

学位論文題名

Algorithmic Learning Theory of
Formal Languages and its Applications

(形式言語の計算論的学習理論とその応用)

学位論文内容の要旨

機械学習，特に帰納的学習は，人工知能の研究において重要な課題のひとつである。本研究では，具体例と質問から形式言語を帰納的に学習する問題を，理論と応用のふたつの観点から考察する。形式言語は，プログラミング言語の設計やオートマトンなどの制御システムの設計などの計算機科学の基礎をなす概念であり，形式言語の機械学習の方法を明らかにすることによって，プログラミング自動合成やシステム自動設計などの能力をもつ知的システムを構築することが可能となる。本研究は，(1)形式言語の効率的で実用的な学習方法に関する新しい理論的結果の提供，(2)学習アルゴリズムの実用的な応用例の提供，の2点を目的とする。

形式言語の学習としては，線形 (linear) 言語，等行列 (equal matrix) 言語，半線形 (semilinear) 集合の学習問題を理論的に考察し，新しい学習方法と結果を示した。まず，文脈自由言語のサブクラスである線形言語の学習に対して，新しい学習方法を提案した。文法の導出を制御集合という概念を用いることによって，線形言語の学習問題は正則言語の学習問題と関係づけられる。すなわち，任意の線形言語は，非終端記号をひとつしか持たない文法（これを骨格 (skeleton) 文法とよぶ）の導出を，正則な制御集合で制御することによって生成されることを示し，この事実から，線形文法は，骨格文法とその正則制御集合を同定することによって学習することができることを明らかにした。これにより，制御集合にもとづく学習方法を提案し，決定性有限オートマトンを学習する任意のアルゴリズムを用いて，学習アルゴリズムを構成することのできる線形文法のふたつのサブクラス，even 線形文法と括弧付線形文法を明らかにした。したがって，これらのサブクラスに対しては，従来の決定性有限オートマトンの効率良い学習アルゴリズムを活用することができる。

また，並列書き換えシステムのひとつである等行列文法の学習も，制御集合によって正則言語の学習問題と関係づけられることを示した。線形文法の場合と同様の制御集合による学習方法によって，等行列文法に対しても，決定性有限オートマトンの任意の学習アルゴリズムを用いて，

学習アルゴリズムを構成することのできるふたつのサブクラス, even 等行列文法と構造化等行列文法を明らかにした。線形言語と等行列言語は正則言語よりも上位の言語クラス, すなわち, 正則言語を真に包含するクラスであり, 制御集合を用いることによって, 従来よく研究されていた決定性オートマトンの学習方法は, 正則言語よりも上位の言語クラスである線形言語, 等行列言語にも適用可能であることを明らかにした。

次に, 並列計算モデルの学習問題に焦点をあて, 等行列文字や可換文法, ペトリネットなどの並列計算モデルの意味論的な基礎となる半線形集合の学習問題を考察した。線形 (linear) 集合とは, 有限の基底から生成される非負整数上の部分半群であり, 半線形集合とは, 線形集合の有限和である。まず, 線形集合は正の具体例だけから学習可能であるのに対し, 半線形集合は正の具体例だけでは学習できないことを明らかにした。そして, 線形集合の所属性判定問題が *NP* 完全であることを示し, この事実をもとに, 具体例から効率良く学習するアルゴリズムを構築することは, 線形集合に対してさえも計算量において難しいことを明らかにした。そこで, 部分性と包含性に関する質問を用いて, 半線形集合のいくつかのサブクラス, 特に線形集合を効率良く学習するアルゴリズムを示した。これらの結果や学習アルゴリズムは, 等行列文法, 可換文法およびペトリネットの学習問題にも適用可能である。

これらの形式言語の学習に関する理論的結果は, さまざまな計算モデル, 特に並列計算モデルの学習問題に, いくつかの肯定的・否定的な結果とともに, 効率の良い実用的な学習アルゴリズムの構成法を与える。

理論的に解明された学習アルゴリズムの実用的な応用として, 図形言語の学習と文法指導型構造エディタのカスタマイズのふたつの応用例を考察した。図形言語は, 統語的パターン認識で用いられる, デジタル図形言語と鎖状コード (chain code) 図形言語のふたつのタイプの言語を考察した。どちらのタイプの言語も等行列文法によって効果的・効率的に表現可能であり, 等行列文法の学習に関する結果と学習アルゴリズムを用いて, 効率良く学習可能な図形の例を示した。たとえば, デジタル表現された数字や鎖状コードで表現された長方形などの図形は, 等行列文法の学習アルゴリズムを用いて, 効率良く学習可能である。また, 複数の概念からなる図形の集合は, 等行列文法で表現した場合, 正の具体例だけから学習できないことを示すことができた。

また, 文脈自由文法の学習アルゴリズムを用いて, 文法指導型構造エディタをカスタマイズする方法を示した。文法指導型構造エディタでは, プログラムなどのテキストに導出木を割り当て, 割り当てられた導出木にしたがって編集を行う。したがって, ユーザは文法的に正しい編集のみが許され, 編集されたプログラムには文法上の誤りはない。しかし, この利点も, 導出木を,

割り当てる文法が、ユーザにとって編集しやすいものである場合にのみ有効であり、そうでない場合には、逆に欠点となってしまふ。たとえば、コンパイラなどで主に用いられる $LR(k)$ 文法などを、導出木割り当てのための文法として用いた場合には、簡単な算術式に対しても、ユーザにとって非常に編集しにくい編集構造を割り当ててしまふ。また、プリティプリントが導出木にしたがって行われている場合には、ユーザにとって非常に読みにくい表示となってしまふ。このような場合、従来の文法指導型構造エディタでは、ユーザ自身が文法をカスタマイズする必要があったが、正確に行うのには、文法に関する専門知識を必要とし、非常に困難であった。そこで、文脈自由文法の学習アルゴリズムを用いて、ユーザと対話的に具体例と質問から文法を効率良くカスタマイズする方法を提案した。そして、EMACS エディタ上に、学習機能をもつ文法指導型構造エディタ ACE を開発した。ACE では、文法の専門知識がなくても、望ましい編集構造をもつ具体例を作成し、それを ACE に与えることによって、編集構造を規定する文法を簡単にカスタマイズすることが可能である。また、ACE を用いて、新しいプログラミング言語の文法を、対話的に設計することも可能である。

最後に、本論文は、線形言語、等行列言語および半線形集合に対して、効率的で実用的な新しい学習方法を明らかにした。また、図形言語の学習と文法指導型構造エディタのカスタマイズのふたつの応用例を示し、形式言語の学習アルゴリズムが知的システム構築に有効であることを明らかにした。

学位論文審査の要旨

主査	教授	宮本衛市
副査	教授	伊達惇
副査	教授	嘉数侑昇
副査	助教授	赤間清

具体的事実から規則性を獲得する帰納的学習のアルゴリズムを開発することは、環境の変化に柔軟に対応した行動をする知的システムを構築するための重要な研究課題である。

本論文では、具体例から形式言語を帰納的に学習する問題を、理論と応用のふたつの観点から考察し、形式言語の学習アルゴリズムを用いて知的システムを構築する方法を検討している。

本論文では、はじめに、正則言語よりも上位の言語クラスである線形 (linear) 言語、等行列

(equal matrix) 言語の学習問題を理論的に考察し、新しい学習アルゴリズムを提案している。そのアルゴリズムの基礎となる考え方は、文法の導出を制御する制御集合という概念を用いることによって、線形言語、等行列言語の学習問題を正則制御集合を発見する問題を含む3つの部分問題に分割して解くことである。このうち正則制御集合を発見する問題は、従来からよく研究されていた決定性有限オートマトンの学習方法を利用することが可能である。この方法で多項式時間学習可能な文法クラスを検討した結果、線形文法のふたつのサブクラス (even 線形文法と括弧付線形文法)、および等行列文法のふたつのサブクラス (even 等行列文法と構造化等行列文法) に対する学習アルゴリズムを提案している。

次に、並列計算モデルの学習問題に焦点をあて、等行列文法や可換文法、ペトリネットなどの並列計算モデルの意味論的な基礎となる半線形集合の学習問題を考察している。まず、半線形集合の真部分クラスである線形集合が正の具体例だけから学習可能であるのに対し、半線形集合は正の具体例だけからでは学習できないことを明らかにしている。そして、線形集合の所属性判定問題が NP 完全であることを示し、この事実をもとに、具体例から効率良く学習するアルゴリズムを構築することは、線形集合に対してさえも計算量の点から難しいことを示している。また、部分性と包含性に関する質問を用いて、半線形集合のいくつかのサブクラス、特に線形集合を効率良く学習するアルゴリズムを与えている。さらに、半線形集合に対するこれらの知見や学習アルゴリズムを拡張して、等行列文法、可換文法およびペトリネットの学習アルゴリズムを提案している。

本論文では、理論的に検討した学習アルゴリズムの実用的な応用として、図形言語の学習と文法指導型構造エディタのカスタマイズのふたつの応用例を考察している。図形言語としては、統語的パターン認識で用いられる、デジタル図形言語と鎖状コード (chain code) 図形言語のふたつのタイプの言語について考察している。これらの図形言語の具体例について考察し、等行列文法によって効果的・効率的に表現可能であり、等行列文法の学習に関する結果と学習アルゴリズムを用いて、効率良く学習可能な図形は何かを検討している。

また、文脈自由文法の学習アルゴリズムを用いて、具体例と質問からユーザと対話的に文法指導型構造エディタをカスタマイズする方法を示している。そして、EMACS エディタ上に、学習機能をもつ文法指導型構造エディタ ACE を開発している。ACE は、ユーザが望ましい編集構造をもつ具体例を作成し、それを ACE に与えることによって、ユーザが文法の専門知識を持たなくても、編集構造を規定する文法を簡単にカスタマイズできるエディタになっている。

これを要するに、本論文は、形式言語のいくつかのクラスに対する新しい帰納的学習アルゴリズムを提案し、それらのアルゴリズムを実際に適用したシステムを開発して有効性を確認し、学

習アルゴリズムを用いて知的システムを構築するための有益な知見を得ており、情報工学の進歩に貢献するところ大なるものがある

よって著者は博士（工学）の学位を授与される資格あるものと認める。