

確率的ニューラルネットワークにおけるハイパーパラメータ設定に関する研究 RBF カーネルにおける半径の自動決定による分類精度向上の試み

Study on the hyperparameter settings of probabilistic neural network: an attempt to automatically determine the radius of the RBF kernels

○高橋克¹, 森田峻平², 保谷哲也³

Katsumi Takahashi, Shunpei Morita, Tetsuya Hoya,

Abstract: In the field of machine learning, a central problem resides in the bloated network scale and high degrees of freedom due to the large number of hyper-parameters. We then focus upon a probabilistic neural network, which has an overwhelmingly low network scale, as well as a low degree of freedom. In this work, we empirically show that a PNN with a simple radius setting of the radial basis functions can still yield a satisfactory classification performance, using a total of 13 publicly available datasets.

1. 研究背景

近年の GPU の高性能化や人工知能ブームも相成り、急速に発展、運用されている機械学習、特に画像認識分野において、計算機が高性能になった故、ネットワークの肥大化やそれに伴うハイパーパラメータの数の多さに起因する高すぎる自由度から、ネットワークモデルの複雑化がより一層極まっていることが指摘されている[1].

本研究は、ネットワークがより単純かつ自由度の少ない確率的ニューラルネットワーク (Probabilistic Neural Network, 以下 PNN) [2]を用い、その唯一のハイパーパラメータである放射基底関数 (Radial Basis Function, 以下 RBF) [3]における半径 σ を、識別させるデータによって自動的に決定する手法を提案し、この問題の解決に向けて一石を投じることを主眼にする。また、本稿では提案手法の妥当性を 13 のデータセットを用いて検証したことについて報告する。

2. PNN [2]

PNN は、RBF-NN[3]の一つであり、多層ネットワークの 1 つに分類される。基本的な PNN の構造は、図 1 に示すように、パターン層、結合層、および出力層の 4 層により構成されている。

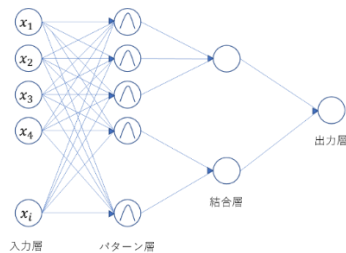


図 1 PNN の構造

パターン層は、放射基底関数 (RBF) のセットにより構成されており、結合層はパターン層によって計算さ

れたもののクラス毎の線形結合である。最後に出力層においては、結合層に蓄積されたユニットの各ターゲットクラスの加重投票を比較し、最大の投票を出力として出す。

このように、PNN のアーキテクチャは単純であるが、強力なマッピング機能を維持しており、主流の BPNN と比べ、分類問題における興味深い代替手法とも考えられる。

2-1. RBF [3]

PNN における放射基底関数 (RBF) はガウスカーネルが用いられ、その式は (1) により与えられる。図 2 に 2 次元入力空間の RBF を示す。

$$f(x, x') = \exp\left(\frac{-\|x - x'\|^2}{\sigma^2}\right) \quad (1)$$

PNN において、パターン層のユニット数は学習データ数に相当し、RBF の中心 x に各学習データをセットすることが学習である。 x' はテストデータにより与えられ、図 2 で示されるように、 $x = x'$ の場合、 $f(x, x')$ は最大の "1" をとる。また、 x と x' の差が大きくなるにつれ、その出力値は減少し、ハイパーパラメータ σ により、その度合いは大きく変わる。 σ の値が小さいほど入力の微小な変化に対する反応が大きくなるため、より高精度の分類が可能になる。ただし、 σ の値が極端に小さい場合、有意領域が狭くなり、適切な分類が行えなくなる。従って、入力データの分布に対する適切な σ の値の決定が必要になる。

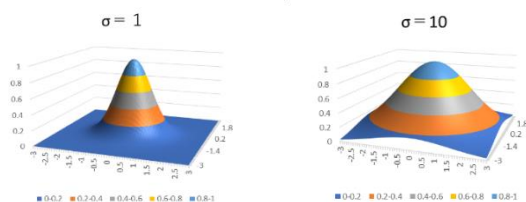


図 2 2次元入力空間における RBF

2-2. 提案手法

上記 RBF のハイパーパラメータ σ において, Haykin は[4]において, 以下(2)式のように決定する手法を提案している. ここで, d_{max} は学習データにおける全ペア間のユークリッド距離最大値であり, M は全学習データ数である.

$$\sigma = \frac{d_{max}}{\sqrt{2M}} \quad (2)$$

次に提案する σ の決定方法を(3)式に示す. ここで, d_{max} は学習データにおける全ペア間のユークリッド距離最大値であり, N は分類するデータのターゲットクラス数である.

$$\sigma = \frac{d_{max}}{N} \quad (3)$$

3. 実験内容と結果

実験に用いたデータセットのクラス数, 学習データ数, テストデータ数を表1に示す. また, 提案手法, (2)式の方法でそれぞれ PNN を用いた分類精度, 更に, 比較のため, k近傍法 (KNN) [5] の分類精度を含めたものを表2に示す.

表1 実験に用いたデータセット

Dataset	クラス数	学習データ数	テストデータ数
Tic-Tac-Toe[6]	2	663	295
wdbc[7]	2	398	171
Inosphere[8]	2	200	151
Balance-Scale[9]	3	438	187
Contraceptive Method Choice[10]	3	1031	442
StatLog (Landsat Satellite)[11]	7	4435	2000
MNIST[12]	10	10000	2000
OptDigit[13]	10	3823	1797
Speech Filing System[14]	10	540	360
PenDigits[15]	10	7494	3498
Yeast[16]	10	890	594
ISOLET[17]	26	6238	1559
LetterRecogniton[18]	26	16000	4000

表2 提案手法, (2)式, kNN による正答率

Dataset	提案手法 (%)	(2)式 (%)	KNN (%)
Tic-Tac-Toe[6]	58.64	83.05	83.73 (k=5)
wdbc[7]	96.49	94.15	98.25 (k=10)
Inosphere[8]	85.43	94.70	94.04 (k=2)
Balance-Scale[9]	73.80	65.24	87.70 (k=8)
Contraceptive Method Choice[10]	45.02	44.80	49.10 (k=8)
StatLog (Landsat Satellite)[11]	80.85	89.35	90.35 (k=3)
MNIST[12]	92.20	20.85	92.10 (k=3)
OptDigit[13]	98.33	94.82	98.00 (k=1)
Speech Filing System[14]	96.94	96.94	97.50 (k=5)
PenDigits[15]	94.25	97.97	97.80 (k=3)
Yeast[16]	54.88	46.63	55.56 (k=7)
ISOLET[17]	88.84	11.55	92.37 (k=9)
LetterRecogniton[18]	96.20	95.80	95.68 (k=1)

4. 考察と課題

表2に示すように, 提案手法と(2)式の正解率を比較した場合, 13個中9個のデータセットにおいて, 提案手法の方が優位または同等の精度となったが, 比較的クラス数の少ないデータセットにおいては, 低い精度となった.

提案手法ではクラス数により d_{max} を除算するため, クラス数が少ない場合, 半径 σ が大きくなることとなり, 精度に影響していると考えられる.

よって今後の課題として, クラス数が少ない場合も考慮した, 半径 σ の決定方法が挙げられる.

5. 参考文献

- [1] Chuan Li : Open AI's GPT-3 Language Model, A Technical Overview, Lambda, <https://lambdalabs.com/blog/demystifying-gpt-3/> (2020/6/30)
- [2] Specht, D. F. : Probabilistic neural networks, Neural Networks, 3, 109-118,(1990)
- [3] Broomhead, D. S.; Lowe, David. : Radial basis functions, multi-variable functional interpolation and adaptive networks (Technical report). RSRE. 4148. (1988)
- [4] Haykin, S. : A Comprehensive Foundation, Neural Networks, Macmillan College Publishing (1994)
- [5] Belur V. Dasarathy : Nearest neighbor (NN) norms : NN pattern classification techniques, IEEE Computer Society Press, (1991)
- [6]Tic-Tac-Toe: <https://archive.ics.uci.edu/ml/datasets/Tic-Tac-Toe+Endgame>
- [7]wdbc: <https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+Diagnostic>
- [8]Inosphere : <https://archive.ics.uci.edu/ml/datasets/ionosphere>
- [9]Balance-Scale : <https://archive.ics.uci.edu/ml/datasets/balance+scale>
- [10]Contraceptive Method Choice: <https://archive.ics.uci.edu/ml/datasets/Contraceptive+Method+Choice>
- [11]StatLog (Landsat Satellite): [https://archive.ics.uci.edu/ml/datasets/Statlog+\(Landsat+Satellite\)](https://archive.ics.uci.edu/ml/datasets/Statlog+(Landsat+Satellite))
- [12]MNIST : <http://yann.lecun.com/exdb/mnist/>
- [13]OptDigits :<https://archive.ics.uci.edu/ml/datasets/optical+recognition+f+handwritten+digits>
- [14]Speech Filing System : <https://www.phon.ucl.ac.uk/resource/sfs/matlab.php>
- [15]PenDigits : <http://odds.cs.stonybrook.edu/pendigits-dataset/>
- [16]Yeast : <http://archive.ics.uci.edu/ml/datasets/yeast>
- [17]ISOLET : <https://archive.ics.uci.edu/ml/datasets/isolet>
- [18]LetterRecogniton: <https://archive.ics.uci.edu/ml/datasets/Letter+Recognition>