射影法を用いて保存則を発見するニューラル常微分方程式

松原 崇
大阪大学 大学院基礎工学研究科

1. はじめに

ニューラルネットワークは画像や自然言語の処理で目覚ましい成果を上げているが、力学系のモデル化にも盛んに研究されている。対象となるのは、物理シミュレーションにおける化学的ダイナミクス、気候変動予測や天気予報のための気候ダイナミクス、自動車やロボットの物理的ダイナミクスなどがある。その歴史は少なくとも 1990 年代に遡り、多くのアプローチが提案されてきた。近年提案された、ニューラル常微分方程式(NODE)は、連続時間ダイナミクスのためのニューラルネットワークを再定義した。対象となるシステムはシステムの状態を用いた常微分方程式(ODE)で記述できる。NODE はベクトル場をニューラルネットワークに置き換え、数値積分を用いて解を求める。

実世界のほとんどのシステムには、時間の経過とともに変化しない量である「保存量」が存在する。あるシステムが保存量を持つとき、初期値に対して、解は保存量の等高線に留まる。これまで、保存量に関する事前知識を取り入れることで、対象システムを正確に学習することが試みられてきた。Greydanus らは、ハミルトン方程式をニューラルネットワークで近似し、ハミルトニアンと呼ばれるシステムのエネルギーを保存するハミルトニアンニューラルネットワーク(HNN)を提案した(1)。Finzi らはグラフニューラルネットワーク等を利用して線形運動量と角運動量を保存するアーキテクチャを提案し(2)、また HNN を拡張してホロノミック制約を持つシステムを扱えるようにした(3)。松原らは、離散化された偏微分方程式(PDE)の質量を保存するモデルを提案した(4)。これらの研究によって、ニューラルネットワークは対象システムの保存量に関する事前知識をより取り入れることで、ダイナミクスをより正確にモデル化できることが証明された。

これまでの研究では、主に既知の保存量を保存することが目的であった。しかし、ニューラルネットワークが未知の対象システムを学習する状況では、対象システムの保存量も未知であることが予想され、上記のどの手法が有効であるかは明らかでない。そこで本稿では、データから保存量を発見し保存する手法を提案する。この手法は、データから様々な種類の保存量を同じ枠組みで発見し、保存しつつ予測できる。また、HNN のような既知の保存量を保存するように設計されたニューラルネットワークと組み合わせることが可能である。

2. 手法

2.1 準備

システム$\frac{d}{dt}=f(u)$を考える。$u\in\mathbb{R}^N$はN次元の状態、fはベクトル場である。ある量$V(u)$が補ｚン料であるとは、任意の解$u(t)$についてその値が一定であることである。つまり$\frac{d}{dt}V(u)=0$である。システムが少なくともK個の局所的に線形独立な保存量$V_1,\dots,V_K$を持つとき、初期値u_0から得られる解$u(t)$は$(N-K)$部分多様体

$$\mathcal{M}=\{u\in\mathbb{R}^N\colon V_k(u)=V_k(u_0)\ \mathrm{for}\ k=1,\dots,K\}$$

上に存在する。よって、点uにおいて、ベクトル場fはこの部分多様体の接空間

$$T_u\mathcal{M}=\{w\in\mathbb{R}^N\colon \nabla V_k(u)^\top w=0\ \mathrm{for}\ k=1,\dots,K\}$$

上にある。このとき

$$\frac{d}{dt}V_k(u)=\nabla V_k(u)^\top\frac{d}{dt}u=\nabla V_k(u)^\top f(u)=0$$

なので、量V_kが保存されることがわかる。

NODE や HNN のような、ニューラルネットワーク等で実装された基盤となるモデル$\frac{d}{dt}u=g(u)$を想定する。対象システム$\frac{d}{dt}=f(u)$の観測

対象システム	ダイナミクスの分類	次元数N	保存量
重力二体問題	正準形式のハミルトン系	8	エネルギー、線形運動量、角運動量
二重振り子	ポアソン系	8	エネルギー、ホロノミック制約
フィッツフュー=南雲モデル	ディラック構造	4	キルヒホッフの法則

表1：検証データセット

データを用いて、この基盤となるモデルを訓練し、数理モデルを得ることが可能である。しかし、その結果は常にモデル化誤差を含む。NODE は一切の保存量を仮定しておらず、HNN もエネルギー保存則しか仮定していない。よって、実際に予測に使うと、モデル化誤差が蓄積し、短期間で全く異なる無意味な結果を導いてしまう。

2.2　提案手法

そこで、ニューラルネットワークによって、データから保存量を同定し、それを保存しつつ予測を行うことを考える。なお、簡単禍のため 2.1 節と同じ変数を用いるが、ここからは対象システムではなくニューラルネットワークについて議論することに注意されたい。

K個の出力を持つニューラルネットワークを導入し、各出力が 1 つの保存量$V_k:\mathbb{R}^N\to\mathbb{R}\ (k=1,\dots,K)$を学習すると想定する。保存量の集まりをベクトル

$$\boldsymbol{V}(u)=\left(V_1(u),V_2(u),\dots,V_K(u)\right)^\top$$

で表現する。提案手法が作るベクトル場をfとする。V_kは保存量なので、$M(u)=\frac{\partial \mathrm{V(u)}}{\partial u}$とすると

$$\boldsymbol{0}=\frac{d}{dt}\boldsymbol{V}\big(u(t)\big)=M(u)\frac{d}{dt}u(t)=M(u)f(u)$$

が拘束条件となる。また、$f(u)$は$g(u)$に近いことが望ましい。つまり

$$\min_f\|f(u)-g(u)\|^2$$
$$s.t.\ M(u)f(u)=0$$

なる最適化問題を解きたい。ラグランジュの未定乗数$\lambda\in\mathbb{R}^K$を用いて

$$\frac{\partial}{\partial f}(\|f(u)-g(u)\|^2+\lambda^\top M(u)f(u))=0$$

より、

$$f(u)=g(u)-M(u)^\top\lambda$$

であり、これを解くと

$$f(u)=P(u)g(u),$$
$$P(u)=I-M(u)^\top(M(u)M(u)^\top)^{-1}M(u)$$

を得る。これは一種の射影法であり、ベクトル場を射影することで保存量が保存されることを保証する。

基盤となるモデルgの訓練と同じように、提案手法fを訓練できる。fはgと$\boldsymbol{V}$の組み合わせでできているため、２つのニューラルネットワークを同時に訓練すると考えることができる。もしくは、fという非常に特殊な構造をした 1 つのニューラルネットワークを訓練すると考えてもよい。いずれにせよ、自動微分を用いることで簡単に勾配を計算することができ、確率的勾配降下法によって訓練できる。

無論、学習された保存量V_kが対象システムの真の保存量と一致するとは限らず、少なくともモデル化誤差を含む。しかし、保存量が存在するという仮定によって、より精度の高いモデル化が可能であると予想される。

3. 実験

3.1　実験設定

表 1 にある保存量を持つシステムを用いて、提案手法と基盤となるモデルの比較評価を行った．2 次元配置空間上の重力二体問題は、正準形式の典型的なハミルトン系である。全エネルギーに加えて、空間の対称性に関係する保存量、つまりx方向とy方向の線形運動量と角運動量を持っている。二重振り子は、極座標においてハミルトン系である。しかし、それを直交座標に変換すると、ハミルトン系ではなくポアソン系になる。二本の振り子の長さと二つの重りの移動方向がホロノミック制約となり、それぞれが対応する保存量を持つ。エネルギーを含めて５つの保存量がある。フィッツフュー＝南雲モデルは、生物の神経細胞を電気回路としてモ

		重力二体問題		二重振り子		フィッツフュー=南雲モデル	
	K	MSE	VPT	MSE	VPT	MSE	VPT
基盤	0	5.17	0.362	0.82	0.110	73.66	0.236
提案手法	1	7.10	0.374	0.75	0.156	54.18	0.127
	2	7.78	0.450	0.73	0.198	37.03	0.437
	3	>1000	0.147	0.69	0.411	>1000000	0.007
	4	>1000	0.101	0.77	0.395		
	5	>1000	0.080	0.80	0.585		
	6	>1000	0.070	12.53	0.005		

表 2：実験結果

デル化したもので、スパイクと呼ばれる急激な電圧変化を示す。電気回路はその回路のトポロジーとキルヒホッフの電流・電圧則によって拘束されたシステムとみなすことができる。インダクタとコンデンサに流れる電流と印加される電圧を状態とすると、４つの状態に対して２つの拘束があり、二つの保存量がある。抵抗器におけるエネルギー散逸のため、ポアソン系ではなく、ディラック構造を持つとみなされる。ただ、ページ数の都合で詳細は省略する。

数値積分は Dormand–Prince 法(dopri5)を用いた。保存量$\boldsymbol{V}$、NODE、HNN を隠れ層二層の全結合ニューラルネットワークで実装した。各隠れ層は 200 ユニットで構成され、tanh 関数を活性化関数として持つ。重みは直交行列として初期化した。二重振り子は二次の常微分方程式であり、位置の時間微分は状態の一部である速度として既知であるため、加速度のみをニューラルネットワークの出力として扱った。

また、最小化する損失関数として、１ステップ誤差を用いた。具体的には、真の状態と前のステップから予測した状態の平均二乗誤差(MSE)である。学習には Adam を用い、バッチサイズ 200、学習率は10^{-3}に初期化し、コサイン波に従ってゼロまで減衰させた(5)。

評価指標として、損失関数と同じ 1 ステップ誤差を用いた。まとめる際に見やすいよう 10^9 倍した。状態やエネルギーの予測誤差は、状態が発散した場合に非常に大きくなってしまったり、位相のずれに鈍感であったりするので、代わりに valid prediction time (VPT)を用いた(6)。これは初期値問題において、予測状態の MSE が閾値を初めて超えるまでの時間を、時系列長 S で割ったものである。VPT を求める前に、状態の各要素を学習用データセットで平均 0、分散 1 となるように正規化し，閾値を 0.01 とした。

3.2　ハミルトン系からの学習

基盤となるモデルとして HNN を用い、提案手法を二体問題データセットからの学習で評価した。予備実験で、提案手法は HNN のハミルトニアンHを保存量V_kの一つとして扱わない方が良い性能を得られることがわかった。5 回の試行の中央値と標準偏差を表 2 左端にまとめた。提案手法は$K = 1$～2で単なる HNN より優れた VPT を達成し、$K = 3$で急に性能が低下した。提案手法が、HNN のハミルトニアンHに加え、２つの保存量を発見したことを示唆している。二体問題はハミルトン系であり、HNN で学習可能なはずであるが、ハミルトニアンH以外の保存量が存在するという事前知識が、より良い学習を導くことが分かる。提案手法を用いると 1 ステップ誤差が悪化したことから、提案手法を用いない HNN は短期的な変化に過適合し

ており、長期的なダイナミクスを予測することが困難であることが示唆される。

3.3　未知のシステムからの学習

その他のデータセットに対し、基盤となるモデルを NODE とした結果も表 2 にまとめた。二重振り子データセットにおいて、提案手法は$K=$1から5の範囲で 1 ステップ誤差と VPT を改善した。特に、$K = 5$では、VPT がベースラインの 5 倍以上となった。二重振り子では、システムエネルギーに加えて、位置に２つのホロノミック制約があり、速度を含む２つの制約を導く。従って、提案手法が$K = 5$で最良の VPT を得、$K > 5$で破綻するのは理に適っている。フィッツフュー＝南雲モデルはエネルギー保存系ではないが、キルヒホッフの電流・電圧則が状態を拘束し、二つの保存量を導く。$K = 2$で 1 ステップ誤差と VPT が非常に良くなった。二重振り子の場合は$K = 5$、フィッツフュー＝南雲の場合は$K = 2$と、提案手法はすべての保存量を見つけることができたと言える。

4. まとめ

本稿では、ニューラルネットワークを用いて対象システムの保存量を同定し、それを用いることでより良いモデル化と予測を実現する手法について紹介した。提案手法は、ニューラルネットワークが保存量を学習すると仮定し、時間発展を定義した部分多様体に射影することで、データから保存量を発見し保存できる。適切な数の保存量の存在を仮定すると、基盤となるモデルよりもはるかに長い時間、将来の状態を正確に予測できる。提案手法は、システムのエネルギー、運動量、制約に関する保存量を全く同じアプローチで発見し保存できる。そのため、提案手法は未知の力学系の性質を明らかにし、科学的な発見に貢献することが期待される。

参考文献

(1) S. Greydanus, M. Dzamba, and J. Yosinski, "Hamiltonian Neural Networks," Advances in Neural Information Processing Systems (NeurIPS), 2019.

(2) M. Finzi, S. Stanton, P. Izmailov, and A.G. Wilson, “Generalizing Convolutional Neural Networks for Equivariance to Lie Groups on Arbitrary Continuous Data,” International Conference on Machine Learning (ICML), pp. 3146–3157, 2020.

(3) M. Finzi, K.A. Wang, and A.G. Wilson, “Simplifying Hamiltonian and Lagrangian Neural Networks via Explicit Constraints,” Advances in Neural Information Processing Systems (NeurIPS), 2020.

(4) T. Matsubara, A. Ishikawa, and T. Yaguchi, “Deep Energy-Based Modeling of Discrete-Time Physics,” Advances in Neural Information Processing Systems (NeurIPS), 2020.

(5) D.P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," International Conference on Learning Representations (ICLR), 2015.

(6) P.R. Vlachas, J. Pathak, B.R. Hunt, T.P. Sapsis, M. Girvan, E. Ott, and P. Koumoutsakos, "Backpropagation algorithms and Reservoir Computing in Recurrent Neural Networks for the forecasting of complex spatiotemporal dynamics," Neural Networks, vol.126, pp.191–217, 2020.