

令和元年6月27日現在

機関番号：22701

研究種目：挑戦的萌芽研究

研究期間：2016～2018

課題番号：16K12398

研究課題名(和文)統計的因果推論の形式化・自動化に向けての挑戦

研究課題名(英文)Formalization and automation for statistical causal inference

研究代表者

汪金芳(Wang, Jinfang)

横浜市立大学・データサイエンス学部・教授

研究者番号：10270414

交付決定額(研究期間全体):(直接経費) 2,700,000円

研究成果の概要(和文): (1) 複数の読影者から得られたクラスターデータに基づいて、2つの画像診断法に対する非劣性検定を提案し、急性くも膜下出血患者に対して実施した動脈瘤診断法から得られたデータに本提案手法を適用し、その有効性を確認した。(2) 複数の読影者による評価データを統合するための多次元変量効果モデルを提案し、それに基づいた感度と特異度の推定法や、感度と特異度の同時信頼区間の構築法を提案した。(3) 定理証明支援系 Coq とその拡張である SSReflect を用いた、Wang(2010) で提案された代数系 Cain、とそれに基づく条件付き独立性の形式化を行った。

研究成果の学術的意義や社会的意義

理論と応用の両面において、統計的因果推論は極めて重要な問題である。本研究では、この問題を正面から挑戦し、独自の代数系を提案し、それに基づく条件付き独立性の形式化を行った。これにより、条件付き独立性に関する操作の半自動化・自動化の可能性を開いた。

研究成果の概要(英文): (1) Proposed non-inferior tests for image-diagnosis based on cluster data and successfully showed the usefulness of the methods by simulation studies and applying the tests to patients with acute subarachnoid hemorrhage. (2) Proposed random effect models for combining data obtained by multiple raters. New confidence intervals are proposed for both sensitivity and specificity. (3) We carried out formalization of the algebraic system of cain proposed by Wang (2010) for manipulating statistical conditional independence using the proof assistant SSReflect.

研究分野: statistics, data science, computer science

キーワード: causal inference conditional independence formalization cain SSReflect

1. 研究開始当初の背景

「統計科学の基礎において、因果推論の概念は確率の概念以上に重要である」(David Cox, 1986, JASA)。因果推論は凄まじい発展を遂げ続けている。Pearl (2000) の出版以降、介入による因果の定式化が大きな影響を及ぼしている。Pearl (2000) は Lewis (1973) の反事実的因果論に基づき、Wright (1971) や Menzies and Price (1993) などの因果操作論を発展させ、因果関係を非巡回有向グラフで表し、構造方程式の操作による因果効果の定量的評価法を提案している。

しかし、Dawid (2000) が指摘するように、(a) 介入の認識論的定義は明確ではなく、また (b) 介入では説明できない因果関係が存在することなど、因果介入論は様々な難点を含む。これらの難点を回避するための試みとして、Dawid (2001) は条件付き独立性の公理系である separoid を提案した。separoid は Pearl & Paz (1987) の graphoid の一般化であるが、graphoid と同様全ての公理が、「A B」の形で構成され、このような公理系は不完全であることが知られており (Studený et al., 2010)、このような公理系は機械証明にも向かない (Garrigue, 2015)。

申請者はこれまで、条件付き独立性の性質の直接的公理化を避け、確率密度関数の代数的性質の公理化し、cain とよばれる汎代数を構築した (Wang, 2010)。cain から separoid を導くことができる。本研究は、(1) cain 代数の形式化、また (2) 条件付き独立性の自動証明、という課題に挑戦する。この挑戦は数学全般を型理論 (type theory) に基づく形式化の一環として捉えることができる (久我, 2015)。

2. 研究の目的

本研究の目的は、公理的立場から統計的因果推論 (causal inference) を行う代数的システムの構築に挑戦することである。研究期間内に全力を挙げ、関連し合う次の 2 つの目標、すなわち、(1) 統計的因果推論の土台となる条件付き独立性の形式化 (formalization) と、(2) 条件付き独立性を計算機により自動証明できるアルゴリズムの構築、を実現させる。具体的には、汎代数の枠組みで構築された cain 代数 (Wang, 2010) を、定理証明支援系 SSReflect を用いて形式化し、目標 (1) を実現させる。また条件付き独立性の判定問題を、汎代数族 (variety of algebra) のメンバーシップ問題 (membership problem) として定式化し、cain 代数の再構築により、目標 (2) を実現させる。

3. 研究の方法

研究期間内に実現可能な目標である、(1) 条件付き独立性の形式化、と (2) 条件付き独立性の自動証明アルゴリズムの構築、の 2 つの課題に焦点を当て研究を進めていく。より具体的に、目標 (1) を実現させるため、cain 代数を、定理証明支援系 SSReflect を用いて形式化を行う。Cain 代数と SSReflect の整合性が非常によいことを後に詳述する。一方、目標 (2) を実現させるため、条件付き独立性の判定問題を、汎代数のメンバーシップ問題として定式化し、cain 代数を再構築し発展させることにより解決する。メンバーシップ問題は、部分代数 (subalgebra) B が部分代数 A の生成された族 (variety)、 $V(A)$ に属するかどうか、すなわち、 $B \in V(A)$ 、を決定する問題に帰着させて研究を進めていく。

4. 研究成果

(1) 複数の読影者から得られたクラスターデータに基づいて、2 つの画像診断法に対する非劣性検定を提案し、コンピュータシミュレーションや、急性くも膜下出血患者に対して実施した動脈瘤診断法から得られたデータに本提案手法を適用し、その有効性を確認した。

(2) 複数の読影者による評価データを統合するための多次元変量効果モデルを提案し、それに基づいた感度と特異度の推定法や、感度と特異度の同時信頼区間の構築法を提案した。

(3) 定理証明支援系 Coq とその拡張である SSReflect を用いた、Wang (2010) で提案された代数系 Cain、とそれに基づく条件付き独立性の形式化を行った。

(4) 連続変量の離散化を行い、頻度として表に纏めることがしばしばある。このような疎表に基づく「セル回帰分析」の手法を開発した。セル回帰分析法は従来の回帰分析法や区間回帰分析法の拡張としてみる事ができる。

(5) 条件付き独立性の公理的な研究の中で、Graphoid と Separoid は代表的なものである。代表者が提案した公理系 Cain はこれらのアプローチとは異なり、確率密度関数や確率関数の代数的な性質に着目し、全ての公理を等式で記述する点でユニークである。このことにより、条

件付き独立性の公理系である Cain を計算機で実装 することができた。用いた言語は Coq の拡張となる Ssreflect である。

(6)統計的、計算的、人間的視点 から、データサイエンスに関する総合的研究を行い、またこれらの理念を横浜市立大学データサイエンス学部の教育研究に活かすように実践的に取り組んだ。

5 . 主な発表論文等

〔雑誌論文〕(計 4 件)

Saeki, H., Tango, T. and Wang, J. (2016). Estimating the diagnostic accuracy from multiple raters based on a bivariate random effects model, 計量生物学, 37, 23--44.

Saeki, H., Tango, T. and Wang, J. (2017). tatistical Inference for Non-inferiority of Difference in Proportions of Clustered Matched-pair Data from Multiple Raters, Journal of Biopharmaceutical Statistics, Vol. 27, 70—83.

Yamaguchi, K. Kin, S. Shimoyama, M. Hagiwara, M. Yamamoto and J. Wang (2017). Formalization of the Conditional Independence using Coq/SSReflect, Technical Reports of Mathematical Sciences, Chiba University, Vol.29 (1), 1--40.

Jinfang Wang and Masataka Taguri (2018). Bringing Up Data Scientists in Japan: Yokohama City University, Proceedings of 10th International Conference on Teaching Statistics.

〔学会発表〕(計 9 件)

Jinfang Wang (2016). Bayesian prediction based on profile-reference data, The 22nd International Conference on Computational Statistics(国際学会), 2016年08月23日~2016年08月26日, Felipe, Oviedo, Spain.

Jinfang Wang (2016). Cell regression, 数理統計ひこね 2016, 2016年12月02日~2016年12月03日
滋賀大学彦根キャンパス(滋賀県・彦根市).

Jinfang Wang (2017). Cell Regression and Reference Priors, 科研費シンポジウム「統計的モデリングと計算アルゴリズムの数理と展開」, 2017年02月16日~2017年02月17日, 名古屋大学 情報科学研究科(愛知県・名古屋市).

Jinfang Wang (2017). On posterior predictive probabilities of diabetes based on comprehensive medical examination data, 2017年度日本計量生物学会年会, 2017年03月16日~2017年03月17日, 央大学後楽園キャンパス(東京都・文京区).

Jinfang Wang (2017).Predictive inference with transferred priors, 2017 Hangzhou International Statistical Symposium(国際学会).

Jinfang Wang (2017).Transfer Regression and Predictive Distributions,10th Conference of the IASC-ARS/68th Annual NZSA Conference(国際学会).

Jinfang Wang and Tomoaki Nakatani (2018). Fostering Data Scientists at Yokohama City University, Japan, 27th South Taiwan Statistical Conference(招待講演)(国際学会).

Jinfang Wang (2018). Cell regressions and transfer priors, 27th South Taiwan Statistical Conference(招待講演)(国際学会).

汪金芳 (2018). 横浜市立大学データサイエンス学部の現状と今後, 滋賀大学データサイエンスフォーラム 2018(招待講演).

〔図書〕(計 1 件)

汪 金芳 (2016). 一般化線形モデル (統計解析スタンダード), 朝倉書店.

〔産業財産権〕

出願状況（計 0 件）

名称：
発明者：
権利者：
種類：
番号：
出願年：
国内外の別：

取得状況（計 0 件）

名称：
発明者：
権利者：
種類：
番号：
取得年：
国内外の別：

〔その他〕

ホームページ等

<http://datascience.jp.com/>

5 . 研究組織

本研究は研究代表者のみによって構成されている。

(1)研究分担者

研究分担者氏名：

ローマ字氏名：

所属研究機関名：

部局名：

職名：

研究者番号（8桁）：

(2)研究協力者

研究協力者氏名：

ローマ字氏名：

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。