2版

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科研費 KAKENHI

科学研究費助成事業　　研究成果報告書

令和　５ 年　 ６ 月 ２７ 日現在

機関番号：１４３０１

研究種目：基盤研究(B)（一般）

研究期間：2018～2020

課題番号：１８Ｈ０３２５２

研究課題名（和文）視覚的分析技術を使ったビッグデータからの偏微分方程式の導出

研究課題名（英文）Derivation of Partial Differential Equations from Big Data using Visual Analytics

研究代表者

小山田　耕二（Koyamada, Koji）

京都大学・学術情報メディアセンター・教授

研究者番号：００３０５２９４

交付決定額（研究期間全体）：（直接経費）　　13,300,000円

研究成果の概要（和文）：ビッグデータから偏微分方程式（PDE）を自動的に導出することは、新しい科学的知見を得るための有力な手段である。本研究では、ニューラルネットワーク（NN）を使用して、PDEモデルを構築し、KdV方程式に対してPDE導出を行った。NNは、ビッグデータから自動的にパターンを学習し、予測を行うことができる。しかし、NNは過学習に陥りやすいため、正則化回帰分析を使用して、適切な偏微分項を選択した。さらに、NNの構造に関して、層数とニューロン数を変化させた場合の誤差を可視化することで、最適なNN構造を探索した。NNモデルによって予測された解と、厳密解の違いを把握することが重要であることが示唆された。

研究成果の学術的意義や社会的意義
本研究は、ビッグデータからのPDE導出において、ニューラルネットワークを有効に使用することを示した。この手法は、物理現象の解明やエンジニアリングの最適化など、様々な分野に応用可能であり、新しい科学的知見を得るための有力な手段となる。また、NNの構造に関する最適な探索方法も提供されている。社会的には、この手法は、エネルギー効率や生産性の向上など、様々な産業分野において、新しい技術革新の促進につながる可能性がある。さらに、医療分野や環境保全などの分野にも応用可能であり、人々の生活向上に貢献することが期待される。総合的に見て、科学技術の発展と社会の発展に貢献することが期待される、重要な成果です。

研究成果の概要（英文）：Automatically deriving partial differential equations (PDEs) from big data is a powerful means of obtaining new scientific insights. In this study, a neural network (NN) was used to construct a PDE model and perform PDE derivation for the KdV equation. NNs can automatically learn patterns from big data and make predictions. However, NNs are prone to overfitting, so regularized regression analysis was used to select the appropriate partial differential terms. Furthermore, the optimal NN structure was explored by visualizing the error when changing the number of layers and neurons. It was suggested that understanding the difference between the solutions predicted by the NN model and the exact solution is important

研究分野：可視化情報学

キーワード：偏微分方程式　可視化「

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９（共通）

1．　研究開始当初の背景

過去10年間にセンサー、データストレージ、および計算資源のコスト低減によって可能になったデータ駆動型発見手法は、実験から生成された高次元のデータを特徴付けるための様々な革新的手法の開発を容易にした。このとき重要なことは、ある興味深い現象の時空間活動を示す時系列データから基礎となる物理法則や支配方程式をどのように発見するかである。基礎的な偏微分方程式（PDE）を導き出すための伝統的な理論的方法は、保存法、物理原理、および/または現象学的挙動に根ざす演繹的なものである。このような背景の元、我々は、空間的位置で収集された時系列データのみに基づいてPDEを導き出す帰納的方法を提案する。本手法を用いると与えられたデータをうまく説明する既存PDEを選択することも可能になる。

研究課題の核心をなす学術的「問い」

2017年8月8日に出版された日本学術会議提言「科学的知見の創出に資する可視化」[4]では、ビッグデータ時代に人間がすぐれた科学的知見を創出するためには、視覚的分析環境の構築・利活用が重要であると述べられている。具体的には、科学的方法の各段階において、前述の提言において明らかにされた**視覚的分析環境**構築に関する要求要件においては、以下項目がリストされている

- 観察・問題設定の段階において、超大規模なデータを如何に俯瞰させることができるか？
- 仮説構築・検証の段階において、潜在因子を如何に探索させることができるか？

ビッグデータからのPDE導出は、一種の科学的知見の創出であり、視覚的分析環境の利活用が期待できるものと考えられる。すなわち、**研究課題の核心をなす学術的「問い」**（Research Question：RQ）は、「視覚的分析システムは、どのようにして、ビッグデータからの偏微分方程式導出に資することができるか？」である。

本研究では、PDEを以下形式の非線形偏微分方程式で表現できるものとする。

$$h_t = N_{\mathbf{x}}^{\lambda} h,\ \ \mathbf{x}, \lambda \in \Omega,\ \ t \in [0, T]\ (1)$$

ここで、$h(t, \mathbf{x})$は、解を表し、$N_{\mathbf{x}}^{\lambda}$はλによってパラメータ化された非線形演算子であり、Ωは、D次元空間$\mathbf{R}^{D}$の部分集合である。 例として、1次元バーガース方程式は、

$$N_{x}^{\lambda} h = \lambda_1 h h_x - \lambda_2 h_x\ ,\ \ \lambda = (\lambda_1, \lambda_2)$$

の場合に対応する。 ここで、添え字は時間または空間のいずれかでの偏微分を示す。本研究では、まず前提として、パラメータλ を定数とする線形偏微分方程式を、次に、追加すべき偏微分項を考えるものとする。

2．　研究の目的

本研究で明らかにしようとすることは、RQ「視覚的分析システムは、どのようにして、偏微分方程式の導出に資することができるか？」に対する仮説の有効性であり、それをどの程度まで行うかについては、**研究方法（仮説検証のための実験）**で述べる。

本研究において、仮説は以下のとおり設定する。

H1　**正則化回帰**を適用すれば、重要な偏微分項を絞ることができ、効果的な偏微分方

程式の導出を可能にする。

H2 **収束的交差写像法**（Convergent Cross Mapping:CCM）を適用すれば、時間偏微分に影響を及ぼす偏微分項を選択することができ、効果的な偏微分方程式の導出を可能にする。

H3 因果グラフ可視化において**潜在因子探索**技術を適用すれば、効果的に追加すべき偏微分項の発見を支援することができる。

正則化回帰はモデルのパラメータの学習に使われ、特に過学習を防ぎ、汎化能力を高めるために使われる機械学習法である。機械学習において最も一般的なのは L1 正則化 と L2 正則化である。CCM は、2 つの時系列データ間の因果関係の統計的検定方法であり、偏微分方程式で記述された偏微分項のように、因果関係の影響が分離不可能である純粋に決定論的なシステムに最も適している[B]。CCM を使って構築される因果グラフを対象にした**潜在因子探索**は、未観測変数が表現する因子をモデル化する技術のことである[A]。

3. 研究の方法

本研究で設定した仮説検証について、これをどのように、どこまで明らかにしようとするのかを述べる。

研究内容１

H1 「正則化回帰を適用すれば、重要な偏微分項を絞ることができ、効果的に偏微分方程式の導出を可能にする。」を検証するために、時系列データが定義される離散点毎に、できるだけ多くの偏微分項から構成される線形偏微分方程式として、時間偏微分項を表現する。

本研究では、疑似計測データを使って、もとの偏微分方程式の偏微分項を含む複数の偏微分項を計算し、正則化回帰を適用し、各偏微分項に対する係数を計算する。正則化には、一般的な Lp ノルム正則化を用いる。非ゼロの係数に対する偏微分項がもとの偏微分項を含んでいるかどうかを確認する。

研究内容２

H2 「CCM を適用すれば、時間偏微分に影響を及ぼす偏微分項を選択することができ、効果的に偏微分方程式の導出を可能にする。」を検証するために、偏微分項の各ペアに対して、時系列データが定義される離散点毎に CCM を用いて因果強度を計算し、これをスカラデータとして可視化すれば、因果強度の空間分布を俯瞰的に理解することができる。

本研究では、疑似計測データを使って、もとの偏微分方程式の偏微分項を含む複数の偏微分項ペアに対する因果強度スカラデータを俯瞰可視化し、もとの偏微分項を選択できるかどうかを確認する。具体的には、偏微分項ペアに対する因果強度スカラデータに対して個別の色付けを行い、俯瞰的可視化手法としてボリュームレンダリング手法を用いて、複数スカラデータの融合的可視化を行う。具体的には、ある偏微分項に対しては、同一の色相で、その値に対して、彩度・値を変化させるような色付けを行う。また、不透明度については、スカラデータの値が不透明度と正の相関を持つように変化させる。さらに、重畳したノイズデータが、また、数値解を使った場合利用した格子解像度が、どの程度偏微分項の特定に影響するかを評価する。

研究内容３

H3 「因果グラフ可視化において潜在因子探索技術を適用すれば、効果的に追加すべき偏微分項の発見を支援することができる。」を検証するために、時間偏微分項を根ノードとする偏微分項ネットワークに対して、構造方程式モデリング手法[1]を用いて潜在因子探索を行い、説明能力を向上させるノード（非定数係数の偏微分項）を追加することができる視覚的分析システムを構築する。このシステムを使えば、偏微分方程式導出に資する偏微分項ネットワークを補完することができる。

本研究では、疑似計測データを使って、もとの偏微分方程式の偏微分項と抽出された偏微分項とを比較し、もとの偏微分項とどの程度一致するかを確認する。

4．研究成果

偏微分方程式導出において、本研究では NN を使ってモデル構築を行った後で、正則化回帰分析を使用した偏微分項の選択を行った。誤差評価に関して、層ごとのニューロン数：N_Nは同じにして、また、離散点数も一定にして、それぞれの誤差を、層数：N_Lと層ごとのニューロン数N_N（N_L, N_N）の関数としてヒートマップ表示したところ、各誤差が最小となる場所に隔たりがあった。この問題を解決するために、二つのパラメータ最適化を統合化した。すなわち、NN モデルの損失関数：E_uとE_{RRE}を加算して、単一の目的関数（損失関数）を設定し、これを最小化する適切な NN 構造を探索する手法を開発した（図 1 参照）。

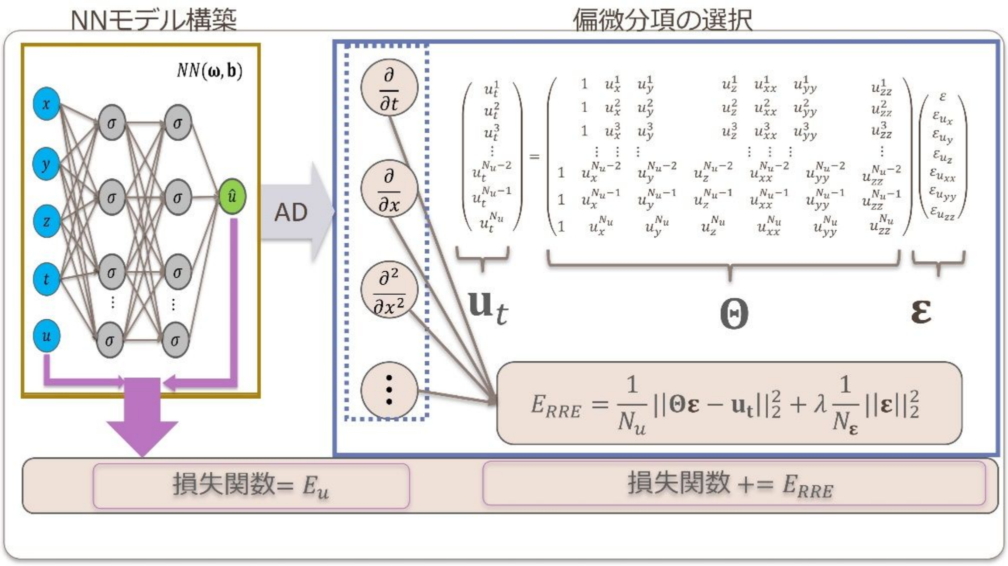


図 1　PDE 導出手法

本手法を厳密解のある KdV 方程式に適用する。偏微分項の候補については、最小構成と拡張構成の二つのケースについて行い、NN パラメータに対して、値を変化させながらその誤差を可視化する。

KdV 方程式は、非線形波動を記述する非線形偏微分方程式の一つである。ソリトン解を有する可積分系の代表的な例として知られる。時間変数 t と空間変数 x をもつ一次元実数値関数 u(x, t) に対して、α, β をゼロではない任意の実定数として

$$u_t + \alpha u u_x + \beta u_{xxx} = 0$$

で与えられる非線形偏微分方程式を KdV 方程式という。また、各変数と u に適当なスケール変換を施せば、係数を α = 6, β = 1 と取りなおすことができる。このとき、各変数に対する偏微分を右下の添え字として表せば、

$$u_t + 6uu_x + u_{xxx} = 0$$

となる。KdV 方程式は、浅水波などの非線形波動現象を記述する。初期条件が与えられた場合、

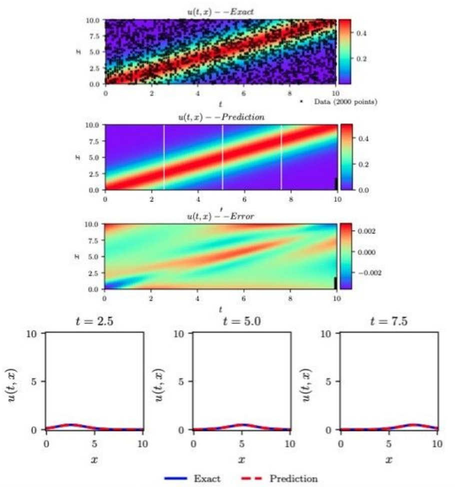
KdV 方程式は次の厳密解を持つ。

$$u(x,t) = \frac{1}{2} sech^2 \left(\frac{1}{2}(x,t) \right)$$

KdV 方程式の厳密解において、x∈[0,10]および t∈[0,10]の定義域をサンプリングして、その座標値に基づいて$u(x,t)$を計算する。 x の定義域を 256 分割し∆t ＝ 0.1 ごとに解を計算する。これらのデータについては、空間観測点の数n_x = 256、時間観測点の数n_t = 101、データセットのサイズn_d = 256856となる。結果を図 4 における最上段のカラーマップ図に示した。偏微分項候補ライブラリは、次のように正解の偏微分項（太字で示す）を拡張して構成した。

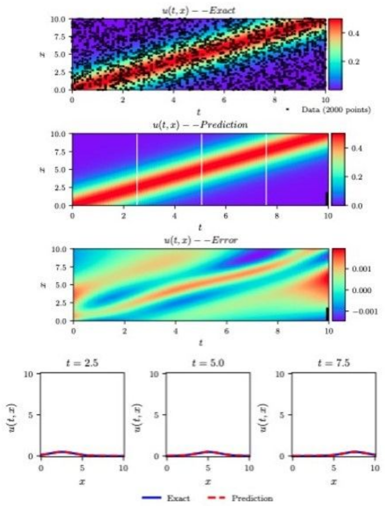
$$\Phi = [u_x, u_{xx}, \boldsymbol{u_{xxx}}, \boldsymbol{uu_x}, uu_{xx}, uu_{xxxx}, u^2u_x, u^2u_{xx}, u^2u_{xxx}]$$

まず、隠れ層ごとに 20 個のニューロンを持つ 9 層 NN を使用して、KdV 方程式の近似解を計算する。活性化関数は Sigmoid 関数 sigmoid（x）とし、学習用データセットの数は 2000 ポイントでランダムに選択された。 NN は、図 1 で定義された損失関数を最小化することによって構築される。結果を図 2 における中段のカラーマップ図に示した。図 2 における下段のカラーマップ図には、NN モデルと厳密解の間の誤差分布（$\hat{u} - u$）を示す。また、時刻を固定したときのデータ分布$\{u(x,t)|\ t$ = 2.5,5.0,7.5$\}$を誤差分布の下に示す。さらに図 2 最下段には導出偏微分方程式を示すが、ライブラリにおける候補偏微分項の数が多かったので、導出誤差は十分ではなかった。



Correct PDE	$u_t + 6uu_x + 1u_{xxx} = 0$
Identified PDE (clean data)	$u_t + 0.92477u_x + 0.40510uu_x + 0.28091u^2u_x + 0.0928930u_{xxx} = 0$

図 2　拡張ライブラリによる導出結果



Correct PDE	$u_t + 6uu_x + 1u_{xxx} = 0$
Identified PDE (clean data)	$u_t + 5.87759uu_x + 0.9571562u_{xxx} = 0$

図 3　最小ライブラリによる導出結果

そこで、ライブラリの候補偏微分項を正解なものだけに最小構成すると導出誤差は向上した。すなわち、偏微分項候補ライブラリを、次のように

$$\Phi = [uu_x, u_{xxx}]$$

と構成した場合、図 3 に、厳密解、NN モデル、誤差分布、時刻固定でのデータ分布、導出偏微分方程式を示した。

５．主な発表論文等

〔雑誌論文〕　計1件（うち査読付論文　1件／うち国際共著　0件／うちオープンアクセス　0件）

１．著者名 Koji Koyamada, Yu Long, Takuma Kawamura, and Katsumi Konishi	４．巻 12
２．論文標題 Data-driven derivation of partial differential equations using neural network model Koji Koyamada, Yu Long, Takuma Kawamura, and Katsumi Konishi	５．発行年 2021年
３．雑誌名 International Journal of Modeling, Simulation, and Scientific Computing	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） 10.1142/S1793962321400018	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5件（うち招待講演　0件／うち国際学会　1件）

１．発表者名 小山田耕二
２．発表標題 ビッグデータからの偏微分方程式導出精度における初期条件網羅性が与える影響について
３．学会等名 横幹連合シンポジウム
４．発表年 2019年

１．発表者名 Masaki Kamikubo, Katsumi Konishi, Koji Koyamada
２．発表標題 MaMultidimensional Signal Smoothing Algorithm based on Locally Low-Rank Approach
３．学会等名 JSST2019（国際学会）
４．発表年 2019年

１．発表者名 リュウ ウ，小山田 耕二，坂本 尚久，水野 翔太
２．発表標題 Data driven derivation of partial differential equations
３．学会等名 日本応用数理学会 2019年度年会
４．発表年 2019年

1．発表者名 水野翔太，小山田耕二，夏川浩明
2．発表標題 偏微分方程式の導出
3．学会等名 日本シミュレーション学会 AI研究委員会研究会
4．発表年 2018年

1．発表者名 小西克己，小山田耕二
2．発表標題 ビックデータからの偏微分方程式導出",
3．学会等名 第46回可視化情報シンポジウム2018論文集
4．発表年 2018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	小西　克巳 (Konishi katsumi) (20339138)	法政大学・情報科学部・教授 (32675)	

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