

氏名	イ 黄	イ 延	カ オウ 蛟	(
学位の種類	博士(情報工学)			
学位記番号	情工博甲第187号			
学位授与の日付	平成18年9月30日			
学位授与の条件	学位規則第4条第1項該当			
学位論文題目	Inductive Inference of the Unions of the Pattern Languages from Positive Data by the Most Fitting Hypotheses (最適仮説によるパターン言語和の正データからの帰納推論)			
論文審査委員	主査	教授	篠原	武
		”	原尾	政輝
		”	梶原	誠司
		助教授	石坂	裕毅
		”	平田	耕一

## 学位論文内容の要旨

パターンとは、定数記号と変数記号からなる有限の文字列である。パターン言語は、パターンの変数を定数記号列で置換えて得られるもの全体の集合である。帰納推論とは、言語の例からその言語を帰納的に推測することをいい、言語の反例が与えられない場合、正データからの帰納推論という。

正データからの帰納推論においては、一般的すぎるものを避けて、何らかの極小性を保証する仮説を求めることが基本的である。しかし、極小な仮説からの推論は、その極小概念によって、適用できる言語族が制限される。包含関係に関する極小性をもつ仮説は *minl* といわれ、その適用範囲は広くあることが分かっている。また、*minl* を効率的に得るための手段として、包摂関係に対する極小性をもつもの (*mmg*) も調べられていた。一般的には、与えられた集合の *minl* は複数存在する。本論文では、複数の *minl* を比較するために、もう一つの極小性を提案している。それは、ある長さ以下の文字列で言語に含まれるものの個数が最小となる極小性である。すべての例を含み、しかもこの極小性をもつ仮説を「最適被覆 (fittest cover)」という。

まず、最適被覆を仮説として用いている場合、推論への影響を調べている。そのために、最適被覆を探す計算を用いた二つの学習者を定義し、それらにより推論することが可能な言語族の特徴づけを行っている。一見すると、その範囲は狭く見えるが、実は、推論可能となる言語族は、本研究の対象言語族のパターン言語およびその和言語の族をすべて含んでおり、かなり広いことを示している。また、*minl* であると同時に最適被覆である仮説がどのようにして得られるかも検討している。特徴集合という概念を用いれば、そうした仮説を得ることができることを示している。いくつかのパターン言語や木パターン言語およびそれらの和の族を対象として、その特徴集合となるものを調べている。さらに、部分文字列パターンを中心に、最適被覆を得るための計算量も調べ、効率的なアルゴリズムを示している。

つぎに、最適被覆という極小性が実際上の価値をもつかどうかを、二つの実験を通して検証している。実験では、生物学的配列からのコンセンサスパターンの発見問題を取り上げている。コンセンサスパターンとは、共通の特徴をもつ生物学的配列の特徴を表すパターンである。一つの実験は、コンセンサスパターンがはっきりと知られている生物学的配列を使い、それらに対して得られた *mmg* 中のパターンと、既知のパターンと比べて、より似たものが最適被覆の極小性をもつかどうかを調べている。もう一つの実験は、一部分の生物学的配列だけ使って *mmg* を探し、その *mmg* を残りの配列や負例(特徴を持たない配列)に照合して *mmg* の精度を測定している。繰り返して得られた多くの *mmg* について、精度の高いものが最適被覆の極小性をもつかどうかを調べている。いずれの実験においても、最適被覆がより望ましい仮説であることを確認している。

## 学位論文審査の結果の要旨

本論文に関し、調査委員から最適被覆と *minl* の仮説の計算可能性に関する違い、最適被覆の有効性を示す実験とその妥当性の根拠、今後の課題と展望などについて質問がなされたが、いずれも著者から満足な回答が得られた。

また、公聴会においても、多数の出席者があり、種々の質問がなされたが、いずれも著者の説明によって質問者の理解が得られた。

以上により、論文調査及び最終試験の結果に基づき、審査委員会において慎重に審査した結果、本論文が、博士（情報工学）の学位に十分値するものであると判断した。