審 査 の 結 果 の 要 旨

氏　名　久保山　哲二

本論文は、「Matching and Learning in Trees（木の照合と学習）」と題し、 木構造データに対するパターン照合と機械学習に関する数理的な性質について研究した結果をまとめたものである。研究にあたっては、他分野にわたり多数提案されている既存の木構造のパターン照合アルゴリズムを数理的な性質から系統的に整理しなおし、その過程で得られた新たな理論的事実を、機械学習アルゴリズムの開発に利用している。また、開発した機械学習アルゴリズムをバイオインフォマティクスにおける糖鎖構造の分類とモチーフ発見に適用し、その有効性を評価している。これらの結果は、以下の９章にまとめられている。

第１章は「Introduction（序論）」であり、本研究の背景および研究課題と、本論文の構成を述べ、その後、研究成果を要約している。研究の背景として、HTML・XML 文書や、自然言語の構文木、RNA の二次構造、糖鎖構造などの木構造として表現可能なデータに対する効率的なパターン照合アルゴリズムの必要性を説いている。次に、編集距離の概念に基づく木構造の近似パターン照合アルゴリズムが、適用範囲の広い一般的なフレームワークであること、また、従来の編集距離に基づく近似パターン照合の研究に、様々な混乱があることを述べている。同様に、サポートベクターマシン（SVM）に代表される学習器を用いたカーネル法による木構造の分類学習についても、同様の混乱があることを説き、これらの混乱の原因が、近似パターン照合の意味を厳密に記述・比較するための理論的なフレームワークがなかったことに起因するとして、その数理的な解明を研究課題として設定している。

第２章は「Approximate Tree Matching（木の近似照合）」と題し、既存の木構造に対する近似パターン照合アルゴリズムを、文字列の近似パターン照合と類比させながら、網羅的にサーベイし、統一的な表記法により厳密に定式化している。その過程で、具体的に既存研究の問題点を洗い出している。

第３章は「Theoretical Foundation of Approximate Tree Matching (木の近似照合の理論的基礎)」と題し、木構造の近似パターン照合を、２つの半順序集合の間の写像とみなすことにより、半順序代数を用いて、既存の編集距離の概念を統一的に記述するための基礎理論を構築している。この理論を用いて、近似パターン照合アルゴリズムの操作的な意味を、代数的な側面から厳密に記述している。

第４章は「Relationship Analysis among Tree Edit Distance Measures （木の編集距離尺度間の関連性の解析）」と題し、既存の様々な木構造の近似パターン照合アルゴリズム間の関係を明らかにしている。その１つとして、文字列のアラインメントの自然な拡張である「木のアラインメント」と、異なるアルゴリズムであると考えられてきた「less-constrained 編集距離」が、実は同じ意味を持つアルゴリズムであることを明らかにしている。その過程で、「木のアラインメント」の代数的な意味づけを行い、これが２つの木を１つに結合するための必要十分条件になっていることを示しており、データ統合などの分野での広い応用が考えられる。その他にも、厳密に等価性が示されていなかったいくつかの木構造の近似パターン照合アルゴリズムに対して、証明を与えている。さらに、多様な木構造の近似パターン照合アルゴリズムを、その数理的な性質に基づいてクラス分けし、これらのクラス間の階層関係を明らかにしている。

第５章では、「Kernel-based Learning for Trees（カーネルに基づく木の学習）」と題して、木構造の分類学習問題を、木構造間の類似度であるカーネル関数（木カーネル）を設計する問題として捉え、既存の木カーネルについて概観している。また、一般的な設計手法である畳み込みカーネルの概念に基づき提案されている最も表現力の高い既存の木カーネルが、実際には畳み込みカーネルのクラスではなく、分類学習に用いるためには理論的な裏づけが必要であることを示している。

第６章は、「Mapping Kernel for Trees（木のマッピングカーネル）」と題し、既存の木

カーネルを、第４章で明らかにした木構造の近似パターン照合のクラス階層の観点から統一的に定式化し、既存の木カーネルの意味づけを行っている。また、従来研究では理論的な裏づけなくカーネルとして用いられてきた木の類似度が、実際にカーネル法による学習問題に適用できる数学的条件を満たすことを厳密に証明している。さらに、木構造の近似パターン照合の各々のクラスに対応する２つの木カーネルを新たに提案し、最も表現力の高かった既存の木カーネルよりも、さらに高い表現力を持つことを示している。また、「木のアラインメント」のクラスに対応する木カーネルが存在しないことも厳密に示している。

第７章は、「Spectrum Kernel for Trees （木のスペクトラムカーネル)」と題し、高速でかつ一般性の高い木カーネルとして木のスペクトラムカーネルを提案している。木のスペクトラムカーネルは、文字列の q グラムの概念を木に拡張した木の q グラムを提案し、２つの木構造に共通して含まれる q グラムを数え上げることにより類似性を測る高速なアルゴリズムを提案している。さらに、木のスペクトラムカーネルを拡張したグラム分布カーネルを提案している。

第８章は、「Application to Glycan Classification （糖鎖の分類への応用)」と題し、前章で提案された２つの木カーネルを、バイオインフォマティクスの分野における糖鎖構造の分類学習、および、糖鎖構造からのモチーフ抽出に実際に適用し、分類学習の性能を評価している。その結果、従来の一般的な木カーネルや、糖鎖専用のカーネルと比べて、分類能力と計算効率の両面において、高い性能を示すことを示している。また、モチーフ抽出では、既存の糖鎖専用のカーネルよりも一般的な構造をもつモチーフを抽出できることを示している。

第９章は、「Conclusion and Future Work（結論と今後の課題)」と題し、本論文の成果を要約し、関連する未解決問題や今後の発展について述べている。

このように、本論文は、木構造の近似パターン照合を数理的に記述するフレームワークを与え、木構造の近似パターン照合のクラス階層として統一的に整理したものである。このようなクラス階層の観点が実際に、木構造の分類学習のアルゴリズムの分類にも適用可能であり、この観点から新たなアルゴリズムが開発できることを示している点で、既存研究の数理的な解析にとどまらず、様々な応用の可能性を持った普遍性の高い研究であり、本研究分野の発展に寄与すること大である。

よって、本論文は博士（工学）の学位請求論文として合格と認められる。