

## 審査の結果の要旨

論文提出者氏名 鶴岡 慶雅

本論文は“Acquisition of Probabilistic Knowledge from Real-World Data and its Application (実世界データからの確率的知識の獲得とその利用)”と題して、コンピュータを用いて実世界の大量データから必ずしも確実とはいえない確率的な知識を獲得する方法と、そのような知識を利用する方法について、基礎理論とアルゴリズムを構築、自然言語処理およびゲームプレイヤを対象にした具体的な適用を通して実証した研究の成果についてまとめたもので、英文で記述された以下の7章よりなる。

第1章“Introduction”においては、研究の背景と目的を述べている。コンピュータによる知的な処理の実現はその草創期から夢見られてきたが、いまだに完全な実現に至っていない。その主要な原因のひとつに、コンピュータに知識を与えることの難しさがあげられる。近年の情報処理技術の普及に伴い文書などの大量のデータがコンピュータで利用可能な形で提供されるようになってきており、コンピュータにこれらから知識を学習させる方式が種々提案されてきているが、実世界のデータは誤りを含む場合も多く、従前の技術はこうした不確実な知識を扱う際には不十分である。そこで本研究では、誤りを含む可能性のある大量のデータから、確実とは限らない確率的な知識を獲得する方法と、獲得した確率的知識を適切に利用する方法について、精度・コストの両面での改善を目的としている。

第2章は“Learning Classifiers”と題し、機械学習により分類器を構成する方式について既存の研究成果を概観、教師つき学習においては大量のラベルつきデータを用意する必要がある問題点を指摘している。

第3章“Estimating Reliability of Hypotheses”においては、確率モデルのパラメータ推定方式である最大尤度方式とベイズ学習を紹介し、ベイズ学習を用いて仮説の信頼度を推定する手法を述べている。またベイズ学習においては、背景知識として適切な事前分布を用いることが重要であること示し、事前分布の選択指針を与えている。

第4章 “Decision Lists Learning Using Bayesian Learning” においては、自然言語処理などで広く用いられる決定リスト分類器において、仮説の信頼度をベイズ学習によって推定する方法と、英語の多義語および日本語の擬似多義語についての実験結果を示し、事前分布の適切な選択により正解率が向上するとともに、信頼度の推定精度が大きく改善できたことを報告している。

第5章 “Learning Classifiers from Uncertain Examples” においては、元のデータに付与されたラベルが不確実で確率的である場合のクラス分類器の学習の方法と、ラベルつきデータからの学習結果である不確実な知識を、ラベルのないデータの上に展開して利用する方法を述べている。さらに、この方式を日本語の擬似多義語の分類器学習に用いた実験結果を報告し、提案方式の適用によりごく少量のデータのラベル付けと大量のラベルなしデータの利用によって、低コストでの分類器学習が可能であることを示している。

第6章 “Game-Tree Search with Probabilistic Knowledge Extracted from Records of Games” においては、将棋の棋譜データから着手、すなわち盤面遷移の確率を学習し、その結果から局面実現確率を求め、コンピュータ将棋プレイヤーの先読みのための資源配分を局面の実現確率に応じて配分する方法を述べている。また、この方式を用いたプレイヤーの実験の結果、これまで標準的に用いられてきた一定深さでの先読み打ち切り方式に比して、大幅に性能を向上できることを示している。

第7章 “Conclusion” では本研究の成果をまとめている。

以上これを要するに、本論文ではコンピュータに知的な働きをさせる前提として重要である知識獲得の人手コストを低減するために、確率的な知識の獲得と利用を可能にする方法を提案し、実験を通じてその有効性を実証しており、その成果は電子情報工学上貢献するところが少なくない。

よって本論文は博士（工学）の論文として合格と認められる。