制限ボルツマンマシンの中身を見てみよう

物理情報工学科　渡辺研究室

# 何をしているか？

手書き数字を覚えた制限ボルツマンマシンに、数字を一瞬だけ見せて、「今見た数字を思い出して書きなさい」と伝えて、出力されたイメージを可視化しています。

QR コード

自動的に生成された説明

# 何がわかるか？

図形 が含まれている画像

自動的に生成された説明制限ボルツマンマシンは、入力されたデータを内部に情報圧縮して蓄えます。その脳内情報から、入力画像を復元させています。学習していない、見たことがない画像を見せられてもうまく復元できません。

QR コード

低い精度で自動的に生成された説明

また、脳内に情報を直接入力することで、画像を出力させることができます。これは、脳に直接電極を刺して、うまく電気信号を与えると、実際には見ていないものを「見た」と誤認させているようなものです。

# 詳しく知りたい方へ

**制限ボルツマンマシンとは？**

2024年のノーベル物理学賞は、プリンストン大学(アメリカ)のジョン・ホップフィールド氏とトロント大学(カナダ)のジェフリー・ヒントン氏が受賞しました。ホップフィールド氏は、「ホップフィールド・ネットワーク」と呼ばれる連想記憶を実現する仕組みを考案し、ヒントン氏はボルツマンマシン、オートエンコーダなど、画期的なモデルの考案者です。ボルツマンマシンは、「0」か「1」のいずれかの値をとるノードが、相互にエッジで結びつけられたネットワーク構造です。各ノード上に定義された「バイアス」と、エッジに定義された「重み」を適切に調整することで、いろんなパターンを学習することができます。バイアスは、そのノードがどれだけ1の値を取りやすいか、エッジは、両端のノードがどれだけ同じ値になりやすいかを決めるパラメータです。このように、ノードが二値の値を取るモデルは「イジングモデル」と呼ばれ、統計力学で広く研究されて来ました。統計力学ではエッジやバイアスの値が与えられた時に、ノードがどのような状態になるかを調べることがほとんどでしたが、ボルツマンマシンは逆に、ノードが望ましい状態になるためには、エッジやバイアスがどのような値になるべきかを調べ、学習によりその値を獲得します。これにより、覚えさせたいパターンをボルツマンマシンに覚えさせることができます。ボルツマンマシンはとても面白い仕組みを持っていますが、学習が極めて難しいという難点がありました。そこで、ノードを「可視層」と「隠れ層」の二層に分けて、同じ層に属すノード間にはエッジを定義しないという制限を課した、制限ボルツマンマシン(Restricted Boltzmann Machine, RBM)が考案されました。制限ボルツマンマシンはボルツマンマシンよりも効率的に学習することができます。このデモでは、そのボルツマンマシンの一種、制限ボルツマンマシンの動作の様子を展示しています。

**制限ボルツマンマシンが覚えるもの**

制限ボルツマンマシンが覚えるのは与えられたパターンです。しかし、どのパターンがどの数字であるかは教えていません。例えば「0」の画像は見せますが、それが「０という数字である」ということは教えません。このようにデータだけが与えられ、正解がない学習は「教師なし学習」と呼ばれます。制限ボルツマンマシンは、勝手に覚えた数字をランダムに出力することもできますが、一般には可視層と隠れ層をうまく使います。例えば可視層に何かパターンを入力すると、そのパターンに対応して隠れ層のノードにパターンが浮かび上がります。そして、逆に隠れ層のパターンから可視層のパターンを再現することができます。これが「一瞬だけ見せた数字をもう一度書かせる」ことに対応しています。みなさんが、一瞬だけ一桁の数字を見せられて、「同じものを書け」と言われたら、同じ数字を書くことができるでしょう。しかし、覚えていない文字、例えばアラビア語を一瞬だけ見せられても、それを再現することはできません。アラビア語の知識がないから抽象化できないためです。制限ボルツマンも同様に、あらかじめ覚えたパターンだけに反応することができます。逆に、知らないパターンが来たら「知らない」と感じるため、この性質を利用して、ノイズ修正や異常検知などに応用することができます。また、構造が単純であることから、「AIは何を考えているのか」を調べるためのツールとして研究が進められています。

![QR コード

自動的に生成された説明]()

https://calc.appi.keio.ac.jp

渡辺研究室

![QR コード

自動的に生成された説明]()

https://www.appi.keio.ac.jp

物理情報工学科