

博士(文学)学位請求論文審査報告要旨

論文提出者氏名	大橋 洸太郎
論文題目	テキストデータの収集の程度を測る指標に関する研究
審査要旨	
<p>調査紙を用いて意見や評価といった回答者の態度を収集した研究事例は、今日まで多くの場面で見られてきた。調査紙には通例、研究者の意図する研究課題を網羅した質問項目が並べられているが、その回答形式としては回答となる選択肢が予め用意されたプリコード型と、自由回答を求める自由記述型の二種類に大別される。これら両者の回答形式を併せ持った調査紙も多くみられるものの、後者の自由記述型の質問項目は、自由な意見をテキストデータとして収集するため、回答の種類はプリコード型よりも非常に多く、どれだけ回答を収集しても新たな意見、知見が無くなるのが稀有であり、そのため新知見が得られ難くなったある時点で知見の収集を打ち切らなければならないという性質を有している。しかしながら、これまで知見の収集を打ち切るための統一的な基準は存在せず、研究者間で知見の収集の十分さの程度についての客観的な評価ができないという現状があった。</p> <p>本論文は、水産資源学における資源量推定法を応用し、知見の収集の十分さの程度を表す捕獲率という客観的な指標の提案を行い、様々なデータに対する応用研究を示していくことで、提案した捕獲率の有効性についての吟味を行っている。本論文は、全8章で構成されている。</p> <p>第1章ではまず、自由記述の収集における問題提起を行い、続いて研究目的について言及している。また質問紙法の回答形式の分類や資源量推定法に加え、KJ法、テキストマイニング、質的研究法の紹介、データマイニング手法といった本論文で用いられる回答処理、得られた知見の更なる分析手法に関する基本的な説明を行っている。</p> <p>第2章では、自由記述データから、資源量推定の方法を用いて知見数の飽和度を示す捕獲率を具体的に推定する方法を提案している。この章では資源量推定の方法としてSchnabel法を用いており、Schnabel法の説明に加え、適用例として標本数が数百程の、ブランドのイメージに関する自由記述の回答を、KJ法によって処理したデータ用いている。本章で提案された手法は、資源量推定の公式を自由記述の収集の飽和度の指標として用いる端緒となるものであり、捕獲率によって、収集した自由記述データから導かれる理論や研究成果がより盤石な基盤をもっていたことを1つの観点からの傍証として示すことができる点で、非常に有益な手段である。</p> <p>第3章では、より大きな標本サイズへの捕獲率の適用を想定し、資源量推定の方法としてDeLury法を用いた捕獲率の提案を行っている。DeLury法の説明に加え、ファントム変数法を用いた共分散構造分析のソフトウェアによる実行方法についても紹介され、適用例として標本数が数千程の、ブランドのイメージに関する自由記述の回答を、テキストマイニングによって処理した名詞と形容詞のデータ用いている。本章における捕獲の対象は知見ではなく、名詞や形容詞といった、イメージのキーワードとなる単語である。このため本章は、データが数千単位の大きな標本サイズとなるにつれ、人の手によるKJ法が難しくなった際に、テキストマイニングという自動コーディング技術を用いて捕獲率を計算する道筋を示した章といえる。</p> <p>第4章では、特に第3章で扱ったような大きな標本サイズについて、捕獲率と共に用いることで自由記述の収集の程度に関して更なる考察を行うことができる遭遇率という指標が提案されている。</p> <p>第3章と同じデータを用い、3種類のパレート分布から最もデータに当てはまりのよいモデルを選択し、単語の出現頻度をモデル化し、そのモデルの累積分布関数を用いることで、未知の単語がまだどれほど得られやすいかを知ることができる遭遇率の計算方法が説明されている。</p>	

第5章では、第3章と第4章で提案された捕獲率と遭遇率を用いて、Web上のデータを対象とした応用研究を示している。特定のテーマに沿って収集したWeb文書の捕獲率と遭遇率を計算し、得られた名詞の解釈を行った応用例と共に、この応用例で得られたWeb文書内に存在していたリンク情報を用い、更に収集したWeb文書を選択的に加えることで、捕獲率の結果を上昇させていく方法を示している。

第6章では、授業評価データを対象とした応用研究を行っている。私立大学の必修授業についての自由記述形式のアンケートから、KJ法によって授業改善に役立つ知見を収集し、これらの知見が手元のデータの枚数で十分に収集されているかどうかを検討している。

第7章では、実際にインタビューを行ったデータを分析対象として捕獲率を計算・応用する研究を試みている。最近2年間に就職活動を行った学生について、就職活動における成功体験を文章化し、KJ法によって就職活動に役立つ知見を収集し、これらの知見が手元のデータの枚数で十分に収集されているかどうかを検討した。また得られた知見を変数とみなし、データマイニング手法を用いて変数間の関係性について更なる分析を行っている。

第8章では、第5章から第7章までで示された3つの応用研究の結果を踏まえ、第2章から第4章までで提案されてきた捕獲率と遭遇率の有効性についての総合考察を行っている。これらの応用研究と、提案を行った各章の適用例の結果から、提案された捕獲率、遭遇率の様々なデータでの有効性を確認し、今後の応用についての展望を述べている。

本論文の成果から、捕獲率、遭遇率は、自由記述の収集の程度を示す指標として、KJ法やテキストマイニングといった回答処理の方法を用いることで、様々な標本サイズ、自由記述型のデータである多様な分野のデータに適用可能であることが示された。この適用範囲の広さに加え、提案された指標によってテキストデータの収集の十分さを客観的に示すことができたという問題解決の観点から、本論文が実用的、学術的に有用であると考えられる。以上より、本論文は博士学位論文の水準を満たすものである。

公開審査会開催日	2013年 11月 25日		
審査委員資格	所属機関名称・資格	博士学位名称	氏名
主任審査委員	早稲田大学文学学術院教授	教育学博士(東大)	豊田秀樹
審査委員	早稲田大学文学学術院准教授	博士(文学・早大)	福川康之
審査委員	多摩大学経営情報学部教授	博士(経営学・法政)	豊田裕貴
審査委員			
審査委員			