

クラスタリング用ニューラルネットの学習方式

5R-3

鹿山 昌宏 阿部 重夫
(株) 日立製作所 日立研究所

1. まえがき

ニューラルネット¹⁾に関する研究が盛んであるが、ネットワークの汎化能力との関連から学習方式の改善を論じた報告は意外に少ない。本報ではクラスタリング問題を対象に、教師信号を繰り返し提示する従来学習法ではネットワークの汎化能力が十分に高められないことを示す。さらに与えられた教師信号の基で汎化能力を最大化する学習方法を検討する。

2. 学習方法

3層で検討した内容を4層以上のニューラルネットに拡張することは比較的容易と考えられるため、本報では図1に示す3層ニューラルネットに限定して以下の検討を進める。

まず従来学習法の問題点を指摘する。クラスタリング問題の場合、通常各出力ニューロンは各クラスタ(カテゴリ)と対応づけられ、学習時にネットワークに提示される教師出力信号は、出力ニューロン数を要素とするベクトルで表される。学習は教師入力に対応したネットワークの出力が、このベクトルに一致するようにシナプスの重みを修正することで行われる。ここで問題は教師出力ベクトルは各クラスタを識別するために便宜上定義された値に過ぎず、ネットワークの出力誤差を最小化する操作と、境界を汎化能力にとっての理想位置(たとえばクラスタの中央)に近づける操作が直接対応しないことである。このため3.2節のシミュレーション結果からも裏付けられるが、ネットワークの出力誤差が小さいとき、これをさらに微小にしても汎化能力は向上しない。

この点を改善するために本学習法では、教師入力信号に適切な振幅の乱数を重畳する^{2,3)}。この手法の場合、

$$(X_1, \dots, X_i, \dots, X_m)$$

の教師入力信号に対して、実際にネットワークに入力される値はその都度発生させた振幅Rの乱数を重畳し、

$$(X_1+r_1, \dots, X_i+r_i, \dots, X_m+r_m)$$

$$\text{但し } |r_i| \leq R$$

となる。例えば乱数を一様に発生する場合、ネットワークの入力信号は、教師入力信号を重心とし一辺の長さRの超立方体の内側に均等に存在する。このためクラスタ境界がこの空

間を横断する確率を低くできる。したがってクラスタ境界の各クラスタへの近接を妨げられ、汎化能力を向上できる。

課題として、最適な乱数の振幅値の決定方法がある。乱数の振幅Rとクラスタ間の距離2pにおいて、一般に $R \gg p$ であると、入力空間に両クラスタの入力信号が重複する領域が生じる。また $R < p$ の場合、境界は超立方体の外に高い自由度で存在できるため、いずれも理想位置から遠ざかる可能性が高い。これより $R \approx p$ に設定するのが適当と推察される。一方クラスタの数が3以上の場合、通常、クラスタ同士のpはさまざまな値となる。このような場合にはpの値を特定できず、したがってRの値も決定できず問題は困難となる。

この点を解決するために本論文では、学習の初期に振幅の大きな乱数を与え、学習回数とともこれを漸減し、最終的に0とすることを検討する。この学習方法の目的は、まず学習の初期に加えた大きな乱数により、距離を隔てた(pの値が大きい)クラスタ同士の境界をほぼ中央位置に確定し、その後加算する乱数の振幅を小さくし、最初に確定した境界を破壊しないように、近接したクラスタ間の境界を中央位置に位置付けることにある。学習が進み、乱数の振幅が小さくなるにつれ、出力誤差、したがってシナプスの重みの更新量は微小となるため、距離を隔てたクラスタ間の境界を大移動させることなく、近接したクラスタ間の境界を中央位置に位置付けることができると期待される。

以下シミュレーションにより本手法の有効性を確認する。

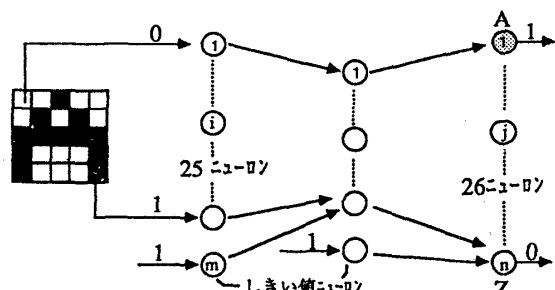


図1 3層ニューラルネットワーク

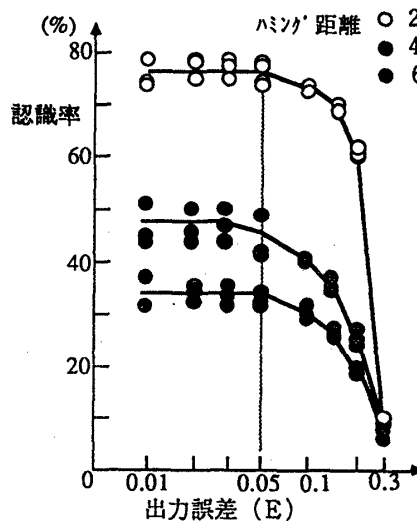


図2 認識率と出力誤差の関係 (中間層ニューロン数9個)

3. 実験結果および考察

3.1 実験システム

シミュレーションは図1のネットワークで行った。画像の各ピクセルに対応した25個の入力ニューロン、A~Zに対応した26個の出力ニューロン、適当数の中間層ニューロンからなる。入力層と中間層にはさらに1個ずつのしきい値ニューロンを設けた。汎化能力(認識率)は、教師入力パターンのビットを適当数反転して作った評価データを各文字あたり50個、計1300データ作り評価した。反転したビット数を以下ハミング距離という。

3.2 出力誤差と汎化能力

まずクラスタリング問題において、ネットワークの出力誤差が小さい領域で、出力誤差の減少が汎化能力向上と対応しないことを示す。以下の解析で出力誤差Eは各出力ニューロンの誤差を平均し、

$$E^2 = (1/2) \cdot (\sum (T_j - Y_j)^2) / n \quad (1)$$

Tk: k番目の出力ニューロンに対応した教師出力

Yk: k番目の出力ニューロンの出力値

n: 出力ニューロン数

で計算する。

図2に出力誤差と、出来上がったネットワークの認識率の関係を示す。各振幅に対して初期値を変えた3つのネットワークを作り認識率を評価した。図より認識率は出力誤差が0.03~0.05より小さくなると、飽和して一定値となる。したがってこのネットワークで出力誤差をさらに減少させる学習を行っても無意味であり、認識率をさらに向上させるためには学習方法自体を改善しなければならないことが分かる。

3.3 乱数重畳学習法

まず、学習中一定振幅の乱数を一貫して重畳した教師入力信号で学習を行った結果を示す。図3は乱数の振幅と認識率の関係である。図より、振幅の大小にかかわらず乱数を重畳すると認識率が向上している。乱数の重畳は汎化能力を向上させ、したがって境界を理想位置へ近づけるのに有効なことが分かる。適当な振幅値については、評価データのハミング距離が2のとき0.5~1.0であるのに対し、ハミング距離が4, 6の場合1.5が最良であった。これより評価データの教師入力信号からの空間距離が大きくなると、これに相当する大きな乱数を加えて学習した方が良いことが分かる¹⁾。

大きな乱数を加えた場合の誤認識文字を詳細に分析したと

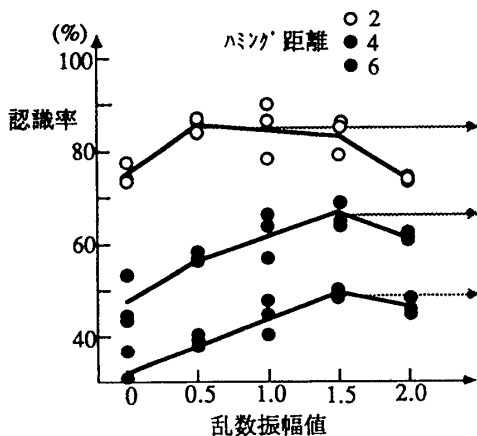


図3 乱数振幅値と認識率の関係

ころ、C, B, Oのような特定文字における認識率の低下が著しい。特定文字に共通しているのは、入力空間で比較的近接した地点に、他のクラスタが存在することである。Cからハミング距離を1隔てた地点にOが存在し、BはE, G, O, P, Sからハミング距離5以下しか隔たっていない。一方、加算した乱数の振幅が0.5のときには、このような特定文字は観察されない。

以上より教師入力信号に乱数を印加した効果には2面あり、境界からクラスタを遠ざけるためには大きな値とすることが望ましいが、クラスタ間の距離に対して過大な値とすると、逆にこれらを識別できないことにより認識率が低下する。したがって、各クラスタがさまざまな空間距離で混在している一般のアプリケーションで、教師入力から種々のハミング距離にある入力をまんべんなく認識する場合、一定振幅の乱数の重畳ではうまく行かない。

3.4 乱数漸減学習法

次に教師入力に加算する乱数の振幅を、学習の進行(学習回数)とともに漸減する手法を検討する。図4に漸減方法を変えて学習したネットワークの認識率を示す。Aは振幅を2.0から学習回数に比例して一様に減少させたもの、Bは学習回数の平方根に、またCは2乗に比例して振幅を漸減したものである。本解析に関する限り、特定の漸減方法が認識率に著しく好影響を及ぼす様子は観察されない。総合的にはA, Cの結果が優れており、評価データのハミング距離にかかわらず図3の最大認識率の水準を確保している。したがって乱数の振幅を適当に漸減することにより前節の問題は解消され、評価データのノイズ値に関係なく汎化能力が最大化される。

4. おわりに

ネットワークの汎化能力を最大化することを目的に、クラスタリング用多層ニューラルネットの学習方法を検討した。開発手法は乱数の重畳でクラスタの形状を仮想的に拡大して学習するもので、教師信号が少ない場合に、これを補い高性能なネットワークを得る手法としてとりわけ効果大きい。

5. 参考文献

- 1) D.E.Rumelhart et al, "Parallel Distributed Processing (PDP)", vol. 1, 2 MIT Press (1986)
- 2) J.Sietsma et al, "Creating artificial neural network that generalize", Neural Networks, vol. 4 No.1 (1991)
- 3) 宮武ほか, "雑音環境下でのニューラルネットを用いた単語音声認識", 日本音響学会講演論文集 (1991)

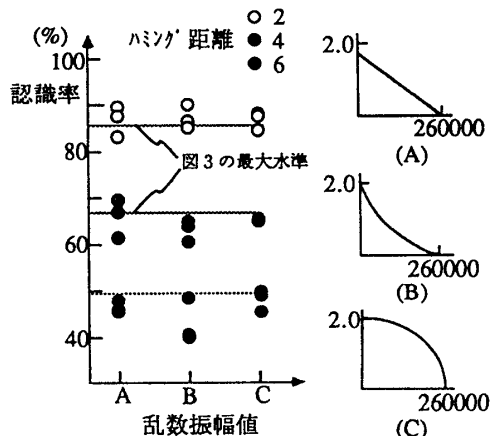


図4 乱数減少方法と認識率の関係