

ベンチマークデータを用いた 時系列勾配ブースティング木の実験評価

Experimental evaluation of Time-Series Gradient Boosting Tree with Time-Series Benchmark datasets

今村 光良 ^{*1*2} 中川 慧 ^{*1*3} 吉田 健一 ^{*3}
Mitsuyoshi Imamura Kei Nakagawa Kenichi Yoshida

^{*1}野村アセットマネジメント株式会社
Nomura Asset Management Co., Ltd

^{*2}筑波大学大学院 システム情報工学研究科
University of Tsukuba Graduate School of Systems and Information Engineering

^{*3}筑波大学大学院 ビジネス科学研究科
University of Tsukuba Graduate School of Business Sciences

In this paper, We evaluated the time-series gradient boosting decision tree method using benchmark data. Our time-series gradient boosting tree has weak learners with time-series and cross-sectional attribute in its internal node, and split examples based on dissimilarity between a pair of time-series or impurity between a pair of cross-sectional attributes. It has been empirically observed that the method induces accurate and comprehensive decision trees in time-series classification, which has gaining increasing attention due to its importance in various real-world applications.

1. はじめに

本論文では、別報 [中川 18] で提案された時系列勾配ブースティング木について、様々な時系列データセットに対して分析を行い、その有効性を評価する。別報 [中川 18] では時系列およびクロスセクションの属性を持つデータセットに対する時系列勾配ブースティング決定木を提案した。勾配ブースティング木などのツリーモデルは一般に可読性に優れており、説明責任という観点から実務への適用がしやすい。勾配ブースティング木は決定木を多段に組み合わせることで確率的な予測値を導出する手法であり、非常に高い精度を実現できる。しかしながら従来の決定木学習手法は時系列属性を想定していないため、時系列データを含むデータ集合に適用する場合、データの前処理が必要となる。最も単純な前処理の方法として、時系列データを計測値の平均値や標準偏差などの時系列データのモーメントで置き換える方法が考えられる。ただし、この方法は時系列データの構造、すなわち形状を無視しており、例えば形が大きく異なる時系列データを同一視してしまう欠点がある。そこで、勾配ブースティング・アルゴリズムに用いる弱学習器として、以上の問題点を克服した時系列データの形を陽に扱い時系列データ全体を対象とする [Yamada 03] の基準例分割テストによる分割を用いた時系列決定木を使用する。時系列決定木は相違度基準として動的時間伸縮法 (DTW) を用いているが、時系列ペアに対し距離が定義されていることから、[Nakagawa 17] のように決定木学習手法ではなく、最近傍法 (k -NN や k^* -NN [Anava 16]) を用いることも考えられる。ただし、最近傍法は怠惰学習 (lazy learning) であるために分類モデルが存在せず、学習結果が分かりにくいという欠点がある。本論文では、85 種類の異なる時系列データセットである UCR データセット [Chen 15] を用いて、DTW を距離尺度とした 1-Nearest Neighbor (NN) と先行研究である [Yamada 03] の時系列決定木をベンチマークに、時系列勾配ブースティング

木の性能を評価する。

2. 先行研究

勾配ブースティングは、ブースティング・アルゴリズムの一種である。ブースティングとは、集団学習の枠組みの一つで、複数の弱学習器を統合して全体の学習器を構成する手法である。弱学習器としては、決定木が使われることが多い。ブースティングにおける弱学習器として決定木を用いることは、データの外れ値に強い、数値変数と離散変数や欠損値などを扱いやすいなどの利点がある。本研究では弱学習器として、[Yamada 03] の時系列決定木に、時系列以外の通常の特徴量も追加した決定木を使用することを提案する。そこで、以下では時系列決定木および勾配ブースティング木についてレビューする。

2.1 時系列決定木

時系列決定木は、内部ノードに基準となる属性時系列を持つ決定木であり、基準例分割テストまたはクラスター分割テストによってサンプルを分割していく [Yamada 03]。[山田 03] の実験からクラスター分割テストより基準例分割テストの結果が良好であることから、本稿では基準例分割テストを分割手法として用いる。

データセット D は、 n 個のサンプル $\{x_i, i = 1, \dots, n\}$ を含み、各サンプル x_i はそれぞれ m 個の属性 $\{a_j, j = 1, \dots, m\}$ とクラスラベルまたは目的変数 $\{y_i, i = 1, \dots, n\}$ を持つ。各属性 a_j は、時間順に値を並べた時系列データ又は系列を持たない通常の連続値、名目属性を持つ。ここで $x(a_j)$ は一つのサンプルであり、基準例と呼ぶ。基準例分割テストは、サンプル x_i の属性 a_j に関する時系列データを $x_i(a_j)$ で表すと、データセット全体を、 $DTW(x(a_j), ex_i(a)) < \theta_i$ を満たすデータセットから構成される集合 $L(x, a_j, \theta_i) = \{(y, x) | DTW(x(a_j), x_i(a_j)) < \theta_i\}$ とそれ以外の集合 $R(x, a_j, \theta_i) = \{(y, x) | DTW(x(a_j), x_i(a_j)) \geq \theta_i\}$ に分割する。ここで $DTW(x, y)$ は、動的時間伸縮法 (DTW) に基づいた類似度を表す。DTW は時系列データにおける一点のデータをもう片方の時系列データにおける複数点

連絡先: 今村 光良, 野村アセットマネジメント株式会社, 〒103-8260 東京都中央区日本橋一丁目 12 番 1 号

のデータに対応づけられるため、時間方向の非線形な伸縮を許容する。このため、長さが異なる時系列データのペアに適応できる上に、結果が人間の直観により合致する。具体的なアルゴリズムは、Algorithm1 の通り。

基準例分割テストはサンプルが基準例 $x(a_j)$ と DTW で計測して、類似しているか、あるいは類似していないかという基準で閾値 θ を変えながら分割する。当該分割の評価基準としては、利得比基準など通常の決定木で用いられている方法が使える。

Algorithm 1 DTW distance

```

1: procedure DTW( $x, y$ )
    ▷ Initialize matrix D

2:   Var  $D[N, M]$ 
3:    $D[1, 1] = 0$ 
4:   for  $i = 2$  to  $N$  do
5:     for  $j = 2$  to  $M$  do
6:        $D[i, j] = \infty$ 
7:     end for
8:   end for
    ▷ Calculate DTW distance

9:   for  $i = 2$  to  $N$  do
10:    for  $j = 2$  to  $M$  do
11:       $D[i, j] = d(x[i - 1], y[j - 1])$ 
         $+ \min(D[i, j - 1], D[i - 1, j], D[i - 1, j - 1])$ 
12:    end for
13:  end for
14:  return  $D[N, M]$ 
15: end procedure

```

2.2 勾配ブースティング木

[Friedman 01] は、ブースティングの各ステップで構築するモデルに回帰ツリーを用いる勾配ブースティング木 (gradient boosting tree) という手法を考案した。これは特徴量 (入力) を x 、ラベルを y として、弱学習器 $f_m(x)$, $m = 0, \dots, M$ に対して全体の学習器 $F(x)$

$$F(x) = f_0(x) + f_1(x) + \dots + f_M(x) \quad (1)$$

を損失関数 $L(y, F(x))$ が最小になるように弱学習器 $f_m(x)$ を逐次的に学習し統合する。つまり、最初に学習器 $F_0(x) = f_0(x)$ が与えられるとし、 m ステップ目の学習では m 個の学習器からなる全体の学習器を損失関数 $L(y, F(x))$ が最小になるように弱学習器を決める。

$$F_m(x) = F_{m-1}(x) + f_m(x) \quad (2)$$

具体的なアルゴリズムは以下の通りである。勾配ブースティング木の特徴は、その時点でのモデルの予測値と結果の残差を目的変数とした回帰ツリーを構築し、そのツリーモデルの予測値による値を加算することにある。

3. 提案手法 - 時系列勾配ブースティング木

前章で確認した [Yamada 03] の基準例分割テストによる時系列決定木において、時系列属性以外の属性、通常の連続値、名目属性も追加する。属性が時系列データの場合には基準例分割テストによって分割を行い、時系列以外の場合には、通

Algorithm 2 Gradient Boosting Tree

```

1: procedure GRADIENT BOOSTING TREE( $y, x$ )
    ▷ Initialize  $F_0$  with a constant

2:    $F_0(x) = \sum_{i=1}^N \arg \min(y_i, c)$ 

3:   for  $m = 1$  to  $M$  do
4:      $r_{im} = -[\frac{L(y_i, F(x_i))}{F'(x_i)}]_{F(x)=F_{m-1}(x_{i-1})}$ 
        ▷ Fit a decision tree to predict targets  $r_{im}$ 
5:      $f(x) = \text{FitTree}(r_{im}, x)$ 
6:      $\rho_m = \arg \min_{\rho} \sum_{i=1}^N L(y_i, F_{m-1}(x) + \rho f(x))$ 
7:      $F_m(x) = F_{m-1}(x) + \gamma \rho_m f(x)$ 
8:   end for
9:   return  $F(x) = \sum_{m=1}^M F_m(x)$ 
10: end procedure

```

常の決定木と同様の基準で分割を行う。具体的なアルゴリズム (StandardExSplit with Cross-Sectional Data) は、Algorithm3 の通り。 $H(\bullet)$ は不純度を表す関数で、本稿では、よく使用される Gini 係数を用いる。

Algorithm 3 StandardExSplit with Cross-Sectional Data

```

1: procedure SPLIT( $\{y_1, \dots, y_n\}, \{x_1, \dots, x_n\}$ )
2:   for each samples  $x_i$  do
3:     for each attribute  $a_j$  do
4:       if attribute  $a_j$  is time-series then
5:         for each samples  $x_k$  do
6:            $L(x, a_j, \theta_i) = \{(y, x) | DTW(x_i(a_j), x_k(a_j)) < \theta_k\}$ 
7:            $R(x, a_j, \theta_i) = \{(y, x) | DTW(x_i(a_j), x_k(a_j)) \geq \theta_k\}$ 
8:            $G(L, R, \theta_i) = \frac{|L|}{|L+R|} H(L) + \frac{|R|}{|L+R|} H(R)$ 
9:         end for
10:        else
11:           $L(x, a_j, \theta_i) = \{(y, x) | x(a_j) < \theta_i\}$ 
12:           $R(x, a_j, \theta_i) = \{(y, x) | x(a_j) \geq \theta_i\}$ 
13:        end if
14:         $G(L, R, \theta_i) = \frac{|L|}{|L+R|} H(L) + \frac{|R|}{|L+R|} H(R)$ 
15:      end for
16:    end for
17:     $\theta^* = \arg \min_{\theta} G(L, R, \theta)$ 
18:    return best split  $\theta^*$ 
19:  end procedure

```

この分割アルゴリズムを用いた時系列決定木を勾配ブースティング木の弱学習器として利用する。ただし、本論文の分析においては、データセットにクロスセクションのデータが存在しないため、クロスセクションのデータは用いない。

4. 実験

提案手法の有効性を確認するために、実データを用いた分析を行う。85 種類の異なる時系列データセットである UCR データセット [Chen 15] を用いて、時系列勾配ブースティング木の性能を評価する。UCR データセットは平均 0、分散 1 に正規化されているため、特別な前処理は行っていない。各データセットはそれぞれ、Image Outline(IO)、Motion Classification(MC)、

Sensor Readings(SR)、Synthetic(Syn) という 4 つのカテゴリに分類できる。表 1 にベースラインである DTW 距離を用いた 1-NN(1-NN DTW) と先行研究で提案された時系列決定木 (TST)、そして提案手法である時系列勾配ブースティング木 (TSGBT) の各データセット毎の Error Rate をまとめた。31 のデータセットについてベンチマークである 1-NN DTW に対して改善がみられた。

勾配ブースティング木は、怠惰学習 (Lazy Learning) である k -NN とは異なり、学習が必要であるため、学習にはある程度のサンプル数が必要であると考えられる。そこで、別報 [中川 18] の金融時系列データの分析と同程度である 400 以上のサンプルが確保できるデータセット 21 個についての各カテゴリの平均 Error Rate を示したのが表 1 である。14 のデータセットについて改善がみられた。

5. まとめ

本論文では、別報 [中川 18] で提案された時系列勾配ブースティング木について、85 種類の異なる時系列データセットである UCR データセット [Chen 15] を用いて、時系列勾配ブースティング木の性能を評価した。ベースラインである DTW 距離を用いた 1-NN(1-NN DTW) と先行研究で提案された時系列決定木 (TST)、そして提案手法である時系列勾配ブースティング木 (TSGBT) の各データセット毎の Error Rate を評価した結果、31 のデータセットについてベンチマークである 1-NN DTW に対して改善がみられた。

また、勾配ブースティング木は、怠惰学習 (Lazy Learning) である k -NN とは異なり、学習が必要であるため、学習にはある程度のサンプル数が必要であると考えられる。そこで、400 以上のサンプルが確保できるデータセット 21 個についての各カテゴリの平均 Error Rate を評価したところ、14 のデータセットについて改善がみられた。

参考文献

- [Anava 16] Anava, O. and Levy, K.: k^* -nearest neighbors: From global to local, in *Advances in Neural Information Processing Systems*, pp. 4916–4924 (2016)
- [Chen 15] Chen, Y., Keogh, E., Hu, B., Begum, N., Bagnall, A., Mueen, A., and Batista, G.: The UCR Time Series Classification Archive (2015), www.cs.ucr.edu/~eamonn/time_series_data/
- [Friedman 01] Friedman, J. H.: Greedy function approximation: a gradient boosting machine, *Annals of statistics*, pp. 1189–1232 (2001)
- [Nakagawa 17] Nakagawa, K., Imamura, M., and Yoshida, K.: Stock Price Prediction using k^* -Nearest Neighbors and Indexing Dynamic Time Warping, in *Artificial Intelligence of and for Business (AI-Biz 2017)* (2017)
- [Yamada 03] Yamada, Y., Suzuki, E., Yokoi, H., and Takabayashi, K.: Decision-tree induction from time-series data based on a standard-example split test, in *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*, pp. 840–847 (2003)

表 1: Error Rate of UCR Time-series Datasets

Name	Category	1-NN DTW	TST	TSGBT
50Words	IO	0.31	0.68	0.40
Adiac	IO	0.40	0.80	0.51
ArrowHead	IO	0.30	0.34	0.39
Beef	SR	0.37	0.63	0.57
BeetleFly	IO	0.30	0.30	0.30
BirdChicken	IO	0.25	0.25	0.30
Car	SR	0.27	0.43	0.40
CBF	Syn	0.00	0.11	0.01
ChlorineConcentration	SR	0.35	0.45	0.41
CinCECGtorso	SR	0.35	0.54	0.44
Coffee	SR	0.00	0.11	0.11
Computers	SR	0.30	0.38	0.32
CricketX	MC	0.25	0.62	0.36
CricketY	MC	0.26	0.61	0.39
CricketZ	MC	0.25	0.66	0.36
DiatomSizeReduction	IO	0.03	0.17	0.14
DistalPhalanxOutlineAgeGroup	IO	0.21	0.24	0.21
DistalPhalanxOutlineCorrect	IO	0.23	0.25	0.23
DistalPhalanxTW	IO	0.29	0.36	0.23
Earthquakes	SR	0.26	0.29	0.23
ECG	SR	0.23	0.21	0.21
ECG5000	SR	0.08	0.08	0.07
ECGFiveDays	SR	0.23	0.32	0.32
ElectricDevices	SR	0.40	0.36	0.25
Face (all)	IO	0.19	0.66	0.28
Face (four)	IO	0.17	0.11	0.35
FacesUCR	IO	0.10	0.54	0.18
Fish	IO	0.18	0.43	0.28
FordA	SR	0.44	0.36	0.26
FordB	SR	0.41	0.36	0.24
Gun-Point	MC	0.09	0.27	0.21
Ham	SR	0.53	0.49	0.33
HandOutlines	IO	0.20	0.19	0.17
Haptics	MC	0.62	0.63	0.60
Herring	IO	0.47	0.41	0.39
InlineSkate	MC	0.62	0.74	0.70
InsectWingbeatSound	SR	0.65	0.69	0.58
ItalyPowerDemand	SR	0.05	0.09	0.09
LargeKitchenAppliances	SR	0.21	0.27	0.19
Lightning-2	SR	0.13	0.26	0.25
Lightning-7	SR	0.27	0.51	0.42
MALLAT	Syn	0.07	0.36	0.16
Meat	SR	0.07	0.10	0.12
MedicalImages	IO	0.26	0.45	0.28
MiddlePhalanxOutlineAgeGroup	IO	0.25	0.26	0.25
MiddlePhalanxOutlineCorrect	IO	0.35	0.37	0.31
MiddlePhalanxTW	IO	0.42	0.40	0.40
MoteStrain	SR	0.17	0.18	0.19
Non-Invasive Fetal ECG Thorax1	SR	0.21	0.71	0.24
Non-Invasive Fetal ECG Thorax2	SR	0.14	0.88	0.18
OliveOil	SR	0.17	0.17	0.20
OSU Leaf	IO	0.41	0.52	0.43
PhalangesOutlinesCorrect	IO	0.27	0.31	0.20
Phoneme (readme)	SR	0.77	0.80	0.77
Plane	IO	0.00	0.21	0.00
ProximalPhalanxOutlineAgeGroup	IO	0.20	0.16	0.18
ProximalPhalanxOutlineCorrect	IO	0.22	0.20	0.17
ProximalPhalanxTW	IO	0.26	0.28	0.23
RefrigerationDevices	SR	0.54	0.53	0.50
ScreenType	SR	0.60	0.57	0.62
ShapeletSim	Syn	0.35	0.45	0.46
ShapesAll	IO	0.23	0.90	0.29
SmallKitchenAppliances	SR	0.36	0.27	0.22
SonyAIBORobot Surface	SR	0.28	0.29	0.13
SonyAIBORobot SurfaceII	SR	0.17	0.21	0.21
StarLightCurves	SR	0.09	0.07	0.05
Strawberry	SR	0.06	0.13	0.09
Swedish Leaf	IO	0.21	0.48	0.18
Symbols	IO	0.05	0.24	0.10
Synthetic Control	Syn	0.01	0.05	0.03
ToeSegmentation1	MC	0.23	0.28	0.26
ToeSegmentation2	MC	0.16	0.27	0.27
Trace	SR	0.00	0.09	0.06
Two Patterns	Syn	0.00	0.03	0.00
TwoLeadECG	SR	0.10	0.20	0.20
uWaveGestureLibraryX	MC	0.27	0.31	0.23
uWaveGestureLibraryY	MC	0.37	0.39	0.31
uWaveGestureLibraryZ	MC	0.34	0.35	0.29
UWaveGestureLibraryAll	MC	0.11	0.39	0.10
Wafer	SR	0.02	0.03	0.02
Wine	SR	0.43	0.48	0.35
WordSynonyms	IO	0.35	0.63	0.47
Worms	IO	0.54	0.48	0.51
WormsTwoClass	IO	0.34	0.34	0.34
Yoga	IO	0.16	0.24	0.19

表 2: Average Error Rate of UCR Time-series Datasets by Category

Category	1-NN DTW	TST	TSGBT
Image Outline	0.23	0.49	0.24
Motion Classification	0.27	0.36	0.20
Sensor Readings	0.24	0.37	0.19
Synthetic	0.00	0.03	0.00

[山田 03] 山田悠, 鈴木英之進, 横井英人, 高林克日己 他: 動的
時間伸縮法に基づく時系列データからの決定木学習, 情報
処理学会研究報告知能と複雑系 (ICS), Vol. 2003, No. 30
(2002-ICS-132), pp. 141–146 (2003)

[中川 18] 中川慧, 今村光良, 吉田健一: 時系列勾配ブースティ
ング木による分類学習金融時系列予測への応用, 人工知能学
会全国大会論文集 2018 年度人工知能学会全国大会 (第 32
回) 論文集社団法人 人工知能学会 (2018)