

科学研究費助成事業 研究成果報告書

令和 6 年 6 月 2 日現在

機関番号：34310

研究種目：基盤研究(C) (一般)

研究期間：2021～2023

課題番号：21K12018

研究課題名(和文) 新しい知識の気づきに導く共非線形性分析手法の開発と応用

研究課題名(英文) Development and Application of Co-nonlinearity Analysis Methods Leading to Novel Knowledge Awareness

研究代表者

大崎 美穂 (Ohsaki, Miho)

同志社大学・理工学部・教授

研究者番号：30313927

交付決定額(研究期間全体)：(直接経費) 3,200,000円

研究成果の概要(和文)：多変数間の複雑な従属関係(共非線形性)を知ることは現象の理解と解明への第一歩である。既知知識の積上げでは気づき難い未知の共非線形性を、多種多様なデータから帰納的に見出す機械学習技術が望まれる。この実現に向けて、本研究では以下の尺度と手法の提案・開発・評価に取り組んだ。ニューラルネットワーク回帰(NNR)とグループラッソ(GL)を組み合わせて共非線形性を検出する尺度NNR-GL。NNR-GLに情報統合(IA)の機能を加えて共非線形変数集合・代表を発見する手法NNR-GLIA。研究成果としてNNR-GLとNNR-GLIAを完成させ、実験によりこれらの原理的・実用的有効性を示すことができた。

研究成果の学術的意義や社会的意義

共非線形性尺度NNR-GLは従来の従属関係尺度の問題(2変数間に限定、設定の難しさ等)を解決し、より高い性能を達成できる。NNR-GLで得た従属関係を集合に集約するNNR-GLIAは共非線形変数集合・代表を発見可能である。断片的な従属関係を検出する尺度や事前に絞り込んだ変数集合の因果構造を推論する手法は存在するが、変数集合を絞り込む手法は見られない。我々が知る限り、NNR-GLIAはこの役割を担う初の手法である。NNR-GLとNNR-GLIAは「まずは仮説にとらわれずに現象の理解の糸口を探す」ことを支援するため、科学や工学に広く貢献すると考えられる。

研究成果の概要(英文)：Knowing the complex dependent relationships (co-nonlinearity) between multiple variables is the first step to understanding and elucidating phenomena. Machine learning techniques are necessary that can inductively discover unknown co-nonlinearities, which are difficult to notice with known knowledge, from a wide variety of data. To realization that, the present study proposed, developed, and evaluated the following measure and method: A measure called NNR-GL that detects co-nonlinearity by the combination of Neural Network Regression (NNR) and Group Lasso (GL). A method NNR-GLIA that discovers the sets of co-nonlinear variables and representatives by adding the function of Information Aggregation (IA) to NNR-GL. As research results, we completed NNR-GL and NNR-GLIA and demonstrated their theoretical and practical effectiveness through experiments.

研究分野：知能情報学

キーワード：知識発見 非線形従属関係 共非線形性尺度 正則化 ニューラルネットワーク グループラッソ

様式 C - 19、F - 19 - 1 (共通)

1. 研究開始当初の背景

現象を構成する変数間の複雑な従属関係を明らかにすることは、現象の理解予測や問題解決に不可欠である。ゆえに科学とその応用では永らく、変数間の従属関係を解明する試みが成されてきた。これらの試みを俯瞰すると、次の3段階が考えられる。第1段階(粗削りな知識の発見): 現象が多変数を持つ、あるいは、複数の現象が混入した状況で従属関係を持つ変数を見出す。第2段階(知識の洗練化): 第1段階で得た変数間の従属関係をドメイン知識と対応付け、定性的に従属関係を解釈する。第3段階(知識の確立): 第2段階で得た解釈可能な従属関係を定式化して、数理的・定量的に記述する。

従来、影響要因を厳密に制御した実験と現象に関する既知の知識に基づき、条件を絞り込んで第2, 3段階に取り組む従属関係分析が主流である。第1段階から始める場合も、既知の知識によって2変数間に線形関係や限定的な非線形関係のみを仮定することが多い。既知の視点からのアプローチは重要であるが、第1段階において、専門家が気づいていない視点を与えるアプローチも望まれる。深層学習によるブラックボックス化した問題解決が普及した現在では、本研究の開始当初以上に、変数同士の関係と現象への影響を知るニーズが高まっている。

2. 研究の目的

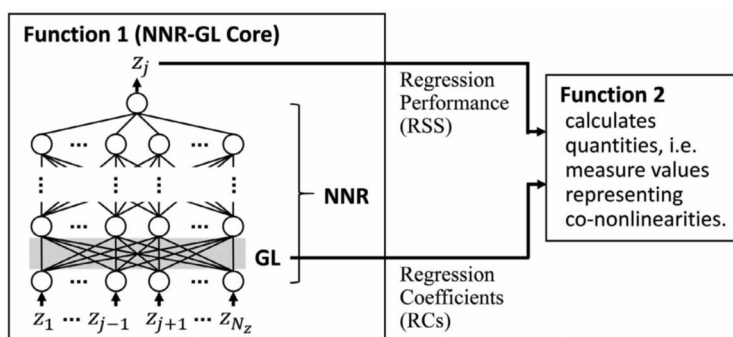
我々は以上の背景を踏まえ、現象を非線形モデル化しつつも、人が理解し得る元の変数空間に立ち返ることが可能な機械学習の仕組みを考案して、第1段階の気づきと第2, 3段階への橋渡しを支援することを考えた。これを実現すべく、「多変数」「非線形」「再現性」「解釈性」を満たす形で、ニューラルネットワーク回帰(NNR)、グループラッソ(GL)、情報統合(IA)を要素技術とする従属関係分析手法の確立を目的とした。より具体的に述べると、多変数間の非線形な従属関係(本研究では共非線形性と呼ぶ)を検出する尺度NNR-GL、および、NNR-GLを活用して共非線形変数集合・代表を発見するNNR-GLIAの提案・開発・評価に取り組んだ。

3. 研究の方法

(1) 共非線形性尺度 NNR-GL

変数間の従属関係の検出のために、これまでに様々な尺度が提案されてきた。最も一般的な尺度は相関係数(CC)であるが、これは2変数間の線形な従属関係しか検出できない。非線形な従属関係を検出できる尺度としては以下が挙げられる。特性関数への変換に基づく距離相関係数(DCC)、カーネル写像に基づくヒルベルト・シュミット独立基準(HSIC)、変数空間の区分化に基づく最大情報係数(MIC)など。これらには対象が2変数であることや、過学習に陥らない設定の難しさの問題がある。そこで我々は、ニューラルネットワークによる非線形写像に基づく尺度NNR-GLを考案するに至った。

NNR-GLの構成を下図に示す(文献[1] M. Ohsaki et al., NNR-GL: A Measure to Detect Co-Nonlinearity Based on Neural Network Regression Regularized by Group Lasso, IEEE Access, 2021より引用)。NNR-GLは、入力層の各変数にL1正則化を課すGLを組み込んだNNRと、回帰性能と変数の重みを統合して尺度値とする処理から成る。NNRは回帰を通して、多変数間に存在する共非線形性をモデル化する。GLは共非線形性の強弱を明確化して、各入力変数の重みに反映する。NNRの回帰性能とGLの変数の重みを統合すれば、複数の入力変数と出力変数の間にある非線形な従属関係の強さが求まる。これらの機能により、NNR-GLは多変数間の共非線形性を同時に検出できる。

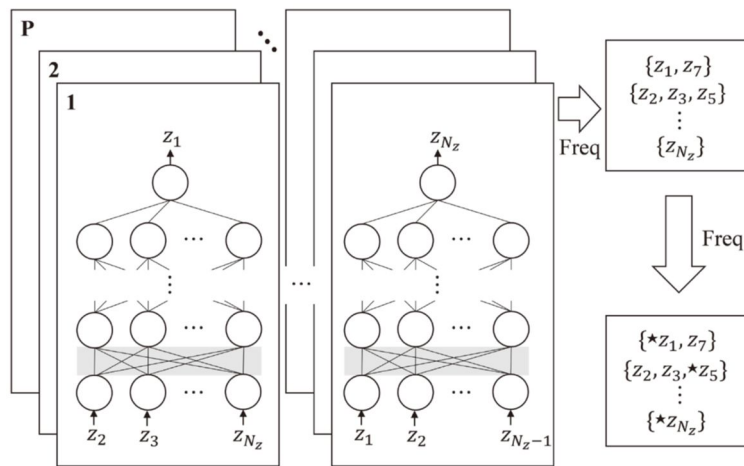


NNR-GLの有効性を検証するため、他の従属関係尺度と比較する形で次の評価実験I, IIを行った。評価実験I(原理的有効性の検証)では、正解の従属関係が分かっている人工データを用い、NNR-GLの検出結果が正解と一致するのかを調べた。人工データの作成では独立な1変数と従属な2変数を考えて、2変数の間には様々な従属関係式を想定した。具体的には線形、指数、2次、3次、正弦波の5種類の関数である。さらに、独立な1変数と従属な10変数において5種類の

従属関係が混在する複雑な条件も用いた。評価実験 II (実用的有効性の検証) では、正解の従属関係が分かっているベンチマークの実データを用い、NNR-GL の検出結果が正解と一致するのかを調べた。糖尿病に関する Pima, 病院の各種検査に関する US-130 Hospital, アワビの生育に関する Abalone, 奨学金と学業に関する Voucher の各データセットを用いた。評価実験の結果については、4. 研究成果の(1)で報告する。

(2) 共非線形変数集合・代表発見手法 NNR-GLIA

NNR-GL は多変数間の非線形な従属関係を検出できるが、変数の集合や因果関係のヒント(原因になり得る代表的な変数)までは導出できない。そこで我々は、検出した断片的な従属関係を集約して、変数集合と代表変数を求める IA を考案した。そして、NNR-GL と IA を組み合わせた共非線形変数集合・代表変数の発見手法 NNR-GLIA を提案した。その構成を下図に示す(文献[2] M. Ohsaki et al., Evaluation of the Neural-network-based Method to Discover Sets and Representatives of Nonlinearly Dependent Variables, IEEE CYBCONF, 2021, 文献[3] 大崎ら, 共非線形性尺度 NNR-GL を用いた変数集合発見とその実験的評価, 情報科学技術フォーラム, 2023 より引用)。NNR-GLIA は初期乱数や出力変数を変化させながら NNR-GL を繰り返し用い、複数の検出結果を求める。そして、変数の出現頻度に応じて IA に結果を集約させる。NNR-GL がモデル化した共非線形写像の凸性に基づき、集約した変数集合内の代表変数も発見する。



NNR-GLIA の性能評価として、原理的有効性を検証する評価実験 III と実用的有効性を検証する評価実験 IV を行った。実験条件には NNR-GL の評価実験と同じ条件、および、本研究の将来的な発展を見越して因果推論・探索手法と比較する条件を採用した。評価実験の結果については、4. 研究成果の(2)で報告する。

4. 研究成果

(1) 共非線形性尺度 NNR-GL

本研究の成果として、共非線形性尺度 NNR-GL を新たに提案し、その技術とソフトウェアを開発できた。そして、3.(1)で述べた評価実験 I, II により NNR-GL の有効性を確認できた。以下に、評価実験 I の結果を抜粋して報告する(文献[1]より引用)。下図は、人工データに対する非線形従属関係の検出結果の一部である。上側は2次関数、下側は3次関数の関係を持つ条件である。横軸は観測時に加わるノイズ率、縦軸は真陽性率、もしくは真陰性率である。NNR-GL の性能は星印を付与した水色の曲線で、従来の尺度の性能は他の曲線でプロットした。右の表には真陽性と真陰性を総合した性能を示す。NNR-GL の曲線は最も高い位置にあり、従来の尺度よりも総合性能が高く、ノイズにも頑強であった。

Parab				CC	0.6568
				DCC	0.7830
				HSIC	0.7839
				MIC	0.8466
				NNR-GL(S)	0.9688
				NNR-GL(M)	0.9681
Cubic				CC	0.7794
				DCC	0.8174
				HSIC	0.7887
				MIC	0.8181
				NNR-GL(S)	0.9500
				NNR-GL(M)	0.9424

(2) 共非線形変数集合・代表発見手法 NNR-GLIA

本研究のさらなる成果として, NNR-GL を活用して変数集合・代表を発見する NNR-GLIA を新たに提案した. その技術とソフトウェアの開発とともに, 3.(2) で述べた評価実験 III, IV により NNR-GLIA の有効性を示すことができた. 人工データを用いた評価実験 III のうち, 複数の従属関係が混在する条件で得た結果を下表に示す(文献[2]より引用). なお, 正解の変数集合・代表は{*x1, *x2}, {*x3, x4}, {*x5, x6}, {*x7, x8}, {*x9, x10}, {*x11}である. *は代表変数, x11は他の全変数と独立な変数である. NNR-GLIA は多くの条件で正解を発見しており(太字の箇所), その性能は従来の尺度よりも高かった.

3000 sample points (1000 for training, 1000 for validating, and 1000 for testing)				
Dependence	Noise rate: 0 [%] Results 1, 2, and 3	Noise rate: 1 [%] Results 1, 2, and 3	Noise rate: 5 [%] Results 1, 2, and 3	Noise rate: 10 [%] Results 1, 2, and 3
Multi-sets	{*x1, *x2}, {*x3, x4}, {*x5, x6}, {*x7, x8}, {*x9, x10}, {*x11}	{*x1, *x2}, {*x3, x4}, {*x5, x6}, {*x7, x8}, {*x9, x10}, {*x11}	{*x1, *x2}, {*x3, x4}, {*x5, x6}, {*x7, x8}, {*x9}, {*x10}, {*x11}	{*x1, *x2}, {*x3}, {*x4}, {*x5, x6}, {*x7, x8}, {*x9}, {*x10}, {*x11}

1500 sample points (500 for training, 500 for validating, and 500 for testing)				
Dependence	Noise rate: 0 [%] Results 1, 2, and 3	Noise rate: 1 [%] Results 1 and 3	Noise rate: 5 [%] Results 1, 2, and 3	Noise rate: 10 [%] Results 1, 2, and 3
Multi-sets	{*x1, *x2}, {*x3, x4}, {*x5, x6}, {*x7, x8}, {*x9, x10}, {*x11}	{*x1, *x2}, {*x3, x4}, {*x5, x6}, {*x7, x8}, {*x9, x10}, {*x11}	{*x1, *x2}, {*x3, x4}, {*x5, x6}, {*x7, x8}, {*x9}, {*x10}, {*x11}	{*x1, *x2}, {*x3}, {*x4}, {*x5, x6}, {*x7, x8}, {*x9}, {*x10}, {*x11}
		Result 2 {*x1, *x2}, {*x3, x4}, {*x5, x6}, {*x7, x8}, {*x9}, {*x10}, {*x11}		

300 sample points (100 for training, 100 for validating, and 100 for testing)				
Dependence	Noise rate: 0 [%] Results 1, 2, and 3	Noise rate: 1 [%] Results 1, 2, and 3	Noise rate: 5 [%] Results 1, 2, and 3	Noise rate: 10 [%] Results 1, 2, and 3
Multi-sets	{*x1, *x2}, {*x3, x4}, {*x5, x6}, {*x7, x8}, {*x9}, {*x10}, {*x11}	{*x1, *x2}, {*x3, x4}, {*x5, x6}, {*x7, x8}, {*x9}, {*x10}, {*x11}	{*x1, *x2}, {*x3, x4}, {*x5, x6}, {*x7, x8}, {*x9}, {*x10}, {*x11}	{*x1, *x2}, {*x3}, {*x4}, {*x5, x6}, {*x7}, {*x8}, {*x9}, {*x10}, {*x11}

150 sample points (50 for training, 50 for validating, and 50 for testing)				
Dependence	Noise rate: 0 [%] Results 1 and 2	Noise rate: 1 [%] Results 1 and 3	Noise rate: 5 [%] Results 1, 2, and 3	Noise rate: 10 [%] Results 1 and 3
Multi-sets	{*x1, *x2}, {*x3}, {*x4}, {*x5, x6}, {*x7, x8}, {*x9}, {*x10}, {*x11}	{*x1, *x2}, {*x3, x4}, {*x5, x6}, {*x7, x8}, {*x9}, {*x10}, {*x11}	{*x1, *x2}, {*x3}, {*x4}, {*x5, x6}, {*x7, x8}, {*x9}, {*x10}, {*x11}	{*x1, *x2}, {*x3}, {*x4}, {*x5, x6}, {*x7}, {*x8}, {*x9}, {*x10}, {*x11}
	Result 3 {*x1, *x2}, {*x3, x4}, {*x5, x6}, {*x7, x8}, {*x9}, {*x10}, {*x11}	Result 2 {*x1, *x2}, {*x3}, {*x4}, {*x5, x6}, {*x7, x8}, {*x9}, {*x10}, {*x11}		Result 2 {*x1, *x2}, {*x3}, {*x4}, {*x5}, {*x6}, {*x7}, {*x8}, {*x9}, {*x10}, {*x11}

次に, 実データを用いた評価実験 IV の結果の一部を下表に示す. 糖尿病に関する Pima データセットに, NNR-GLIA, および, 従来の尺度と IA の組合せを適用した結果である. 正解の従属変数集合は, {妊娠回数 x1, 年齢 x8}, {グルコース濃度 x2, 血圧 x3, 三頭筋の厚さ x4, インシュリン量 x5, BMI x6}, {糖尿病の血統 x7} の 3 つである. 表中の太字は正解との一致を表す. NNR-GLIA のみが全正解を発見し, かつ, 不正解を出さなかった. ここには掲載していないが, 他の実データについても従来の尺度より NNR-GLIA の方が高性能であった.

Measure	Detected sets of dependent variables
CC	{x1, x3, x8}, {x2, x5}, {x4, x6}, {x7}
DCC	{x1, x3, x8}, {x2, x5}, {x4, x6}, {x7}
HSIC	{x1, x3, x8}, {x4, x5, x6}, {x2}, {x7}
MIC	{x1, x8}, {x4, x6}, {x2}, {x3}, {x5}, {x7}
NNR-GLIA	{x1, x8}, {x2, x3, x4, x5, x6}, {x7}

以上より, ニューラルネットワーク回帰 (NNR), グループラッソ (GL), 情報統合 (IA) を要素技術として組み合わせて, 多変数間の非線形従属関係を分析する手法 NNR-GL, NNR-GLIA を実現したと言える. 研究期間中に得られた成果は学会発表と学術論文で公表した. 今後の研究では, 「変数間に何らかの関係がある(従属)」を発見する本分析手法を, 「何が原因で結果が生じたのか(因果)」の発見へと発展させる予定である.

5. 主な発表論文等

〔雑誌論文〕 計3件（うち査読付論文 3件/うち国際共著 2件/うちオープンアクセス 1件）

1. 著者名 Ohsaki Miho, Kishimoto Naoya, Sasaki Hayato, Ikeura Ryoji, Katagiri Shigeru, Ohnishi Kei, Sebastian Yakub, Then Patrick	4. 巻 9
2. 論文標題 NNR-GL: A Measure to Detect Co-Nonlinearity Based on Neural Network Regression Regularized by Group Lasso	5. 発行年 2021年
3. 雑誌名 IEEE Access	6. 最初と最後の頁 132033 ~ 132052
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/ACCESS.2021.3111105	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 Ohsaki Miho, Sasaki Hayato, Kishimoto Naoya, Katagiri Shigeru, Ohnishi Kei, Sebastian Yakub, Then Patrick	4. 巻 -
2. 論文標題 Evaluation of the Neural-network-based Method to Discover Sets and Representatives of Nonlinearly Dependent Variables	5. 発行年 2021年
3. 雑誌名 IEEE International Conference on Cybernetics CYBCONF-2021	6. 最初と最後の頁 101 ~ 106
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/CYBCONF51991.2021.9464151	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 Obata Rina, Ohnishi Kei, Fukumoto Makoto, Ohsaki Miho	4. 巻 -
2. 論文標題 Experiment to Investigate Awareness of Tastes for Users in Interactive Evolutionary Computation	5. 発行年 2022年
3. 雑誌名 Joint International Conference on Soft Computing and Intelligent Systems and International Symposium on Advanced Intelligent Systems SCIS&ISIS-2021	6. 最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/SCISISIS55246.2022.10002067	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6件（うち招待講演 0件/うち国際学会 0件）

1. 発表者名 岳一斗, 大西圭, 福本誠, 大崎美穂
2. 発表標題 インタラクティブ進化計算の同一タスクを連続実行するユーザに起こる変化の分析
3. 学会等名 第37回ファジィシステムシンポジウム
4. 発表年 2021年

1. 発表者名 小畑里菜, 大西圭, 福本誠, 大崎美穂
2. 発表標題 インタラクティブ進化計算におけるユーザの好みの自覚度と評価する解の質の関係
3. 学会等名 第38回ファジィシステムシンポジウム
4. 発表年 2022年

1. 発表者名 大崎美穂
2. 発表標題 小中規模データ活用と説明可能性のための機械学習と知識発見
3. 学会等名 同志社大学主催, 第9回「新ビジネス」フォーラム
4. 発表年 2023年

1. 発表者名 大崎美穂, 田儀樹, 佐々木捷人, 岸本真弥, 片桐滋, 大西圭
2. 発表標題 共非線形性尺度NNR-GLを用いた変数集合発見とその実験的評価
3. 学会等名 情報科学技術フォーラムFIT2023
4. 発表年 2023年

1. 発表者名 田儀樹, 大崎美穂, 片桐滋, 大西圭
2. 発表標題 混同行列に基づく不均衡データ分類ニューラルネットワーク
3. 学会等名 情報科学技術フォーラムFIT2023
4. 発表年 2023年

1. 発表者名 中屋敷楓, 小畑里菜, 口石和真, 大西圭, 福本誠, 大崎美穂
2. 発表標題 インタラクティブ進化計算ユーザの好みの自覚に関する実験
3. 学会等名 進化計算シンポジウム2023
4. 発表年 2023年

〔図書〕 計0件

〔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
研究分担者	大西 圭 (Ohnishi Kei) (30419618)	九州工業大学・大学院情報工学研究院・教授 (17104)	
研究分担者	片桐 滋 (Katagiri Shigeru) (40396114)	同志社大学・理工学部・教授 (34310)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関			
オーストラリア	Charles Darwin University			
マレーシア	Swinburne University of Technology			