

氏 名	はら ぐち かず や 原 口 和 也
学位(専攻分野)	博 士 (情 報 学)
学位記番号	情 博 第 258 号
学位授与の日付	平 成 19 年 3 月 23 日
学位授与の要件	学 位 規 則 第 4 条 第 1 項 該 当
研究科・専攻	情 報 学 研 究 科 数 理 工 学 専 攻
学位論文題目	Studies on Classifiers Based on Iteratively Composed Features (反復構成特徴に基づいた分類器に関する研究)
論文調査委員	(主 査) 教 授 永 持 仁 教 授 太 田 快 人 教 授 福 嶋 雅 夫

論 文 内 容 の 要 旨

「機械にどう学習させるか」という問いは情報学の様々な場面において現れ、実世界に多くの応用を持つ。本論文は、学習問題の中でも特に重要とされる分類問題を中心に扱っている。分類問題では入力としてデータが事例の集合として与えられ、各事例にはオラクル関数によって真偽いずれかのクラスが付されている。出力として、オラクル関数に（近似的に）等価な関数を求めなければならない。分類問題に対する一般的なアプローチでは関数を具体的な構造物（分類器）として処理するための表現モデルが必要である。分類器の意志決定過程が人間に解釈可能な新しい表現モデルとして、本論文では ICF (Iteratively Composed Features, 反復構成特徴) モデルを提案し、分類器構成アルゴリズムの設計とその実験的評価を主たる内容としている。本文 6 章から成り立っている。

第 1 章は序論であり、分類問題の定義とそれに対する一般的なアプローチ、および従来の研究について触れている。また、従来モデルである決定木モデルと本論文で提案する ICF モデルの用語・記法を導入し、論文全体の構成についても述べている。

第 2 章は ICF モデルと決定木モデルの潜在能力の比較を、理論および実験の両側面から議論している。前者では、オッカムの剃刀に基づいた学習の成功を定義し、その定義の下で ICF モデルが決定木モデルより優れていることを示している。後者では、それぞれのモデルによって高性能な分類器を得ることができるかについて実験を行い、多くのデータに対して ICF モデルの方が優れていることを示している。

第 3 章は ICF モデルによる分類器の構成アルゴリズムを提案している。部分的小および全体的な知識を表現する特徴のそれぞれについて、分類器の構成アルゴリズムを設計している。適切に定められたパラメータの下、多くの場合、前者は後者のみならず決定木モデルにおける従来アルゴリズム (C4.5) より優れた分類器を構成することを実験によって示している。

第 4 章は 3 値データに対して設計された提案アルゴリズムを、数値や記号で記述された一般のデータに拡張している。このため、一般のデータを 3 値データに変換する離散化スキームを提案している。パラメータを適当に選んだとしても、多くの場合、拡張されたアルゴリズムは C4.5 より優れた分類器を構成することを実験によって示している。

第 5 章は学習におけるデータのサイズについて議論している。とりわけルールと呼ばれる構造に着目し、その学習に用いるデータの十分量及び必要量を導いている。特に必要量については、ランダム性に基づいた独自の仮説を立ててアプローチしており、その正当性を実験によって検証し、離散確率論に基づいて上界の導出も試みている。

第 6 章は結論であり、本論文のまとめと今後の課題を述べている。

論 文 審 査 の 結 果 の 要 旨

本論文は機械学習に関する話題から、分類問題およびルール学習に用いるデータのサイズを中心に考察している。得られた結果は以下の通りである。

1. 分類問題では、分類器の意志決定過程が人間に解釈可能であることが知識発見の観点から重要である。従来モデルの決定木が命題論理を記述するのに対し、提案モデル（ICF）では抽象的な概念がいくつかの具体的な概念から反復構成されるプロセスを記述している。この知識表現は概念の階層構造と言われ、近年人工知能や知識発見等の分野で注目を浴びている。ICFではこれを分類器の表現モデルとして取り入れており、優れた汎化能力を持つICF分類器の存在が理論と実験の両側面から示されている。これにより、分類問題に対するICFの有望性をうかがうことができる。

2. 本論文が提案するICF分類器の構成アルゴリズムはいくつかのパラメータを持つ。パラメータの値を指定された範囲の中で適当に定めたとしても、多くの場合において、従来の決定木構成アルゴリズムより優れた分類器を構成できることが実験によって示されている。知識発見の現場では、分類器構成アルゴリズムの適切なパラメータの調整は難しいとされるが、この結果によって提案アルゴリズムの実用性を期待することができる。

3. ルール学習はデータマイニングにおける主要な手法の一つであるが、どの程度の量のデータを学習ツールに与えるべきかは重要な問題である。この問題に対し、本論文ではルール学習に用いるデータの十分量及び必要量を離散確率論に基づいて導いている。特に必要量についてはランダム性に基づいた独自の仮説を立て、その正当性が実験によって示されている点は興味深い。そのアイデアは機械学習における他の問題にも応用できることが期待される。

以上のように、本論文は、分類問題に対する新たな表現モデルICFの提案、およびICFに基づいた分類器構成アルゴリズムの開発や、ルール学習に必要および十分なデータの量の提案を行ったものであり、得られた成果は学術上および応用上寄与するところが少なくない。よって、本論文は博士（情報学）の学位論文として価値あるものと認める。また、平成19年2月16日に実施した論文内容とそれに関連した試問の結果合格と認めた。