Phase transition encoded in neural network

須賀勇貴

最終更新日：2023 年 11 月 12 日

これは，柏，菊池，富谷先生による論文”Phase transition encoded in neural network”を日本語でまとめたものである．

Abstract

相転移検出を目的としたニューラルネットワークの一側面について論じる．そのために，まず温度でラベル付けされたイジング模型と Potss 模型の配位データで温度を測定させるようにニューラルネットワークを学習させる．ここで，私たちは，機械に配位が秩序相にあるか無秩序相にあるかは明示的に与えていないことに注意．にもかかわらず，学習したニューラルネットワークはパラメータ (重みをバイアス) から臨界温度を特定することができる．私たちは，温度教師付きニューラルネットワークがどのような量を学習するかに注目することで，相転移の情報をどのように捉えるか理解しようと試みる．私たちの詳細な分析により，機械は訓練の程度に応じて異なる物理量を学習していることが明らかになった．この研究の主な観察は，学習済みニューラルネットワークの重みが温度に加えて，相転移の情報をどのように持つかということである．

1　はじめに

物質の相を探索することは，基礎となる微視的な物理システムの赤外線構造を明らかにするための最も重要なタスクである．これらの相は，理論の持つ対称性に基づいて分類される．いくつかの異なる相を含む理論では，相転移はそれらの境界で発生する．それらのうち，二次相転移の性質は次元の数によってのみ決定され，微視的な詳細とは独立した，大域的対称性の基礎理論，つまり，普遍性クラスによって分類される．しかしながら，実際のところ，微視的な理論データに基づいて位相を解析したり，相転移を検出したりすることは一般的に難しい課題である．なぜなら，それらの問題を正確に解決する，もしくは対応する赤外理論を特定することが難しいためである．したがって，これらの問題を解明するためには，数値的なアプローチが大量に取り組まれてきた．明白で主要な障害は，自由度の数が増加するにつれて，数値解析がますます困難になることである．

機械学習はコンピュータサイエンスの分野で急速に発展し，パターン認識，画像処理などで顕著な成功を収めてきた．最近では，機械学習が物理学のさまざまな分野でも適用されていることが目

撃されている．相転移の検出は，機械学習が新たな進展を遂げる可能性がある興味深い例の一つであり，すでにスピンシステムなどの単純なモデルでいくつかアプローチが提案され，試験されている．これらの研究はでは，教師あり学習では [3-18]，教師なし学習では [19-24] などで行われている．ここでは，入力データはニューラルネットワークやそのトレーニングプロセスとは独立して準備されてる．例えば，モンテカルロシミュレーションや興味を持つ物理系に基づく実験データが，必要な入力データを提供する．

物理系の相境界を検出する一つのアプローチは，教師ありバイナリ分類である，ここでは，ニューラルネットワークがトレーニングされ，それが秩序だったり無秩序だったりする相を区別できるようになる．実際に，このアプローチは生データからいくつかのモデルの相転移を合理的に検出する [3]．最近では，潜在変数と物理量との間の暗黙のつながりに関する研究が広く行われ，機械学習技術のブラックボックス性を取り除く究極の目標に向けて進展が見られている [3, 12, 15, 18]．また，物理的な洞察に基づく特徴エンジニアリングの支援を受けて，トポロジカル相転移や BKT 相転移などの非標準的な相転移も検出に成功した [10, 12, 14, 16, 25, 26].

別の興味深いアプローチが [4] で提案され，相転移を検出するために，秩序パラメータの情報がトレーニングの結果としてニューラルネットワークの重みにエンコードされていると仮定している．彼らは教師あり機械学習を用いて，2 次元の Ising モデルの臨界温度を特定しようとした．完全に連結したネットワークと畳み込みニューラルネットワークは，入力のスピン構造の目標温度を正確に予測できるようにトレーニングされた．驚くべきことに，トレーニング中に相転移に関する直接の情報を与えていないにもかかわらず，彼らは重みから相転移温度を抽出することに成功した．これは，ネットワークが温度の教師付き学習の過程で相転移を自発的に捉え，機械パラメータにエンコードしていることを意味しています．ただし，その結果の基本的なメカニズムは未解決のままである．

この記事の目的は，この相転移検出のメカニズムを理解することである．このアプローチが捉えるものを理解することは，それを他の未知のシステムに適用しようとする際に役立つ．異なる物理的メカニズムによって引き起こされる場合，例えば準長距離秩序（BKT 転移）や位相など、正確に相転移を検出できない可能性があるからである，実際，この方法は，ニューラルネットワークのトレーニング方法によって異なる物理量，つまり入力構成の特徴を捉えることが分かる．アイデアを説明するために，温度予測の本質を具現化した簡略化されたニューラルネットワークの構造に注目する．

2　臨界温度予測

2 次元イジング模型を考える．ハミルトニアンは

$$H(\boldsymbol{\sigma}) = -J \sum_{\langle i,j \rangle} \sigma_i \sigma_j \tag{1}$$

ここで，J は結合定数，$\sigma_i \in \{-1, 1\}$ は，サイズが $L \times L$ の正方格子のサイトに存在するスピンの自由度 (スピン変数) で，周期的境界条件が課されている．和は最近接に対するサイトにわたる．

結合定数をボルツマン重みに吸収させて，e^{-KH}/Z，$K = \beta J$ とすることで，ハミルトニアンを次のように再定義する．

$$H(\boldsymbol{\sigma}) = -\sum_{\langle i,j \rangle} \sigma_i \sigma_j \tag{2}$$

ここで，$Z = \sum_{\boldsymbol{\sigma}} e^{-KH(\boldsymbol{\sigma})}$ は分配関数である．

我々は，その 2 次相転移に関連する臨界温度を検出しようと試みている．しかしながら，最初は，教師あり順伝播ニューラルネットワークを用いて温度測定器を構築することを行う．その後，訓練されたニューラルネットワークの重みとバイアスを調べ，2 次元イジング模型と 3 状態のポッツ模型の臨界温度を特定しようと試みる．

メトロポリス・ヘイスティングス法を用いて，各温度で 2000 個のイジングスピン配位を生成する．これらは，20 の目標温度と一緒にニューラルネットワークに供給される．我々は，TENSORFLOW をバックエンドに使用した KERAS パッケージを用いて，全結合ニューラルネットワークと畳み込みニューラルネットワークの 2 種類のアーキテクチャを実装した．

全結合ニューラルネットワーク：

詳細を書く．入力の自由度は $\{\sigma_i\}$ $(i = 1, 2, \ldots, L \times L)$ で，イジングモデルの場合スピンに対応する．隠れ層のユニット x_a $(a = 1, 2, \ldots, N_b)$ は

$$x_a = \mathrm{softmax}(w^{(1)}_{ai}\sigma_i + b^{(1)}_a) \equiv \frac{e^{w^{(1)}_{ai}\sigma_i + b^{(1)}_a}}{\sum_a e^{w^{(1)}_{ai}\sigma_i + b^{(1)}_a}} \tag{3}$$

ここで，アインシュタインの縮約記法を使用していることに注意．2 つ目の層は，重み $w^{(2)}_{\alpha a}$ とバイアス $b^{(2)}_\alpha$ に置き換えたものである．最終層の変数 y_K $(K = 1, 2, \ldots, N_o)$ は隠れ層と同じ形をとる

$$y_k = \mathrm{softmax}(w^{(2)}_{ai}\sigma_i + b^{(2)}_a) \tag{4}$$

出力 $\{y_K\}$ に基づいて，入力配位の温度は

$$K^{\mathrm{output}} = \underset{K}{\mathrm{argmin}}(y_K) \tag{5}$$

で決定される．つまり，温度 α は確率分布 y_K の中から最も値が高い成分採用されるということである．訓練は，Adam という最適化手法を用いて，交差エントロピー

$$E(y_K, \mathbf{1}_{K=K_i^{\mathrm{target}}}) = -\frac{1}{N_o}\sum_i \mathbf{1}_{K=K_i^{\mathrm{target}}} \ln y_k \tag{6}$$

という誤差関数を最小することによって実装した．ここで，i は入力配位のラベルを示す．インジケーター関数は次のように定義される．

$$\mathbf{1}_{K=K_i^{\mathrm{target}}} = \begin{cases} 1 & (K = K_i^{\mathrm{target}}) \\ 0 & (\mathrm{otherwize}) \end{cases} \tag{7}$$

畳み込みニューラルネットワークは 3 つの畳み込み層と最後の全結合層からなる式 (**??**).

手順を示す．

1. マルコフ連鎖モンテカルロ法より，配位を生成：
2. ニューラルネットワークを温度計としてトレーニングする．入力はスピン配位，出力は予測温度.
3. ニューラルネットワークに現れる学習済みの重みとバイアスを分析する．マシンパラメーターに相転移の情報がどのように含まれるかについて説明する.

2.1　2 次元イジングモデル

最初に，[4] で既に研究されている 2D Ising モデルを見ていく．このモデルは，100 の目標温度で調査され，その重みは秩序変数のように振る舞い，すなわち自発的な磁化を示す．私たちの主な目的は臨界温度を定量的に推定することではなく，そのメカニズムを理解することなので，目標温度の数を 20 に減少させる．具体的には，$K = 0.05, 0.1, 0.15, \ldots, 1.0$ のような値である．ニューラルネットワークより教師あり機械学習を行った.

図 1 2D Ising モデルの場合，全結合層の重みおよびそれらのトレーニング後の平均について説明する．横軸は入力構成の温度 K を表し，左のパネルの縦軸はニューラルネットワークの隠れユニットに接続された成分に対応している．右のパネルの縦軸は，各 K に対する重みの平均を示している．重みの平均値は，正確な臨界温度 $K_e^{\text{exact}} \approx 0.4407$ の周りで著しく変化している.

格子サイズは $L = 16$ であり, 隠れユニットの数は $N_h = 80$, そして $N_o = 20$ はそれぞれ 20 の目標温度に対応している．臨界温度は正確に知られており，$K_c^{\text{exact}} = \frac{1}{2}\ln\sqrt{2} + 1 \approx 0.4407$ である．訓練後の 2 層目の重みが図 1 である ([4] より)．臨界温度は重みの和を $c_1 \tanh\left[c_2(K - K_c)\right]$ でフィッティングしてパラメータ c_1, c_2, c_3 を導くことで予想する．実験では格子サイズ $L = 8, 16, 32$ で行った．実際，最終的な重みの平均は秩序変数のように振る舞うようです（図 1 の右パネル）．次のセクションで重みの詳細な構造について議論します.

2.2　2 次元 3 状態ポッツ模型

2 次元イジング模型の臨界温度の学習メカニズムの詳細に入る前に，別の例として 2 次元 3 状態ポッツモデルを見ていく．ハミルトニアンは以下のように表される.

$$H(\{\Phi_i\}) = -\sum_{\langle i,j \rangle} \delta(\Phi_i, \Phi_j) \tag{8}$$

ここで，Φ_i は三つの値を取り, これは Ising スピン σ_i の一般化である．したがって，温度 K でラベル付けされた構成 $\{\Phi_i\}$ がニューラルネットワークの入力である．2 次元 3 状態ポッツモデルは，単純なランダウ理論とは異なり，ゆらぎの影響で $K_c \approx 1.0050$ で 2 次の位相転移を示すことが知られている [29-31].

3 DISCUSSION

これまでに，第二層の重みの値の変化が位相転移を示していることを観察してきた．これは，臨界温度の情報がなんらかの形で訓練されたニューラルネットワークにコード化されていることを示唆している．以下では，訓練された全結合型／畳み込み型のニューラルネットワークを注意深く検証し，それらが入力構成の特徴として抽出する（物理的な）量が何であり，それが温度の予測や臨界温度の検出とどのように関連しているかを理解しようとする [15, 17].

3.1 ニューラルネットワークでエンコードされた磁化

位相転移の秩序変数が 2D Ising モデルにおいて自発的な磁化であることから，トレーニング後にはその情報がニューラルネットワークにエンコードされていることは自然なことと言える．定量的な論拠を示すために，2D Ising モデルの場合においてトレーニングされたニューラルネットワークの重みとバイアスを調査して，簡略化されたモデルを構築する．最初に、簡略化のために（2）の中の隠れユニットの数を 80 から 3 に減少させる．図 3 に示されているように，それでも臨界温度を捉えていることに気づく．

第二層をモデリングする前に，まず最初の層の特性を調査する．図 4 は，式（3）の第一層の出力 $\tilde{x}_a \equiv w^{(1)}_{ai}\sigma_i + b^{(1)}_a$ と，入力の Ising スピン構成の磁化密度との相関を示している．この観察から，我々はこれらの出力から，図 4b に示すように，磁化 $m(\{\sigma\})$ に対して線形な三つのラインでモデリングする．

$$\tilde{x} = \begin{pmatrix} \tilde{x}_0 \\ \tilde{x}_1 \\ \tilde{x}_2 \end{pmatrix} = \begin{pmatrix} -m \\ \epsilon \\ m \end{pmatrix} \tag{9}$$

ここで，$\epsilon > 0$ は定数である．さらに，活性化関数として，我々の目的のために softmax の代わりに最大値関数を使用している．この変更は最終結果に影響を与えない．$x_a = \max(\tilde{x}_a)$ は次のベクトルを生成する．

$$m < -\epsilon : x = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix},\ -\epsilon \le m < \epsilon : x = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix},\ \epsilon \le m : x = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \tag{10}$$

パラメータ ϵ は、強磁性相と常磁性相を分離する閾値磁化と解釈できる [3].

三つの隠れユニットの磁化依存性を理解したら，次に第二層を分析する．この層の重みは図 3 に示されている．温度を低温，臨界温度，高温の三つの部分に分割する．それぞれ以下のベクトルで表される．

$$\text{Low K} : \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix},\ \text{Critical K} : \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix},\ \text{High K} : \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \tag{11}$$

No-dimensional output space において．この手法により，出力次元 N_o を実質的に 3 に削減す

る．図 3 に従って，我々は以下のように重みをパラメータ化する．

$$w^{(2)} = \begin{pmatrix} -\Delta & 0 & -\Delta \\ -\delta & -\delta & -\delta \\ 0 & -\Delta & 0 \end{pmatrix} \tag{12}$$

ここで，$\Delta > \delta > 0$. 行列 $w^{(2)}$ の Ka 成分は $w^{(2)}_{Ka}$ に対応する．バイアスは重みよりもはるかに小さいため，無視する．厳密なパラメータ化は，以下の議論には必要ない．このとき，$y_K = \max(\tilde{y}_K) = \max(w^{(2)}_{Ka} x_a) + b^{(2)}_K$ は以下の出力を生む.

$$m < -\epsilon : y_K = \max(w^{(2)}_{K0}) \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \tag{13}$$

$$-\epsilon \leq m < \epsilon : y_K = \max(w^{(2)}_{K1}) \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \tag{14}$$

$$\epsilon \leq m : y_K = \max(w^{(2)}_{K2}) \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \tag{15}$$