

京都大学教育研究振興財団助成事業
成 果 報 告 書

2023年 6月 7日

公益財団法人京都大学教育研究振興財団

会 長 藤 洋 作 様

所 属 部 局 医学部附属病院

職 名 助教

氏 名 阿部寛康

助 成 の 種 類	令和4年度 ・ 研究活動推進助成			
申請時の科研費 研究 課 題 名	情報複雑性を考慮したリサーチクエスト新発見ソフト開発と 健康状態パターン把握			
上記以外で助成金 を 充 当 した 研 究 内 容	特になし			
助成金充当に関 わる共同研究者	(所属・職名・氏名) 特になし			
発表学会文献等	(この研究成果を発表した学会・文献等) 日本計算機統計学会第36回シンポジウム(富山国際会議場, 2022年11月27日) 統計数理研究所共同研究 複雑な構造をもつデータに対する多変量解析法に関する研究会(統計 数理研究所, 2023年3月10日)			
成 果 の 概 要	研究内容・研究成果・今後の見通しなどについて、簡略に、A4版・和文で作成し、 添付して下さい。(タイトルは「成果の概要／報告者名」)			
会 計 報 告	交付を受けた助成金額	1,000,000	円	
	使用した助成金額	891,148	円	
	返納すべき助成金額	108,852	円	
	助成金の使途内訳	費 目	金 額	
		大規模コンピュータ	735,900	
		書籍等資料購入	111,394	
		学会発表用費用	4,000	
英文校正代費用		39,584		
その他	270			
当財団の助成に つ いて	(今回の助成に対する感想、今後の助成に望むこと等お書き下さい。助成事業の参考にさせていただきます。) 本助成金によって研究の足掛かりを得ることができました。			

成果報告書および成果の概要は、財団に郵送(あるいは持参、または成果報告書に押印してPDFでメール送信も可)するとともに、Excel・Wordファイル(印鑑不要)でメール送信して下さい。

成果の概要/阿部寛康

1. 研究内容

大規模な医療情報から臨床研究におけるリサーチクエスションとなりうる知見を獲得するための行列分解型の多変量解析手法について検討した。医療情報は多種多様な形式のデータ（変数）が混在し、それぞれに合わせた確率分布を仮定すべきであるが、現場レベルでは、その操作の簡単さから、カテゴリカル変数をダミー化した後に主成分分析や因子分析といった全てが量的変数であることが前提となっている手法が用いられる現状があると考えられる。少し専門的になると、量質混在データに対する同様の解析手法として、FAMD(Factor Analysis of Mixed Data)という手法が存在し、これは統計解析ソフト R の有名なパッケージである FactoMineR にも搭載されている。しかしながら、この手法ではカテゴリカル変数に対する因子負荷量の重みが量的変数と比べて小さくなる傾向があり、カテゴリカル変数に関わる知見発見を見逃すおそれがある。これは前者の単純な方法でも起きうる。そこで本研究では量的変数に対しては正規分布を仮定し、カテゴリカル変数に対しては二項分布あるいは多項分布を仮定した行列分解型多変量解析法について、データを用いたパラメータ推定を行うためのアルゴリズムを開発し、数値シミュレーションによる既存手法との推定の正確さの比較を行った。

2. 研究成果

まず、足掛かりとして、カテゴリカル変数が全て二値変数である場合について検討した。本来であれば例えば 3 カテゴリ以上の変数も含めて検討すべきであるが、二値変数に当てる二項分布を用いる場合は比較的単純な数学的操作によりに推定アルゴリズムを導出できるため、この条件からまずは検討した。結果、提案手法の方が正確にパラメータを推定できる（つまり量質の重みの隔たりがない）ことが数値シミュレーションによって確認された。次に多項分布を用いた場合について検討した。多項分布を用いる場合の推定アルゴリズム導出において利用できる補助関数法が存在することが文献調査の結果判明したため、それを応用した行列分解手法を検討した。結果、3 カテゴリ以上の変数がある場合においても、安定した推定が可能であることを確認した。

3. 今後の見通し

まずは上記の多項分布を用いた場合の手法について、実際の様々な医療データへ適用していく必要がある。これに加えて、生存時間変数やカウント変数のような正規分布を当てることが不適切と考えられる量的変数や、順序尺度であるカテゴリカル変数が混在するデータでの行列分解法について検討を進めていきたい。