

ERATO 湊離散構造処理系プロジェクト事後評価（予備評価）報告書

【研究総括】湊 真一（北海道大学大学院情報科学研究科／教授）

【評価委員】（敬称略、五十音順）

浅野 哲夫（北陸先端科学技術大学院大学情報科学研究科／研究科長）

安達 淳（委員長；国立情報学研究所／副所長）

今井 桂子（中央大学工学部情報工学科／教授）

宇野 裕之（大阪府立大学理学系研究科／准教授）

久光 徹（日立製作所中央研究所ライフサイエンス研究センタ／主管研究員）

評価の概要

計算機は、産業プロセスの最適化や解析、マーケティング、バイオインフォマティクスなど、様々な情報処理に活用されている。近年の爆発的に増大している大規模データを処理するためには、計算機ハードウェアの高速化だけでなく、膨大な離散構造データを数学的に簡約化し効率よく計算する「アルゴリズム技術」の重要性が高まっている。ERATO 湊離散構造処理系プロジェクトは、基本的な離散構造の1つである論理関数を処理する BDD と、さらにその進化形である ZDD の技法を発展させ、多様な離散構造を統合的に演算処理する技法を体系化し、システム検証や最適化、データマイニング、知識発見などを含む分野横断的かつ大規模な実問題を高速に処理するための技術基盤の構築を目指したものである。

本プロジェクトの研究体制は、離散構造処理基盤グループ、機械学習・充足応用グループ、統計・マイニング応用グループの3つの研究グループで構成され、湊研究総括の強力なリーダーシップのもと、離散構造処理の基盤研究から実装・応用研究まで、各グループが有機的に連携を図り研究を推進している。

プロジェクトが発足して約4年が経過し、社会に大きなインパクトがある数々の興味深い研究成果が出ている。様々な問題の解を列挙するだけでなく、全ての解集合を計算機のメモリ内で索引化し ZDD を一括生成するフロンティア法の開発を行い、従来の手法から計算空間の大幅な削減を実現した。本手法は「圧縮－列挙－索引化」という新しい離散構造処理の世界を今、真に切り拓くものであり、さまざまな社会インフラの解析等への応用が期待できる。本手法をスマートグリッド電力網へ応用し、天文学的に膨大なスイッチ構成の組合せを調べ、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計算することに成功した。また、科学データの統計検定において、従来よりも格段に正確な補正P値を計算できる統計手法を開発した。近年、膨大な観測データをもとにした解析が研究の重要な手段となっている場合が増えており、本手法を実験解析結果の検定に適用することで、従来見過ごされてきた要因を発見することが可能となる。本研究成果は、実験系のあらゆる学問に貢献するものであり、今後、世界中で広く利用されることが期待されている。また、ZDD を用いたパス列挙のための高速アルゴリズムの開発を行い、グリッドグラフにおけるパス列挙の数え上げの世界記録を更新した。この成果は、日本科学未来館のメディアラボで「フカンギの数え方」として展示を行い、一般の方にアルゴリズム技術のおもしろさと社会的重要性をわかりやすく示した。

以上のとおり、ERATO 湊離散構造処理系プロジェクトは、卓越した研究水準を示し、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」に資する成果が得られるであろうと評価できる。

1. 研究プロジェクトの設定および運営

1-1. プロジェクトの全体構想

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。およそ計算機が扱うあらゆる問題は、単純な基本演算機能を要素とする離散構造の処理に帰着される。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。応用分野としては、ハードウェア/ソフトウェアの設計、大規模システムの故障解析、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、バイオインフォマティクス、web 情報解析など多種多様であり、現代社会に対する大きな波及効果を持つ。

そのような背景のもと、本プロジェクトの構想は、従来は個別独立であることが多かった「応用領域に依存しない普遍的・抽象的な計算の基礎理論」と「個別の工学的応用領域に特化したエンジニアリング」の研究領域を有機的に繋ぐ実装技術の体系が存在すべきであるという信念に基づくものである。この領域では、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が、各応用分野の固有の問題にアドホックに対応するのではなく、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る。これは既存の研究分野としては存在しなかった新しい視点であり、大いに独創性が認められる。

本プロジェクトでは、BDD (Binary Decision Diagram:二分決定グラフ) およびその進化形である ZDD (Zero-Suppressed BDD; ゼロサプレス型 BDD) の技法を基盤とし、様々な問題に適用可能な枠組みの検討を行い、これらを整理・体系化するとともに、より高次の汎用的な演算処理系を構成することを目指している。最適化の研究は世界中で多く行われているが、列挙系の研究では本プロジェクトが世界の最先端を進んでおり、さらに多くの応用領域で重要な技術となる可能性が高いと期待できる。我が国の産業競争力という観点から、重要な項目の一つであることは間違いなく、本研究プロジェクトの全体構想は ERATO にふさわしいものである。

1-2. プロジェクトの目標・計画

本プロジェクトでは基盤：離散構造処理の体系化と基盤ソフトウェアの整備、応用：実例題への適用と効果の提示、普及：人材育成とコミュニティの確立の3つの目標を設定しており、それぞれにアプローチ方法を設定している。

基盤では、さまざまなデータ構造や目的に応じた処理系を検討し、さらに BDD/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・普及活動を行うことを計画した。

ここでは、科学と工学の中間に位置するアート層を BDD/ZDD を核とする「離散構造処理系」と定義し、「圧縮-列挙-索引化」の一連の処理から成る枠組みであると考え、さまざまな実用分野で実際に動作するプログラムやそれらの構成要素であるライブラリ群を作り提供することを目標に設定している。この目標はきわめて妥当であり、社会にも有益であると考えられる。

応用では、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をとって研究を進め、訴求力のある大規模な実例題を基にして、その効果を発見力の大幅な改善等、定量的に示すことを計画した。バイオ・ゲノム応用や、ビッグデータ向けのアルゴリズム研究開発、ビジネス・

統計マイニング応用の研究の推進に大いに期待が持てる設定である。

普及では、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ことを計画した。ワークショップや、セミナーなどを通して、新鮮なアイデアを真剣に議論でき、また時には新しい共同研究に発展することが考えられる。ただし、この目標を達成するための計画は、やや具体性や詳細度を欠いていたと感じられる。もう少し詳細に決めておいた方が良かったと考える。

1-3. プロジェクトの運営

本プロジェクトは、BDD/ZDD を基盤として、様々な問題に適用可能な枠組みを検討し、汎用的なアルゴリズムを開発し、プロジェクト内の研究グループに情報を発信する中心拠点である「離散構造処理基盤グループ」と、バイオ・ゲノム応用や、ビッグデータ向けのアルゴリズム研究開発を行う「機械学習・充足応用グループ」と、ビジネス・統計マイニング応用と非専門家向けソフト開発を行う「統計・マイニング応用グループ」で構成されている。3つの拠点で、異なる分野をカバーするとともに、各拠点間は高品質 TV 会議システムで常時接続されており、研究テーマに応じて密に連携して研究推進できる仕組みを取り入れている。

また、意欲的な若手研究者の育成が重要な柱になっており、研究コミュニティの活性化に寄与するように研究運営体制が作られている。ERATO 雇用メンバーとしては、当該分野あるいは周辺分野の若手の有力研究者を集めている。またすでにポストを得ている若手の有力研究者は共同研究者として密な連携を取っている。特に春と秋に開催しているプロジェクト主催のワークショップに参加して、相互の研究内容に関して議論を交わしているほか、前述の TV 会議システムをフルに活用し、定期的開催される ERATO セミナーなど個別の研究打ち合わせに活用している。また、統計・マイニング応用グループにおいては、他分野や企業の研究者との交流・連携が行われている。

研究費の執行状況は、研究員の人件費が大きな割合を占める。主要機器としては大規模グラフデータ等を用いた計算実験用として大容量メモリーサーバーの導入を行い、グループ間でも有効に活用されている。

2. 研究の達成状況および得られた研究成果

2-1. 離散構造処理基盤グループ

離散構造処理基盤グループは、BDD や ZDD をはじめとした離散構造処理の基盤的技術に関して、フロンティア法と呼ばれる効率的な探索手法や、系列集合・順列集合に関する研究開発を行い、社会的に重要な様々な実問題に対して実験を進め本手法の有効性を示している。

BDD/ZDD は、グラフを列挙索引化するためのデータ構造としても有用である。さまざまな制約条件をもつグラフ構造から、その解集合を表現する ZDD を高速に構築する方法として Knuth が近年示したパス列挙アルゴリズム Simpath が知られているが、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。本手法は様々な問題の解を列挙するだけでなく、全ての解集合に対して索引化構造を構築するものであり、「圧縮・列挙・索引化」という新しい離散構造処理の世界を今まさに切り拓きつつある。

当プロジェクトで開発したフロンティア法をもとに、スマートグリッド上の効率的な計算手順の開発に取り組み、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) の組合せを列挙することに成功した。その結果、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。

電力網の最適化や安定化の問題は、社会経済に与える影響が極めて大きく、潜在的な波及効果が極めて大きい研究成果である。また本成果は、社会的な意義と共に、「圧縮・列挙・索引化」と「最適化」を融合するものである点で大きな意味があると考えられる。さらにフロンティア法は、電力網だけでなく、水道、ガス、鉄道、道路、通信、避難所配置、選挙区割りなど、様々な社会インフラの解析に応用できることから今後の活用が期待されている。

また、フロンティア法の応用事例の一つとして、ZDD を用いたパス列挙のための高速アルゴリズムの開発を行っている。ここではグリッドグラフでのパス列挙について、ZDD を用いた数え上げ手法の最適化に取り組み、 $n=21(21 \times 21)$ のグリッドグラフでの列挙で世界記録を達成した。その後 $n=26$ までの数え上げに成功している。本手法は、将来的には、スケジューリング問題や設計問題、信頼性評価等への応用が期待できる。

Bryant 教授が 1986 年に発表した BDD の計算量に関する予想は、これまで 25 年間、正しいと信じられていたが、本プロジェクトで初めて反例を発見した。BDD の根幹に関わる理論への貢献により、本プロジェクトの技術レベルを世界にアピールした成果であるといえる。本成果においては、どのような状況の時に否定的な結果になるのかを示すことができれば、実問題を取り扱う時の参考になると考える。

系列集合の処理アルゴリズムの開発では、文字列データなどの順序関係を持つ系列集合（シーケンス）に対する BDD/ZDD の拡張を行った。また組合せ集合だけでなく、順列集合を対象とした拡張を検討し、 π DD として実験システムを構築した。今後、フロアプラン（配置問題）への応用や、制約付き整列問題への応用展開が期待される。

ZDD の基本パッケージについては、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 として公開している。今後はパッケージを周知し、多くのユーザに使ってもらうための工夫が必要である。

2-2. 機械学習・制約充足応用グループ

機械学習・制約充足グループは、バイオインフォマティクス分野で整備されつつある化合物データや大規模なシーケンス配列データを対象として、大規模離散構造処理技術として検索技術やソーティング技術、また確率を統合した論理推論技術の開発を行っている。また、離散構造処理基盤グループや共同研究者と連携し、BDD/ZDD の並列化技術の実装・実験や電力ネットワーク等の実レベルの大規模データに対する組合せ列挙、最適化技術の開発に取り組んでいる。

自然科学さらには社会科学分野の研究において、新しい現象を見つけたときには、その結果の信頼性を統計的に担保する必要があるが、誤発見の確率を示す検定値（P 値）が広く用いられてきた。しかし、従来の統計検定手法は観測できる対象が増えれば増えるほど、発見の基準を厳しくする必要があり、その結果、観測対象が増えたのに、科学的発見が減るといった現象が起きることがあり、有意義な実験結果が不当に低く評価されることがあった。

本研究グループでは、出現頻度の低い組合せは誤発見率を変化させないという数理的性質に着目し、従来法より格段に正確な補正 P 値を算出する新手法 LAMP（Limitless-Arity Multiple testing Procedure）を開発した。この手法を、乳がん細胞株の増殖・分化に関与している転写因子の研究に利用したところ、既存の遺伝子発現データから新たな組合せ因子を発見することに成功した。この新手法は、様々な実験の結果の検定に適応することにより、従来分からなかった要因を再発見することに寄与する可能性もある。また、新たに行う実験からの組合せ要因の発見に大いに寄与すると期待される。本手法はビックデータから新たな科学的発見をもたらす統計手法として、米国科学雑誌「米国科学アカデミー紀要(PNAS)」に掲載された。このことは他分野においても、本プロジェクトの研究成果が認められたことを意味し評価に値する。

また、遺伝子データ等の超大規模データを実時間で分類・検索、ソーティング、マイニング処理するための技術開発を行った。具体的には、テキストデータや画像データを対象にした効率的なソーティング技術や、ソーティングによって高速に全ペア類似度検索を行う複合ソート法を提案し、数千万オーダーの化合物からなるデータや大規模配列データに対して適用実験を行ない、Web 上にその結果をデータベースとして公開した。これらの技術はバイオデータのみならず、大規模なデータを扱う離散構造処理技術にとって基盤的な要素技術と位置付けられる。

さらに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ている。ここでは、代謝ネットワーク等で使用されるベイジアンネットワークにおいて組合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。確率推論に関して、背景知識の論理式を BDD で表現し、その上で推論を行う手法は、これまで組合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大規模確率推論に対して、効率的な計算方法を与える。本手法は因子分析など広汎な応用が考えられ、産業界と連携して実例をもとに推進することを期待する。

モンテカルロ木探索では、膨大な組合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している。ここでは本手法を用いた大規模並列化計算により囲碁プログラムの大幅な強化を実現し、他分野への応用が可能な基盤技術として期待される。

以上、いわゆる「きれいなデータ」を対象としてアルゴリズムの精緻化を求めるという、理論家に閉じたアプローチではなく、個別研究分野の問題を解くために、当該分野の研究者と密に交流しその分野の知識の理解なども含めてアルゴリズムの面から協力し、斬新なアルゴリズムの提案や処理の高速化など、個別分野で従来のままのアプローチでは見いだすことができなかった解法を実現したことを評価する。

2-3. 統計マイニング応用グループ

統計・マイニング応用グループは、離散構造処理系の基盤技術を、データマイニングと知識発見、機械学習と自動分類、統計解析・モデリング、最適化への応用を行っている。

行動履歴やログデータ等、インターネットをはじめとしてさまざまな産業データ、あるいは個人データの蓄積が進んでいる。こうした膨大な非定型データから因果関係を発見し、活用していくことは次世代の情報活用基盤として極めて重要な意味を持っている。このような観点から、単なるフラットなデータ構造ではなく、相互に因果関係に代表される関係を持つようなデータを木構造あるいはグラフ構造として表現し、これらを効率的に取り扱う仕組みを開発しているほか、実際のビジネスデータを用いたマイニング検証実験、離散構造処理基盤グループの成果を活用して開発されたソフトウェアの開発・公開にも取り組んだ。

離散構造処理基盤グループで開発したアルゴリズムを実装し、ソフトウェアライブラリ NYSOL として公開した。主なアプリケーションを以下に示す。Ekillion は JR の大都市近郊区間内の経路を対象に、指定条件に合わせて全パスを高速列挙するアプリケーションである。ここではパス列挙には離散構造処理基盤グループで開発した Graphillion を使っている。また、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ている。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っている。さらに、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。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、現在公開に向けた開発を続けている。最近では、増々大規模データの解析の必要性が叫ばれているが、その利用が専門の研究者に限られているのが現状である。NYSOL ライブラリにて統一的インタフェースで実装された湊プロジェクトのアルゴリズム

ムは、数理系の研究者だけではなく、社会科学系の研究者も手軽に利用できるようになってきている。このことは研究成果の社会還元の見点から非常に重要である。

機械学習・統計・データマイニングの応用研究については、物事の間接性を表すのに適したデータ構造であるグラフ系列から知識を発見するために、膨大なグラフ系列から、重要なグラフ構造の変化を発見する様々な技術に取り組んでいる。また、因果構造探索のためのデータ解析法の開発では、データに潜む因果構造を推定するための統計的方法を開発している。さらに、千～数十万次元といった極めて高次元なデータを扱う場合、事象の組合せが莫大になり、実用的な計算が不可能になってしまうが、データが持つ離散的な構造、特に劣モジュラ性（離散凸性）を用いることにより、この計算を可能とするような効率的なアルゴリズムの開発に取り組んでいる。何れの研究も従来の技術的枠組みを改良するものではなく、世界的に取り組まれてこなかったテーマやアイデアによるものである。これは、特にゲノムワイド相関解析に適用できる新しい手法であり新たな国際的研究課題を創出するものである。

さらに、本研究グループでは、離散構造処理基盤グループで開発した基礎アルゴリズムをマーケティング・金融分野に適用した。ここでは、ビジネスデータを用いたマイニング検証実験として、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ている。金融への応用として市場心理に基づく株価予測を行い、その成果は、関西学院大学の経済学の専門家らと連携し、産業界での実用化を視野にいれて検討を行っている。またマーケティングへの応用では、ブログや Twitter を解析することで、ホテル評価サイトの解析、オリンピック関連 tweet の解析、TV 番組関連 tweet からのトピック検知を行った。ビジネスなどの個別応用分野で、従来のアプローチにない解法を、統計・データマイニングから提案し、具体的な手法を示し、新たな問題解決の道を開拓したことを評価する。

以上に基づき研究成果を俯瞰すると、BDD/ZDD を核に、この手法を精緻化して様々な問題に適用できるようなレベルに高度化したこと、アルゴリズムを現実の個別分野の問題に適用し、当該ドメインで従来実現できなかった問題に解を与えたこと、そして科学実験の組み合わせ要因の発見のための新しい多重検定法という形で、広い範囲の研究に寄与する方法論を開拓したこと、の三つの具体的成果が得られている。特に、「離散構造データの圧縮・列挙・索引化」という考え方の有効性を実際に示した点での功績は大である。列挙自体は従来から様々な研究が行われているが、本プロジェクトでの成果は、これらを索引化し、それに基づいて演算処理をするという斬新なアイデアによるものである。国際的にも評価の高い ZDD の提案を端緒として、高速データ探索法、統計・機械学習手法での活用など、関連するアルゴリズムや適用方法を具体的に提示し、広い問題に適用できる環境を実現した。今後さらに社会的重要性が高い問題について応用分野にて活用を図り、有効性を示してほしい。

また、離散構造処理系に関して、アルゴリズムの精緻化にとどまることなく、容易にアルゴリズムを使うことができるようなソフトウェアの実装と普及、現実問題への適用、他分野研究者の持つ問題を従来にない方法で解決するなど、様々な問題解決を行った。公開ソフトの使い方のマニュアルによって、普及するかどうかが決まるところがあるので、プロジェクト終了後の対応についても配慮が必要だと思われる。

論文および学会・会議での発表においては、発表件数は多くないが、応用領域においてインパクトファクターの高い専門論文誌に発表しており、今後も引き続き努力して頂きたい。さらに成果を学術と産業へ還元するために、前記の俯瞰に基づいて知識を体系化し、専門書の執筆を進めていただきたい。また、それと並行して、理想的には非専門家でも抵抗なく使えるレベルのツールを作成し、効果的な例題と共に公開して頂きたい。最終年度には、是非これらの実現を目指して欲しい。

3. 研究成果の科学技術、社会・経済への貢献

3-1. 科学技術への貢献

論理関数、組合せ集合を効率的に扱うことの出来る BDD/ZDD に関する研究の重要性は、これまでも理論計算機分野では知られてきていたが、それを実用分野に適用し、その有効性を示すことが出来た点は高く評価できる。また、本プロジェクトの研究成果が科学技術全般におけるツールとして使用できる可能性を示した。

最適化技術は世界的に競争が激しいが、列挙に関する研究はあまり行われておらず、伝統的に日本が強い研究分野である。ドイツの大型科研費「Algorithm Engineering」のコミュニティや、google/IBM/MS などの研究所、DIMACS などは実用的な問題に関心があるが、最適化が中心で列挙に注目する人はまだ少ない。また、これまでの列挙に関する研究は数学的興味に偏っていたが、本プロジェクトの圧縮列挙索引化技術により、科学技術全般におけるツールとして使用できる可能性が示された。

応用分野では、データ工学、知識発見とマイニング、機械学習、確率推論、人工知能、充足解探索、システム設計などの分野があるが、日本から分野横断的な組織作りを行い海外に対して優位性を示す必要がある。

特筆すべき成果は、スマートグリッドを実現するための電力網の最適設計方法を見いだしたことである。これは当該ドメインではなかった全く新しいアプローチの提案であるだけでなく、問題解決の新しい局面を切り開くものである。社会的にも電力網最適化に大いに貢献すると期待される。

次の大きな成果は、科学実験での組合せ要因の発見に寄与する新しい多重検定法の提案である。これは、従来有意であると判定されなかった実験結果を有意なものとして再発見する可能性を与えるものであり、また今後様々な実験結果の検定に活用されるものとして期待される。科学的手法の進歩に大いに貢献する研究成果といえる。

アルゴリズムの精緻化のみならず、高速化や大規模化など関連するアルゴリズムを合わせて開発し、しかも現実問題に適用するためのソフトウェアも開発した。これは単にアルゴリズムの普及のみならず、当該ドメインでの新しい問題解決の道を与えるもので、ビッグデータに関係する広い分野に影響を与えると期待される。

3-2. 社会・経済への貢献

最適化アルゴリズム等が、電力網、バイオインフォマティクス分野などで使われることにより、スマートグリッドの進展や新しい病気の発見手法につながり、社会的に大いに貢献することが期待される。電力網に関する結果は、電力が経済基盤になっている現代では経済的関心も高く、また社会一般で電力の問題が話題になっているので、これに貢献できたことは社会的課題の解決に貢献できることを示したと言える。また、LAMP 手法についても、大量データからの有効パターン発見の分野で、今後大きな役割を果たす事が期待できる。ただし、上記の貢献が実現するためには、教科書のような形での成果の整理と、ツール群の公開が不可欠である。本プロジェクトでは、後者に関する取り組みは既に実施されており、前者にも是非着手して頂きたい。また、列挙しなければ解けない問題と、そうでない問題の見極めについて、今回のプロジェクトで得られた知見があれば、それをまとめることも重要であると考えられる。

本プロジェクトは探索手法や最適化手法という数理的な手法を基盤としており、その手法を適用して解析できる分野は広い。研究期間終了前の評価実施段階では、具体的な応用での成果がまだ多く出ているという段階ではないため、研究取りまとめに当たって、今後さらに多くの適用分野が開拓され

るような工夫をすることが望まれる。

4. その他特記すべき事項

4-1. 若手研究者支援

アルゴリズム開発能力の高い若手人材を育てることに尽力しており、また、流動性を高めるために真摯に研究指導しており、その成果が若手の成長として実証されている。若手人材は、基礎的な研究領域に閉じこもることなく、他分野の研究者と協力し問題解決を行うような志向性を涵養されている。ビッグデータ処理が様々な問題に適用されていることが必要とされている現在、本プロジェクトで育成された気鋭の人材の今後の活躍が大いに期待される。研究者の雇用にあたっては、敢えて BDD/ZDD の非専門家を選択し、実問題の解決を通して育成し、研究職の獲得につなげている。また、多くの若手が参加する定期的なセミナーを開催しており、人脈の育成は行われていると考える。海外展開が国内活動に比べて不足気味なのと、学会・研究会としての組織化がまだ充分でないことは、現状の課題であると考えられる。

4-2. アウトリーチ活動

「フカシギの教え方」について、日本科学未来館をはじめとして、北大文化祭、北大博物館での展示を行い、多数の来場者に対して問題意識やプロジェクト成果の宣伝活動を行った。本プロジェクトで扱っている抽象的な問題を、一般の方に容易に理解してもらえるようにするアウトリーチ活動を積極的に行っている。最もアウトリーチしにくい分野でのこの努力は大いに評価できる。特に、YouTube を活用して、「フカシギの教え方」によりアルゴリズムの有効性に関する啓蒙を行い、140 万を超えるビューを達成したことは特筆すべきである。

真の意味での技術の浸透は、分かり易い教科書の出版と、ツール群の公開があって初めて可能となる。本プロジェクトはそもそもそれを指向しており、今後とも是非それに注力して頂きたい。

5. 総合評価

本プロジェクトは、基本的な離散構造の 1 つである論理関数を処理する BDD と、さらにその進化形である ZDD の技法を発展させ、多様な離散構造を統合的に演算処理する技法を体系化し、システム検証や最適化、データマイニング、知識発見などを含む分野横断的かつ大規模な実問題を高速に処理するための技術基盤の構築を行った。アート層の創出・開拓というプロジェクトの理念、および一般利用可能なプログラム群の構築という目標は、これまで強調されてこなかった視点であり独創的である。特に、提案した手法を、スマートグリッドにおける電力網設計問題や超高速アルゴリズムを用いた新たな統計検定手法の二つに代表される重要課題に対して適用し、当該ドメインで顕著な成果を上げることができた。これは、本プロジェクトの研究成果である BDD/ZDD およびその発展技術による「圧縮・列挙・索引化」手法が、社会的に重要な様々な応用をもたらす重要な基盤技術であることを明確に示しており、今後のビッグデータ処理時代のアナリティックスを担う一領域を開拓すると大いに期待するものである。

また、一般の理解を得にくい抽象的な学問領域において、一般の方へのアウトリーチ活動を積極的に行い、このような問題の重要性と本来的な困難さの理解を向上するとともに、これら研究分野がどのように役立つかの直感的理解を増大させたことは大いに評価できる。

今後の課題として、開発したソフトウェアツールの一層の整備・公開と、専門書等の執筆が挙げられるが、これらは社会への成果還元観点から非常に重要である。是非とも優先して行ってほしい。また、理論・応用ともにプロジェクトの研究成果が外部にて、さらに多く参照・引用されるように、プロジェクトの残りの研究期間で、他分野を巻き込んだ研究成果の普及が期待される。

以上のとおり、ERATO 湊離散構造処理系プロジェクトは、卓越した研究水準を示し、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」に資する成果が得られるであろうと評価できる。