制限ボルツマンマシンの中身を見てみよう

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。ここでは、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見てみましょう。

# このデモは何をしているか？

正方形 が含まれている画像

自動的に生成された説明

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。書いた数字は見せられた数字と全く同じ筆跡ではありませんが、誰がみても同じ数字だと認識することができます。これは、みなさんが見せられた数字を筆跡も含めて完璧に覚えたわけではなく「数字の1だ」というように抽象化して記憶し、自分の記憶にある数字の1を書いているからです。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。

さて、このデモではMNISTと呼ばれる0から9までの手書き数字を制限ボルツマンマシンにあらかじめ覚えさせています。この状態で、マシンに何かパターンを見せて、それを自分の記憶を頼りに再構成させた結果を表示させています。覚えたもの(数字の0から9)に近ければ再構成できますが、そうでなければ再構成することができません。

# ボルツマンマシンとは？

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。

# 制限ボルツマンマシンとは？

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(visible layer)」と「隠れ層(hidden layer)」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Contrastive-Divergence 法と呼ばれる学習法により、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。

# 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入力されたパターンです。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。その意味で正解ラベルがないので「教師なし学習」の一種となっています。

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。この性質を利用して、ノイズ修正や異常検知に応用することができます。