結果、探索の各ステップでn(Z)とc(Z)を記録しておき、両者の数値を参照することで、安全な枝刈りができることが分かった。その結果、47都道府県の2016年の自殺者データで実験した結果、枝刈りをしない方法で探索すると、40日程度かかる計算を0.05秒に削減でき、100万倍以上の高速化となる。もう1つのノースカロライナ100郡の乳児突然死での実験では、対数尤度比の閾値67.7で2件ヒットし、計算時間0.51秒であった。対数尤度比の閾値を67.0に下げるとほぼ同じ計算時間で1582件ヒットした。100領域規模の例題で、素朴な計算だと 1 兆年かかるレベルの計算を世界で初めて現実的な時間で厳密最適解を求めることができた。

2.2.2 機械学習・制約充足応用の発展

LAMPソフトウェアの拡張を図り、ゲノム関連解析のためのLAMPLINK、並列処理を可能にしたMP-LAMPを開発した。また、マテリアル・インフォマティクス、バイオ・インフォマティクスの分野への研究に発展し、ベイズ最適化の応用として熱電材料の設計に応用された。また、医療品の処方や製造方法の最適化にも適用した。さらに、モンテカルロ木探索によって材料探索ができるMDTSソフトウェアを開発し、熱伝導の最適化問題やナノフィルムでの界面熱輸送の問題に適用した。バイオ・インフォマティクス分野のRNAの逆折りたたみのアルゴリズムMCTS-RNAを開発した。さらに、有機材料の自動設計アルゴリズムChemTSを開発し、公開した。さらに、分子の動きの軌跡を検索するアルゴリズムTS-MDの開発にも成功した。

(1)LAMP法の発展

LAMP法の拡張であるLAMPLINK[25]とMP-LANP[26]を開発した。

LAMPLINKはゲノムワイド関連解析(Genome-wide association studies(GWAS))[75]における、個々の一塩基多型(SNP)[76]のスクリーニングを行う際に使う手法として開発した。スクリーニングを行うとき、SNP の組み合わせ効果が重要である。そのためゲノム関連研究で広く使われている PLINK に LAMP を組み込んだ LAMPLINK を開発した。LAMPLINK は、フィッシャーの直接確率検定[77]またはカイ 2 乗検定[78]を使用して、ゲノム関連データの特定の表現型に関連する統計的に有意の組み合わせを列挙する。その時、LAMP 法の特徴である出現頻度の低い組み合わせを除去することによって、意味のある組み合わせの列挙を可能にした。

これらの理論上の利点にもかかわらず、ゲノムワイド関連解析への LAMP の適用はまだいくつかの研究に限定されていた。マイナー対立遺伝子頻度(Minor Allele Frequency、MAF)[79]の高い SNP がデータに含まれている場合、LAMP はかなり多くの計算時間を必要とすることが知られている。その問題の 1 つの解決法としては、複数のコンピュータを並列に使うことである。最近使われだした Amazon Web Service(AWS)などのクラウドプラットフォームや企業内のコンピュータ群に展開できるように改善したのが MP-LAMP である。AWS は並列処理フレームワークとして Map-Reduce というソフトウェアを提供しているが、アルゴリズムは不均衡なツリー検索を使用して設計されているため、LAMP を Map-Reduce だけで並列化することはできない。そのため、新たな並列処理のソフトウェアを実装したのが MP-LAMP である。2 つの大きなデータセットで実験した結果、MP-LAMP を超並列クラスターで実行すると、CPU コアの数に比例したスピードアップが実現できた(図 2-11)。

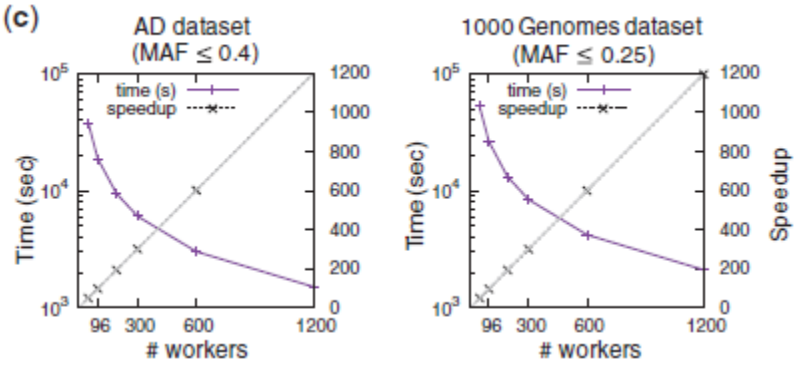


図 2-11　クラウドクラスターを使ったときの個々の計算(workers)の増加に伴う実行時間とスピードアップの実現[26]

[75] ゲノム全体をほぼカバーするような、50 万個以上の一塩基多型(single nucleotide polymorphism: SNP)の遺伝子型を決定し、主に SNP の頻度(対立遺伝子や遺伝子型)と、疾患や量的形質との関連を統計的に調べる方法論である。

[76] ある生物種集団のゲノム塩基配列中に一塩基が変異した多様性が見られ、その変異が集団内で 1%以上の頻度で見られる時、これを一塩基多型(Single Nucleotide Polymorphism:SNP)と呼ぶ。

[77] (主に)2 行 2 列の分割表(クロス集計表)におけるふたつの要因が互いに独立であるかどうかを推定する検定法。

[78] 検定統計量が漸近的にカイ 2 乗分布に従うような統計的検定法の総称である。

[79] 遺伝子型間の頻度の違いは、マイナー対立遺伝子頻度(Minor Allele Frequency, MAF)と呼ばれる指標で表わされる。MAF は集団における遺伝子の頻度であり、[0.05,0.5]の値を取る。

(2)マテリアル・インフォマティクス、バイオ・インフォマティクスへの展開

本研究はモンテカルロ木探索やベイズ最適化、機械学習等の手法をバイオ・インフォマティクスやマテリアル・インフォマティクスへ適用する研究へと進展している。材料設計や新たな材料発見を実現するために、候補となる多くの材料から目的とする性質を最適化するように材料組成やその構造を選択することを効率的に行うことが求められている。材料候補の数は材料の性質を決める変数に対して指数関数的に比例する可能性があるため、容易ではない。以下に、適用する手法別の成果を述べる。

①ベイズ最適化の応用

ベイズ最適化のために従来より効率的なプロトコルを設計し、COMBO(COMmon Bayesian Optimization library)と呼ばれるオープンソースのPythonライブラリとして実装した[27]。COMBOを利用した応用を以下に述べる。

原子グリーン関数とベイズ最適化を組み合わせた手法により、Si-Si、Si-Ge の界面熱コンダクタンス(ITC)を最小化および最大化することを目標に検討した[28]。その結果、6万を越える構造から数%の最適構造を見出すことができた。結論は想像を越え、衝撃的であった。最小の ITC は非周期的な超格子構造の場合であり、周期構造の最小値より 50%も小さかった。その結果、構造と ITC の関係が明らかになり、本手法が熱伝導に関するナノ構造の設計に有効であることが分かった。

また、ベイズ最適化と熱電磁界計算を交互に繰り返す手法により、超狭帯域波長選択性熱放射器を計算によって設計した[29]。特定の目的とする赤外線波長に対して、複数のコンポーネント(Si、Ge、および SiO_2)で構成される80億を超える多層膜の候補から最適な構造を効率的に特定することができた。得られた最適化された構造は、Q係数が273の高い鋭い放射率を示す非周期的な多層メタ材料であった。実際に設計された材料を作成し、Q係数が188で、最適な性能が実現できた。これは過去に経験的に設計および作成された構造物よりも高い性能であった。この成果によって、ベイズ最適化の有用性が実証できた。

さらに、ベイズ最適化に基づく結晶構造予測手法を提案した[30]。ベイズ最適化による結晶構造予測では、機械学習技術を使用して、より少ない数の検索試行で多数の候補構造から最も安定した構造を効率的に選択できた。ベイズ最適化とランダム検索を組み合わせた結晶構造予測を、NaCl や Y_2Co_{17} などの既知の結晶に適用した結果、最小構造を見つけるために必要な検索試行回数を、純粋なランダム検索と比較して計算コストを30〜40%と大幅に削減できた。

この手法は医薬品の処方や製造方法を最適化することにも適用された[31]。口腔内で崩壊する錠剤の処方とプロセスパラメータを最適化する目的で行った。医薬品の基準を同時に満たすためのスコアを定義し、パフォーマンス測定を使用して、実験した結果、ベイズ最適化は、実験の数を効率的に減らし、最適な定式化とプロセスパラメータを取得できた。その

結果、不要な実験の排除を実証しただけでなく、さまざまな条件のシミュレーションにより、依存関係を示すことができた。これは、実際の応用で役立つ。ベイズ最適化は、個々のスキルと経験への依存を減らし、最適化タスクの効率と有効性を高め、医薬品開発における処方と製造研究を促進することが期待される。

② モンテカルロ木探索(MCTS)の応用

バイオ・インフォマティクスやマテリアル・インフォマティクスの手法の1つとして、モンテカルロ木検索(MCTS)が挙げられる。モンテカルロ木探索を使った成果を以下にまとめた。

(ｉ)MDTSによる材料探索

モンテカルロ木探索によって材料の探索ができるオープンソースプログラム(MDTS)を開発した[32]。本研究では、2つのSiリード間のGe合金部分について熱伝導最大、最小になる合金を得ることを例に研究した。図2-12に本研究の実施例を示した。この図では原子の位置の数NはN=16である。

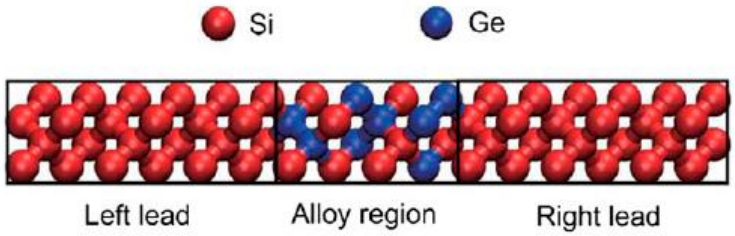


図2-12　2つのSiリード間のSi-Ge合金構造。この図は16原子の例[32]

各位置の原子SiあるいはGeのどちらを割り当てるかによって熱伝導の最適化の問題を解決する。原子の種別をxとし、f(x)を目標とする評価値とし、f(x)を最大にするxの組み合わせをMDTSによって求める。比較のためにベイジアン最適化による結果と比較したのが図2-13である。図2-13(a)の結果から、ベイジアン最適化の時間は、Nが増加するにつれて指数関数的に増加する。図2-13(b)ではN=22の2つの方法の効率は同等であり、N<22の場合、ベイジアン最適化の方が、N>22の場合はMDTSが効率的であった。

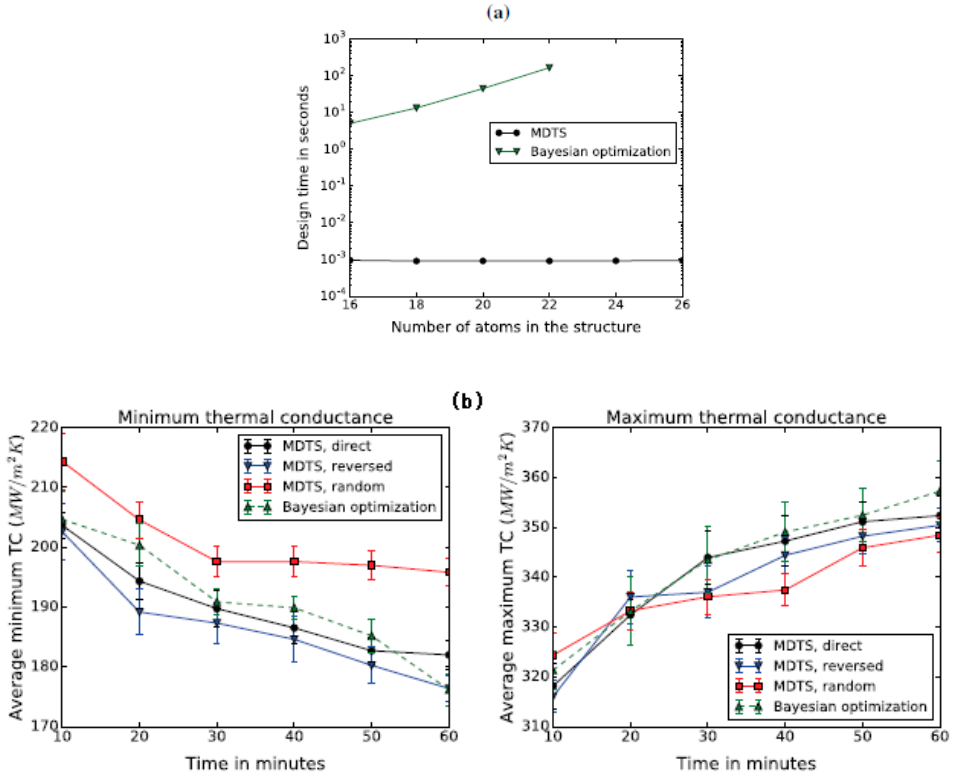


図 2-13　MDTS とベイジアン最適化による結果の比較。(a)N 個の原子数に対して候補の構造を選択するための設計時間。(b)設計時間とシミュレーション時間の両方を含む合計計算時間に対する最適な観測熱伝導率(最小および最大)(N=22)。[32]

(ⅱ)ナノフィルムでの界面熱輸送

また、同様な計算として Si と Ge ナノフィルムでの界面粗さの熱輸送の最適化の検討を行った[80]。この設定では、粗さはフィルムの層の数と粗さの程度になり、粗さの構成の総数は指数関数的に増加する(例えば、10 層と粗さの程度が 4 では総数は 1,048,576 個になる)(図 2-14)。

[80] Ju S., Dieb T.M., Tsuda K., Shiomi J., "Optimizing Interface/Surface Roughness for Thermal Transport(Poster)", Machine Learning for Molecules and Materials NIPS 2018 Workshop (2018).

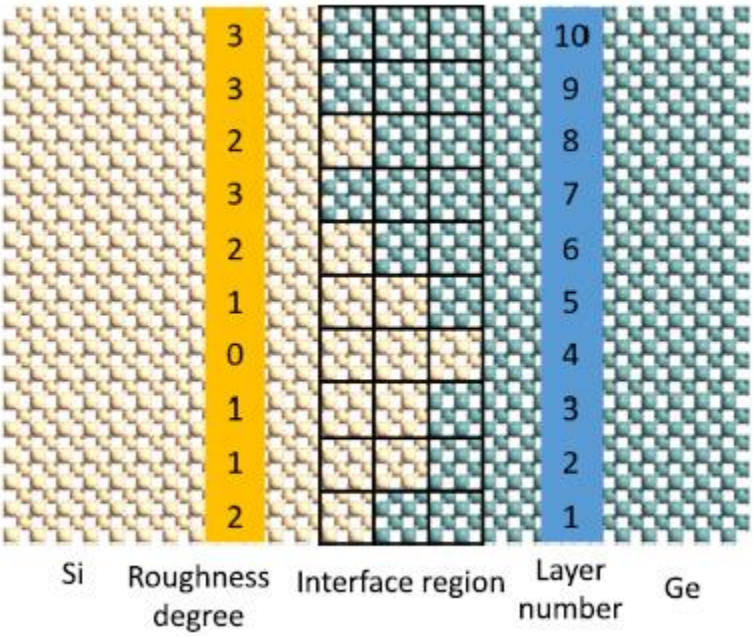


図 2-14　Si-Ge の 2 層ナノフィルムの界面粗さ。可能な構成の数は、層の数と粗さの程度とともに指数関数的に増加する。[33]

この検討の結果では、界面熱伝導率の最適なものは、平坦と非常に粗いものの中間範囲であった。この検討は大規模な検索範囲に適用できることを示すとともに、効率の良さが確認された。

(iii)RNA の逆折りたたみのアルゴリズム MCTS-RNA の開発[34]

モンテカルロ木探索の手法はバイオや化学の分野にも応用された。バイオへの応用として、RNA の逆折りたたみ[81]のアルゴリズムを開発した。RNA から転写した蛋白質等の転写物の機能は、RNA の主に二次構造[82]によって決定される 3 次元分子構造に関係している。RNA 逆折りたたみアルゴリズムでは、折りたたまれた RNA 構造とターゲット構造の間の類似性を測定する目的関数を使用して、生成された RNA 配列を評価する。さらに、RNA 分子の機能に決定的に影響を与える GC 含量(グアニンとシトシンの割合)などの他の配列特性も考慮に入れた。この研究では、MCTS を使う MCTS-RNA という新しいアルゴリズムを開発した。

[81] RNA の逆折りたたみ：RNA の立体構造から RNA の配列を導くこと。
[82] RNA の二次構造：一次構造の RNA 配列をもとに、分子内で塩基対を形成するように配列を平面表示した構造。
https://www.weblio.jp/content/RNA%E3%81%AE%E4%BA%8C%E6%AC%A1%E6%A7%8B%E9%80%A0

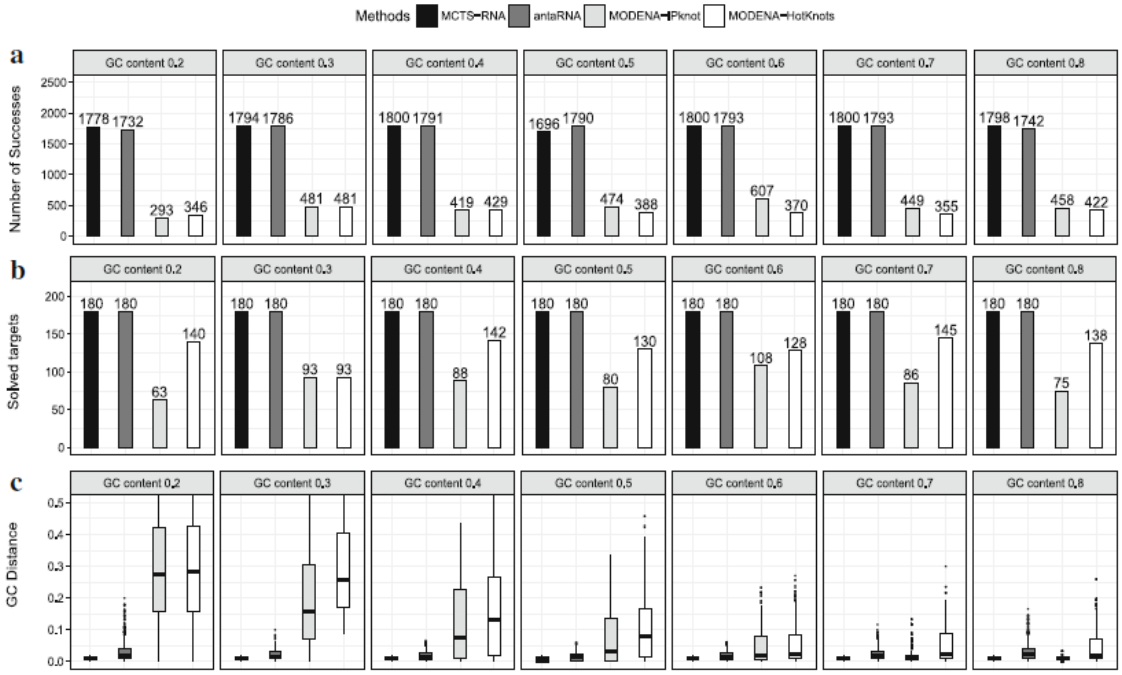


図 2-15　シュードノット[83]構造を含み GC 含量の異なる目標値での MCTS-RNA、antaRNA、MODENA の実験結果。(a)249 のターゲット構造で正常に設計された配列の総数。(b)解決されたターゲット構造の数。(c)GC 含量の誤差の分布。[34]

図 2-15 に本研究で開発した MCTS-RNA と今までに開発されたアルゴリズムによる探索結果の比較を示す。この比較では計算では難しいシュードノットを含む計算結果を示した。これらの結果から、GC を含む検索において、MCTS-RNA が既存ツールより優れていることが分かった。

(iv)有機材料の自動設計アルゴリズム ChemTS の開発

さらに、有機材料の自動設計にモンテカルロ木探索のアルゴリズム ChemTS を開発した[35]。望みの物性を持つ有機分子化合物を自動的に発見することはこれまで難しいこととされてきた。近年、Variational Autoencoder(VAE)[84]などの深層学習を用いた手法が続々と提案されており、実現可能性が高まっている。本研究では、Recurrent neural network[85]と、分子

[83] 少なくとも 2 つのステムループ構造を含む核酸の二次構造で、一方のステムの片側が他方のステムの間に位置している。シュードノットの構造形態は、コンテクスト依存性や塩基対形成領域の重複のため、計算生物学的手法による検出には適していない。
https://ja.wikipedia.org/wiki/%E3%82%B7%E3%83%A5%E3%83%BC%E3%83%89%E3%83%8E%E3%83%83%E3%83%88

[84] VAE(Variational Autoencoder、変分自己符号化器)はディープラーニングによる生成モデルであり、変分ベイズ推定法の一種である。https://ai-kenkyujo.com/term/what-is-vae/

[85] 回帰型ニューラルネットワーク(RNN)は、ノード間の結合が配列に沿った有向グラフを形成する人工ニューラルネットワークのクラスである。これによって、時系列のための時間的な動的振る舞いを示すことが可能となる。
https://ja.wikipedia.org/wiki/%E5%9B%9E%E5%B8%B0%E5%9E%8B%E3%83%8B%E3%83%A5%E3%83%BC%E3%83%A9%E3%83%AB%E3%83%8D%E3%83%83%E3%83%88%E3%83%AF%E3%83%BC%E3%82%AF

の表現としてSMILES[86]と呼ばれる文字列をノードとしたモンテカルロ木探索を用いた(図2-16)。

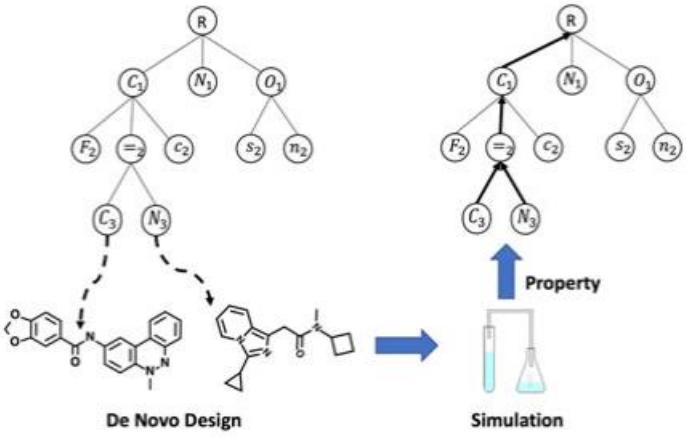


図2-16　ChemTSによる新規化合物の設計[35]

ChemTSとその他の手法による、実験結果を表2-10に示した。ChemTSは最大の評価関数値を達成するとともに、既存のソフトウェアOnly RNNと同等の数の生成分子数を生成することができた。

表2-10　計算時間2、4、6、8時間での評価関数Jの最大値と1分あたりの生成分子数[35]

Method	2 h	4 h	6 h	8 h	Molecules/Min
ChemTS	4.9 ± 0.4	5.4 ± 0.5	5.5 ± 0.4	5.6 ± 0.5	41 ± 1.6
RNN+BO	3.5 ± 0.3	4.5 ± 0.2	4.5 ± 0.2	4.5 ± 0.2	8.3 ± 0.0
Only RNN	4.5 ± 0.3	4.6 ± 0.3	4.8 ± 0.3	4.8 ± 0.3	41 ± 1.4
CVAE+BO	−30 ± 27	−1.4 ± 2.2	−0.6 ± 1.1	−0.0 ± 0.9	0.1 ± 0.1
GVAE+BO	−4.3 ± 3.1	−1.3 ± 1.7	−0.2 ± 1.0	0.3 ± 1.3	1.4 ± 0.9

Notes: The rightmost column shows the number of generated molecules per minute. The average values and standard deviations over 10 trials are shown.

(ｖ)分子の軌跡を検索するアルゴリズムTS-MDの開発

また、さらに分子のConformation(配座)について、その配座への分子の動きにより配座へ到達する軌跡を検索する手法を開発した[36]。原子および分子はある時間の間相互作用することが許される。これによって生じる原子や分子の動的発展の様子(動き)を計算によって導く手法として分子動力学法(Molecular Dynamics：MD)が用いられる。本研究ではMDと木探索を組み合わせた木探索 MD(TS-MD)を使って分子の動きを計算する手法を開発した。TS-MDでは、タンパク質の構造をノードとし、計算された軌跡をエッジで表す(図2-17)。

[86] SMILES記法とは、分子の化学構造をASCII符号の英数字で文字列化した、構造の曖昧性の無い表記方法である。SMILES文字列は多くの種類の分子エディタにおいてインポート可能で、二次元の図表あるいは三次元のモデルとして表示することができる。
https://ja.wikipedia.org/wiki/SMILES%E8%A8%98%E6%B3%95

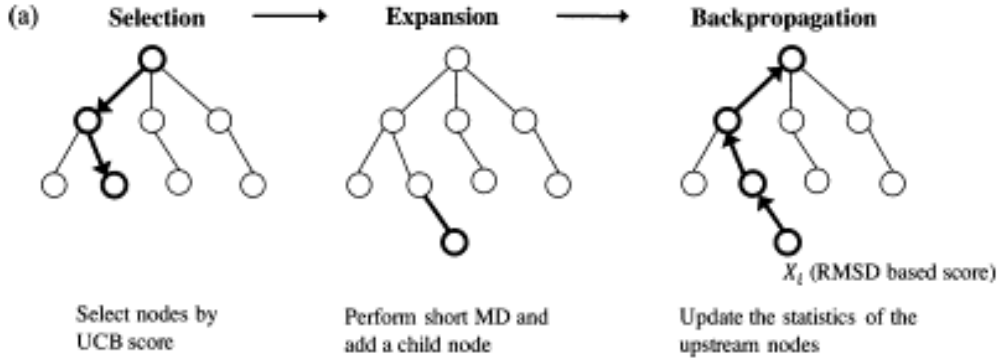


図 2-17　木探索による計算の軌跡[36]

小さな蛋白質としてチグノリン(Chignolin)の折りたたみについて、既存の手法 PaCS-MD と開発した TS-MD を例に実験した。図 2-18 は、既存の PaCS-MD および TS-MD の検索木の例として、reactant(反応物)、product(生成物)、および中間構造を示している。図 2-18(a)は PaCS-MD で失敗したケースの典型的な例を示し、図 2-18(b)は TS-MD が誤って折りたたまれた状態にトラップされるのを回避したことを示している。

TS-MD は、UCT アルゴリズムと短い MD シミュレーションの組み合わせに基づいて、さまざまな状態を効果的にサンプリングし、局所的な安定した構造から脱出することができた。この方法は、ドメインの動きやタンパク質-リガンド結合など、あらゆるタンパク質のコンフォメーション変化に適用できる。TS-MD には、現在、ターゲット構造が必要である。ただし、報酬を適切に変更すれば、TS-MD を汎用に使用することも可能である。さらに、検索パラメータの改善、並列化などの加速手法の適用、大規模なリゾチームなど分子への適用など、将来的な改善、検証の余地があることを明らかにした。

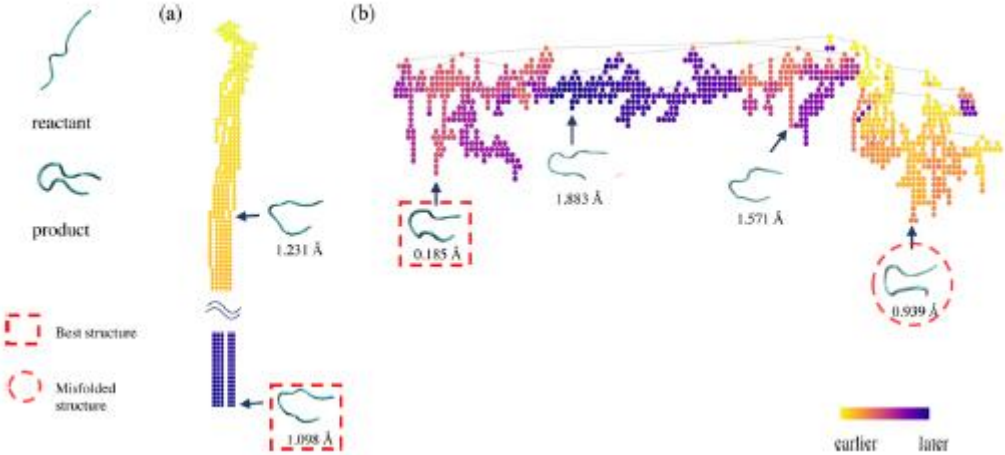


図 2-18　チグノリンの折りたたみでの(a)PaCS-MD および(b)TS-MD の検索ツリーの例[36]

図 2-18 の各ノードの色は、ノードが木に追加される順序を示す。一部のノードの中間構造も、生成物との平均二乗偏差(RMSD)の値で示されている。赤い点線の四角の構造は、各試

行で最小の RMSD を持ち、赤い点線の円の構造は、誤って折りたたまれた構造を示す。図 2-18(a)PaCS-MD では RMSD が生成物の値と大きく異なり、誤った折りたたみを示している。図 2-18(b)TS-MD では、最初に、誤って折りたたまれた構造への経路が優先的にサンプリングされたが、しかし、誤った極小値から脱出し、最終的に生成物に到達した。

以上の結果、ベイズ最適化は、候補の数が限られているケースに対処し、最適な候補をできるだけ早く見つけるのに適しており、一方、モンテカルロ木探索は、候補の数が非常に多いか無制限の場合に適していることが分かった。ベイズ最適化を使用すると、候補全体の数パーセントのみを計算することで最適な構造が得られ、計算リソースを大幅に節約できることを示した。ベイズ最適化と比較して、モンテカルロ木探索アルゴリズムは、大域的に最適候補ではなくても、それに近い候補を得ることに優れていた。

2.2.3 統計・マイニング応用の発展

広範で大規模な非線形系について、非線形回帰残差から変数間の因果関係を一意に推定できることを明らかにした。ナノギャップ、ナノポアセンサによる抵抗・電流パルス分析について機械学習を応用し、鷲尾は高精度で単一分子を特定する技術を大阪大学の谷口と共同で開発した。この技術はベンチャー企業(アイポア株式会社)の創設に結び付いた。また、国立研究開発法人物質・材料研究機構(NIMS)の MEMS センサへ、機械学習を適用してガス種の識別を迅速に識別する技術を開発した。

(1)非線形系についての大規模因果推論原理・手法の研究

従来、因果推論は線形で非ガウスノイズを有する大規模系でしか、実用的解析原理・手法は知られていない。本研究では、広範で大規模な非線形系についての実用的原理・手法を研究した[38]。

本研究では、ノイズがガウスか非ガウスか、変数とノイズが独立かなどに係わらず、変数間の因果の向きに沿った非線形回帰残差が逆向きの残差より小さいことを、数学的に証明した。すなわち、理論では、対象系が非線形ならば、ノイズの性質や独立性、交絡変数の有無などに依らず、非線形回帰残差から変数間の因果関係を一意に推定できることが明らかとなった。予想される予測誤差は 2 つの変数間の因果方向に大きく依存するという関係、すなわち、原因から効果を予測すると、効果から原因を予測するよりも誤差が小さくなる関係により因果関係を推測する方法を開発した。

人工的なデータの結果を図 2-19 に示した。人工的に非線形性とノイズ量を制御するパラメータ b と α を導入した。人工的なデータセットにおいて、本手法は他の手法より優れた推測精度を示した[37]。

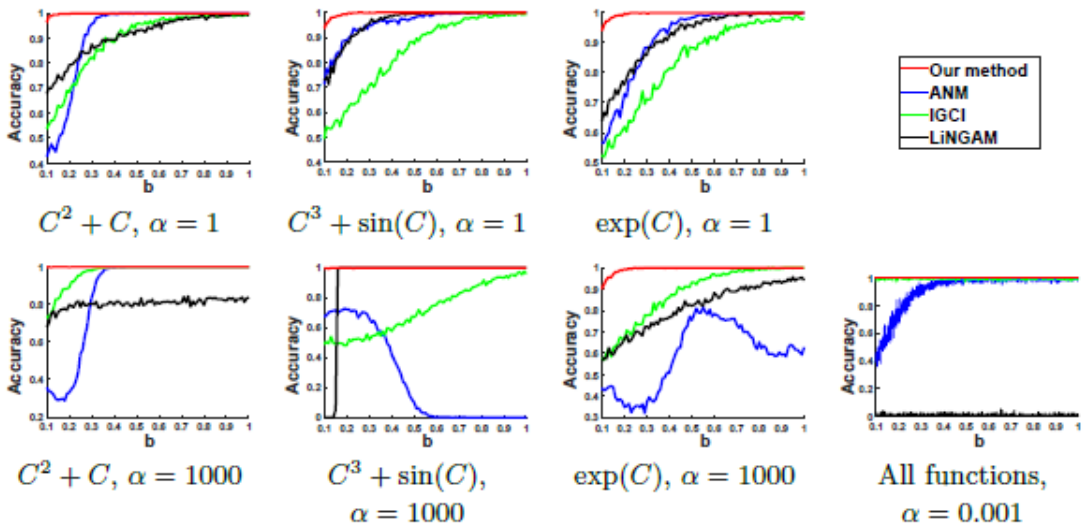


図 2-19　人工データセットにおけるさまざまな因果推論方法の精度の概要[37]

基になる関数 ϕ を原因変数 C の関数で示している。パラメータ α はノイズの量を示す。α が大きいほど、小さいノイズになる。パラメータ b は、ϕ の非線形性を制御する。

実際のデータについて評価した結果を図 2-20 に示した。他の手法と同等の結果を示したが、まだ改善の余地があることが分かった。

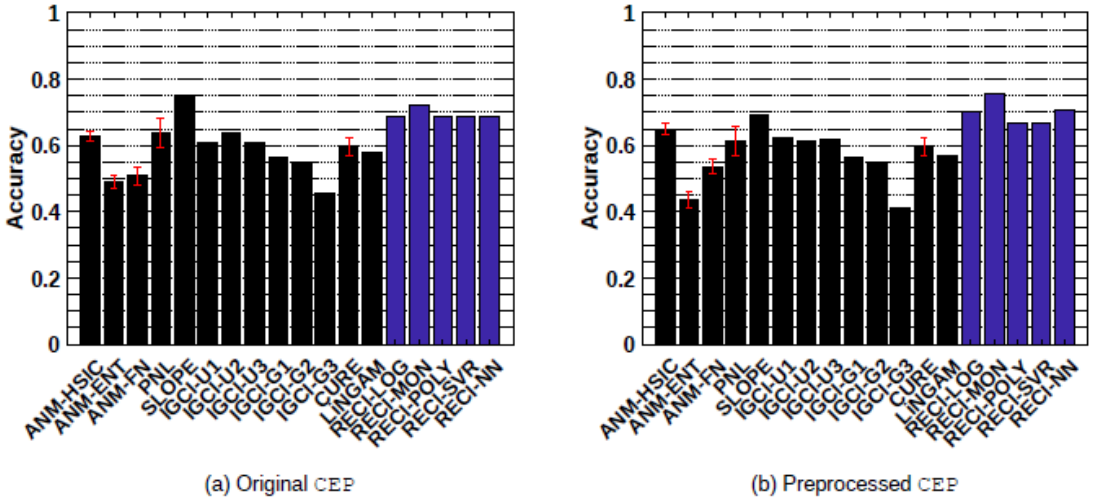


図 2-20　実際の一般的に使われる cause-effct pairs(CEP)データセットのすべてのメソッドの評価結果。RECI が本方法であり、ハイフン以後は回帰手法の名称である。(a)は元のデータの評価結果、(b)低密度の点を削除した前処理バージョンの結果を示した。[38]

(2)ナノギャップ、ナノポアセンサの抵抗・電流パルス計測に対する機械学習適用

ナノギャップ、ナノポアを通過する際のイオン電流-時間波形を 250KHz 以上の高時間分解能で計測し、鷲尾は高精度で単一分子を特定する技術を大阪大学の谷口と共同で開発し

た[39],[40],[41],[42],[87]。多様な環境で正確な計測結果を得るために、ノイズなどのあらゆる外乱に対応する推定原理を検討した。従来は、事前知識がない場合には、信号強度がノイズ強度より大きい(SN 比が良い)ことを前提とし、信号弁別やフィルタリングでノイズ除去を行うしかなかった。本研究では、機械学習により信号とノイズのパターンの違いから効果的なノイズ除去を行う、一般的かつ新しいノイズ除去原理の検討を行い、実験的に有効性を確認した。図 2-21 にその模式図を示した。塩基配列の計測において、ノイズ除去用の最尤推定による PU(正例と Un-labeled 事例)分類器[88]とクラス分類器を組み合わせた手法をナノギャップ計測の 2 塩基ポリマー分類に適用し、高精度な分類性能を確認した。

本成果は、ベンチャー企業のアイポア株式会社設立に繋がった。

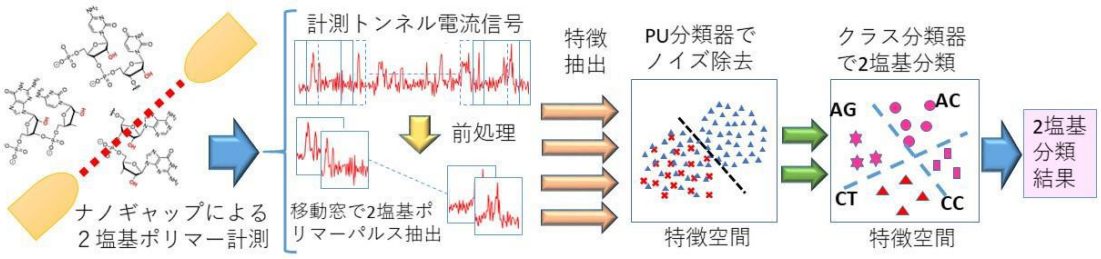


図 2-21　ナノギャップ計測に機械学習ノイズ除去を適用した高精度 2 塩基識別[89]

(3) ガスセンサ・嗅覚センサへの応用

国立研究開発法人物質・材料研究機構(NIMS)と共同でガス吸着センサ(MSS)を開発した。NIMS が MEMS の技術を使って 4 つの Si センサ膜を形成し、MEMS 膜の上に様々な機能性材料の膜を塗布した感応膜を持つ構造を持ち、ガスの吸着に伴う膜の応力による変形を高感度に感知するセンサである(図 2-22)。

[87] CREST 研究領域「計測技術と高度情報処理の融合によるインテリジェント計測・解析手法の開発と応用」、平成 28 年度採択課題「機械学習と最先端計測技術の融合深化による新たな計測・解析手法の展開」

[88] Positive-Unlabeled Classification の略。正例データと、負例の代わりにラベルなしデータによる分類。

[89] https://www.jst.go.jp/kisoken/crest/evaluation/nenpou/h30/JST_1111092_16815521_2018_PYR.pdf

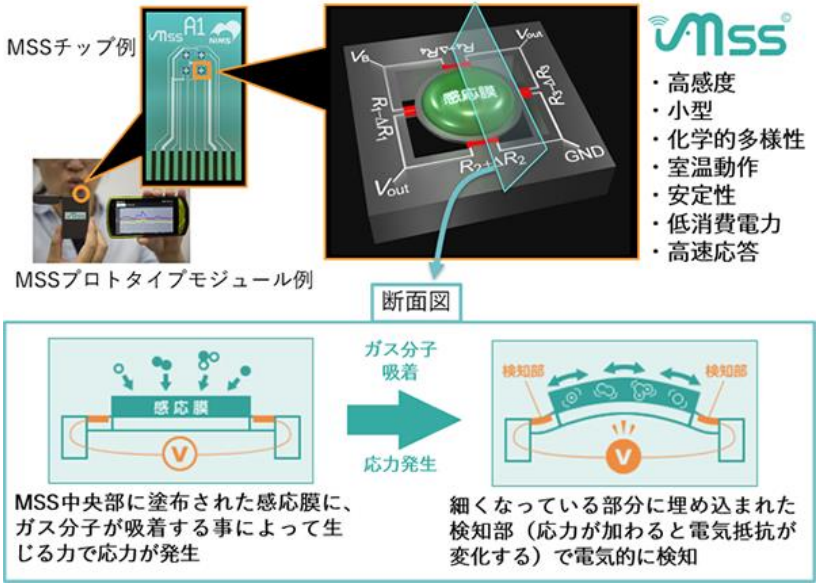


図 2-22　ガス吸着センサ(MSS)の構造と検知原理[90]

得られたデータから、ガスや匂いの分類を実現するための機械学習モデルを開発した[43],[44],[45]。蒸気(臭気)を分類する機械学習モデルを開発するための基本的な手順を図 2-23 に示した。最初に、トレーニングデータセットを標準化し、続いて次元削減を行う。次に、クラスラベル付きのデータに基づいて分類器をトレーニングし、予測モデルを開発した。テストデータセットを使って予測モデルを検証した。この研究では、ネストされた相互検証(CV)を通じて機械学習モデルを開発した。

「外側のループ」の角の丸い長方形は、蒸気の 4 つの測定ファイルで構成される個々の測定データセットを表す。予測モデルを得るために 7 つのデータセットを使って開発と最適化を行った。残りの 1 つのデータセットにより開発したモデルを評価した。トレーニングデータセットとテストデータセットのすべての組み合わせ(外部ループ)に対して、このようなモデルの構築と評価のプロセスを実行した。

外側のループのトレーニングデータセットからトレーニングセットとテストセットをランダムに分割して内側のループのデータセットを作成した。この内側のループによって予測モデルを最適化した。

この研究では、主成分分析(principal component analysis：PCA)を次元削減アルゴリズムに使用し、6 つの分類アルゴリズムを分類器として使用した(線形カーネルを備えたサポートベクターマシン(SVM)、放射状基底関数(RBF)カーネルを備えた SVM、ロジスティック回帰(LR)、決定ツリー(DT)、ランダムフォレスト(RF)、および多層パーセプトロン(MLP))。

[90] https://mss-forum.com/about/

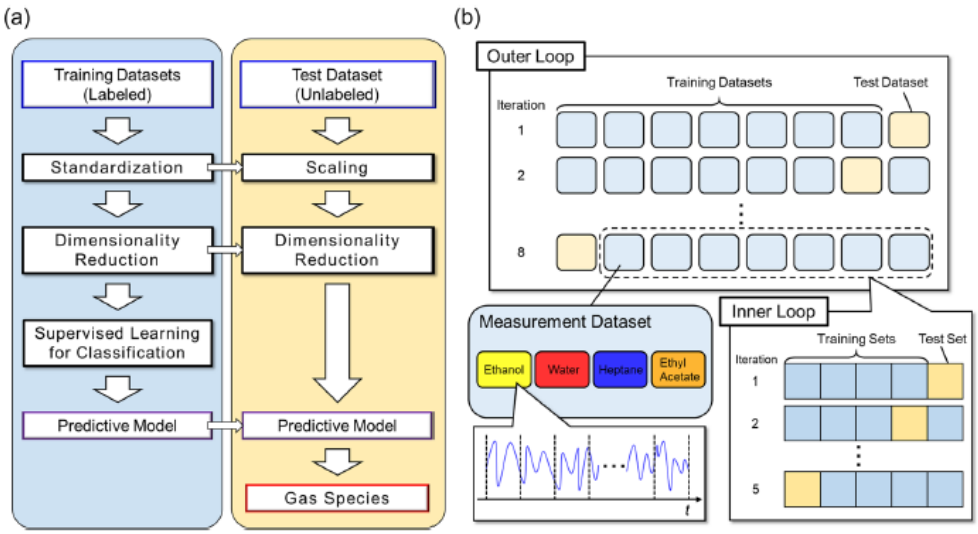


図 2-23　(a)分類モデルを開発するための基本的な手順。(b)ネストされた CV の図[45]

「外側のループ」の角の丸い長方形は、蒸気の 4 つの測定ファイルで構成される個々のデータセットを表している。外側のループのトレーニングデータセットは、ランダムに 5 つのデータセット(「内側のループ」の長方形)に分割される。

分類モデルの測定結果を表 2-11 にまとめた。これらの分類器で約 0.85 の分類化精度を達成した。特に、ランダムフォレストに基づく分類モデルは、6 つのモデルの中で最高の精度(0.90±0.15)を得た。これらの結果によって、ガス種を高精度で識別できることを示した。ガス種を識別するために必要な測定時間はわずか 3.0 秒であり、迅速な(実質的にリアルタイムの)ガス識別ができた。

表 2-11　MSS センサを使った溶液ガスの分類モデルの測定結果[45]

Classifier	Optimized Parameters	Accuracy
Support Vector Machine (Linear Kernel)	Number of PCs: 40 C: 0.1	0.82±0.15
Support Vector Machine (RBF Kernel)	Number of PCs: 80 C: 10.0 γ: 0.01	0.83±0.18
Logistic Regression	Number of PCs: 20 C: 10.0	0.82±0.16
Decision Tree	Maximum depth: 10	0.81±0.20
Random Forest	Number of estimators: 1024	0.90±0.15
Multilayer Perceptron	α: 1.0 Hidden layer size: (128,128,64)	0.84±0.16

2.3　プロジェクト参加研究員の活動状況

本プロジェクトは 3 グループあり、その拠点も北海道、東京、大阪の 3 箇所に分かれて研究が行われた。各グループのグループリーダーを除いた研究員の人数とキャリアアップし

た人数をまとめた(表 2-12)。企業からの研究者を除いた全員がキャリアアップしていることが分かった。統計・マイニング応用グループは本プロジェクトで雇用された研究者はいなかった。

5 人の技術員については所属グループが不明のため、表からは除いた。5 人の技術員の中、3 人は企業から参画し、終了後その企業へ戻った。1 人はベンチャー企業の役員になった。1 人は、終了直後 CREST に参画していたが、その後不明である。

表 2-12　研究員の人数とキャリアアップ数

グループ	研究者数	キャリアアップ人数	キャリアアップした人以外の動向
離散構造処理基盤	8	4	1 人が企業から参画し終了後戻った。 2 人がベンチャー企業へ参画。 1 人は外国人、海外の大学へ。
機械学習・制約充足応用	8	6	2 人が企業から参画し終了後戻った。
統計・マイニング応用	0	0	
合計	16	10	

これら研究員の中で、様々な大学の准教授へとキャリアアップしたのは、吉仲亮、戸田貴久、川原純、斎藤寿樹、また理化学研究所(理研)のユニットリーダーになったのは、美添一樹、田部井靖生、産業技術総合研究所の主任研究員となったのは、櫻井祐子である。

本プロジェクトの正式の研究員ではなく、グループリーダーの研究室の研究者として本プロジェクトに協力した研究者が多数存在する。特に統計・マイニング応用グループはグループリーダーの研究室の研究者が中心となって研究が行われた。

統計・マイニング応用グループについて、マイニング論文等から、本研究を推進し貢献したと思われる研究者について調査した。特に顕著なキャリアアップした研究者は以下の通りである。統計・マイニング応用グループの清水昌平は大阪大学で助教(2009 年度～2012 年度)から准教授(2012 年度～2015 年度)になり、その後滋賀大学の教授(兼、理化学研究所革新知能統合研究センターチームリーダー)になっている。因果推論の分野での著書[91]もあり、この分野で活躍している。河原吉伸も大阪大学の助教(2009 年度～2012 年度)から准教授(2012 年度～2018 年度)になり、その後九州大学の教授(2019 年度～)(兼、理化学研究所革新知能統合研究センターチームリーダー)になっている。杉山磨人は大阪大学の助教(2013 年度～2016 年度)の後、国立情報学研究所(NII)の准教授(2017 年度～)になっている。

2.4　第 2 章まとめ

終了後の研究成果をグループごとにまとめると、以下の通りとなる。

[91] 清水昌平、「統計的因果探索」(機械学習プロフェショナルシリーズ)、講談社(2017)

(1)離散構造処理基盤グループ

ZDD 技術の改良・発展として、論理関数を表すデータ構造 ZSDD と Zd-DNNF を開発、厳密被覆問題に使われる Knuth の DLX の ZDD による改良等を研究した。また、ZDD の適用範囲を、ベイジアンネットワーク、バンディット問題、ナップザック問題等へ拡大した。さらに、通信ネットワークシステムの信頼性設計、配電ネットワーク設計、ホットスポットクラスター検出等の応用を研究した。

①ZDD の改良・発展

(A)論理関数を扱う SDD の発展版の ZSDD、Zd-DNNF を開発

(B)Knuth のアルゴリズム DLX の ZDD による改良

②ZDD の適用範囲拡大

代表的な離散数学の問題への ZDD 適用による範囲拡大を行った。

(A)ベイジアンネットワークへの適用

(B)バンディット問題への適用

(C)ナップザック問題への適用等

③各種応用の実現

以下の各種応用問題に取り組んだ。

(A)通信ネットワークの信頼性設計(NTT との共同)

(B)配電ネットワーク設計(早稲田大学との共同研究)

(C)ホットスポットクラスター検出(岡山大学、奈良先端科学技術大学院大学との共同研究)

(2)機械学習・制約充足グループ

LAMP ソフトウェアの拡張を図り、ゲノム関連解析のための LAMPLINK、並列処理を可能にした MP-LAMP を開発した。また、マテリアル・インフォマティクス、バイオ・インフォマティクスの分野への研究に発展し、ベイズ最適化の応用として熱電材料の設計に応用された。また、医療品の処方や製造方法の最適化にも適用した。さらに、モンテカルロ木探索によって材料探索ができる MDTS ソフトウェアを開発し、熱伝導の最適化問題やナノフィルムでの界面熱輸送の問題に適用した。バイオ・インフォマティクス分野の RNA の逆折りたたみのアルゴリズム MCTS-RNA を開発した。さらに、有機材料の自動設計アルゴリズム ChemTS を開発し、公開した。さらに、分子の動きの軌跡を検索するアルゴリズム TS-MD の開発にも成功した。

①LAMP 法の発展

(A)ゲノム関連研究に広く使われている PLINK に LAMP を組み込んだ LAMPLINK の開発

(B)LAMP を並列化により高速化した MP-LAMP の開発
を行った。

②マテリアル・インフォマティクスへの展開

(A)ベイズ最適化への応用
　ベイズ最適化のための COMBO を開発した。これにより、
(ｱ)界面熱伝導の最小化、最大化設計の実現
(ｲ)超狭帯域波長選択性放射のための多層膜最適化構造実現
(ｳ)結晶構造予測手法の提案、医薬品の処方や製造方法の最適化
を行った。
(B)モンテカルロ木探索(MCTS)の応用
モンテカルロ木探索によって材料を探索できるプログラム MDTS を開発した。
これを使って、
(ｱ)Si-Ge 合金の熱伝導が最大、最小となる合金を研究
(ｲ)Si と Ge のナノフィルムでの界面粗さと熱輸送最適化の検討
を行った。

③バイオ・インフォマティクスへの展開
(A)MCTS-RNA というアルゴリズム開発による RNA 逆折りたたみの研究
(B)有機材料の自動設計に MCTS を使うアルゴリズム ChemTS の開発と有効性の確認
(C)蛋白質分子における分子が配座へ落ち着くまでの軌跡を検索する手法(TS-MD)の開発
を行った。
これらの研究の結果、ベイズ最適化は候補の数が限られている場合の最適化解を得るのに優れており、候補の数が多い時に適しているモンテカルロ木探索(MCTS)には、大域的に最適ではなくてもそれに近い候補を得ることに優れることが分かった。

(3)統計・マイニング応用グループ

①非線形系の大規模因果推論原理・手法の研究
因果推論について非線形系の実用的原理・手法に成果を挙げた。変数間の因果の向きに沿った非線形回帰残渣が逆向きの残渣より小さいことを証明し、改善の余地があるものの実際のデータで確認した。

②ナノギャップ、ナノポアセンサの抵抗・電流パルス計測に対する機械学習適用

大阪大学の谷口らと共同で、抵抗・電流パルスを使った計測に機械学習を適用し、ナノギャップ・ナノポアを通過するイオン電流の波形解析から単一分子を特定する計測法を確立した。この成果はベンチャー企業、アイポア株式会社の設立に繋がった。

③ガスセンサ・嗅覚センサへの応用

NIMS との共同研究により、ガスセンサ・嗅覚センサの開発に機械学習を適用し、リアルタイムのガス識別に成功した。

(4) 人材育成について

人材育成について、本プロジェクトは研究者総数 16 名に対して、参画した企業へ戻った人 3 名、ベンチャー企業へ参画 2 名、外国へ戻った研究者 1 名(計 6 人)以外、すべての研究者(10 名)がキャリアアップを実現した。また、本プロジェクトに登録されていない研究者の中にも、3 名がキャリアアップを実現している。

第3章　プロジェクト成果の波及と展望

3.1　科学技術への波及と展望

科学技術への波及と展望はアルゴリズムの研究、応用研究、研究者の育成に分けることができる。アルゴリズムの研究については、論文の検索から、湊の ZDD 論文自体の引用、及び ZDD 関連キーワード論文の投稿など関係者以外の近年の ZDD 関連論文の国内外の広範な増加が明らかとなった。また、アルゴリズム関連論文 7 件についても被引用数の指標である FWCI の値から平均以上に引用されていることが分かった。応用研究では、マテリアルズ・インフォマティクス(MI)とバイオ・インフォマティクス(BI)の分野における、日本の研究戦略[92]の流れに沿った成果が得られており、今後も MI/BI 領域での大きな波及効果が期待できる。また、ナノポアセンサと機械学習の組み合わせによる分子同定の技術は、ベンチャー企業創設に繋がっただけでなく、今後の発展や他分野への応用が期待できる。さらに、有用な人材を輩出した研究者育成の成果もこの分野に大きな波及効果をもたらしている。

3.1.1　アルゴリズムの研究

離散構造はグラフやネットワーク構造と親和性が高く、それをどう扱い、応用問題を解くかが本プロジェクトの中心テーマである。解き方には離散数学をそのまま使う方法と応用にあたって連続数学を適用する方法もある。本プロジェクトの「離散構造処理基盤研究」は離散数学を主に扱い、「機械学習・制約充足応用研究」と「統計・マイニング応用研究」は主として連続数学を手段とした研究であった。離散数学は IT(情報技術)や AI の基礎技術であることはよく知られており[93]、大学で広く教育カリキュラムに取り入れられている。離散構造処理基盤研究は、離散数学を使って応用問題を解く研究、すなわち、離散構造とそれを使って解くためのアルゴリズムの研究であった。

(1)ZDD の展開と波及

一般的に、微分積分等を代表とする連続数学の分野では、その社会的応用は多岐に渡り、物理学や工学の基礎、医学や薬学への応用(MRI 診断、新薬開発等)、マクロ経済学理論等がすぐに挙げられる。一方、離散数学とは連続数学では扱えない問題を扱う所に特長がある。連続数学ほど範囲は広くはないが、この分野の社会的な応用としてデジタル論理回路の設計、ネットワークフロー解析等が知られている。特に、現在複雑化しているコンピュータデジタル回路の設計への応用は重要である。

そのような状況の中で、本プロジェクトでは離散構造として研究総括が提案した ZDD を使い様々なアルゴリズムを開発した。ZDD は二分木の表現方法を改善することにより、デー

[92] https://www.jst.go.jp/sip/k03.html
https://www.meti.go.jp/policy/mono_info_service/connected_industries/index.html
[93] 茨木俊秀、「AI 時代の離散数学」、オーム社(2020)

タ構造の簡略化とそれを利用した高速なアルゴリズムに特長がある。ベイジアンネットワーク、バンディット問題、ナップザック問題等へ適用し、従来より優れた性能を得た。

湊が 1993 年に初めて発表した ZDD 論文[2]の被引用数は 2020 年末で 461 件であった。1993 年～2020 年までの ZDD 論文の被引用数について、年ごとの推移を調べた(図 3-1)。この図では被引用数の表示を、国内、海外及びその全体(国内＋海外)に分けて示した。参考のために、その中から湊が著者になっている論文数を棒グラフで示した。

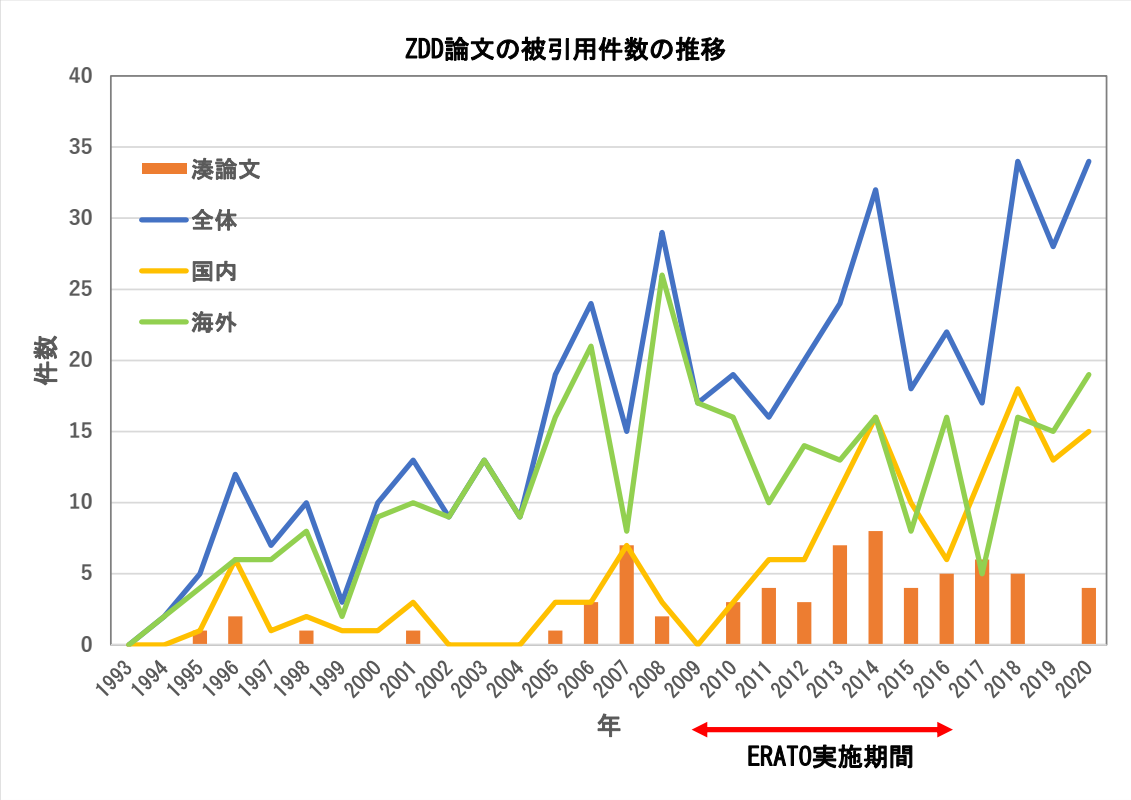


図 3-1　ZDD 論文の被引用数の推移(1993 年～2020 年)(DB:Scopus、検索日：2021 年 1 月 13 日)

この図から下記のことが分かった。

(A) ZDD 論文の発表から 2010 年頃(本プロジェクトの開始頃)まで日本の研究はほぼ湊の研究のみであった。その間、外国の研究が増えている。

(B) 海外論文からの被引用数は 2008 年をピークに以後平均すると 10 件～15 件が続いている。

(C) 日本からの論文数が 2014 年頃より増えだしている。一方、湊の論文は、2015 年以後ほぼ 5 件以下ということが分かった。

(D) 全体として ZDD 論文から 27 年経過しているにも拘わらず、今なお 30 件/年程度の被引用数を示し、現在も全体としては増加傾向である。

これらの結果が ZDD 関連の研究そのものと同じ傾向を持つのかを確認するために、ZDD 関連の研究を表すキーワードを選び、そのキーワードがタイトル、抄録、キーワードに使われている論文を検索した。キーワードとして、zero suppressed、zerosuppressed、zero-suppressed や ZDD、ZDDs、ZSDD を選んで検索した。その結果を図 3-2 に示した。

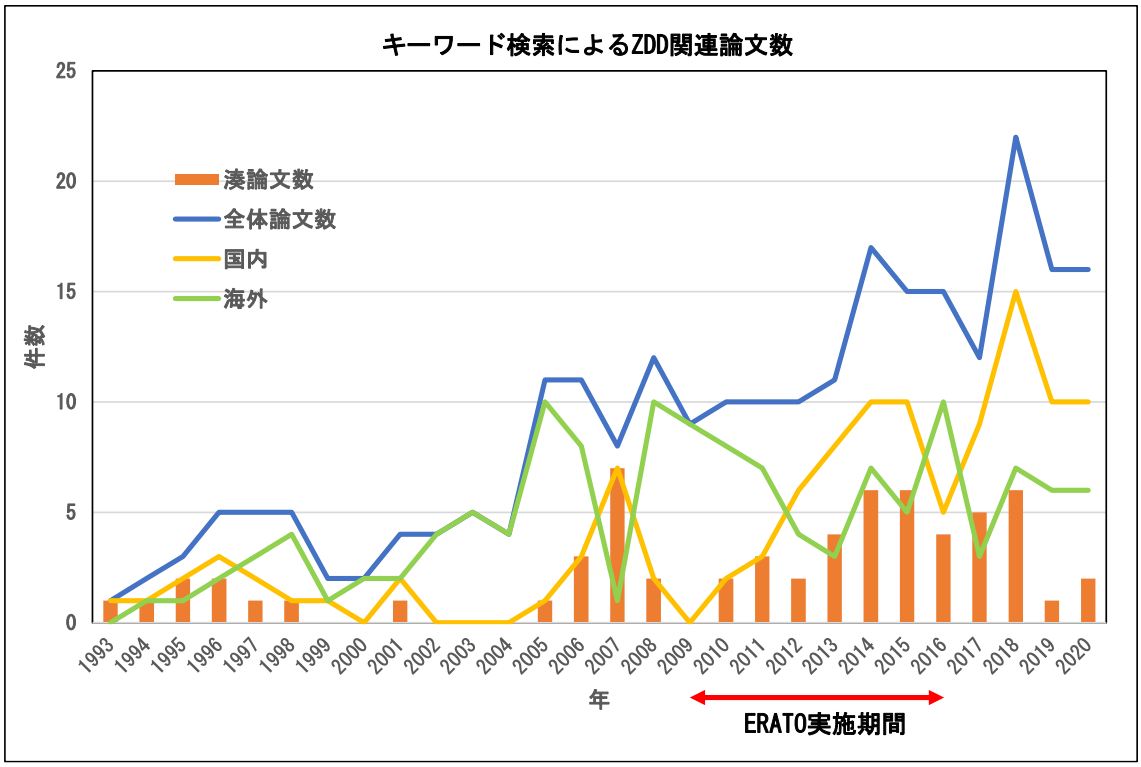


図 3-2　キーワード検索による ZDD 関連の論文数の推移(DB:Scopus　検索日:2021 年 1 月 12 日)

この結果、全体の論文数は 247 件と全体の被引用数 461 件より少なかったが、図 3-1 の特徴(ⅰ)～(ⅳ)をほぼ踏襲している。

これらの論文数の検索から、海外の論文数はほぼ一定の論文が発表されていること、日本からの論文数が 2013 年頃より次第に増えてきている。ゆっくりではあるが、着実に世界と日本の ZDD 関連の研究に波及している。初出の ZDD 論文が今も、世界の論文に引用される基本的な論文であることが明らかになった。

日本では、九州大学、関西学院大学、広島市立大学、明治大学、高知大学、神奈川大学、新潟大学、日産自動車、島根大学、埼玉大学等に広がっている。海外では、ルーヴァン・カトリック大学(ベルギー)、ニコシア大学(キプロス)、ユタ大学(米国)、暨南大学(中国)、中山大学(中国)、桂林工学院(中国)、黔南民族师范学院(中国)、南メソジスト大学(SMU)(米国)、スイス連邦工科大学ローザンヌ校(EPFL)等の研究者が ZDD 関連の研究を発表していることが分かった。最先端の研究として、SMU と EPFL の共同研究では量子キュービット回路

を最適にマッピングする技術として ZDD を使う研究を行っている[94]。また、Internet of Things(IoT)への応用が提案されている[95]。離散数学会において著名なカーネギーメロン大学の R.E.Bryant 教授が論文 2 報(2018 年、2020 年)を発表している。

(2)「機械学習制・約充足応用研究」と「統計・マイニング応用研究」のアルゴリズム

一方、「機械学習・制約充足応用研究」と「統計・マイニング応用研究」においては、開発したアルゴリズムは多岐に渡っている。以下の表 3-1 は開発された主なアルゴリズム論文の引用数をまとめたものである。この表の FWCI 指標からすべて平均より多く引用されていることが分かる。COMBO、πDD、DirectLiNGAM は、それぞれ被引用数は異なるものの、それぞれの分野において平均より多く引用されていることが分かった。特に、COMBO は上位 1.8%に入る被引用数であった。これらの結果は、本プロジェクトで開発されたアルゴリズムの多くが注目されていることを示している。

表 3-1　開発されたソフトウェア論文の被引用数と FWCI[96]

発表年	ソフトウェア名称	グループリーダー	被引用数 2020/10 時点	FWCI	上位%
2011	πDD	湊	15	3.28	6.4%
2011	DirectLiNGAM	鷲尾	116	3.99	4.6%
2013	LAMP	津田	55	1.16	25.0%
2016	Graphillion	湊	19	1.53	18.2%
2016	COMBO	津田	87	6.11	1.8%
2017	MDTS	津田	16	1.08	26.8%
2017	ChemTS	津田	44	2.91	7.2%

3.1.2　応用研究の展開

(1)離散構造処理基盤技術の応用への展開

[94] Smith K., Thornton M., Soeken M., Schmitt B., de Micheli G., “Using ZDDs in the mapping of quantum circuits”, Electronic Proceedings in Theoretical Computer Science, EPTCS, 318, 106-118 (2020)

[95] Viktoros S.A., Michael M.K., Polycarpou M.M., “Compact Fault Dictionaries for Efficient Sensor Fault Diagnosis in IoT-enabled CPSs”, Proceedings - 2020 IEEE International Conference on Smart Internet of Things, SmartIoT 2020, 9192014 (2020)

[96] 「Field-Weighted Citation Impact(相対性被引用インパクト)」：この論文が類似の論文(同じ出版時期、文献タイプ、分野)と比較してどの程度引用されているかを示す。FWCI が“１”を上回る論文は、平均よりも多く引用されていることを意味する。

本プロジェクトでは、社会におけるネットワーク構造を持つスマートグリッド電力網への応用、通信ネットワークシステムの信頼性設計、配電ネットワーク設計、ホットスポットクラスター検出等について、ZDD 構造を使った列挙アルゴリズムによる成果を挙げた。

特に、スマートグリッド電力網の研究は、湊の共同研究者であった早稲田大学の林と東京電力パワーグリッド株式会社が中心となって、スマートグリッド実現に向けた配電網の電力損失最小化の実証試験のプロジェクト(CREST：[EMS]分散協調型エネルギー管理システム構築のための理論及び基盤技術の創出と融合展開(研究総括藤田政之)、研究領域の「協調エネルギー管理システム実現手法の創出とその汎用的な実証および評価の基盤体系構築」(研究代表者林泰弘)2015 年度～2020 年度実施)に展開された[97]。その結果、235 スイッチの配電ネットワークについて最適化設計を実現した。

また、湊と共同研究した NTT の研究者は、列挙アルゴリズムの成果として、①複数のネットワークの候補から最も信頼性の高い構造を自動で求められること、②計算できる拠点が従来は 10 程度であったが、100 まで増やすことができ実際のネットワーク設計時に活用できるようになった、③予算や信頼性の目標値といった制限条件を入力、どのようなリンクを張れば制限をクリアできるかを求めることができる点であると述べている[98]。②の拠点の数 100 の制限下において、①や③の項目については列挙アルゴリズムによって最適化設計できることが分かった。これら 2 つの応用研究から、ZDD による列挙手法により 100～200 程度のノード数において、最適設計が可能であることを実証した。今後、他の社会的なネットワークについても同様な研究が行われることが期待される。

ノード数の壁は、コンピュータの発展だけでは容易に乗り越えられない壁であり、さらなる理論やそれに基づくアルゴリズムの改良、現実的な工夫等が求められている。ホットスポットクラスター検出は、枝刈りの手法を開発したことにより、実用的な時間で解を得ることができた。このような手法は今後の開発に道を拓くものである。

(2)機械学習・制約充足応用グループにおける応用研究の発展と波及

LAMP 法は CREST(「ビッグデータ統合利活用のための次世代基盤技術の創出・体系化」研究領域の「離散構造統計学の創出と癌科学への展開」2015 年度～2020 年度実施)に展開した。これは国内 3 箇所の有力癌研究機関と連携し、小児癌の神経芽腫の原因究明を目指した。その成果として、ゲノム解析を組み込んだ LAMPLINK や並列処理を組み込んだ MP-LINK を開発した。また、ベイズ最適化アルゴリズム COMBO やモンテカルロ木探索のアルゴリズム MCTS、MDTS、ChemTS、TS-MD 等の多くのアルゴリズムを開発し、そのアルゴリズムをマテリアルズ・インフォマティクス(MI)やバイオ・インフォマティクス(BI)の分野へ適用し、熱電素子、薄膜の熱伝導や RNA の逆折りたたみ、望みの物性をもつ有機分子化合物の自動発見、分子の配座へ到達する軌跡の検索手法開発等、多彩な成果を挙げた。その結果、ベイズ最適

[97] https://www.jst.go.jp/pr/announce/20160420/
[98] https://www.rd.ntt/research/CT99-323.html

化、モンテカルロ木探索が材料開発、薬品開発、RNA の解析等に有効であることを示した。これらの研究のプレスリリース(表 3-2)は ERATO 終了後(2017 年)～現在(2020 年)の 4 年間で 10 件発表されており、特に 2018 年は 4 件がプレスリリースとして発表されている。このことは研究の質の高さを示し、この分野における研究をリードしていることを表している。

表 3-2　津田が関係するプレスリリース

発表日	表題	発表大学・研究所
2020.10.21	機械学習により薄膜作製プロセスの高速化を実現 ～外部データなしで試料作製回数を大幅に低減、材料開発コスト削減に期待～	東京大学、物質・材料研究機構
2020.05.28	「例外」を発見する AI「BLOX」の開発 －AI を用いた革新材料の開発に新たな道筋－	理化学研究所、横浜市立大学、物質・材料研究機構
2019.03.19	機械学習を用いて熱電材料の大幅な出力向上に成功 ～従来の実験では探索範囲外の組成で実現 汎用元素による熱電材料の実用化加速に期待～	科学技術振興機構、東京大学、物質・材料研究機構
2019.01.23	機械学習により世界最高クラスの熱放射多層膜を設計し、その実証に成功 ～約80億の候補から最適構造を探索 省エネルギー社会への貢献に期待～	理化学研究所、科学技術振興機構、東京大学、新潟大学、物質・材料研究機構
2018.09.07	人工知能が専門家の約2万倍の速さでスペクトルを解釈 ～知識や職人技なしで、物質の性質を明らかに～	科学技術振興機構、東京大学
2018.08.31	人工知能でタンパク質を自動設計 ～様々な機能性タンパク質開発の加速に期待～	理化学研究所、東北大学、産業技術総合研究所
2018.08.27	AI による有機分子の設計とその実験的検証に成功 － 有機エレクトロニクスなど機能性分子の設計に道筋 －	理化学研究所、科学技術振興機構、東京大学、物質・材料研究機構
2018.03.16	材料設計における AI の有用性を実証 高イオン伝導率を有する全固体リチウムイオン電池用固体電解質の開発を効率化	理化学研究所、富士通
2017.11.15	人工知能が「繰り返し成長すること」で計算コストを 1/3600 に削減 ～界面構造を高速に決定し、高性能な物質開発を加速～	科学技術振興機構、東京大学
2017.04.18	機械学習により熱流を制御するナノ構造物質の最適設計に成功	科学技術振興機構、東京大学、物質・材料研究機構
2016.03.08	ビッグデータを活用して界面構造の決定スピードを 100 倍以上高速化！-触媒や電池の開発を加速-	東京大学

発表日	表題	発表大学・研究所
2015.11.17	マテリアルズ・インフォマティクス手法により超低熱伝導物質を高効率に多数発見 －材料科学と情報科学の融合研究に革新的成果－	京都大学

さらに、開発したソフトウェアを誰でも利用できるように公開することで、例えば、COMBO、ChemTS が実際にベンチャー企業 MI-6 で使われるなど大きな波及効果を生んでいる。

(3)統計・マイニング応用グループの応用研究の発展と波及

CREST([計測技術と高度情報処理の融合によるインテリジェント計測・解析手法の開発と応用]研究領域の「機械学習と最先端計測技術の融合深化による新たな計測・解析手法の展開」2016 年度～2021 年度)に発展し、鷲尾が大阪大学産業科学研究所の谷口と協力して、ナノギャップ、ナノポアを通る分子を電流パルスによって検知、同定する技術を開発した。この技術は機械学習の技術を駆使することによって、分子検出のみならず、分子の同定を可能にした。その結果は、リアルタイムのガス検知のみならず、コロナウイルスを含むウイルス検出等にも繋がる可能性がある[99]。また、DNA シーケンサ[100]や神経伝達物質検出[101]などに波及しており、今後のさらなる発展・波及が期待される。

さらに、NIMS との共同研究により、臭気センサの開発も行い、機械学習により臭気の分別を実現した。この NIMS による臭気センサ[102]は膜型表面応力センサ(MSS)と呼ばれ、MEMS による梁構造とそれに塗布された感応膜によって、臭気による撓みを検知する単純な構造である。感応膜を変えたセンサを複数同時に用いて、機械学習の技術によって臭気分別を実現している。単純だが、高感度な構成のセンサから、機械学習を使えば高度な弁別が可能であることを示し、センサの概念を一変させた。今後、このようなコンセプトのセンサと計測方法が増えることが期待できる。

機械学習の研究は、教師ラベル無しビッグデータからの高速高精度分類器学習手法の探求へと発展している[103]。

3.1.3 研究者の育成

米国や中国に比べて、日本の情報科学(AI を含む)関連分野の人材は不足しており、研究論文数等で差をつけられている。そのため、本プロジェクトやその後の ERATO 河原林巨大グラフプロジェクトにおいても人材の育成が大きな目標として掲げられている。本プロジェ

[99] https://resou.osaka-u.ac.jp/ja/research/2018/20181121_1
[100] https://www.jst.go.jp/kisoken/crest/evaluation/nenpou/h28/JST_1111092_16815521_2016_PYR.pdf
[101] https://www.jst.go.jp/pr/announce/20200709/pdf/20200709.pdf
[102] https://mss-forum.com/
[103] https://kaken.nii.ac.jp/grant/KAKENHI-PROJECT-20K21815/

クトでは、2.3 項で述べたように多くの優秀な研究者が育った。本プロジェクト研究員 16 名中 10 名がキャリアアップしており、高い割合で人材育成の目標を果たした。その中でも、AI の拠点の 1 つである理研革新知能統合研究センター(AIP)では汎用基盤技術研究グループの 17 チームの中で本プロジェクトに関係する研究者 3 名(河原吉伸、田部井靖生、清水昌平)がチームリーダーあるいはユニットリーダーとなっている。グループリーダーであった津田は AIP の目的指向基盤技術研究グループの 1 つの分子情報科学チームのチームリーダーとして統括している。津田は NIMS でも情報統合型物質・材料開発イニシアティブ(Material Research by Information Integration Initiative:MI2I)のデータ科学グループを率いていた(2020 年 3 月まで)。このように、人材を輩出したのは本プロジェクトの科学技術への大きな波及効果である。

3.2 社会への波及と発展

3.2.1 研究成果の社会への波及と展開

社会への応用として、ZDD 構造の研究成果は社会インフラであるスマートグリッド等へ展開された。この応用の成果から、他の重要な社会インフラネットワークへの応用が期待される。マテリアルズ・インフォマティクス(MI)とバイオ・インフォマティクス(BI)の 2 つの分野に向けても、本プロジェクト成果と研究者が高い評価を受けており、今後の発展が期待できる。また、微小な隙間を通過する物体の検出と同定の技術開発はベンチャー企業の創業に繋がる成果を挙げた。さらに、NIMS との共同研究により、臭気センサの開発に成功し、その実用化が期待されている。さらに、本プロジェクトの研究者の多くがキャリアアップを実現し、一部は日本をリードする研究者となっている。日本科学未来館のメディアラボで、グリッドグラフを対象としたパスの数え上げに関する展示が大反響を呼び、海外からのコメントも多数寄せられた[104]。

(1)ZDD 構造の社会への応用

現代の社会においては、社会的な重要なインフラとして通信ネットワーク、下水道ネットワーク、電気供給ネットワーク等様々な巨大ネットワーク構造が存在し、故障からの迅速な復旧等による信頼性や経済性の確保が重要な課題であり、この解決に ZDD 構造等を使った列挙アルゴリズムが注目されている。本プロジェクトとその後の研究で、NTT との共同研究による 100 ノードの通信ネットワークや早稲田大学との 235 スイッチを持つ配電ネットワークについて、ZDD 適用による列挙の効率化を狙った最適配置や設計問題の研究が行われた。これらの研究は現段階における最大のノード数の条件での実用化のための実験的研究と見ることができる。現在現場のシステムにすぐに応用できるまでには課題があるが、これら課

[104] 終了報告書より

題の解決で、安全で経済的な通信ネットワークや配電ネットワークシステムの実現まで技術レベルは上がっており、今後の波及・発展が期待される。

(2)マテリアルズ・インフォマティクス(MI)、バイオ・インフォマティクス(BI)への波及

素材関連の事業における材料探索は、従来多くの材料作成と実験によって行われてきた。そのため材料研究の歴史が長い日本は、現在世界に比べて優位性を保っている。しかし、最近の MI 発展とともに、その優位性が揺らごうとしている。例えば、世界の素材に関連する国際特許の件数は、ここ 20 年程度で 3 倍強に増加している[105]。中でも、従来少なかった中国、韓国の公開特許件数がそれぞれ約 140 倍、約 43 倍と飛躍的に増加している(日本は約 8 倍、米国は約 2 倍)。今後、MI が普及すれば、米国や日本との素材開発競争が一段と激化する可能性がある。

日本はそのために、MI への取り組みを強化した。日本での MI の取り組みを図 3-3 に示した。日本ではこの図で示すように、国家プロジェクトとして NIMS や産総研などが中心に MI による材料開発に取り組んでいるだけでなく、材料企業やベンチャー企業の取り組みも多くみられるようになった。

図 3-3　日本と米国の MI に関する取り組み[106]

一方、バイオ・インフォマティクス(BI)においても、MI と同様に膨大なデータを扱い、情報科学により研究を促進する動きが世界的に行われている。日本においても文部科学省(文科省)や国立研究開発法人、様々な大学を中心に人材育成や研究が行われている。しかし、バイオ・インフォマティクスの国際会議での中国の存在感の増大と日本人の参加者の先細りを指摘する声がある一方で、ビッグデータをフルに活用して、創薬に革新をもたらすことへの期待も表明されている[107]。

---

105 https://www.smbc.co.jp/hojin/report/investigationlecture/resources/pdf/3_00_CRSDReport090.pdf#page=6&zoom=auto,-15,71

106 https://www.smbc.co.jp/hojin/report/investigationlecture/resources/pdf/3_00_CRSDReport090.pdf

107 https://bio.nikkeibp.co.jp/atcl/column/16/112500015/123100051/

津田自身、NIMS において 2015 年 7 月から始まった情報統合型物質・材料開発イニシアティブ(Mi^{2}i)において、データ科学グループリーダーとして研究をリードするとともに、理研の革新知能統合研究センターの分子情報科学チームのチームリーダーとしてバイオとマテリアルの両方の分野に人工知能技術を適用する研究を行った。本プロジェクトで取り組んだ熱伝導やナノ積層薄膜の研究や、RNA 分子の逆フォールディングのアルゴリズム研究等は機械学習等の手法を駆使して、両方のインフォマティクスの流れに沿った研究であった。また、ベイズ最適化やモンテカルロ木探索の手法を駆使して、現実の課題を解決するアルゴリズムである ChemTS や MCTS-RNA、TS-MD など、多くのソフトウェア開発と公開を行い、MI、BI の研究者へ提供することによって MI、BI 研究に貢献している。このようなアルゴリズム開発だけでなく、他の専門家との共同研究を通して、有機材料の自動設計や分子動力学に基づく分子の動きの計算等有益な知識を生み出すなどの成果を生み出している。また、日経産業新聞の 2016 年 4 月 14 日に「東京大学教授津田宏治氏――機械学習でビッグデータ解析、医学から材料開発貢献(次世代の先導者)」との記事が掲載された。さらに、後述するベンチャー企業 2 社のアドバイザを務め、日本に於ける MI と BI の発展に寄与している。個々の研究の成果が具体的な商品等に結びついた例は見いだせなかったが、本プロジェクトの研究と研究者が、近年、MI・BI の分野へ大きな波及効果を与えていることが分かった。

(3)機械学習によるセンシングデバイスの実現

ナノポアを通る分子を電流パルスによって検知、同定する技術を鷲尾は大阪大学の谷口と共同で開発した。その結果、谷口によるアイポア株式会社の創設に繋がった。アイポア株式会社と株式会社アドバンテストとの共同で微粒子計測器「nanoSCOUTER」を 2019 年に発売した。医薬品やバイオテクノロジーなどの研究開発に利用できるだけでなく、将来は医療を含む幅広い分野への応用を目指している。

さらに、NIMS との共同研究により開発された臭気センサ(MSS)は、NIMS によって様々な企業との共同による実用化研究に進展している(MSS フォーラム)。このセンサは BS フジ「ガリレオ X」(2017 年 11 月 26 日 11:30～12:00)で取り上げられ、紹介された。臭気センサは、人間が持つ五感の中で最も開発が遅れていたセンサであり、このセンサが実用化されれば、ガスの臭気検知、果物の熟成の検知等様々な応用が期待できる。

3.2.2 ベンチャー企業の創出

本プロジェクトの技術の利用によって創出されたベンチャー企業は、株式会社 Magne-Max Capital Management(MMC)[108]、MI-6 株式会社(MI-6)[109]、アイポア株式会社(アイポア)[110]の 3

[108] https://www.magne-max.com/
[109] https://mi-6.co.jp/
[110] https://aipore.com/

社である。これらの他にグループリーダー(津田)がアドバイザを務めるベンチャーが 1 社ある。

(1)MMC

MMC は、2011 年 6 月 20 日に設立された(資本金 9500 万円)。CEO の岡田克彦は関西学院大学の教授も兼務している。現在ヤフー株式会社の傘下として活動している。MMC は行動ファイナンスの研究者、データマイニングの研究者、機械学習の研究者、強化学習の研究者、深層学習の研究者、アルゴリズムの実装技術者、金融の実務家で構成される投資助言会社である。金融市場には毎時刻々と大量のニュースが流れている。これに加えて個人投資家の動向を示すソーシャルなデータなどの情報は膨大である。このような大規模データの取り扱いは人間には不向きであるため、MMC 社では、自然言語処理のアルゴリズムや機械学習アルゴリズムを実装し、これらの大規模データを解析している。MMC には本プロジェクトの技術員であった東と研究員であった倉井が役員として、研究員であった植野が Chief Research Officer として参画しており、湊がアドバイザを務めている。

(2)MI-6

MI-6 は 2017 年 11 月 17 日に設立された(代表：木嵜基博)。起業に当たり木嵜が本プロジェクトのグループリーダーの津田を共同創業者に迎えた[111]。現在、津田はアドバイザである。MI-6 は IT を活用した材料開発の研究、開発、コンサルティング、プロジェクトマネジメントをする事業を展開している。業務内容は、材料開発に関するチーム構成支援、材料実用化までの各種プロセスの研究、開発、コンサルティングである。

2018 年、キシダ化学株式会社と共同で、東京大学津田研究室、三井物産株式会社の協力を得て、リチウムイオン電池の性能向上のための電解液組成予測にマテリアル・インフォマティクス(MI)を活用し、電池の性能発現に有効な難燃性電解液組成を発見した[112]。その他、2018 年度の NEDO 助成金(AI 分野)の獲得、Tech Planter2018 竹中工務店賞受賞、イノベーションリーダーズサミット Top20 選出、未来 2019 ファイナリスト、平成 30 年度起業家万博ファイナリスト、GRASSHOPPER DAY 2019 WINTER 2 位入賞、HONGO AI 最優秀賞受賞などの評価を受けている。

(3)アイポア

アイポアは 2018 年 9 月 20 日に大阪大学ベンチャーキャピタルからの資金により、人工知能を活用した IoT ナノ粒子センサの開発および販売を行う会社として設立された。大阪大学の教授であった谷口が創業者の一人として設立した。谷口、鷲尾が参画した ImPACT プロジェクトから発生したベンチャー企業である。本報告の 2.2.3(2)項で述べた「機械学習

[111] https://media.startup-db.com/interview/mi-6-kizaki
[112] https://mi-6.co.jp/files/pr180517.pdf

によるナノギャップ・ナノポアセンサの抵抗・電流パルス分析」の成果を基に、アイポアが独自に開発したセンサと AI の技術を組み合わせたアイポア AI 粒子識別センサを用いることで、ウイルス等の生態粒子だけでなく工業用途等のナノ粒子を 1 個単位で識別することを可能にした。また測定された結果はアイポア AI システムのサーバで解析され、即座に識別結果が分かる仕組みとなっている。アイポア AI 粒子識別センサは幅広い産業に応用することが可能となっており、大きな期待がかかっている。アイポアは大阪大学が中心となり開発したナノ粒子識別センサ技術の事業化に向けたプラットフォーム開発に取り組んでいくとしている。

さらに、CREST「計測技術と高度情報処理の融合によるインテリジェント計測・解析手法の開発と応用」と内閣府の革新的研究開発推進プログラム(ImPACT)「進化を越える極微量物質の超迅速多項目センシングシステム」[113](大阪大学の谷口、鷲尾が両方に参加)において開発した技術を基礎に、アイポアが開発した AI 粒子識別システムとの組み合わせにより、ごくわずかな電流波形の差異を判別し、粒子の種類を速やかに特定することができた。株式会社アドバンテストからその成果を取り入れた微粒子計測器「nanoSCOUTER」が 2019 年に発売された。谷口はその後非常勤取締役となっている。また、鷲尾が顧問となっている。

(4)モジュラス株式会社[114]

上記 3 社の他に津田が技術アドバイザを務めているのは、ベンチャー企業モジュラス株式会社(Modulus Discovery Inc.)(モジュラス)である。同社は東京とボストンに拠点を置く前臨床ステージの技術を基盤にした創薬企業である。2020 年 5 月の資金調達により 35 億円以上の資本金になっていると思われる。モジュラスのコアバイオロジーについての知見、最新鋭の創薬プラットフォーム、グローバルなネットワーク型研究体制を統合することで、独自の低分子創薬ポートフォリオと様々な研究開発提携を推進していくとしている。製薬企業やバイオテクノロジー企業において幅広い創薬・計算科学研究の経験を積んだチームが運営する、ネットワーク型創薬企業である。アステラス製薬株式会社との提携、東京大学や東京工業大学との共同研究も進めている。

3.2.3 アウトリーチにおける成果と波及効果

日本科学未来館のメディアラボで、第 11 期展示として、湊プロジェクトの成果展示(グリッドグラフを対象としたパスの数え上げに関する展示)を行った。「フカシギの数え方」と題した本展示は予想外の反響があり、2012 年 8 月 1 日から 2013 年 4 月 15 日までの期間中、20 万人を超える来場者数となった。また、展示のコンテンツとして作成した「フカシギの数え方」の動画(格子グラフを対象として組みあわせ爆発の様子をわかりやすく示したもの)は YouTube 上で現在 295 万アクセスを超えたヒットとなった。ニコニコ動画にもアップ

[113] https://www.jst.go.jp/impact/hp_miyata/organization.html
[114] https://modulusdiscovery.com/ja/%E3%83%88%E3%83%83%E3%83%97/

されている。「アルゴリズム技術」に関しては、従来なかなか一般人に理解してもらえるようなコンテンツはなかったが、本展示で開発したコンテンツはその意味で画期的なものであったと評価できる。YouTube 動画に対しては海外からのコメントも数多く寄せられており、国際的にも日本の技術をアピールできた。

【引用文献】

番号	文献名
[1]	Inoue Y., Toda T., Minato S.-I., “Implicit generation of pattern-avoiding permutations by using permutation decision diagrams”, IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, E97-A(9), 1171-1179 (2014)
[2]	Inoue T., Takano K., Watanabe T., Kawahara J., Yoshinaka R., Kishimoto A., Tsuda K., Minato S.-I., Hayashi Y., “Distribution loss minimization with guaranteed error bound”, IEEE Transactions on Smart Grid, 5(1), 102-111 (2014)
[3]	Tabei Y., Tsuda K., “SketchSort: Fast all pairs similarity search for large databases of molecular fingerprints”, Molecular Informatics, 30(9), 801-807 (2011)
[4]	Denzumi S., Kawahara J., Tsuda K., Arimura H., Minato S.-I., Sadakane K., “DenseZDD: A compact and fast index for families of sets”, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 8504 LNCS, 187-198 (2014)
[5]	Ito J.-I., Tabei Y., Shimizu K., Tsuda K., Tomii K., “PoSSuM: A database of similar protein-ligand binding and putative pockets”, Nucleic Acids Research, 40(D1), D541-D548 (2012)
[6]	Ishihata M., Sato T., Minato S.-I., “Compiling Bayesian networks for parameter learning based on shared BDDs”, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 7106 LNAI, 203-212 (2011)
[7]	Tabei Y., Kishimoto A., Kotera M., Yamanishi Y., “Succinct interval-splitting tree for scalable similarity search of compound-protein pairs with property constraints”, Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Part F128815, 2487637 (2013)
[8]	Inokuchi A., Washio T., “FRISSMiner: Mining frequent graph sequence patterns induced by vertices”, IEICE Transactions on Information and Systems, E95-D, 1590-1602 (2012)
[9]	Shimizu S., Inazumi T., Sogawa Y., Hyvärinen A., Kawahara Y., Washio T., Hoyer P.O., Bollen K., “DirectLiNGAM: A direct method for learning a linear non-gaussian structural equation model”, Journal of Machine Learning Research, 12, 1225-1248 (2011)
[10]	Kawahara Y., Washio T., “Prismatic algorithm for discrete D. C. programming problems”, In Advances in Neural Information Processing Systems, 24, 2106-2114 (2011)
[11]	Ting K.M., Washio T., Wells J., Liu T., “Density estimation based on mass”, Proc. of IEEE Int. Conf. on Data Mining(ICDM11), 715-724 (2011)
[12]	Wells J.R., Ting K.M., Washio T., “LiNearN: A new approach to nearest neighbour density estimator”, Pattern Recognition, 47(8), 2702-2720 (2014)
[13]	Sogawa Y., Ueno T., Kawahara Y., Washio T., “Active learning for noisy oracle via density power divergence”, Neural Networks, 46, 133-143 (2013)
[14]	Nishino M., Yasuda N., Minato S.-I., Nagata M., “Zero-suppressed sentential decision diagrams” In AAAI, 1058-1066 (2016)
[15]	Nishino M., Yasuda N., Minato S.-I., Nagata M., “Compiling graph substructures into sentential decision diagrams”, 31st AAAI Conference on Artificial Intelligence, AAAI 2017, 1213-1221 (2017)

[16]	Nishino M., Yasuda N., Minato S.-I., Nagata M., “Dancing with decision diagrams: A combined approach to exact cover”, 31st AAAI Conference on Artificial Intelligence, AAAI 2017, 868-874 (2017)
[17]	Nishino M., Yasuda N., Minato S.-I., Nagata M., “Efficient algorithm for enumerating all solutions to an exact cover problem”, NTT Technical Review, 15(11), 1-5 (2017)
[18]	Gao S., Ishihata M., Minato S.-I., “Separate compilation of Bayesian networks for efficient exact inference”, Transactions of the Japanese Society for Artificial Intelligence, 33(6) (2018)
[19]	Sakaue S., Ishihata M., Minato S.-I., “Efficient bandit combinatorial optimization algorithm with zero-suppressed binary decision diagrams”, International Conference on Artificial Intelligence and Statistics, AISTATS 2018, 585-594 (2018)
[20]	Suzuki H., Minato S.-I., “Fast enumeration of all Pareto-optimal solutions for 0-1 multi-objective knapsack problems using ZDDs”, IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, E101A(9), 1375-1382 (2018)
[21]	Nishino M., Inoue T., Yasuda N., Minato S.-I., Nagata M., “Optimizing Network Reliability via Best-first Search over Decision Diagrams,” Proc. of the IEEE International Conference on Computer Communications (INFOCOM 2018), 1817-1825, Honolulu, HI, USA, Apr. 2018.
[22]	Nishino M., Inoue T., Yasuda N., Minato S.-I., Nagata M., “Network reliability optimization by using binary decision diagrams”, NTT Technical Review, 16(11), 25-29 (2018)
[23]	Takenobu Y., Yasuda N., Minato S.-I., Hayashi Y., “Scalable enumeration approach for maximizing hosting capacity of distributed generation”, International Journal of Electrical Power and Energy Systems, 105, 867-876 (2019)
[24]	Minato S.-I., Kawahara J., Ishioka F., Mizuta M., Kurihara K., “A fast algorithm for combinatorial hotspot mining based on spatial scan statistic”, SIAM International Conference on Data Mining, SDM 2019, 91-99 (2019)
[25]	Terada A., Tsuda K., “Multiple testing tool to detect combinatorial effects in biology”, Methods in Molecular Biology, 1807, 83-94 (2018)
[26]	Yoshizoe K., Terada A., Tsuda K., “MP-LAMP: Parallel detection of statistically significant multi-loci markers on cloud platforms”, Bioinformatics, 34, 3047-3049 (2018)
[27]	Ueno T., Rhone T.D., Hou Z., Mizoguchi T., Tsuda K., “COMBO: An efficient Bayesian optimization library for materials science”, Materials Discovery, 4, 18-21 (2016)
[28]	Ju S., Shiga T., Feng L., Hou Z., Tsuda K., Shiomi J., “Designing nanostructures for phonon transport via Bayesian optimization”, Physical Review X, 7(2), 21024 (2017)
[29]	Sakurai A., Yada K., Simomura T., Ju S., Kashiwagi M., Okada H., Nagao T., Tsuda K., Shiomi J., “Ultranarrow-Band Wavelength-Selective Thermal Emission with Aperiodic Multilayered Metamaterials Designed by Bayesian Optimization”, ACS Central Science, 5(2), 319-326 (2019)
[30]	Yamashita T., Sato N., Kino H., Miyake T., Tsuda K., Oguchi T., “Crystal structure prediction accelerated by Bayesian optimization”, Physical Review Materials, 2(1), 13803 (2018)

[31]	Sano S., Kadowaki T., Tsuda K., Kimura S., “Application of Bayesian Optimization for Pharmaceutical Product Development”, Journal of Pharmaceutical Innovation, (2019)
[32]	M. Dieb T., Ju S., Yoshizoe K., Hou Z., Shiomi J., Tsuda K., “MDTS: automatic complex materials design using Monte Carlo tree search”, Science and Technology of Advanced Materials, 18(1), 498-503 (2017)
[33]	Dieb T.M., Ju S., Shiomi J., Tsuda K., “Monte Carlo tree search for materials design and discovery”, MRS Communications, 9(2), 532-536 (2019)
[34]	Yang X., Yoshizoe K., Taneda A., Tsuda K., “RNA inverse folding using Monte Carlo tree search”, BMC Bioinformatics, 18(1), 468 (2017)
[35]	Yang X., Zhang J., Yoshizoe K., Terayama K., Tsuda K., “ChemTS: an efficient python library for de novo molecular generation”, Science and Technology of Advanced Materials, 18, 972-976 (2017)
[36]	Shin K., Tran D.P., Takemura K., Kitao A., Terayama K., Tsuda K., “Enhancing Biomolecular Sampling with Reinforcement Learning: A Tree Search Molecular Dynamics Simulation Method”, ACS Omega, 4(9), 13853-13862 (2019)
[37]	Blöbaum P., Shimizu S., Washio T., “A novel principle for causal inference in data with small error variance”, ESANN 2017 - Proceedings, 25th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, 347-352 (2017)
[38]	Blöbaum P., Janzing D., Washio T., Shimizu S., Schölkopf B., “Analysis of cause-effect inference by comparing regression errors”, PeerJ Computer Science, 2019(1), e169 (2019)
[39]	Tsutsui M., Yoshida T., Yokota K., Yasaki H., Yasui T., Arima A., Tonomura W., Nagashima K., Yanagida T., Kaji N., Taniguchi M., Washio T., Baba Y., Kawai T., “Discriminating single-bacterial shape using low-aspect-ratio pores”, Scientific Reports, 7(1), 17371 (2017)
[40]	Tsutsui M., Yoshida T., Tanaka M., Yokota K., Arima A., Tonomura W., Taniguchi M., Okochi M., Washio T., Kawai T., “Multimodal resistive pulse analysis using a low-aspect-ratio nanopore”, 22nd International Conference on Miniaturized Systems for Chemistry and Life Sciences, MicroTAS 2018, 2, 754-757 (2018)
[41]	Taniguchi M., Ohshiro T., Komoto Y., Takaai T., Yoshida T., Washio T., “High-Precision Single-Molecule Identification Based on Single-Molecule Information within a Noisy Matrix”, Journal of Physical Chemistry C, (2019)
[42]	Tsutsui M., Yokota K., Yoshida T., Hotehama C., Kowada H., Esaki Y., Taniguchi M., Washio T., Kawai T., “Identifying Single Particles in Air Using a 3D-Integrated Solid-State Pore”, ACS Sensors, 4(3), 748-755 (2019)
[43]	Washio T., Imamura G., Yoshikawa G., “Machine learning independent of population distributions for measurement”, Proceedings - 2017 International Conference on Data Science and Advanced Analytics, DSAA 2017, 2018-January, 212-221 (2017)
[44]	Imamura G., Shiba K., Yoshikawa G., Washio T., “Analysis of nanomechanical sensing signals; physical parameter estimation for gas identification”, AIP Advances, 8(7), 75007 (2018)
[45]	Imamura G., Shiba K., Yoshikawa G., Washio T., “Free-hand gas identification based on transfer function ratios without gas flow control”, Scientific Reports, 9(1), 9768 (2019)