

論文の内容の要旨

Geometrical Structures Embedded in High Dimensional Data Sets and Deep Learning

: Analysis and Application to Dynamical Systems

(高次元データセットに潜む幾何構造と深層学習

: その解析と大自由度力学系への応用)

氏名：本武 陽一

1 序論

近年の情報技術の急速な発展に応じて、これまでにない大規模で高次元なデータの収集が可能となってきた。一方で、我々の高次元データそのものに対する知見はまだ十分にあるとは言い難い。

本研究の目的は、画像データセットのような高次元データ（図1のように1pixelを1次元と考えた場合）の集合が持つ分布構造を、近年急速にその学習アルゴリズムが発展してきたDeep Neural Networks（以下、DNN）の内部状態の解析を通して探ることである（図2）。DNNは、画像を中心とする多様なデータセットに対して、他の学習機と比較して非常に高い識別能力を持っており、そこにはデータについての精緻な情報が埋め込まれていることが予想される。

本研究の主たる貢献は、DNNからデータセットの幾何構造についての情報を取得する手法の提案とその妥当性の定量的な検証を行ったこと、その結果に基づき、DNNによるデータセットの観察を行い、大規模画像データセットの分布が多様体構造を持つことや、画像のラベルの階層的概念構造とデータセットの幾何構造が関係することを示唆する結果を得たこと、そして、これらの結果を応用して、鳥や魚等の群れの集団行動を再現するボイドモデルで生成した大規模で複雑な時系列データから「群れ」構造を抽出することを試み、それに成功したことである。

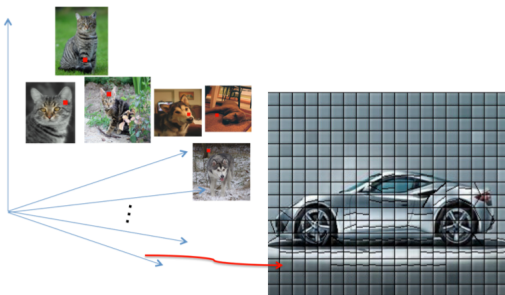


図1: データセットの幾何構造を考える空間：1pixelを1次元とし、画像が1点で表される空間。

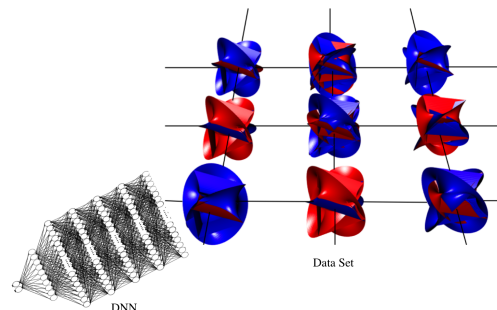


図2: データセットの構造の観測のイメージ：天体を望遠鏡で観測するように、データセットの高次元空間中での分布構造（右）をDNNを用いて観測する（左）（図はカラビ・ヤウ多様体を3次元空間に写像したもの。）

2 高次元データと多様体仮説

本研究では、データセットの構造についてのある仮説を仮定した上で、その詳細な構造を検討していく。その仮説とは、「機械学習における多様体仮説」である。本研究では、[Rifai 11]に基づき、多様体仮説を以下のように定義する。

- 仮説.1 高次元空間に存在する実世界のデータは、非常に低次元の非線形多様体付近に集中している。

仮説.1は、次のように説明される。実世界に存在するデータは、視点の変化や物体自身の運動、手書き文字データ等であれば個人差などによって、少しずつ変化したものの集合となる。その変化が連続的であれば、図3のように、同一クラスのデータが曲線や曲面上に分布すると考えられる。この曲面や曲線が多様体になると考えられるのである。

これに基づいて開発された多くの機械学習アルゴリズムが、有効に機能することから、仮説の妥当性が経験的に支持されている。一方で、直接的にこの仮説の妥当性の検証に取り組んだ研究はこれまでにあまりなく、仮説成立の成否についての答えはまだ得られていない。

3 多様体仮説とDNN

それでは、このようなデータセットの幾何構造はどのような手法で観測できるだろうか。本研究では、データセットの幾何構造の観測装置として、近年注目される機械学習装置であるDNNを用いた。DNNは、以下の仮定が満たされる場合、前節で説明したデータセットの多様体構造を観測する装置として利用できる。

- 仮説.2 パフォーマンスの高い学習済みDNNは、データセットの多様体構造を抽出し、多様体と同じ次元のユークリッド空間（図4参照。以降、「大域的な座標系」と呼称）へ写像する機能をもつ。

[入江 90] や [Hinton 06] では、中間層のノード数を多様体の次元に一致させたDeepAutoEncoderを用いて、その中間層の状態を可視化することで部分的な検証を試みている。しかしながら、それらは可視化による検証に留まっており、高次元な多様体を持つ場合にどうなるか等、この仮説が妥当であるかについての明確な答えはまだ得られていない。

そこで本研究では、仮説.1を満たすデータセットを用いて仮説.2の検証を行なった上で、構造が未知なデータセットの幾何的構造の観察を行なった。

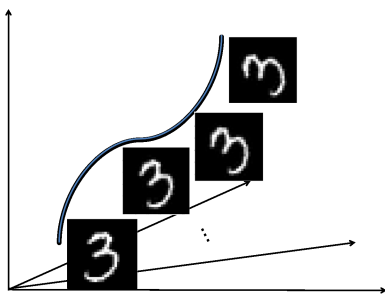


図 3: 回転に対する普遍性による多様体の形成

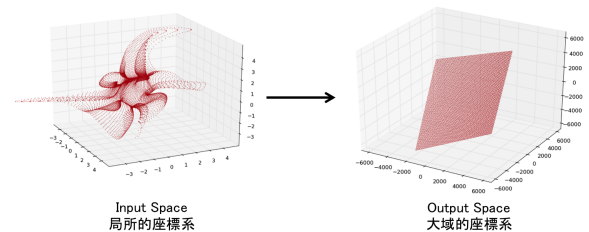


図 4: 大域的な座標系：図の変換では、局所的な座標系の集合として表現される多様体（左図赤点）が、位置によらず、同じ座標系を共有する大域的な座標系へ変換されている（右図赤点）

4 観測装置としてのDeep Neural Networksの性能検証

4.1 手法

DNNが獲得した関数を、データセット多様体を大域的な座標系へ写像する関数だとみなすと、その関数を解析することで、元の多様体の性質を知ることが可能となる。なぜならば、多様体から多様体への写像の微分は、以下で定義される多様体の接空間を定義し、そこから多様体の次元や接ベクトル等の情報を得ることができるからである。ここで考えている入力空間は、図1にあるような1pixelを1次元とする空間である。

DNNの写像関数のヤコビアン行列の特異値・特異ベクトルのうち、0より大きな特異値に対応する特異ベクトルが多様体の接線方向を、0の特異値に対応するベクトルが多様体の垂線方向をあらわす。従って、0でない特異値の数から多様体の次元がわか

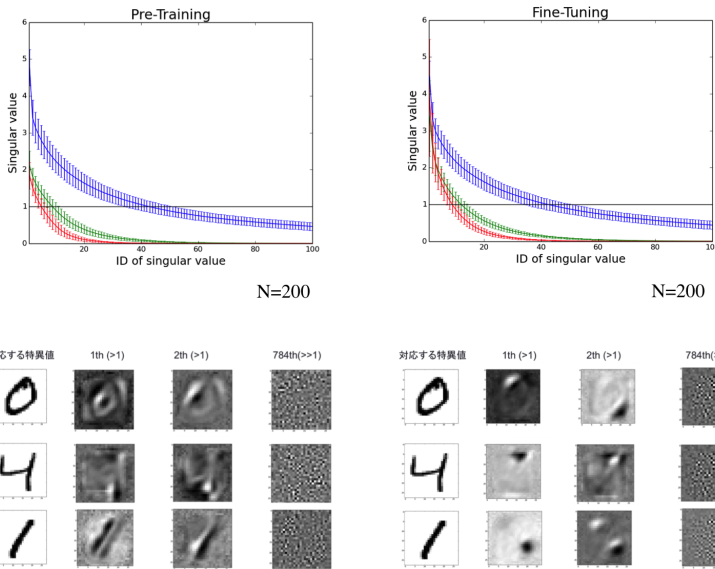


図 5: 上段左: pre-training 後の特異値分布. 上段右: fine-tuning 後の特異値分布. 下段左: pre-training 後の右特異ベクトル. 図 6 に類似したベクトルとなっている. 下段右: fine-tuning 後の右特異ベクトル. エラーバーは標準偏差.

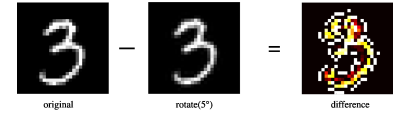


図 6: 回転に対する普遍性による多様体の形成: 元画像ベクトルから 5° 回転した画像ベクトルを引いた差分ベクトル (右図)

る. また, 特異ベクトルには, 右と左があり, 右特異ベクトルが入力空間で表現された多様体の接線・垂線ベクトルをあらわし, 左特異ベクトルは出力空間で表現された多様体の接線・垂線ベクトルをあらわす.

具体的に DNN の各層の写像関数は以下で定義される. $f(x)$ としては, よくシグモイド関数が用いられる.

$$h_j(t+1) = f\left(\sum_i h_i(t) \cdot W_{ij}(t) + B_j(t)\right) \quad (1)$$

ここで, $h_i(t)$ は t 層の隠れ層のノード状態を, $W_{ij}(t)$ は, t 層から $t+1$ 層の間の重み行列を, $B_j(t)$ は, 第 t 層のバイアス値を表すものとする. また, i, j は, 各層でのノードのインデックスになっている. 従って, 第 t 層から $t+1$ 層の写像に対するヤコビアン行列は,

$$J(t) = \begin{pmatrix} \frac{\partial h_1(t+1)}{\partial h_1(t)} & \cdots & \frac{\partial h_1(t+1)}{\partial h_N(t)} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_N(t+1)}{\partial h_1(t)} & \cdots & \frac{\partial h_N(t+1)}{\partial h_N(t)} \end{pmatrix} \quad (2)$$

となる. ネットワーク全体でのヤコビアン J_{all} は, 各層のヤコビアンの積として, 次式のように表される.

$$J_{all} = J(0) \cdot J(1) \cdots J(T-1) \quad (3)$$

4.2 実験と結果

本研究では, MNIST データセットによってトレーニングされた Deep Belief Networks[Hinton 06] を分析した. MNIST データセットは, $O(1)$ 程度の次元の多様体構造を持つと推測されているデータセットである.

MNIST データセットでトレーニングされた DNN を分析した結果, 高階層において, 少数の大きな特異値と, 大多数のほぼ 0 の特異値という急峻な特異値分布が見られた (図 5 左上). また, 多様体の次元に対応する 1 以上の特異値の数は, 第 3 層において $O(1)$ 程度であり, これは先行研究から予想された次元であり, DNN が多様体構造を抽出していることを支持する.

また, 右特異ベクトル (図 5 左下) は, 回転・平行移動に対応する多様体の接線方向 (図 6) に類似した画像が得られた. さらに, 左特異ベクトルの分析から, 出力層が大域的座標系になっていることも確認された. これらの結果は仮説 2 を支持する.

また, クラス毎の特異値分布に着目して分析したところ, クラス毎に違う特異値分布を持ち, 特にクラス「1」の持つ多様体の次元が他より低いとわかった. このことは, 「データセットがクラス毎に別々の部分多様体に埋め込まれること」を示唆する.

一方, Back Propagation (以下, BP) による fine-tuning 後の特異値・特異ベクトルをみると, pre-training 後と比較してそれらが変化することが観察された (図 5 右). これは, ラベルに基づく BP が教師なし学習とは違う作用を持つことを示唆する.

本研究ではさらに, 多様体構造を持つデータを人工的に作成し検証を行い, 同様に仮説を支持する結果を得た.

5 Deep Neural Networks による大規模画像データセットの観測

本研究では、Krizhevsky らによって開発された、畳み込みや pooling, drop out 等の技術を組み込んだ DNN [Krizhevsky 12] を用いて、Imagenet 大規模画像データセット [Deng 09] の観測を試みた。ImageNet データセットは、単語の定義や同義語のグループとグループ間の関係性が記述された英語の概念辞書（意味辞書）である WordNet のオントロジーに従って、各名詞に対応する画像を収集したもので、現在約 1,500 万枚の画像が登録されている。特異値・特異ベクトルは、SVD (singular value decomposition) における上位 500 番目までを計算した。具体的には、公開されている重みデータ (DeCAF [Donahue 14]) を用い、ヤコビアン行列とその特異値・特異ベクトルを算出した。

計算の結果、高階層において MNIST と同様、少数の大きな特異値と大多数のほぼ 0 の特異値という急峻な特異値分布が観察された (図 7 参照)。また、特異ベクトルをみた結果、特異値の大きいベクトルほど、空間的に局所的な構造を持ち、一方で特異値の小さなベクトルほど、空間的に広く分布した構造をもっていることもわかった (図 8 参照)。これは、MNIST の結果と類似している。

本研究ではさらに、DNN に多様体の接線方向と垂直方向にそれぞれ摂動を加えた画像を入力した際の出力の頑強性をみることで、データセットに多様体構造があることをより直接的に確認している。

また、多様体の次元や曲率などの情報しかもたない特異値分布の分布構造をクラス毎にみたところ、MNIST のときと同様にクラス毎に分布構造が違うことが確認された。さらにそのクラス間の相違は、データの階層的な概念構造に対応するかのようになり、低次の概念は低層から、高次の概念は高層から生じ始めることが確認された。このことは、概念の階層構造と DNN の各層で抽出されるデータの多様体構造の間に関係性があることを示唆する。

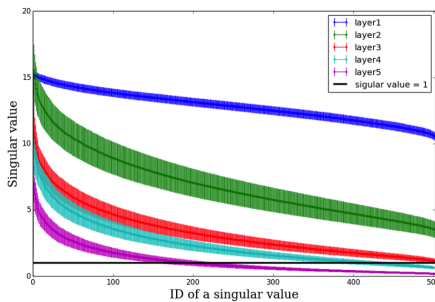


図 7: 特異値分布: それぞれの色に対応する線が、1 層から 5 層までの特異値分布 (複数の入力画像に対する平均) を表す。薄い色は標準偏差を表す。

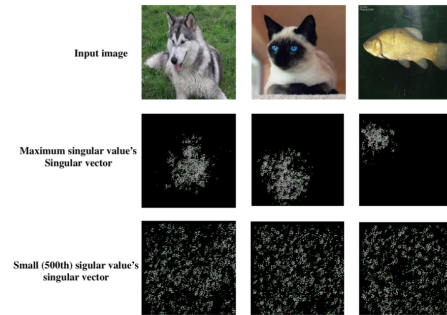


図 8: 右特異ベクトル: 1 段目が入力画像を、2 段目が特異値が最大となる特異ベクトルを、3 段目が計算した中で特異値が最小となる特異ベクトルを表す。

6 大自由度力学系のもつ多様体構造と深層学習

次に、ここまで得られた知見の応用を考える。本研究では、画像以外に多様体構造を持つ大規模データセットとして、鳥や魚等の群れの集団行動を再現するボイドモデルに着目する。特に分析対象としたのは、これまでの研究であまり対象とされなかった、複数の違う種類の群れが相互作用するような大規模な個体数の系であり、そこには「そもそも群れをどのように定義し抽出するか」という本質的な問いかけが生起する。本研究ではこの課題に対して、「群れ行動」を「まるで 1 匹の個体であるように、低い自由度に制限された集団運動」と捉え、そこから「群れは時系列データ中にある多様体構造として抽出される」と考え、DNN を用いて「群れ」抽出を試み、それに成功した。

7 議論と結論

本研究では、高次元データセットが低次元な多様体構造をもち、さらに DNN によってそれが観測可能であることを支持する結果を得た。さらに DNN の各層で抽出されるその多様体構造と、概念の階層構造が関係することも示唆された。この結果を説明できる 1 つの仮説として、データセットが階層的幾何構造をもち、DNN がそれを段階的に展開しているというものが考えられる。この仮説が正しい場合、DNN が高いパフォーマンスを持つことは、そのネットワーク構造自体がデータをモデル化しているためと説明される。

本研究は今後、機械学習分野のみならず、物理学や認知科学・脳科学分野等、多様な分野に有用な知見を提供していけるものと考えられ、位相幾何的な観測や多様な DNN モデルの使用等、さらなる研究を行なっていく必要がある。