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

Classification by Time-Series Gradient Boosting Tree

中川 慧 *1*2 今村 光良 *1*3 吉田 健一 *2 Kei Nakagawa Mitsuyoshi Imamura 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 ビジネス科学研究科 University of Tsukuba Graduate School of Business Sciences

*³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

We propose a time-series gradient boosting decision tree for a data set with time-series and cross-sectional attributes. 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 cross-sectional attributes. Dissimilarity between a pair of time-series is defined by dynamic time warping method or in financial time-series by indexing dynamic time warping method. Experimental results with stock price prediction confirm that our method constructs interpretable and accurate decision trees.

1. はじめに

株価を予測するために様々な方法が提案されている。主に時 系列解析に基づくもの、クロスセクショナルなデータを特徴量 として使用した、いわゆる機械学習に基づくものの2種類が 代表的である。時系列分析手法は、データ自身の過去の変動パ ターンを線形方程式でモデル化するという意味で恣意性の低い 方法である。一方で、機械学習による予測は時系列解析と異な り、特徴量をうまく選択する必要がある。従来のファイナンス分 野で、株価を予測するための特徴量としてはバリューやモメン タム等がある [Fama 92, Fama 93, Asness 13]。 最近ではアナ リストレポートやニュースといったテキストデータを、特徴量 として用いる試みが模索されている [Takahashi 06, 工藤 17]。 本論文では、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 する。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、株価のような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。さらに時 系列属性以外の通常の属性も取り込んだ時系列決定木を弱学習 器として使用することを提案する。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、我々は金融時

連絡先: 中川 慧, 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 系列の分割の際には、時系列データの類似度を計測する手法として、[Nakagawa 17] で提案された Indexing Dynamic Time Warping(IDTW)を使用する。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。以上の時系列決定木を弱学習器として用いた時系列勾配ブースティング木を用いて、日本株式を対象に実証分析を行う。また金融時系列以外の時系列データセットについての実験評価については[今村 18] で議論し、本論文では手法の提案と金融時系列への応用を報告する。

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, ..., 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, ..., n} とク ラスラベルまたは目的変数 { y_i , i = 1, ..., 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 $R(x, a_j, \theta_i) = \{(y, x) | DTW(x(a_j), x_i(a_j)) \ge \theta_i\}$ 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 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。

Alg	gorithm 1 DTW distance	
1:	procedure $DTW(x, y)$	
		\triangleright 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	
		\triangleright 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 $f_m(x), m = 0, ..., M$ に対 して全体の学習器 F(x)

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

$$F_m(x) = F_{m-1}(x) + f_m(x)$$
(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。

Algorithm 2 Gradient Boosting Tree				
1: procedure Gradient Boosting $Tree(y, x)$				
\triangleright Initialize F_0 with a constant				
2: $F_0(x) = \sum_{i=1}^{N} \arg\min(y_i, c)$				
3: for $m = 1$ to $M \operatorname{do}^{c}$				
4: $r_{im} = -\left[\frac{L(y_i, F(x_i))}{F(x_i)}\right]_{F(x) = F_{m-1}(x_{i-1})}$				
\triangleright Fit a decision tree to predict targets r_{im}				
5: $f(x) = FitTree(r_{im}, x)$				
6: $\rho_m = \arg \min \sum_{i=1}^N L(y_i, F_{m-1}(x) + \rho f(x))$				
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				

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、通常 の決定木と同様の基準 (不純度) 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($\{y_1,, y_n\}, \{x_1,, x_n\}$)			
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: $L(x, a_j, \theta_i)$			
7: $= \{(y,x) DTW(x_i(a_j), x_k(a_j)) < \theta_k\}$			
8: $R(x, a_j, \theta_i)$			
9: $= \{(y,x) DTW(x_i(a_j),x_k(a_j)) \ge \theta_k\}$			
10: $G(L, R, \theta_i)$			
11: $= \frac{ L }{ L+R }H(L) + \frac{ R }{ L+R }H(R)$			
12: end for			
13: else			
14: $L(x, a_j, \theta_i) = \{(y, x) x(a_j) < \theta_i\}$			
15: $R(x, a_j, \theta_i) = \{(y, x) x(a_j) \ge \theta_i\}$			
16: end if			
17: $G(L, R, \theta_i) = \frac{ L }{ L+R } H(L) + \frac{ R }{ L+R } H(R)$			
18: end for			
19: end for			
20: $\theta^* = \arg\min_{\theta} G(L, R, \theta_i)$			
21: return best split θ^*			
22: 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。その際に金融時系列が 時系列属性として与えられた場合、時系列データのペアに関す る相違度計算法には [Nakagawa 17] で提案された、株価予測



 \boxtimes 1: Time-Series Tree with Cross-sectinal Data

に有効な Indexing Dynamic Time Warping(IDTW) を用い る。IDTW は金融実務の観点から、月末の値を1として、日々 の価格推移を月末との比を使って表した系列に対して DTW を適用する (Algorithm4)。

Algorithm 4 IDTW distance	
1: procedure $IDTW(x, y)$	
	\triangleright Scaling data
2: Var Ix, Iy	⊳ Initialize Ix,Iy
3: $Ix[1] = 1, Iy[1] = 1$	
4: for $i = 2$ to N do	
5: $Ix[i] = Ix[i-1]\frac{x[i]}{x[i-1]}$	
6: end for	
7: for $j = 2$ to M do	
8: $Iy[j] = Iy[j-1] \frac{y[j]}{y[j-1]}$	
9: end for	
	\triangleright Apply DTW
10: return $DTW(Ix, Iy)$	
11: end procedure	

構築される時系列決定木のイメージが図1である。時系列 に対して、通常の特徴量による分類だけでなく、時系列が自身 の過去の形状と類似しているか、類似していないか、という形 状に基づいた分類が行われている。そのため、どのような分類 が行われるのか直感的な理解が可能となる。

4. 実証分析

前章で提案した時系列勾配ブースティング木 (TS-GBTree) を用いて、日本株を対象に実証分析を行う。使用するデータは TOPIX 指数およびクロスセクションの情報として、指数外の 情報である為替レートのドル円、米国株式指数 SP500、商品 総合指数である SPGCSI 指数の月間収益率を用いる。使用し たデータの期間は、1985年1月から 2018年1月までであり、 学習期間は始点を 1985年1月に固定し、これより予測を実施 する直近の月までとして、2005年1月から 2018年1月まで を検証期間とした。分析にあたっては、時系列勾配ブースティ ング木を用いて、毎月末に予測を行い、上昇予測の場合は月末 に1単位 TOPIX を買い、下落予測の場合には1単位 TOPIX を売り、これを1か月毎に繰り返し予測を行った。比較対象は 時系列決定木 (TS-Tree)、先行研究の IDTW+k*-NN(k*-NN) を用いた。



⊠ 2: Total Return of TOPIX,k*-NN,Time-Series Decision Tree and Time-Series Gradient Boosting Tree.

表 1: Precision, Recall and F-Score of k*-NN

	precision	recall	f1-score	support
Down	0.62	0.52	0.57	65
Up	0.69	0.77	0.73	91
avg/total	0.69	0.77	0.73	156

表 2: Precision, Recall and F-Score of Time-Series Decision Tree(with CS)

	precision	recall	f1-score	support
Down	0.49	0.65	0.56	65
Up	0.67	0.52	0.58	91
avg/total	0.60	0.57	0.57	156

表 3: Precision, Recall and F-Score of Time-Series Gradient Boosting Decision Tree(with CS)

	precision	recall	f1-score	support
Down	0.55	0.65	0.59	65
Up	0.71	0.62	0.66	91
avg/total	0.64	0.63	0.63	156

表 4: Accuracy of Each Method

	k*-NN	TSTree	TSGBTree
TS	66.67%	54.49%	66.02%
TS+CS	-	57.05%	62.82%

時系列勾配ブースティング木と比較対象である k*-NN、時系 列決定木と TOPIX の累積収益率の推移が図 2 である。時系列 勾配ブースティング木は検証期間前半は TOPIX 指数を下回る ものの、リーマンショック以降の 2009 年から安定して TOPIX を上回っている。収益性については、時系列勾配ブースティン グ木は k*-NN に対しては 20%程度、時系列決定木に対して 100%程度上回った。一方で、表 1 が k*-NN の、表 2 が時系 列決定木の、表 3 が時系列勾配ブースティング木の Precision、 Recall および F スコアである。表から勾配ブースティング木と 時系列勾配ブースティング木で、Precision は 0.04%、Recall は 0.06%、F スコアは 0.06%の性能差があった。一方で、勾配 ブースティング木は k*-NN との比較では、Precision、Recall、 F スコア全て下回った。

表4は時系列決定木に対して、勾配ブースティングを用いる こと、およびクロスセクションのデータ(CS)を時系列決定木 に組み込むことによる正答率の改善を表している。勾配ブース ティングにより、精度が約11%改善し、クロスセクションの データを組み込むことで時系列決定木が約2%改善している。 しかしながら勾配ブースティング、クロスセクションのデータ ともに組み込むと、約8%の改善にとどまった。当然ながら、 TOPIX指数に対して説明力を持つであろう特徴量は上記以外 に多数あると考えられ、それらを組み込むことでさらなる精度 改善が見込まれる。

5. まとめ

本論文では、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・アルゴリズムに用いる弱学習器とし て、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 列決定木を使用した。加えて、時系列属性以外の通常の属性も 取り込み、金融時系列データの類似度の計測に [Nakagawa 17] の IDTW を用いた。TOPIX 指数を対象とした実証分析によ り次のような結果が得られた。

- 時系列決定木に対して勾配ブースティングを用いること で予測精度が11%程度改善。
- ・時系列決定木に対してクロスセクションのデータを組み込むことで2%程度予測精度が改善。
- 収益性については比較手法をすべて上回った。

一方で、クロスセクションのデータを時系列勾配ブースティング木に組み込むと、約8%の改善にとどまった。当然ながら、 TOPIX 指数に対して説明力を持つであろう特徴量は上記以外に多数あると考えられ、それらを組み込むことでさらなる精度改善が見込まれる。また、TOPIX 以外の株式指数や個別銘柄、あるいは債券や為替などの他のアセットへの適用が今後の 課題である。

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)
- [Asness 13] Asness, C. S., Moskowitz, T. J., and Pedersen, L. H.: Value and momentum everywhere, *The Journal of Finance*, Vol. 68, No. 3, pp. 929–985 (2013)

- [Fama 92] Fama, E. F. and French, K. R.: The crosssection of expected stock returns, the Journal of Finance, Vol. 47, No. 2, pp. 427–465 (1992)
- [Fama 93] Fama, E. F. and French, K. R.: Common risk factors in the returns on stocks and bonds, *Journal of financial economics*, Vol. 33, No. 1, pp. 3–56 (1993)
- [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)
- [Takahashi 06] Takahashi, S., Takahashi, M., Takahashi, H., and Tsuda, K.: Analysis of stock price return using textual data and numerical data through text mining, in *International Conference on Knowledge-Based* and *Intelligent Information and Engineering Systems*, pp. 310–316Springer (2006)
- [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 *Proceed*ings of the 20th International Conference on Machine Learning (ICML-03), pp. 840–847 (2003)
- [工藤 17] 工藤秀明, 永島淳, 宮崎義弘: 自然言語処理技術を 用いたアナリストレポートの実証分析: センチメントの変 化と株式市場の反応について, 証券アナリストジャーナル
 = Securities analysts journal, Vol. 55, No. 9, pp. 66–77 (2017)
- [今村 18] 今村光良,中川慧,吉田健一: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,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)
- [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)