

# 予測ベースの極性指標に基づく暗号 資産アセットアロケーション戦略の 提案

田口 怜<sup>1</sup>， 和泉 潔<sup>1</sup>， 村山 友理<sup>1</sup>

<sup>1</sup>東京大学大学院 工学系研究科 システム創成学専攻

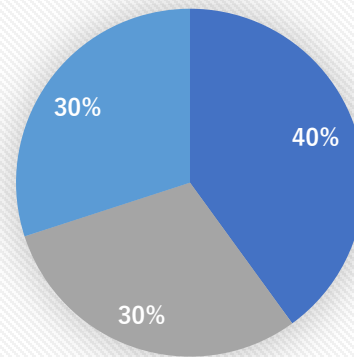
# 1. アセットアロケーションとは

ポートフォリオとは、分散投資を目的とした金融資産の組み合わせを指します。

ポートフォリオのアセットアロケーションとは、投資家の目標やリスク許容度などに応じて最適な資産の配分を決定することを指します。

Portfolio

■ Asset Class 1 ■ Asset Class 2 ■ Asset Class 3



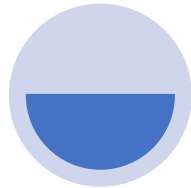
## 2. 現代のアセットアロケーションにおける課題

各先行研究に代表されるアルゴリズムは、数理的には美しいが、現実世界のリーマンショックやコロナショックなど急激な市場の変化に耐えられるのだろうか？

運用の全期間同じアルゴリズムを使っているが、市況に適した資産運用になっているのだろうか？

先行研究	内容概略
平均-分散モデル	ポートフォリオ収益率の平均と分散を考慮した最適化モデルです。資産運用の実務でも多用されます。
CVaR最適化	リスク指標Conditional Value at Risk (CVaR)に基づく最適化モデルです。最適化の条件として期待損失を最小化する必要があります。資産運用の実務でも多用されます。
Yunら (2020)	2段階深層学習モデルを用いた予測に基づくポートフォリオ最適化モデルです。
Chenら (2021)	XGBoost と改良型ホタルアルゴリズムを用いたポートフォリオ最適化モデルです。
Maら (2021)	平均-分散モデルと深層学習を組み合わせたポートフォリオ最適化モデルです。

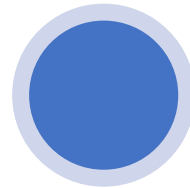
# 3. アセットアロケーションの課題を解く ために着目した技術



## オルタナティブデータ

《理由》

近年注目を集めており、伝統的な金融経済データからは得られない情報が引き出せるといった研究成果が多くあるからです。



## レジーム・スイッチ

《理由》

主に金融時系列解析で用いられる用語であり、時系列データの急激な変動をとらえることができるからです。

アセットアロケーションの対象に暗号資産を選択した理由は、伝統的な資産クラスにおける類似のアプローチの研究がすでにあるからです。

[Asset Allocation Method Based on Sentiment Signals and Causal Information using Multi-asset Classes](#)

Rei Taguchi, Hiroki Sakaji, Kiyoshi Izumi, Yuri Murayama  
International Journal of Smart Computing and Artificial Intelligence 7(2) 2023年12月14日

[Constructing Sentiment Signal-Based Asset Allocation Method with Causality Information](#)

Rei Taguchi, Hiroki Sakaji, Kiyoshi Izumi, Yuri Murayama  
New Generation Computing 41(777-794) 2023年9月11日

# 4. オルタナティブデータとは

- オルタナティブデータとは、投資判断に使われるデータのうち、伝統的に用いられる財務諸表など一般的な公開情報「以外」のデータ群の総称です。

## オルタナティブデータの例

POSデータ



経済ニュースの記事



クレジットカードの利用データ



SNSの投稿



衛星画像



ウェブサイトのトラフィック



携帯端末の利用ログ



# 5. レジーム・スイッチとは

経済・金融情報を含む時系列データは、好況や不況 など「背後にある市場の見えざる状態」が存在しています。

- このような状態を**レジーム (regime)** と呼び、
- この状態の変化を**レジーム・スイッチ (regime switch)** と呼び、
- この状態の変化を数理的に表現したモデルを**レジーム・スイッチングモデル (regime switching model)** と呼びます。

オレンジ/緑で  
レジームがス  
イッチしている

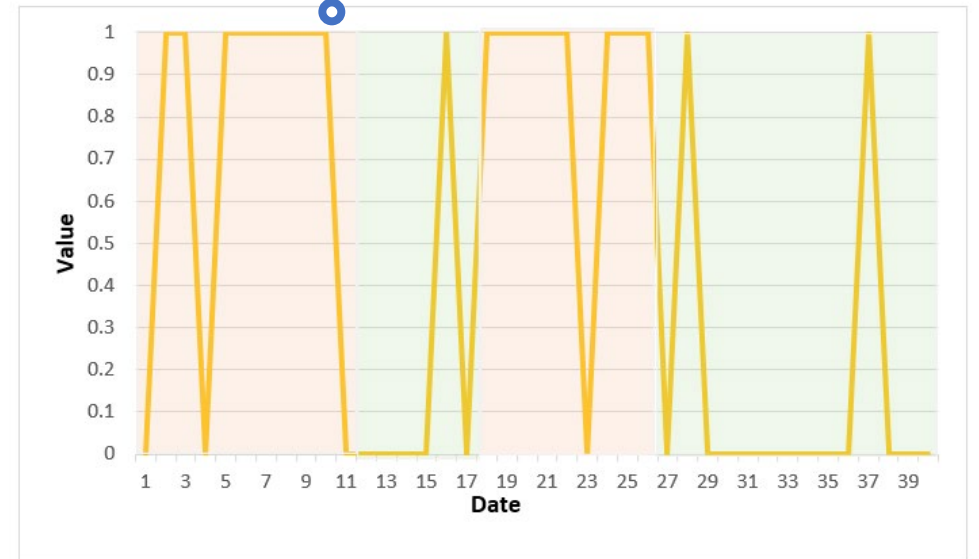


図1.1 レジームスイッチの例

# 6. アセットアロケーションの枠組みに、オルタナティブデータとレジーム・スイッチを用いる際の課題

## レジーム・スイッチを用いる際の課題

- 先行研究において、算出された市況の変化点が、投資戦略の中に効果的に織り込まれていないことです。
- 先行研究において、他のポートフォリオ最適化戦略と比べて、レジーム・スイッチを織り込んだ投資戦略の有用性が十分に実証できていないことです。

## オルタナティブデータを用いる際の課題

- そもそもオルタナティブデータをアセットアロケーションの文脈で使用した先行研究がほとんど存在しないことです。
- アセットアロケーションでは金融時系列データの組み合わせる際の比率を計算するので、オルタナティブデータを何らかの方法で最適化アルゴリズムの外からフレームワークに組み込む必要があることです。

# 7. 課題解決のためのアプローチ

## Itoら (2021)

- 市場環境に応じて投資戦略を切り替える手法に関する研究です。
- 状況に応じてアルゴリズムをスイッチする手法についてのユースケースを得ました。

## Taguchiら (2021)

- テキストデータから作成した極性指標が、既存の経済指標に対してグレンジャー因果性もつことを実証した研究です。
- 作成した極性指標が、既存のいくつかの経済指標に対して先行するという示唆を得ました。

## Taguchiら (2022)

- 極性指標とレジーム・スイッチングモデルが、株式のペアトレーディングシミュレーションに有効であることを示した研究です。
- 極性指標の変化点をリバランスのシグナルとして使用できるという示唆を得ました。
- **本研究では、Taguchiら (2022) をより実践的に拡張し、機動的かつ効果的なアセットアロケーション戦略を構築することを目指します。**



# 8. 本研究の目的



オルタナティブデータから算出した変化点が、リバランスのシグナルとして有効であると示すこと。

多期間において、下方リスクを動的に最小化するフレームワークになっていること。

因果推論の結果が、アセットアロケーションに対して一定の効果があることを示すこと。

# 9. はじめに

- 本研究では、暗号資産を用いた戦術的なアセットアロケーションにテキストデータが有用であることを実証します。
- **本研究の仮説**：極性指標の変化点に応じて運用戦略を切り替えることで、暗号資産ポートフォリオは、ある程度のリスクをヘッジしながら高いリターンを確保することができます。
- **本研究の貢献**：テキストデータを用いた極性指標の変化点推定がアクティブ運用に有効であることを実証し、暗号資産に対する表現力の高いアセットアロケーションのフレームワークを提案します。

# 10. アーキテクチャ

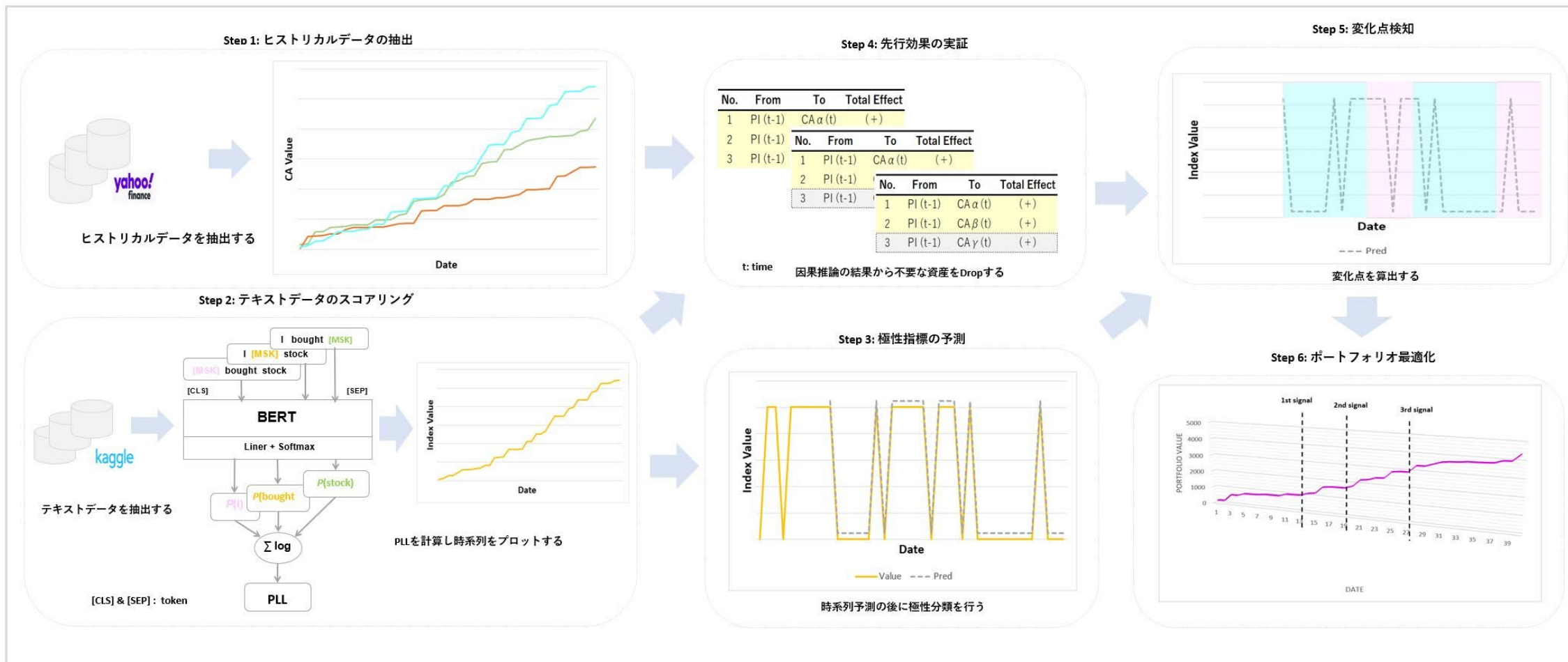


図1 アーキテクチャ

本章ではでは提案手法をSentiment Signal-based Asset Allocation Method with Causal Information using Crypto Assets (CA-SSAAM)と呼称します。

# 11. 提案手法 (1/8)

-テキストデータのスコアリング-

- 本研究では、極性指標の作成にPseudo-log-likelihood scores (PLL) を用いました。PLLはSalazar et al.が提唱した確率的言語モデルに基づくスコアです。

$t$  は時間

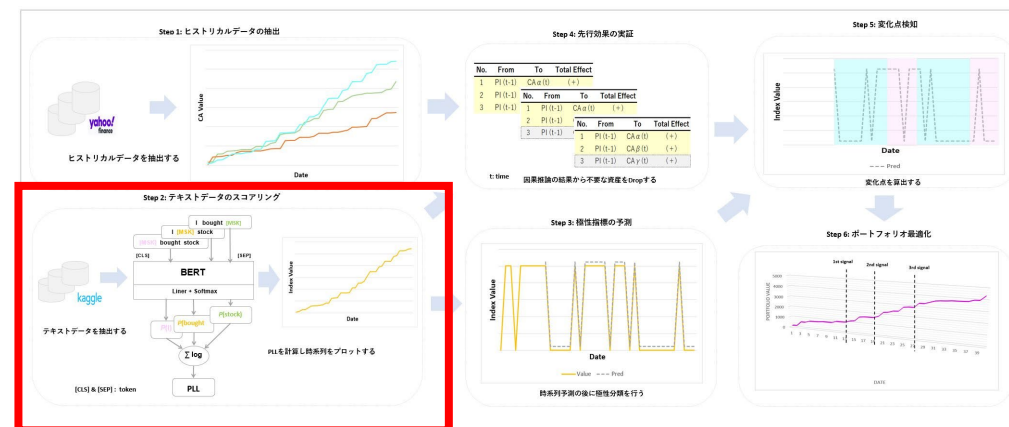
$\Theta$  はモデルパラメータ

$\delta_\eta$  はトークン

$\Delta_{\setminus \eta} = (\delta_1, \delta_2, \dots, \delta_{|\Delta|})$  は過去と現在のトークン

$P_{\text{MLM}}(\cdot)$  は各文章のトークンの確率

$$\mathcal{PLL}(\Delta) := \sum_{t=1}^{|\Delta|} \log_2 P_{\text{MLM}}(\delta_\eta | \Delta_{\setminus \eta}; \Theta)$$



# 11. 提案手法 (2/8)

-極性指標の予測-

- informerとは、Long sequence time-series forecasting (LSTF)に対応したTransformerベースの機械学習モデルです。「ProbSparse」というSelf-Attention手法で特徴抽出性能を維持しつつ計算量を押さえました。

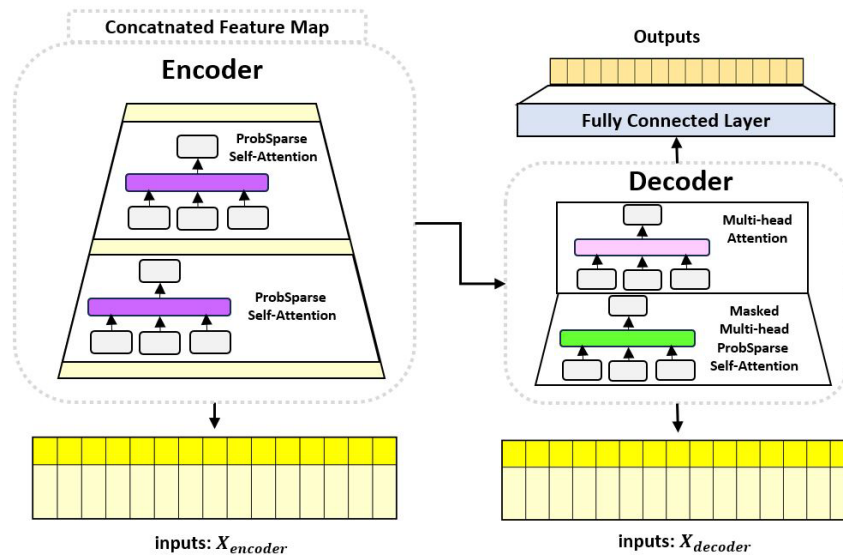
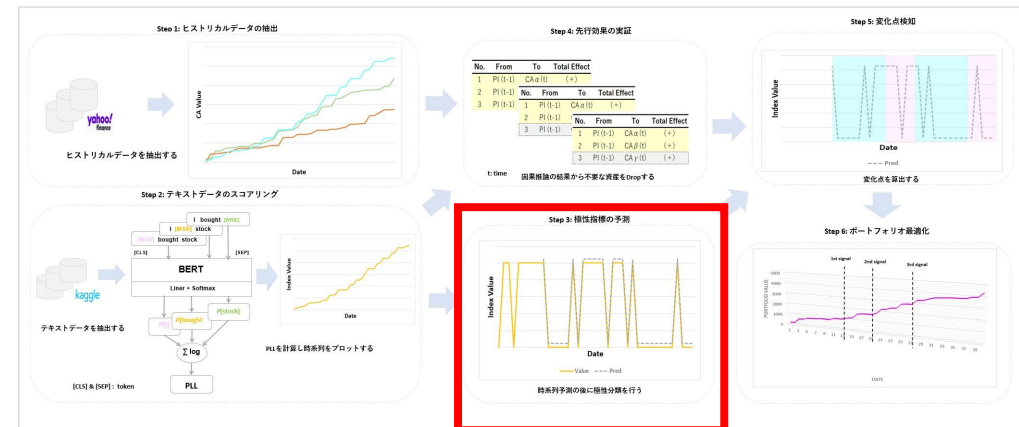


図2 informerのアーキテクチャ



# 11. 提案手法 (3/8)

-極性指標の予測-

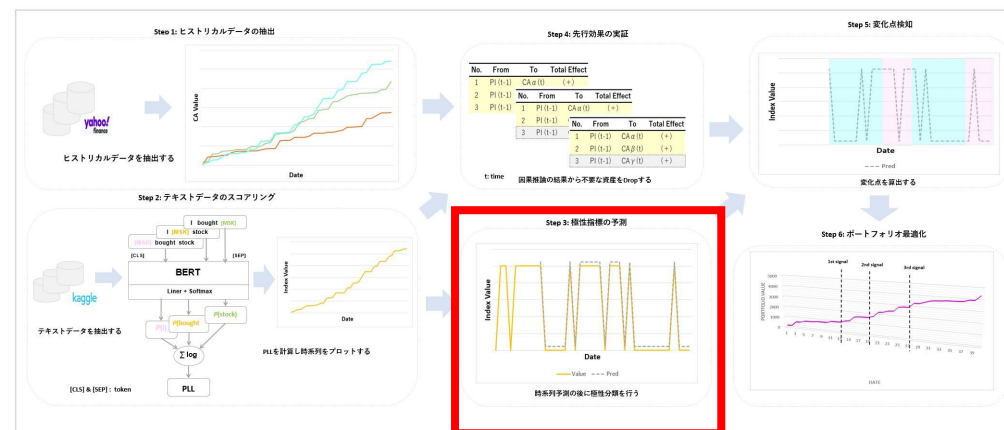
- informerでアウトプットされたスコア値に対し、テキストをPLLで1文ずつスコアリングします。表は極性分類の方法を示しています。

表1 極性分類の方法

分類条件	対応する値
3rd 四分位数 > PLL	1, {positive}
1st 四分位数 ≤ PLL ≤ 3rd 四分位数	0, {neutral}
1st 四分位数 < PLL	-1, {negative}

25% 第一四分位数  
50% 第二四分位数  
75% 第三四分位数

四分位範囲  
= 第三四分位数 - 第一四分位数



# 11. 提案手法 (4/8)

-先行効果の実証-

$t$  は時間  
 $\mathcal{X}(t)$  は変数ベクトル  
 $\tau$  は遅延時間  
 $T$  は満期日  
 $\mathcal{B}_\tau$  は係数行列  
 $\mathcal{X}(t - \tau)$  は係数行列  
 $\mathcal{E}(t)$  は外乱項

- ここでは、VAR-LiNGAMを用いて先行性を実証しました。VAR-LiNGAMは Hyvärinen et al. (2010) が提案した統計的因果推論モデルです。

$$\mathcal{X}(t) = \sum_{\tau=0}^T \mathcal{B}_\tau \mathcal{X}(t - \tau) + \mathcal{E}(t)$$

手順1

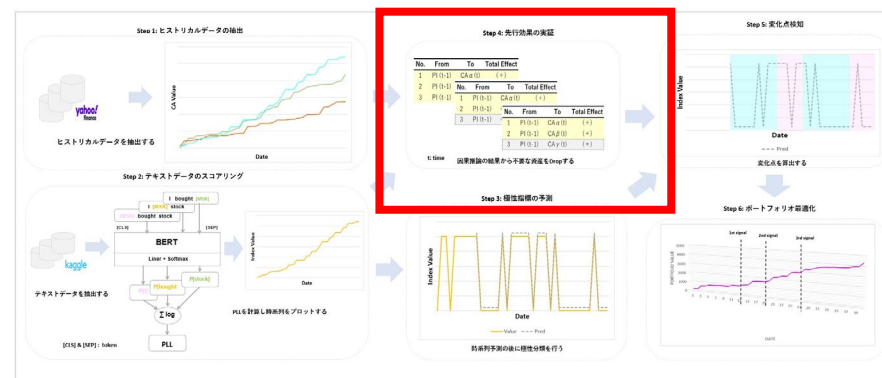
•ラグタイムから現在までの変数間の因果関係に対して、VARモデルを適用します。

手順2

•現在時刻の変数間の因果関係について、上記のVARモデルの残差を用いてLiNGAM推論を行います。

- 本章における「先行性がある」とは、「テキストから生成した極性指標→金融時系列」の方向に因果関係があり、かつ因果推論の総合効果のスコアがプラス (+) であることを指します。

この場合、先行する時間スケールは1ポイント先（直近の未来）である。



# 11. 提案手法 (5/8)

-変化点検知-

- ChangeFinderは、AutoRegressive (AR) モデルの学習に、オンライン学習と忘却機能を追加したSequentially Discounting AutoRegressive (SDAR) モデルを利用した変化点検知アルゴリズムです。

手順1

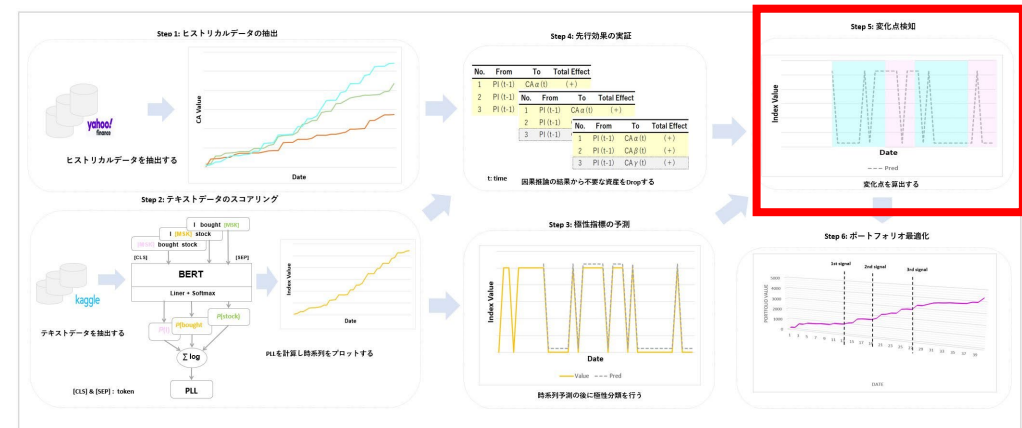
- 第一段階学習として、入力されたデータから ARモデルを定義し、SDARモデルでパラメータの学習を行い、外れ値スコアを算出します。

手順2

- 求めた外れ値スコアに対し、期間ごとに平滑化処理を行います。

手順3

- 平滑化後、再度SDARアルゴリズムにより算出したスコアを変化点スコアとします。

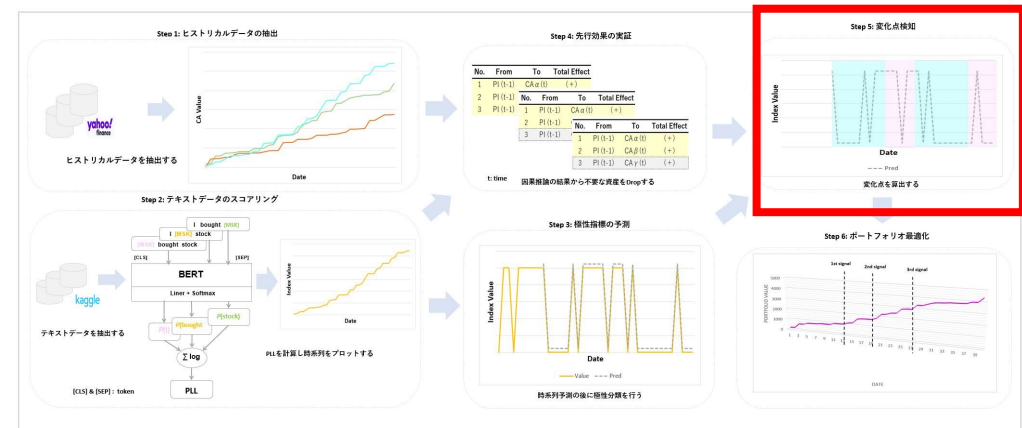




# 11. 提案手法 (6/8)

-変化点検知-

- 本研究における変化点の判断基準は以下の通りです。
  - 変化点スコアの後半1/2 を変化点検知の対象とします。
  - 変化点スコアの閾値は20 とします。
  - 変化点スコアがなだらかな山状の分布を持つ場合、山頂の値を変化点とみなします。



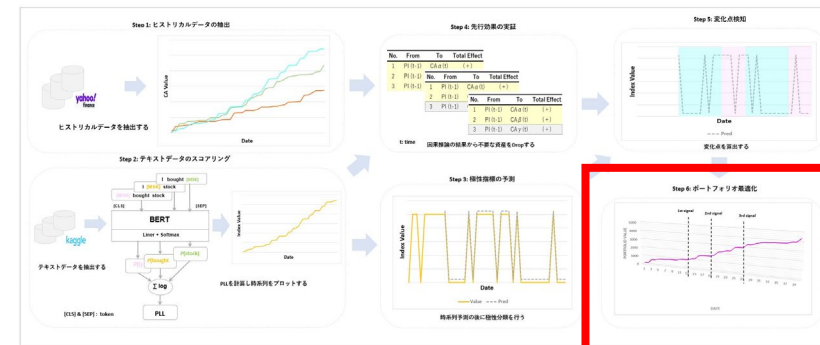
# 11. 提案手法 (7/8)

-ポートフォリオ最適化-

- リスク・パリティ・ポートフォリオは、すべての資産のリスク寄与度が等しいポートフォリオです。
- Relative Risk Contribution (RRC) は、ポートフォリオ全体のリスクに対する対象資産のリスク寄与度の比率として定義されます。

$$RRC_i = \frac{w_i(\sum w)_i}{w^T \sum w} \quad w \text{ は} n \text{次元のポートフォリオ}$$

- このRRCが各資産で等しくなるもの ( $= 1/n$ ) がリスク・パリティ・ポートフォリオとなります。



# 11. 提案手法 (8/8)

-ポートフォリオ最適化-

- 最適化問題は以下のように定式化されます。

$$\min_{w>0} f(w) = w^T \Sigma w - \sum_{i=1}^n \ln w_i$$

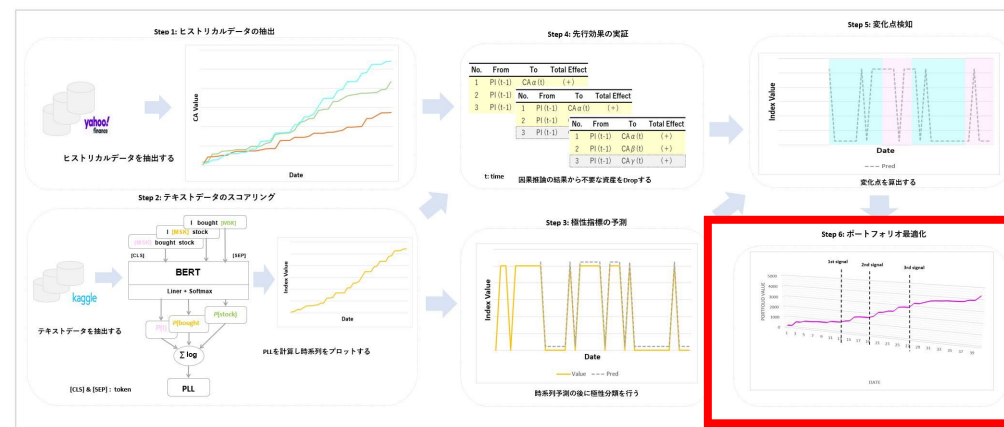
- 最適化問題の最適化条件は以下の通りです。

$$\nabla f(x) = 2\Sigma w - w^{-1} = 0$$

$\ln w_i$  は対数バリア関数

## 本研究の新規性 -因果推論とポートフォリオ最適化の融合-

- ① スタートから最初のセンチメントシグナルに到達するまでのアルゴリズムは、通常のリスク・パリティ・ポートフォリオとなる。
- ② 最適化の際に因果推論の総合効果がマイナスの銘柄について、リバランスのタイミングでドロップする。



# 12. 使用するデータ

**【選択理由】**

- ・ 出版目的が投資と比較的關係が薄いソースを選択し、傾向を判断したいため
- ・ 日次で並んでおり、ある程度分量が多いデータを選択したかったため

Type	Name	Details
News	AKBR	Amazon Kindle Book Review
	HFN	Historical Financial News Titles
	ONI	Onion Article Titles
CA	LTC	Litecoin
	NMC	Namecoin
	PPC	Peercoin

**【選択理由】**

- ・ 広く普及しているため
- ・ 値動きが異なり分散投資しやすいため

表2 データの概要

# 13. 実験における前提条件

- 定期的なリバランスは30日ごとに行われます。
- CA-SSAAM は、定期リバランスとセンチメントシグナルを組み合わせた戦略です。

# 14. 使用するアルゴリズム

手法名	概要
CA-SSAAM (proposed)	本研究で提案する方法です。
BM (RPP)	RPPモデルです。これは、CA-SSAAMから因果推論と変化点検知の影響を取り除いたベンチマークモデルを表現しています。
MV	平均-分散モデルに基づいています。
CVaR	Conditional Value at Risk (CVaR)に基づいています。CVaRはある確率水準におけるポートフォリオの平均的な損失を表すリスク指標です。最適化の条件として期待損失を最小化する必要があります。
EVaR	Entropic Value at Risk (EVaR)に基づいています。EVaRはVaRとCVaRの上限を示すリスク指標です。
CDaR	Conditional Drawdown at Risk (CDaR)に基づいています。CDaRはドローダウンに基づくリスク指標です。最適化の枠組みはCVaR最適化と同様です。

# 15. 評価指標

指標名	概要
Total Return (TR)	ある金融資産への投資から一定期間内に得られたリターンの合計を指します。
Annualized Return (AR)	1年間に発生した投資収益率です。
Annualized Volatility (AV)	1年間に発生した金融資産の価格変動率を表します。
Maximum Drawdown (MDD)	最大資産からの下落率を指します。
Sharpe Ratio (SR)	投資対象が取ったリスクに見合ったリターンを上げたかどうかを測る指標です。
Calmar Ratio (CR)	高リターンが高リスクと関連しているかどうかを示す指標です。
Tracking Error (TE)	ポートフォリオ・リターンとベンチマーク・リターンの差を表します。
Information Ratio (IR)	TEで考慮されるリスクに対して、どれだけ多くのリターンが得られるかを表します。

# 16. 結果と考察

各銘柄と極性指標の因果推論の結果をポートフォリオのリバランスの判断基準として採用することで、統計的因果推論が資産運用において有用であることを示しました。

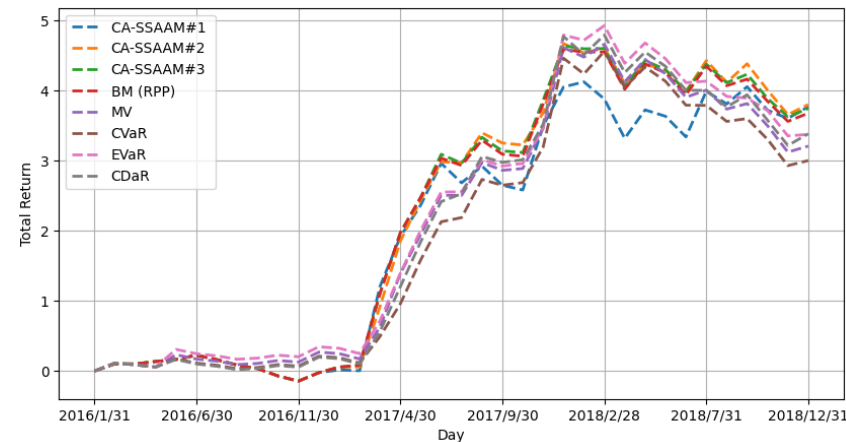
提案手法の全体的な傾向は、下落率が低く、比較的安定的な戦略だということがわかりました。

時系列予測モデルを用いることで、即時性を持たないテキストデータがポートフォリオ最適化のシグナルとなりえることを示しました。

暗号資産ポートフォリオの運用に限っては、株価の情報よりもニュース記事のほうが有用です。

図3 Total Return

SRを見ると、#1-#3がすべて万能ではなく、選択するテキストソースによって結果がばらつくことがあることがわかりました。



Type	Name	Parameter	AR [%]	AV [%]	MDD [%]	SR	CR	TE	IR
Proposed	CA-SSAAM #1	text = akbr	22.60	117.16	92.48	0.19	0.24	46.22	48.89
	CA-SSAAM #2	text = hfn	46.21	100.99	<b>90.79</b>	0.46	0.51	33.43	138.22
	CA-SSAAM #3	text = oni	<b>51.48</b>	95.32	91.25	<b>0.54</b>	<b>0.56</b>	22.16	<b>232.32</b>
Benchmark Portfolio Theory	BM (RPP)	weight_bounds = (0.01, 1)	47.09	96.95	91.22	0.49	0.52	-	-
	MV	significance_level = 0.05	39.73	<b>91.19</b>	92.67	0.44	0.43	48.18	82.47
	CVaR	significance_level = 0.05	31.78	93.01	93.16	0.34	0.34	<b>63.25</b>	50.25
	EVaR	significance_level = 0.05	43.92	92.36	92.54	0.48	0.47	57.95	75.80
	CDaR	significance_level = 0.05	39.42	96.49	93.53	0.41	0.42	60.00	65.69

表3 バックテスト

ポートフォリオ理論と比較した場合、提案手法はローリスク・ハイリターンとなります。



# 17. 本研究の目的は実現できたのか？

オルタナティブデータから算出した変化点が、リバランスのシグナルとして有効であると示すこと。

- レジーム・スイッチを考慮した3資産以上のポートフォリオ最適化戦略を構築することに成功しました。

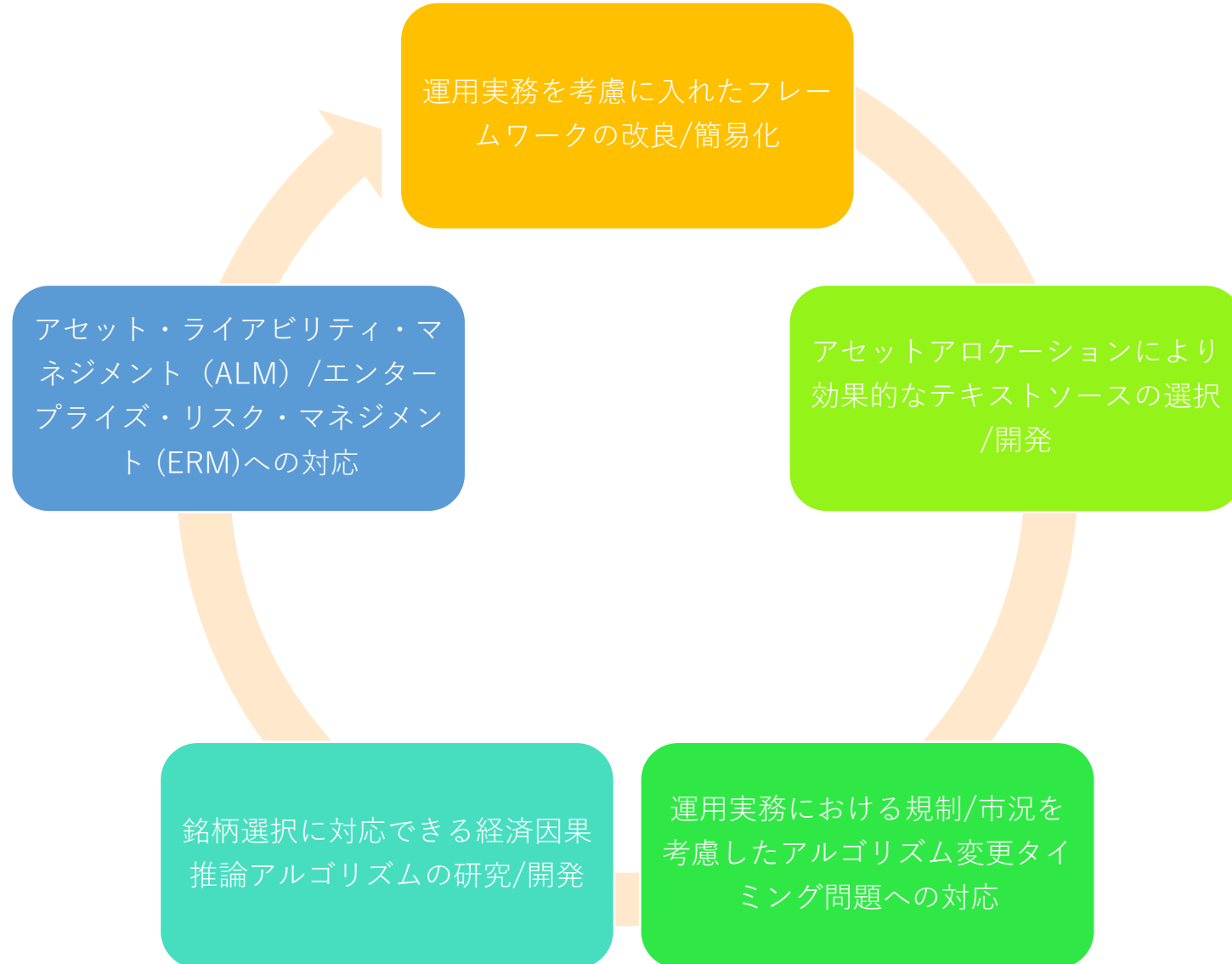
多期間において、下方リスクを動的に最小化するフレームワークになっていること。

- 構成されたポートフォリオに対して8つの評価指標を導入し、多期間において高リスクヘッジ効果を維持できることを示しました。

因果推論の結果が、アセットアロケーションに対して一定の効果があることを示すこと。

- 因果推論のスコアが銘柄選択に有効であることを示しました。

# 18. 今後の展望



ご清聴ありがとうございました