

エッジコンピューティングのための自己組織化ニューラル木立 Self-Organizing Neural Grove for Edge Computing

橋本 依樹[†] 井上 浩孝[‡]
Ibuki Hashimoto Hirotaka Inoue

1. 緒言

近年、畳み込みニューラルネットワーク(CNN)などの深層学習モデルが分類精度を向上させるために実用化されているが、層数に比例して学習時間が増加する問題がある。一方、自己生成ニューラル木(SGNT)に基づく複合分類システム(MCS)の学習時間は短い。我々は、ニューラルネットワークアンサンブルを効率的に分類する新たな剪定手法を提案し、このモデルを自己組織化ニューラル木立(SONG)と名付けた[1]。本論文では、エッジコンピュータの一つである Raspberry Pi 3 上で、剪定 MCS と非剪定 MCS、C4.5 ベースの MCS、k 近傍法を比較し、SONG の性能を比較検討する。

2. 自己組織化ニューラル木立

SONG は、SGNT を基盤とした MCS であり、分類精度を向上させると同時に計算コスト(メモリ使用量と計算時間)を削減することを目的としている。

冗長な葉を剪定することで、モデルの構造を最適化し、メモリ使用量を削減する。

2.1 構築方法

2.1.1 オンライン剪定

学習中にクラスラベルを使用して冗長な葉を削除し、計算コストを削減する。

2.1.2 オフライン剪定

学習後に冗長な葉をさらに剪定し、汎化能力を向上させる。この段階では 10 分割交差検証を用いて剪定の閾値を最適化する。

3. 実験

3.1 実験方法

今回の実験では UCI 機械学習リポジトリにある 10 個のベンチマーク問題[2]に対して、バギングを用いた SONG の計算コスト(メモリ容量と計算時間)と分類精度を測定する。表 1 にデータセットの概要を示す。表 1 において N はデータ数、 m は特徴量の数を表す。

表 1 データセット概要

データセット	N	m	クラス数
balance-scale	625	4	3
breast-cancer-w	699	9	2
glass	214	9	6
ionosphere	351	34	2
iris	150	4	3
letter	20000	16	26
liver-disorders	345	6	2
new-thyroid	215	5	3
pima-diabetes	768	8	2
wine	178	13	3

10 個のベンチマーク問題に対して 10 分割交差検証を行い SONG がどのように剪定されるかを評価する。本実験では、SONG と k 近傍法に以下のような修正ユークリッド距離尺度を用いる:

$$d(x, y) = \sqrt{\sum_{i=1}^m a_i \cdot (x_i - y_i)^2}$$

$$a_i = \frac{1}{\max_j - \min_j} \quad (1 \leq j \leq N)$$

ここで、 m は入力データ(特徴量)の次元数を、 N はデータ数を表す。SONG は取り扱うデータセットごとにオフライン選定の基準となる最適なしきい値 α があるため、0.5 から 1 までの異なる α を設定する。 $\alpha = [0.5, 0.55, 0.6, \dots, 1]$

SONG 内の SGNT の数を 25 とし、各トレーニングセットのサンプリング順序を変えて 100 回の試行を行う。本節の実験はすべて、1.4GHz 64bit quad-core ARM Cortex-A53 CPU、1GB RAM を搭載した Raspberry Pi 3 (Model B+) で行った。

3.2 実験結果

表 2 は、SONG の 100 回の施行における平均必要メモリと分類精度を示している。必要メモリは、SGNT の根、ノード、葉を合計したユニット数である。SONG を最適化することで、平均必要メモリ量は 65% から 96.6% 減少し、分類精度は 0.1% から 2.9% 向上した。これは SONG が効率的にメモリを利用し、データセットに対して高い分類能力を持つことを示している。表 3 では、SONG の性能を評価するため、C4.5 ベースの MCS との比較結果を示している。分類器数 $K=25$ 、バギングを適用した両手法の比較により、SONG は 10 個のデータセット中 6 つで C4.5 ベースの MCS より高精度を達成し、iris データセットを含め全問題で精度が向上した。これにより、SONG はスケーラビリティとノイズ耐性に優れた効率的な MCS であることが確認された。表 4 は SONG と k 近傍法との比較を示している。SONG は 10 個のデータセット中 8 つで平均 1.1% の精度向上を達成し、全ケースでメモリ使用量を抑えた。計算時間では、SONG は大規模データでは k 近傍法より高速で、特に letter データセットでは役 2.33 倍高速だった。SGNT の再利用と並列実行が可能な点も含め、SONG はエッジ環境下での大規模データマイニングにおいて、C4.5 ベースの MCS よりも実用的かつ効率的な手法だと分かった。

SONG は各分類器が独立して動作するために並列処理が可能であり、エッジ環境における大規模データマイニングにおいて、k 近傍法や C4.5 ベースの MCS に比べて実践的かつ効率的な手法であると結論づけられる。

[†] 呉工業高等専門学校 専攻科

[‡] 呉工業高等専門学校 電気情報工学科

表 2
SONG におけるバギングされた SGNT の 100 回試行の平均メモリと分類精度

データセット	平均メモリ			分類精度		
	剪定	非剪定	比率	剪定	未剪定	比率
balance-scale	112.38	861.12	13.1	0.87(5.69)	0.847(7.82)	+2.3
breast-cancer-w	28.59	897.43	3.19	0.972(2.24)	0.968(2.6)	+0.4
glass	104.11	297.77	35.0	0.721(11.7)	0.716(12.8)	+0.5
ionosphere	54.3	472.14	11.5	0.893(7.48)	0.868(7.79)	+2.5
iris	15.11	208.68	7.24	0.964(4.55)	0.961(4.74)	+0.3
letter	6225.08	27028.56	23.0	0.956(0.8)	0.955(0.72)	+0.1
liver-disorders	154.58	471.73	32.77	0.625(14.12)	0.608(16.97)	+1.7
new-thyroid	49.55	298.23	16.61	0.952(6.32)	0.949(6.76)	+0.3
pima-diabetes	205.9	11045.18	19.7	0.75(7.27)	0.73(8.71)	+2.0
wine	14.31	239.04	5.99	0.965(4.95)	0.96(5.57)	+0.5
平均	696.39	3181.99	16.8	0.867	0.856	+1.1

表 3
バギングを用いた剪定された MCS と C4.5 ベースの MCS の性能向上

データセット	SGNT ベース MCS			C4.5 ベース MCS		
	SGNT	MCS	比率	C4.5	MCS	比率
balance-scale	0.782	0.87	+8.8	0.795	0.827	+3.2
breast-cancer-w	0.957	0.972	+1.5	0.946	0.963	+1.7
glass	0.641	0.721	+8	0.664	0.757	+9.3
ionosphere	0.853	0.893	+4	0.897	0.92	+2.3
iris	0.949	0.964	+1.5	0.953	0.947	-0.6
letter	0.879	0.956	+7.7	0.880	0.938	+5.8
liver-disorders	0.58	0.625	+4.5	0.635	0.736	+10.1
new-thyroid	0.935	0.952	+1.7	0.93	0.94	+1
pima-diabetes	0.7	0.75	+5	0.749	0.767	+1.8
wine	0.95	0.965	+1.5	0.927	0.949	+2.2
平均	0.823	0.869	+4.4	0.837	0.874	+3

表 4
最適に剪定された SONG と k-NN の分類精度、メモリ要件、および計算時間 (100 回試行)

データセット	分類精度		メモリ使用量		計算時間 (s)	
	SONG	K 近傍法	SONG	K 近傍法	SONG	K 近傍法
balance-scale	0.883	0.899	114.46	562.5	0.55	0.85
breast-cancer-w	0.976	0.973	26.67	629.1	0.79	0.51
glass	0.756	0.706	115.97	192.6	0.25	0.02
ionosphere	0.909	0.857	53.87	315.9	1.53	0.21
iris	0.973	0.96	13.49	135	0.09	0.02
letter	0.958	0.961	6214.2	18000	167.27	389.76
liver-disorders	0.664	0.647	199.37	310.5	0.35	0.2
new-thyroid	0.968	0.967	53.57	193.5	0.17	0.01
pima-diabetes	0.766	0.753	188.46	691.2	1.26	1.28
wine	0.983	0.977	10.86	160.2	0.2	0.05
平均	0.884	0.873	699.09	2119.1	17.25	39.29

4. 結言

本研究では、SONG は、エッジコンピューティング環境での大規模データセットの分類において、計算コスト削減と分類精度向上を実現する実用的な手法であることが分かった。

今後の課題として、SONG の並列分散処理を研究し、さらに大規模データセットへの適用を目指す。

参考文献

- [1] 大北正昭, 徳高平蔵, 藤村喜久郎, 榎田英功, "自己組織化マップとそのツール," 第 6 章, 丸善出版, 2012.
- [2] M.Kelly, R.Longjohn, K.Nottingham, The UCI Machine Learning Repository, <https://archive.ics.uci.edu>