

自己組織化マップを用いた動径基底関数ネットワークの性能評価 Performance Evaluation of RBF Network Constructed on the Basis of SOM Training Result

山下 和彦[†] Goutam Chakraborty[‡] 馬淵 浩司[‡] 松原 雅文[‡]
Kazuhiko Yamashita Goutam Chakraborty Hiroshi Mabuchi Masafumi Matsuhara

1. はじめに

パターン認識手法の一つに人工ニューラルネットワーク(ANN)が広く使われており、そのネットワークパラダイムの一つに、動径基底関数ネットワーク(RBFネットワーク)がある。RBFネットワークは、非線形の問題を解決することができるが、RBF係数を計算する際に、擬似逆行列を利用しているため、分類数や隠れユニット数が増加した際に、計算負荷が高くなるという問題点が挙げられる。そこで、本稿では、自己組織化マップ(SOM)を用いたRBF係数の計算手法を提案し、計算コスト、認識率の性能評価を行う。また、SOMの学習結果を利用した効率の良いRBFネットワークの隠れユニットの決定手法も提案する。そして、既存のアルゴリズムによって学習されたRBFネットワークと比較し、提案手法の有効性を示す。

2. RBF ネットワーク

2.1 概要

RBFネットワークは、入力層、隠れ層、出力層から構成される3層の階層型のANNであり、パターン認識に広く利用されている。隠れ層には、RBFというGaussian型の基底関数が設けられており、その隠れ層からの出力を出力層へ線形結合することでパターンの分類を行う。基底関数は、式(1)により表わされる。

$$\phi_j = \exp\left(-\frac{\|x - \mu_j\|^2}{2\sigma_j^2}\right) \quad (1)$$

2.2 Two-Stage 学習

Two-Stage学習[1]は、隠れユニット数の決定と、出力層の結合荷重の計算から構成される。隠れユニットの選択には、ランダム選択やk-means法等が用いられる。また、出力層の結合荷重は最小二乗法により一括学習を行う。

結合荷重を行列 W^T とした時、出力層の結合荷重は、式(2)より求めることができる。ここで、行列 T は、教師信号を表している。

$$W^T = \Phi^+ T \quad (2)$$

式(2)の Φ^+ は、行列 Φ の擬似逆行列である。ここで、式(2)は、式(3)に展開することができる。

$$W^T = (\Phi^T \Phi)^{-1} \Phi^T T \quad (3)$$

また、入力層を x 、中間層を y 、出力層を z とした時、出力層の出力は、式(4)で求めることができる。

$$z_k = \sum_{j=1}^N W_{kj}^T y_j \quad (4)$$

[†]岩手県立大学大学院ソフトウェア情報学研究科
[‡]岩手県立大学ソフトウェア情報学部

このように、結合荷重 W^T を基に、隠れ層からの出力を出力層へ線形結合することでパターンの分類を行う。

しかし、一般的にRBFネットワークでは、より多くの隠れユニットが必要となり、RBFパラメータを求める際、擬似逆行列を利用しているため、分類数や隠れユニット数が増加すると計算負荷が高くなる。

そこで、本稿では、計算負荷が高くなるという既存の問題を解決する為に、SOMを用いたRBF係数の計算手法を提案する。

3. 提案手法

3.1 概要

本稿では、SOMの学習結果を利用した、新しいTwo-Stage学習を提案する。提案手法の流れを、図1に示す。提案手法は、First StageとSecond Stageの二つの処理から構成される[2]。First Stageでは、SOMの学習を行い、隠れユニットを決定する。また、Second Stageでは、結合荷重の決定と関数の広がり方を計算する。

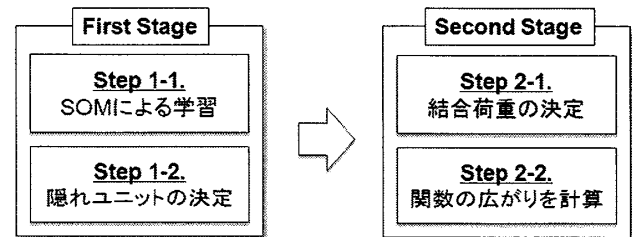


図1: 提案手法の流れ

3.2 First Stage

First Stageでは、SOMによる学習を行い、その学習結果を利用し、隠れユニットの選択と出力層の結合荷重を決定する。RBFネットワークの隠れユニットには、勝利回数が1以上であるSOMの出力ユニットを利用する。したがって、RBFネットワークの隠れユニット数は、SOMの勝利ユニット数となり、SOMの学習結果で可変となる。ゆえに、本手法を利用すれば、隠れユニットを自動的に決定することが可能である。

3.3 Second Stage

RBFネットワークの入力層から隠れ層の結合荷重には、SOMの入力層から出力層の結合荷重を利用する。また、隠れ層から出力層の結合荷重には、出力ユニットの勝利回数を利用することで、擬似逆行列の計算を回避することが可能である。

関数の広がり方は、式(5)より求める。ここで、 σ_i は関数の広がり、 d_j は中心ユニットと近傍ユニットのユークリッド距離を、 N は、近傍ユニット数を表している。また、関数の広がり設定方法を、図2に示す。

$$\sigma_i = \frac{\sum_{j=1}^N d_j}{N} \quad (5)$$

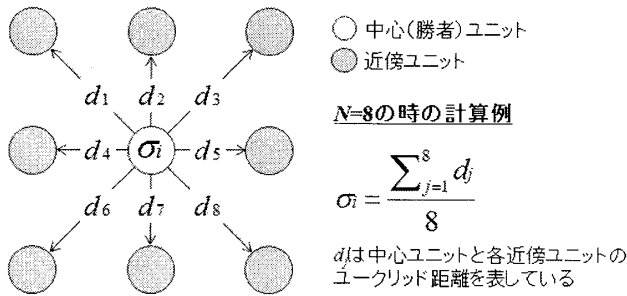


図 2: 関数の広がりの設定方法

4. 評価実験

4.1 実験概要

本実験では、Wisconsin の肺がんデータを利用し、隠れユニット数が増加した際の、認識率と処理時間の比較を行った。このデータセットは、699 人分の患者データがあるが、欠損情報を含む 13 人を除外し、683 人分の患者データを実験に利用した。また、分類数は 2 であり、患者がガンであるか無いかの判別を行う。属性情報の詳細を、以下に示す。

- 凝集塊の厚み
- 細胞の大きさの均一性
- 細胞の形の均一性
- 細胞編部の癒着の度合い
- 上皮細胞の大きさ
- むきだしの細胞核の数
- 障害のない染色質の数
- 通常の細胞の数
- 細胞の間接核分裂の度合い

実験パラメータを、表 1 と表 2 に示す。なお、隠れユニット数の増加が、認識率と処理時間に与える影響を調査する為、隠れユニット数を、36, 49, 64, 81, 100 と増加させ、性能評価を行った。

表 1: SOM のパラメータ

マップサイズ	試行回数	学習率	近傍半径
単調に増加	1,000	0.1	3

表 2: 既存手法のパラメータ

隠れユニット数	関数の広がり
単調に増加	2.0

4.2 実験結果および考察

認識率の推移を、図 3 に示す。既存手法は、隠れユニット数の増加に伴い、汎化能力が落ちていることが分かる。一方、提案手法は、隠れユニット数を増加させても、汎化能力が落ちていないことが確認できた。隠れユニットの選択に、SOM の学習結果を利用することにより、適切な隠れユニットが選択されている可能性が高い。また、提案手法では、SOM のマップサイズを大きくして実験を行うことにより、汎化能力の高いネットワークを自動的に構築することが可能である。

処理時間の推移を、図 4 に示す。図 4 より、提案手法の方が既存手法より高速な学習が可能であることが分かる。既存手法では、隠れユニット数が増加すると、擬似逆行列の計算負荷が高くなり、処理時間が増加した。

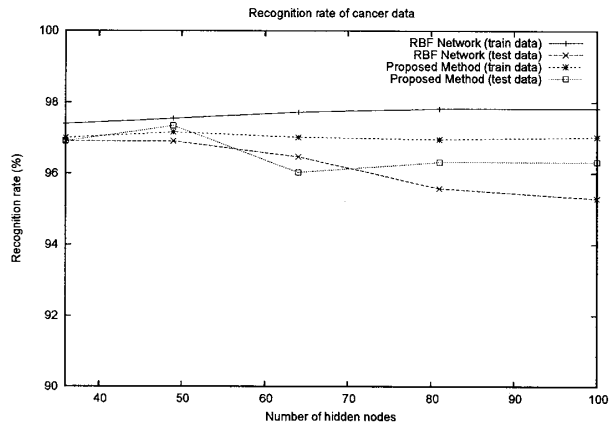


図 3: 認識率の推移

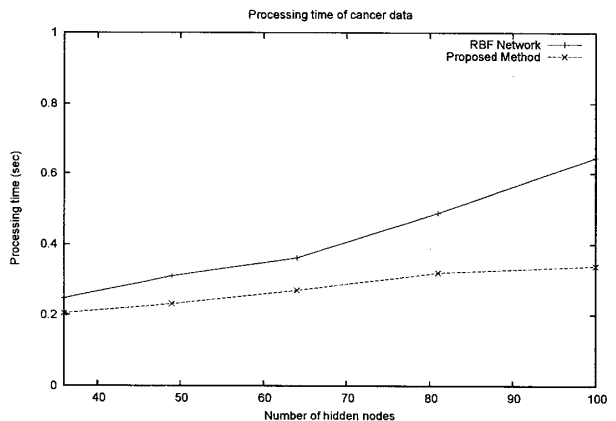


図 4: 処理時間の推移

5. おわりに

本稿では、RBF 係数の計算負荷を低減する為に、SOM を用いた計算手法を提案し、性能評価を行った。評価実験の結果から、既存手法より高速に RBF 係数を計算することが可能であることが分かった。また、隠れユニット数を自動的に決定することが可能であるため、既存手法よりも汎化能力が高いネットワークを構築することができた。

今後は、他の種類のベンチマークデータを用いて実験を行い、本手法の有効性を示す予定である。

参考文献

- [1] Robert J. Howlett, Lakhmi C. Jain: Radial Basis Function Networks 2: New Advances in Design, Physica-Verlag Heidelberg, 2007.
- [2] 山下和彦, Goutam Chakraborty, 馬淵浩司, 松原雅文: 自己組織化マップを用いた動径基底関数ネットワークの構築, 情報処理学会第 71 回全国大会講演論文集 (2), 6Q-6, pp.157-158, March 2009.