様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　２ 年　 ５ 月 １６ 日現在

機関番号：１４２０１
研究種目：基盤研究(C)（一般）
研究期間：2016～2019
課題番号：１６Ｋ００００５

研究課題名（和文）異質性や非定常性のあるデータにおける未観測交絡変数を許す因果構造推定法と応用

研究課題名（英文）Causal discovery in the presence of hidden confounding variables for data with heterogeneity

研究代表者
清水　昌平（Shimizu, Shohei）

滋賀大学・データサイエンス学部・教授

研究者番号：１０５０９８７１

交付決定額（研究期間全体）：（直接経費）　　3,600,000円

研究成果の概要（和文）：LiNGAMモデルは連続変数のみを扱う。異質性を表現するために，LiNGAMモデルが離散変数を扱えるように拡張することを試みた。まずは，離散変数と連続変数の関係が非巡回有向グラフであると仮定したモデルを開発した。また、離散変数を扱うことのできる機械学習モデルと因果モデルを組み合わせることを考えた。未観測共通原因への対応としては，非ガウス性と独立性を利用することで操作変数法の拡張を行なった。さらに，未観測共通原因がどこにありそうかを推測する方法をLiNGAMモデルの枠組みで提案した。

研究成果の学術的意義や社会的意義
LiNGAMモデルは因果探索の標準的な方法の一つとして注目を集めているが，離散変数が混在する状況を扱えるようにすることでさらに応用範囲を広げることができた。また機械学習モデルと因果モデルを組み合わせたモデルについては,制御への応用が期待される。操作変数は広く用いられているが，非ガウス性と独立性を利用した操作変数法については,従来よりも多くの情報を抽出することができることがわかった。未観測共通原因がどこにありそうかを推測する方法については，条件付き独立性を用いる因果関係推測法の枠組みでは，そのような方法が提案されているが，LiNGAMモデルの枠組みでは対応する方法がなかった。

研究成果の概要（英文）：LiNGAM model handles only continuous variables. To represent heterogeneity, we tried to extend the LiNGAM model so that it can handle discrete variables. We developed a model assuming that the relationship between discrete variables and continuous variables is a non-cyclic directed graph. We also considered combining a causal model with a machine learning model that can handle discrete variables. To deal with unobserved common causes, we extended instrumental variable methods by making use of non-Gaussianity and independence. In addition, a method to infer where the unobserved common cause is likely to be is proposed within the framework of the LiNGAM model.

研究分野：統計科学

キーワード：因果探索　因果構造　観察データ　未観測共通原因

様　式　C－１９，F－１９－１，Z－１９（共通）

１．研究開始当初の背景

実質科学の主目的は，因果関係の解明である。例えば，脳領域間の因果構造の解明は，病気の理解や治療法の開発に役立つ。アルツハイマーやパーキンソン病は脳領域をつなぐ神経細胞ネットワークの異常が原因の１ つと言われている。介入を伴う実験ができれば分析はシンプルになるが，ヒトの脳が対象であれば介入することは簡単ではない。そのため，実際に介入する前に，因果構造ネットワークを推定するための統計解析法が望まれている。例えば，「脳領域 A の活動を低下させると，脳領域 B の活動が低下する」といった関係を推定したい。そこで，介入のないデータから因果関係に関する仮 説を探索する研究（因果構造推定）が盛んに行われ始めている。例えば，米国では，NIH が Big data to Data Knowledge Initiative の一環として，Center for Causal Modeling and Discovery of Biomedical Knowledge from Big Data をカーネギーメロン大学に立ち上げている。

介入のないデータによる因果分析の最大の困難は，未観測交絡変数による疑似相関である。そこで 本提案では，疑似相関の問題に対処しつつ，異質で非定常なデータを用いて因果関係に関する情報を 抽出するための統計解析法を研究開発することを目指した。特に，POS データのような顧客ごとに性質が異なりうる 異質性のあるデータや脳活動計測データや気象データのように非定常性のある時系列データに基づく 因果構造推定法を研究開発する。既存の因果構造推定法（Pearl，2009; Spirtes et al. 2001）の多くは，同質性と定常性を仮定しており，このような異質性と非定常性のあるデータに十分対処できていない [Ramsey et al.，2010, NeuroImage]。

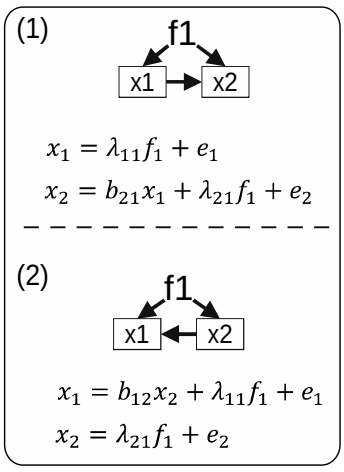
従来の因果構造推定法は通常，陰に陽にガウス分布の仮定に基づいていた。しかし，このガウス分布の仮定が原因で，因果の方向などの因果構造を一意に推定できない場合が多いという欠点があった。一方，信号処理の分野で発展した独立成分分析は，データの非ガウス性を利用して一意に原信号を同定する統計解析法である。独立成分分析と同様に，因果構造推定にデータの非ガウス性を利用することで， 従来法では不可能だった分析が可能になる場合があることを我々は明らかにしてきた。この新しい方法は LiNGAM 法と呼ばれ，発展を続けている。しかし，依然として 同質性と定常性を基礎仮定としている。実際に LiNGAM 法を用いて，マーケティングサイエンスの研究者や神経科学の研究者，気象学の研究者と共同研究を始めると，同質性と定常性の仮定を緩める必要が出てきた。それは，異質性や非定常性に頑健にするということもあるが，むしろ異質性や非定常性自体が重要な情報になるからである。例えば，顧客ごとに購買行動の因果構造に違いがあるとすれば，その違いをプロモーション活動に生かすことができる。また，脳活動の因果構造が実験刺激の変化に応じてどのタイミングでどう変化するかがわかれば，脳活動を理解する手掛かりになる。つまり，「どのような条件下で」何が原因で何が結果かを推定するデータ解析法が望まれている。

２．研究の目的

スーパーマーケットの購買データや脳機能イメージングデータ，そして気象データのような異質性や非定常性のあるデータから，観測変数間の因果構造を推定するための統計解析法を研究開発することをねらった。 これまで多くの推定法が提案されてきた。しかし，未観測の交絡変数（第三変数，共通原因）がある場 合は，観測変数間の因果構造についてほとんど情報を得られなかった。未観測交絡変数による疑似相関を見抜くことが難しかったからである。そこで本研究では，世界に先駆けて，未観測交絡変数の存在を許しつつ因果構造を推定可能にすることを目的とした。さらに，開発した推定法を用いて，マーケティングサイエンス，神経科学，気象学の観点から興味深い因果仮説を見つけることを目指した。

３．研究の方法

同質な集団であり未観測交絡変数がない場合のモデルとして，LiNGAM モデルを開発済みである(Shimizu et al. 2006, Journal of Machine Learning Research)。未観測交絡変数がある場合は，Latent variable LiNGAM モデルを開発済みである(Hoyer et al, 2008, International Journal of Appxoimate Reasoning)。これらのモデルを基礎にして，異質性や非定常性のあるデータに対処するモデルを考えた。



LiNGAM モデルは連続変数のみを扱う。異質性を表現するために，まず LiNGAM モデルが離散変数を扱えるように拡張することを試みた。

まずは，離散変数と連続変数の関係が非巡回有向グラフであると仮定し，離散変数とその親の関係をロジスティックモデル，連続変数とその親の関係を LiNGAM モデルと同様に線形モデルを仮定した。そして，誤差変数は非ガウス連続分布に従うとした。非ガウス性の仮定が LiNGAM モ

デルの特徴である。

次に，離散変数を扱うことのできる機械学習モデルと因果モデルを組み合わせることを考えた。組み合わせることにより，予測に特化した機械学習モデルと変化を推測する因果モデルの特徴を合わせもつモデルをつくる。機械学習モデルに基づく予測に基づいて制御を行うことがあるが，制御するためには因果構造を知る必要があり，そのためにも機械学習モデルと因果モデルを結合することには意味がある。

未観測共通原因への対応としては，まず操作変数法の拡張を試みた。従来の操作変数法は，基本的に非ガウス性と独立性を利用することはしない。そこで，非ガウス性と独立性を利用することで，従来よりも情報を抽出することができるかを数学的に調べた。

また，未観測共通原因がどこにありそうかを推測する方法を LiNGAM モデルの枠組みで開発することを試みた。条件付き独立性を用いる因果関係推測法の枠組みでは，そのような方法は FCI アルゴリズムが知られているが，LiNGAM モデルの枠組みでは対応する方法がなかった。

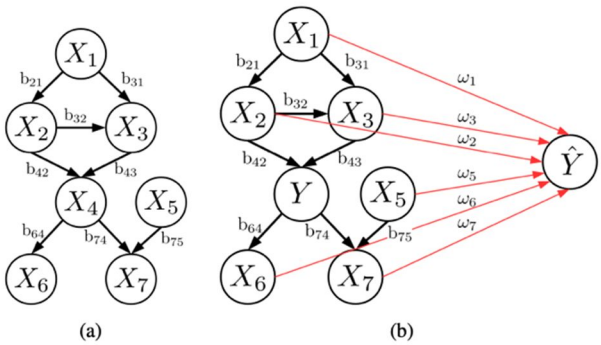
これらの方法を組み合わせて使用することで，異質性や非定常性があるデータに対して，従来よりも効果的に対処できると考えられる。

４．研究成果

(1) 離散変数と連続変数が混在する場合の因果探索法の検討
観察データから因果モデルを推定することは，データ解析において重要な課題である。Shimizu et al. (2006, Journal of Machine Learning Research)は，連続値のデータに対して，データ生成過程を理解するために線形非巡回非ガウスモデルを提案し，サンプルサイズが十分に多い場合にモデルが識別可能であることを示している。しかし，同じ問題で連続変数と離散変数が混在する状況は，実際にはよくあることである。既存のほとんどの因果探索法は，離散変数を除いて連続変数用のアルゴリズムを適用するか，またはすべての連続データを離散化してから離散変数用のアルゴリズムを適用するかのいずれかである。これらの手法では，離散変数を無視したり，離散化による近似誤差が生まれると，重要な情報が失われる可能性がある。本成果では，連続変数と離散変数の両方からなる新しいハイブリッド因果モデルを定義した。このモデルでは，以下のような仮定をしている。(1)連続変数の値は親変数の線形関数に非ガウスノイズを加えたものであり，(2)各離散変数はロジスティック変数であり，分布パラメータは親変数の値に依存する。さらに，モデル選択のための BIC スコアを導出した。新しい因果探索アルゴリズムは，離散化せずに連続データと離散データが混在したデータから因果構造を学習する。シミュレーションを通じて，本手法の効果も検討した。プレプリントととして次のようにまとめ，現在論文改訂中である。

C. Li and S. Shimizu. **Combining linear non-Gaussian acyclic model with logistic regression model for estimating causal structure from mixed continuous and discrete data**. *Arxiv preprint arXiv:1802.05889*, 2018.

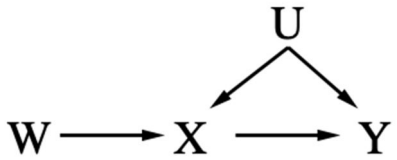
(2) 機械学習モデルと因果モデルを組み合わせたモデルの提案
機械学習が社会実装される中で，予測問題における予測メカニズムの解釈可能性が注目集めている。いくつかの研究では，解釈しやすいパラメータをもつ予測モデルを開発することに焦点が当てられてきたが，説明変数と予測値との間の因果関係は考慮されてこなかった。ここでは，データ生成プロセスの背後にある因果構造と予測メカニズムの因果構造を結びつける。図 2 の(a)が因果モデルで，(b)がそれを機械学習モデルと結びつけたモデルである。そのために，予測に最も大きな因果効果をもつ特徴を特定し，望ましい予測を得るために必要な特徴の因果介入を推定するフレームワークを提案した。このフレームワーク自体は，線形性の仮定を必要とはしないが，ここでは線形モデルを例にして議論を行った。フレームワークの適用可能性については，人工データおよび実世界のデータを用いて検討を行った。その成果は次の論文として出版し，当該国際ワークショップにおいて発表を行った。



P. Blöbaum and S. Shimizu. Estimation of interventional effects of features on

prediction. In Proc. 2017 IEEE International Workshop on Machine Learning for Signal Processing (MLSP2017), pp. 1--6, Tokyo, Japan, 2017.

(2) 非ガウス性と独立性を利用した操作変数法
観察データから因果関係を学習するには，強い仮定が必要である。これは一般的に背景知識によって正当化される。なんらかの仮定の下で，ある変数が目的の因果関係 X → Y に対して，因果グラフの構造を基に操作変数として利用可能であるかを調べることが可能である。しかし，これらの仮定がどの程度一般的なものであるか，また，解である可能性がある同値類をどのように表現するかについては，方法が十分に発展していない。本成果では，完全な因果グラフを再構築することなく，操作変数を定義する局所的なグラフ基準を導き出すことなく，どのような因果効果の集合が発見できるかあるいはできないかを体系的に特徴づける操作変数探索法を導入した。また，非ガウス性の仮定を利用した初の方法を紹介し，識別可能性の問題と解決策を提案した。このようなモデルを有限データから推定することは簡単ではないため，実際の推定を改良する方法も検討した。この結果は次の機械学習のトップジャーナルに掲載された。

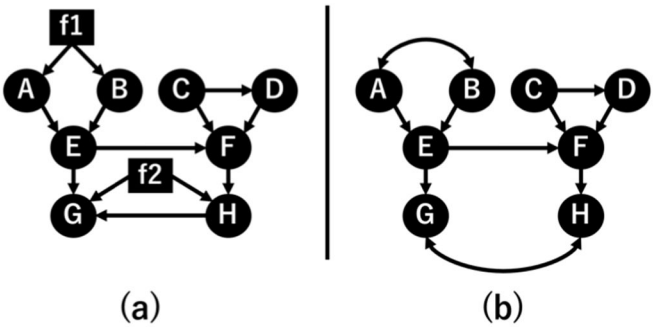


R. Silva and S. Shimizu. **Learning instrumental variables with structural and non-Gaussianity assumptions**. *Journal of Machine Learning Research*, 18: 1--49, 2017.

(3) 非線形因果探索法
2 変数間の因果関係の方向を推論する問題を，それぞれの方向で回帰した時の予測の最小二乗誤差を比較することによって解くことを考えた。原因と結果に関連する関数，条件付きノイズ分布，原因の分布の間に独立性があるという仮定の下で，両方の変数が等しく尺度化され，因果関係が決定論的に近い場合に，因果方向の誤差が小さくなることを示す。これに基づいて，可能な因果関係の両方向への回帰と誤差の比較のみを必要とする実装が簡単なアルゴリズムを提案した。このアルゴリズムの性能を，人工データと実データを用いて，様々な関連する因果探索法と比較した。この結果は次の機械学習の国際会議に採択された。

P. Blöbaum, D. Janzing, T. Washio, S. Shimizu, B. Schölkopf. **Cause-Effect Inference by Comparing Regression Errors**. In *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics (AISTATS2018)*, 2018.

(4) 未観測共通原因に頑健な因果探索法
潜在共通原因がある場合の因果探索法を研究開発することは，重要かつ困難な課題である。関数因果モデルに基づいたアプローチは，潜在共通原因の影響を受ける変数がどこにあるかを探すためには使われていないが，条件付き独立性に基づく方法で制約ベースの方法にはそのような機能を持つものもある。本成果では，潜在共通原因がある場合に観測変数の因果構造を探索するための関数因果モデルに基づく手法を提案した。この手法は,少数の観測変数間の因果方向の推論を繰り返し，その関係が潜在共通原因の影響を受けているかどうかを判定する。そして、最終的に因果グラフを構築する。下図のように双方向の矢印は同じ潜在共通原因を持つ変数のペアを示し，有向の矢印は潜在共通原因の影響を受けない変数ペアの因果方向を示している。人工データと実データを用いた数値実験では，変数間の潜在共通原因のある場所と因果方向を推定するのに有効であった。この結果は次の機械学習の国際会議に採択された。



T. N. Maeda and S. Shimizu. **RCD: Repetitive causal discovery of linear non-Gaussian acyclic models with latent confounders**. In *Proc. 23rd International Conference on Artificial Intelligence and Statistics (AISTATS2020), Palermo, Sicily, Italy*, 2020.

５．主な発表論文等

〔雑誌論文〕　計10件（うち査読付論文　10件／うち国際共著　5件／うちオープンアクセス　4件）

１．著者名 Blobaum Patrick、Janzing Dominik、Washio Takashi、Shimizu Shohei、Scholkopf Bernhard	４．巻 5
２．論文標題 Analysis of cause-effect inference by comparing regression errors	５．発行年 2019年
３．雑誌名 PeerJ Computer Science	６．最初と最後の頁 e169～e169
掲載論文のDOI（デジタルオブジェクト識別子） https://doi.org/10.7717/peerj-cs.169	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

１．著者名 Li Weimin、Zhou Xiaokang、Shimizu Shohei、Xin Mingjun、Jiang Jiulei、Gao Honghao、Jin Qun	４．巻 7
２．論文標題 Personalization Recommendation Algorithm Based on Trust Correlation Degree and Matrix Factorization	５．発行年 2019年
３．雑誌名 IEEE Access	６．最初と最後の頁 45451～45459
掲載論文のDOI（デジタルオブジェクト識別子） https://doi.org/10.1109/ACCESS.2018.2885084	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

１．著者名 Li Weimin、Zhu Heng、Zhou Xiaokang、Shimizu Shohei、Xin Mingjun、Jin Qun	４．巻 1
２．論文標題 A Novel Personalized Recommendation Algorithm Based on Trust Relevancy Degree	５．発行年 2018年
３．雑誌名 Proc. DASC/PiCom/DataCom/CyberSciTec2018	６．最初と最後の頁 418-422
掲載論文のDOI（デジタルオブジェクト識別子） https://doi.org/10.1109/DASC/PiCom/DataCom/CyberSciTec.2018.00084	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

１．著者名 Patrick Bloebaum, Dominik Janzing, Takashi Washio, Shohei Shimizu, Bernhard Schoelkopf	４．巻 84
２．論文標題 Cause-Effect Inference by Comparing Regression Errors	５．発行年 2018年
３．雑誌名 Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics (AISTATS2018), PMLR	６．最初と最後の頁 900-909
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

1．著者名 Shimizu Shohei	4．巻 20
2．論文標題 Non-Gaussian Methods for Causal Structure Learning	5．発行年 2018年
3．雑誌名 Prevention Science	6．最初と最後の頁 431～441
掲載論文のDOI（デジタルオブジェクト識別子） https://doi.org/10.1007/s11121-018-0901-x	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

1．著者名 Ricard Silva, Shohei Shimizu	4．巻 18
2．論文標題 Learning instrumental variables with structural and non-Gaussianity assumptions	5．発行年 2017年
3．雑誌名 Journal of Machine Learning Research	6．最初と最後の頁 1-49
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

1．著者名 Blobaum Patrick、Washio Takashi、Shimizu Shohei	4．巻 44
2．論文標題 Error asymmetry in causal and anticausal regression	5．発行年 2017年
3．雑誌名 Behaviormetrika	6．最初と最後の頁 491～512
掲載論文のDOI（デジタルオブジェクト識別子） https://doi.org/10.1007/s41237-017-0022-z	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

1．著者名 Blobaum Patrick、Shimizu Shohei、Washio Takashi	4．巻 1
2．論文標題 A novel principle for causal inference in data with small error variance	5．発行年 2017年
3．雑誌名 n Proc. 25 th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN2017),	6．最初と最後の頁 347,352
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

1．著者名 Blobaum Patrick、Shimizu Shohei	4．巻 1
2．論文標題 Estimation of interventional effects of features on prediction	5．発行年 2017年
3．雑誌名 Proc. 2017 IEEE Machine Learning for Signal Processing Workshop (MLSP2017)	6．最初と最後の頁 1,6
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

1．著者名 S. Shimizu	4．巻 -
2．論文標題 Non-Gaussian structural equation models for causal discovery	5．発行年 2016年
3．雑誌名 Statistics and Causality: Methods for Applied Empirical Research	6．最初と最後の頁 153-184
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

〔学会発表〕　計11件（うち招待講演　10件／うち国際学会　5件）

1．発表者名 Shohei Shimizu
2．発表標題 Causal discovery, prediction, and control
3．学会等名 Causal Modeling and Machine Learning (CaMaL) Workshop, Guangzhou, China.（招待講演）（国際学会）
4．発表年 2018年

1．発表者名 Shohei Shimizu
2．発表標題 Causal discovery, prediction mechanisms, and control
3．学会等名 he 5th meeting of the Institute of Mathematical Statistics (IMS) meeting series, the IMS Asia Pacific Rim Meeting (IMS-APRM), Singapore（招待講演）（国際学会）
4．発表年 2018年

１．発表者名
清水昌平

２．発表標題
因果探索、予測、そして制御

３．学会等名
2018年度 統計関連学会連合大会，東京．応用統計学会企画セッション: 「統計的因果推論―基本的なアイデアから最近の発展まで―」（招待講演）

４．発表年
2018年

１．発表者名
清水 昌平

２．発表標題
統計的因果推論への招待 - 因果構造探索を中心に -

３．学会等名
システム制御情報学会・計測自動制御学会 チュートリアル講座2017（招待講演）

４．発表年
2017年

１．発表者名
Shohei Shimizu

２．発表標題
Causal discovery and prediction mechanisms

３．学会等名
France/Japan Machine Learning Workshop（招待講演）（国際学会）

４．発表年
2017年

１．発表者名
清水 昌平

２．発表標題
因果探索への招待

３．学会等名
電子情報通信学会IA(インターネットアーキテクチャ)/IN(情報ネットワーク)併催研究会（招待講演）

４．発表年
2017年

１．発表者名
清水 昌平

２．発表標題
因果探索入門

３．学会等名
日本行動計量学会 第20回春の合宿セミナー（招待講演）

４．発表年
2017年

１．発表者名
S. Shimizu

２．発表標題
A non-Gaussian model for causal discovery in the presence of hidden common causes

３．学会等名
A non-Gaussian model for causal discovery in the presence of hidden common causes（招待講演）（国際学会）

４．発表年
2016年

１．発表者名
S. Shimizu

２．発表標題
A non-Gaussian approach for causal structure learning in the presence of hidden common causes

３．学会等名
CRM Workshop: Statistical Causal Inference and its Applications to Genetics（招待講演）（国際学会）

４．発表年
2016年

１．発表者名
清水昌平

２．発表標題
因果構造探索の基本

３．学会等名
研究集会：因果推論の基礎（招待講演）

４．発表年
2017年

１．発表者名 芳賀麻誉美，清水昌平
２．発表標題 関係流動性と消費者自民族中心主義の因果構造分析～非ガウス性を使った因果推論
３．学会等名 日本マーケティング・サイエンス学会 第100回研究大会
４．発表年 2016年

〔図書〕　計2件

１．著者名 清水 昌平	４．発行年 2017年
２．出版社 講談社	５．総ページ数 192
３．書名 統計的因果探索	

１．著者名 黒木学，　清水昌平，　湊真一，　石畠正和，　樺島祥介，　田中和之，　本村陽一，　玉田嘉紀，　鈴木譲，　植野真臣	４．発行年 2016年
２．出版社 共立出版	５．総ページ数 292
３．書名 確率的グラフィ カルモデル	

〔産業財産権〕

〔その他〕

―

６．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考