

入出力関係の解釈が可能なニューラルネット集団の構築と応用

P i t o y o [独立行政法人 国立大学 / 准教授]

背景・目的

現在ニューラルネットワークの応用は工学に限らず、医学、経済学、社会学などの分野で盛んに行われている。数多くのニューラルネットワークの中で最も多く利用されているは Multilayered Perceptron (MLP) である。その理由は MLP の扱い易さと汎化能力にある。しかし、一旦学習が終わると、MLP は入力と出力を結ぶための一種の Black Box としか用いることができない。つまり MLP の中でどのようなルールで入力を出力に変換するかを人間の分かる形であらわすことができない。このことは、安全性などにかかわる問題へのニューラルネットワークの適用を難しくする理由の一つとなる。

内容・方法

本研究は高い学習能力と汎化能力をもつと同時に、学習の結果として得た知識（または関数）を人間が理解しやすい論理ルール（if-thenルール）として解釈できるニューラルネットワークの開発と初期考察を目的とする。

本研究ではいくつかの改良型線形パーセプトロンから構成されるニューラルネットワークのアンサンブル（集団）を開発する。各パーセプトロン（以降、メンバーと呼ぶ）の出力層には confidence ニューロンを追加する。このニューロンは通常の入力に対する信頼度（confidence value）を表す値を出力する。入力に対し、全てのメンバーが confidence value を計算した後に、アンサンブルは常に confidence value の最も高いメンバーの出力を最終的な出力として採用する。学習過程では、confidence value が高く教師信号に対して誤差の小さいメンバーを学習させ、confidence value も増すように訓練を行うと同時にそれ以外のメンバーは confidence value が減る方向に訓練を行う。これにより、学習時にメンバー間の競争が起き、その結果、学習空間の分離が効率的に行われ、各メンバーは学習空間に割り当てられることとなる。各メンバーが線形パーセプトロンであるため、実現される関数は線形関数である。線形関数を人間が解釈しやすい論理関数に変換することが容易であるため、提案したアンサンブルからのルール抽出は容易に行うことができる。

結果・成果

開発したアンサンブルの概念を図1に示し、学習アルゴリズムを図2に示す。

本研究で提案したアンサンブルとその競合学習アルゴリズムの妥当性を検証するため幾つかの実験を実施し、評価を行った。

予備実験として、まずアンサンブルを2入力1出力の論理

関数 XOR に適用した。本問題は代表的な非線形識別問題の一つであり、線形識別のみ実行できるパーセプトロンでは対応することができない。それに対し、2つのメンバーから構成されるアンサンブルを適用することにより、この非線形関数を2つの線形問題に分解し、各々の線形関数を一つのメンバーに割り当てることができることを確認できた。それに加え、3つ以上のメンバーを持つアンサンブルを用いる場合、学習の過程ではこの問題に必要な数だけのメンバーが有効なメンバーとして選ばれることが確認できた。ニューラルネットワークが学習で得た知識を人間に分りやすいルールに変換する場合には最も簡潔な形で表現することが重要であるため、有効なメンバー数を成るべく少なく抑えられる能力は非常に有意義である。アンサンブルの学習後、各メンバーで実現した線形関数を既存の方法で人間が理解しやすい if-then ルールに変換することができることも確認できた。

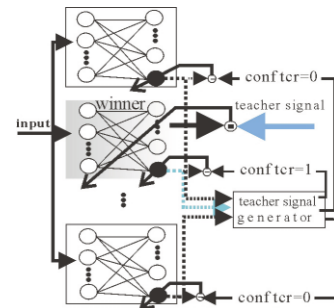


図1. アンサンブルの概念

Learning Process of ELE

```
1: select a training sample
2: run(all_members)
3: select a winner
4: if (performance(winner) < threshold)
5:   train(winner)
6:   increase_confidence(winner)
7:   decrease_confidence(losers)
8: else
9:   decrease_confidence(winner)
10:  increase_confidence(losers)
11:  train(losers)
12: endif
```

図2. 学習アルゴリズム

その他に、アンサンブルを7つの benchmark 問題に適用し、学習能力と汎化能力に関して MLP との性能比較も行った。その結果、本研究で対案したアンサンブルは MLP と様々な問題に関して同程度の能力を持ち、MLP と比較しより分りやすいルールを人間に提供することができることを明らかにした。

今後の展望

本研究では、論理ルールとして解釈可能な線形区分法を実現できるニューラルネットワークアンサンブルを提案し、その基本的な性能に関して評価を行ったことができた。

今後、さらに提案したアンサンブルのより正確な振る舞いを数学的に考察することが必要と考える。

それと同時に、実世界データ（特に医学データ）にアンサンブルを適用し、データから入出力関係を学習によって形成するだけでなく、抽出したルールから人間に新しい知見を提供できるようなシステムとしての利用も検討する。