

2021年 2月 26日

豊橋技術科学大学長 殿

情報・知能工学 専攻  
学位審査委員会  
委員長

南 哲人



### 論文審査及び最終試験の結果報告

このことについて、博士学位論文審査を実施し、下記の結果を得ましたので報告いたします。

学位申請者	菊地 真人		学籍番号	第 143313 号
申請学位	博士（工学）	専攻名	大学院工学研究科博士後期課程 情報・知能工学 専攻	
博士学位論文名	統計量の保守的な推定に関する実証的研究 (Empirical Study for Conservative Estimation of Statistics)			
論文審査の期間	2021年 1月 14日 ～ 2021年 2月 19日			
公開審査会の日	2021年 2月 19日	最終試験の実施日	2021年 2月 19日	
論文審査の結果※	合格		最終試験の結果※	合格
<p>審査委員会(学位規程第6条)</p> <p>学位申請者にかかる博士学位論文について、論文審査、公開審査会及び最終試験を行い、別紙論文内容の要旨及び審査結果の要旨のとおり確認したので、学位審査委員会に報告します。</p> <p>委員長</p> <p>委員</p> <p>北岡 教英 </p> <p>梅村 恭司  渡辺 一帆 </p> <p>印</p> <p>印</p> <p>印</p>				

※論文審査の結果及び最終試験の結果は「合格」又は「不合格」の評語で記入すること。

## 論文内容の要旨

本論文は、情報源から得た事象の観測頻度をもとに統計量を推定することに関し、不偏推定量の問題に着目し、頻度に応じて推定量を保守的に見積もる枠組みを論じたものである。扱う統計量としては、条件付き確率と尤度比の二つを推定の対象としている。前者は関係マイニングや確率的言語モデル、後者は多値分類や統計検定などで広く用いられる統計量である。まず、第2章では、本論文で使用している研究内容を概説している。第3章では、条件付き確率の保守的な推定法を提案している。この手法は、確率分布の信頼区間を構築し、その下限値を推定値とする。提案手法を用いることで高・低頻度の両方を効果的に扱い、関係マイニングに於いて多くの関係を発見できることを示している。第4章では、最適化の枠組みによって正規化を導入し、尤度比を保守的に推定する手法を論じている。この手法を使って、10個の科学雑誌名と科学ニュース記事から雑誌名を特定する実験を行い、提案方法の有効性を示している。第5章では、データに存在しないゼロ頻度のNグラムにも推定値を付与する手法を提案している。この手法では、最適化の枠組みの中で、Nグラム自体の頻度に加え、それを構成する文字や単語に基づく頻度も利用することで、ゼロ頻度のNグラムに対処している。固有表現の左Nグラムを尤度比で予測する実験により、提案手法の有効性を確認した。第6章では、本論文の研究内容を総括し、今後の展望を述べている。

## 審査結果の要旨

情報源から得た事象の観測頻度をもとに統計量を推定することは、データを確率的に処理するときの基本操作である。そして、その推定法は、データを用いた工学的応用での性能を左右する重大な要因になる。本論文は、この推定に関する基礎的な問題を扱っている。推定には不偏推定量がよく用いられているが、事象の観測が低頻度の場合、不偏推定量は二つの問題を抱えている。第一に、不偏推定量は推定の不確実性が大きい。第二に、不偏推定量は偽の事象を真と誤る第一種過誤、真の事象を偽と誤る第二種過誤を同じ損害とみなすが、実際は一方が他方よりも大きな損害を持つことが多い。そこで本論文では、過誤による損害が小さくなるよう、推定量を見積もる枠組みを提案した。この枠組みを統計量の保守的な推定として論じている。この提案は、不偏推定を行わないということであり、独自性のある提案である。

第3章では、推定量の観測後の事後分布を考え、そこから、小さめに見積もる損害よりも、大きめに見積もる損害が大きい状況において、片側信頼区間を構築し、その区間の下限値を推定に用いることを提案している。その提案の有効性を、実用的な価値の高いタスクである関係マイニングのタスクにおいて示している。第4章では、誤差最小の最適化の枠組みに正規化を導入する方法で、第3章と同じ振る舞いをする推定方法を示した。この方法を、名前の抽出のブートストラップ法に適用した。その結果、ブートストラップのステップごとに確実な名前を選んでいくことができていることを示した。第5章では、誤差最小の最適化の枠組みの中に、周辺分布を利用することを提案している。その方法を、直接は観測されないNグラムに対して、文脈に関する尤度比を推定することに適用し、観測されているNグラムの推定の性能を落とすことなく、観測されないNグラムに推定値を与える方法を示している。このように、不偏推定量に対しての問題提起を行い、信頼区間の構築というアプローチでの解決を示し、さらに誤差最小の最適化問題の定式化というアプローチでの解決方法を示した上で、それをさらに発展させている。この成果は、データを用いた工学応用において重要な基盤となるものであり、本論文は博士(工学)の学位論文に相当するものと判定した。

(各要旨は1ページ以上可)