Vol.2020-CSEC-91 No.15
Vol.2020-SPT-40 No.15
Vol.2020-EIP-90 No.15
2020/11/25

クラウドソーシングを用いた資産運用手法開発におけるセキュリティリスクの考察

新倉 広子[1,a)]　今村 光良[1]　面 和成[2]

概要：資産運用業務における運用戦略の開発は，従来の理論モデルをベースとする方法からデータ駆動型の方法に注目が集まっている．一方で，市場分析用途のデータは従来の財務データだけでなく非財務のオルタナティブデータなど供給過多となっており，データ解析需要に応える人的リソースの確保が課題となっている．この課題を解消する解決策として，インターネットを通して不特定多数の人に作業を依頼するクラウドソーシングを活用したアプローチが注目されている．しかしながら，クラウドソーシングには，成果物の品質管理や媒体の設計だけでなく，各要素に絡むセキュリティリスクに配慮する必要がある．そこで本稿では資産運用手法にクラウドソーシングを用いるフレームワークについて整理し，サービス運用におけるセキュリティリスクについて検討する．

キーワード：クラウドソーシング，データ保護

1. はじめに

資産運用ビジネスにおける投資戦略の開発は，Markowitz の報告したポートフォリオ構築に関する理論 [1] を出発点に，理論モデルベースのアプローチから体系化される．平均分散アプローチは現代ポートフォリオ理論の仮定の１つであり，投資家の意思決定が資産のリターンの平均と分散に基づいて行われるという仮定である．このアプローチに基づきポートフォリオが組まれた場合の，金融資産の価格形成を説明する代表的な理論モデルとして CAPM(Capital Asset Pricing Model) が知られている [2], [3], [4]．しかしながら，平均分散アプローチに基づくモデルは期待値 (リターンの平均) の推計が難しい点などの課題があり，標準偏差などのリスク尺度に基づいたアプローチ [5], [6] が提案されている．

近年においては，リアルタイム性の改善による情報の非対称性に基づいた投資機会の獲得に注目があり，ニュースや Web 情報のテキストマイニング，衛星画像の解析など非財務データに分類されるオルタナティブデータを活用した投資戦略の有効性が報告されている [7]．さらに，データ特徴量に基づく機械学習の成功 [8] は，データ活用が増加した投資戦略の需要と親和性が高く，データ駆動型のアプローチとして研究されている [9]．

一方で，データ駆動型の投資戦略開発では，データとモデルの組み合わせにより投資機会となる戦略を広範囲に渡り探索することが必要となる．金融分野に関連する結果の判断には，データから適切な情報を引き出すための金融経済に関する専門的な知識が不可欠であり，モデルを開発する人材を制限する要因である．そのため，戦略開発における人員不足による律速の改善が課題である．

一つの解決策として注目されているのが，クラウドソーシングを活用したタスクの分散化によるアプローチである [10]．このアプローチは，前処理として分析データを専門的な知識が不要な形式に整理することで，データあたりの分析者を分散化する．この分析者による探査範囲の拡大により律速の改善を図っている．

しかしながら，クラウドソーシングベースの開発には，不特定多数の分析者に依頼者の内部データを配布する必要があり，また，分析者と依頼者を仲介する媒体や分析に対する報酬の受け渡しなどの安全な実施には，セキュリティリスクの検討が必要となる．

そこで本稿では，資産運用手法開発におけるクラウドソーシングサービスのプロセスに着目し，代表的なサービスのフレームワークを整理することで，サービス運用におけるセキュリティリスクを検討する．

本稿の構成は以下の通りである．はじめに 2 章では，ク

[1] 野村アセットマネジメント株式会社
Nomura Asset Management Co., Ltd.
[2] 筑波大学システム情報系
Faculty of Engineering, University of Tsukuba
[a)] hr-niikura@nomura-am.co.jp

Vol.2020-CSEC-91 No.15
Vol.2020-SPT-40 No.15
Vol.2020-EIP-90 No.15
2020/11/25

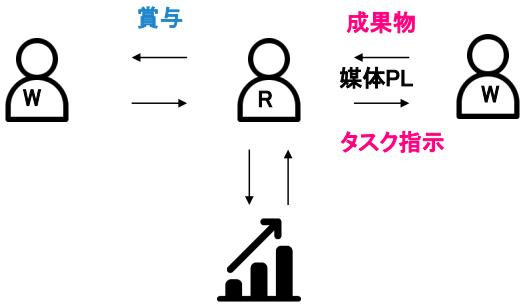

図 1 クラウドソーシングサービスの概念図

ラウドソーシングサービス上の処理やセキュリティリスクの整理を行なう．3 章では投資運用手法開発に関わる処理の整理と課題の切り分けを行い，クラウドソーシングサービスから享受できるメリットを議論する．4 章では，クラウドソーシングによる投資運用手法開発に関連するセキュリティリスクを議論し，5 章でまとめと展望を述べる．

2. クラウドソーシング処理とリスク

クラウドソーシング (Crowdsourcing) サービスは，インターネットを介して不特定多数の作業者にタスクを分散する仕組みを指す．インターネットによる広域への分散が可能となったことで，クラウドソーシングサービスの活用は，目的の結果を得るまでの単純かつ繰り返しとなるプロセスを細分化し，大規模な分散効果により，これまで不可能だった規模での労働力の調達が期待されるメリットを持つ [11]．クラウドソーシングサービスは，情報収集や課題解決，アイデア生成など，様々な目的で普及してきており，企業の労働モデルや雇用形態にも影響を与えつつある．学術的にも，コンピューターと人間の両方を適切に組み合わせて問題解決を計る，ヒューマンコンピューテーションの分野において，重要な労働力の供給源として議論されている [12]．

クラウドソーシングサービスが様々な事業領域で普及する一方で，従来の労働市場でみられなかった，クラウドソーシング特有の問題が顕在化してきている．例えば，クラウドソーシングサービスの依頼者は従業者の成果物を報酬として受け取るが，悪質な従業者が多く存在すると，集約シグナルの品質低下が生じる可能性が考えられる．そこで本章では，クラウドソーシングサービスにおける一般的な処理プロセスと，既知のセキュリティリスクを俯瞰する．とりわけ処理プロセスに関わる構成要素レベルでのセキュリティリスクの切り分けを行なう．

2.1 クラウドソーシングサービスの処理プロセス

クラウドソーシングサービスに関わる処理プロセスは，タスク割り当て，タスク指示，成果物の提出に分類される．図 1 に処理の流れを示す．各プロセスの概要は以下の通りである．

表 1 クラウドソーシングサービスにおけるタスク処理プロセス

処理プロセス	処理母体	処理母体の種類	アウトプット
タスク割り当て	媒体 PL	機械	割り当て処理
タスク指示	依頼者	人間	成果物
成果物提出	従業者	特定なし	特定なし
		機械	品質管理された成果物

- タスク割り当てプロセス
  タスク割り当てプロセスでは，依頼者が個々の従業者に依頼するタスクを決める処理が行なわれる．従業者に割り当てられるタスクは，従業者間で同一の場合と，個々のスキル等にあわせて差別化される場合がある．後者の場合は従業者とタスクの特徴を収集し，業務割り当てをきめる処理が行なわれる．
- タスク指示のプロセス
  タスク指示のプロセスでは，依頼者から従業者にタスクを指示する処理が行なわれる．依頼者はデータや解析リソースを従業者に共有し，従業者は目的のタスクに取り組んで成果物を作成する．
- 成果物提出のプロセス
  従業者が依頼されたタスクを遂行すると，得られた成果物の提出が行なわれる．これを元に依頼者は最終成果物を得る．タスクの見返りとして，依頼者から従業者に賞与が付与される．

上述のクラウドソーシングの各プロセスに関わる処理実体の役割を表 1 に示す．クラウドソーシングサービスに関わる実体は 3 つで，目的のタスクを依頼する依頼者 (R)，タスクに取り組む従業者 (W)，そして依頼者と従業者の意思疎通等を媒介する媒体プラットフォーム (PL) である．処理プロセスや成果物の形式の定義有無により，各プロセスを担当する実体や処理目的が異なる [13]．

2.2 クラウドソーシングサービスの課題

以下ではクラウドソーシングサービスに伴うセキュリティリスクを俯瞰する．本稿では先行研究 [14] を参考に，CIA トライアドに基づき，機密性，完全性，可用性の 3 つの情報セキュリティの観点から議論する．なお，これ以外のセキュリティリスクとしては，依頼者が提示するタスク自体が倫理的に正しくないリスクが挙げられ，依頼者が悪用に加担する例である．実例として，顧客の意思決定に影響を与えるレビューサイトにおいて，依頼者が対象の評判を操作するために，偽のレビューライティングを斡旋する手段として利用された例が報告されている [15], [16]．

各フレームワークにおいて外部通信と内部通信の要因を分けて，対象となるセキュリティリスクを整理すると表 2 の通りとなる．詳細は下記の通りである．

- 機密性 (Confidentiality)
  機密性とは，認可されていない実体やプロセスが機密情報を利用できない状況にする特性である．機密性に

Vol.2020-CSEC-91 No.15
Vol.2020-SPT-40 No.15
Vol.2020-EIP-90 No.15
2020/11/25

表 2 クラウドソーシングサービスにおけるセキュリティリスク

CIA-triad	通信	攻撃例	セキュリティリスク
機密性	外部通信	ハッキング	Web 経由の個人情報抽出
	内部通信	データ/成果物の抽出	分析データ内の個人情報抽出
完全性	外部通信	Web サービスの改ざん	誤タスク/データへの誘導
	内部通信	Spammer	成果物の品質低下
可用性	外部通信	DDoS	サービスを利用不能状態にする

関わるリスクには，権限のないデータへのアクセスが該当する．まず外部通信によるセキュリティリスクとして，プラットフォームに対するハッキングが挙げられる．例えば，第三者がプラットフォームの脆弱性や従業者の登録情報を利用して内部情報を抜き取るリスクなどが該当する．

一方，内部通信によるセキュリティリスクとしては，作業者に関するデータと成果物が漏洩するリスクが挙げられる．例えば，悪意を持った依頼者が従業者のパスワードなどの個人の秘密情報を収集するリスクが考えられる [17]．また，モバイルクラウドソーシングのように従業者の位置情報がタスク割り当てプロセスに利用されるサービスでは，従業者の居住地のプライバシー等が脅かされる可能性が議論されている [18]．

- 完全性 (Integrity)
  完全性とはサービスの正確さや完全さを保護する特性であり，サービスの改ざんなどがセキュリティリスクに該当する．外部通信におけるリスクとしては，従業者が利用する Web サイトが適切に利用できなくなる場合が考えられる．例えばプラットフォームを改竄して誤ったタスク情報を配信することで，従業者が正しいタスクを行なえない可能性や誤ったデータに誘導して従業者が適切なデータにアクセスできない場合が挙げられる．
  一方，内部通信におけるリスクとしては，従業者が提出する成果物が従業者により意図的に内容が変更される場合が挙げられる．代表的なリスク要因として Spammer と呼ばれる従業者の存在がある．Spammer は根拠のない結果を意図的に提出することで，タスクに対して金銭的な見返りを狙う役割を行なう．Spammer による攻撃例としては，インターネットボット (Bot) を使って無秩序な結果を自答的に生成する例などが知られている [11]．また，悪意ある依頼者が特定の従業者の成果物を意図的に変更することを考えられる．従業者の意図せぬ結果が反映されることで不利益を被る．
- 可用性 (Availability)
  可用性とは，認可されている実体が必要なときにサービスを利用できる特性である．セキュリティリスクとしては，クラウドソーシングのシステムが DDoS 攻撃をうける場合が挙げられる．この場合，従業者が依頼者が提供するサービスにアクセスできなくなるため，サービスが成り立たなくなる可能性が懸念される．

3. 資産運用手法におけるクラウドソーシングの役割可能性

データ駆動型の投資手法開発が普及するにつれて，資産運用ビジネスにおけるクラウドソーシングの活用が議論されている [10]．データ駆動型の資産運用手法開発では，過去データに対して良好なパフォーマンスが出せるようなモデル構築がゴールとなる．しかしながら，旧来のデータ駆動型の運用手法開発では，人数と開発期間に制限があるため，分析者の過学習バイアスの混入や高価なファイナンスデータから生み出される戦略数の少なさなどの課題が提起されていた．クラウドソーシングはこれらの課題とも相性がよく，新たな解決策としての有効性が議論されている．その一方，依頼者のタスク依頼から最終成果物の獲得までのプロセスには，クラウドソーシング特有のセキュリティリスクが混入する懸念がある．

本章ではまず，データ駆動型の資産運用手法開発にかかわる諸プロセスを整理する．そしてセキュリティリスク検証のためのフレームワークを提起し，フレームワークにおける課題を述べる．

3.1 資産運用手法開発プロセス

本節では，データ駆動型の資産運用手法開発に関わる一般的な処理プロセスを提示する．資産運用手法開発のプロセスは，先行研究で提案されている構成 [10] を参考に以下の P1 から P6 の 6 段階に切り分けを行なった．これらの開発プロセスに関わるセキュリティリスクの対応を表 3 で提示する．

P1 データ前処理
金融データの分析において扱うデータは，企業の財務諸表に関する時系列データ，マーケット指標，マクロ経済変数などと多岐にわたる．そのため，これらのデータを構造化するには，企業財務データ，経済，および金融システムに関する専門知識が必要となる．そこで，依頼者は，データのモデル化とパターンの特定に注力するために，金融データ固有の煩雑さをデータから取り除く処理を行なう．データの前処理としては，分析データの選択や整理，および正規化や標準化によるバイアス調整が行なわれる．従業者もまた，後述のモデル開発に際して，モデル作成の前処理として入力データからさらに中間データを作成することが考えられる．

P2 データ配布
金融関連データの場合，多くはデータベンダーが著作権を保持しており，契約範囲外の分析者に対してはそのままの形での配布が許可されない．旧来の社内開発では，社内の分析担当者が生データに直接アクセスできるが，分析タスクをアウトソースする場合には，

表 3　データ駆動型の資産運用手法開発に関わるプロセスの一覧。

番号	プロセス名	関係者	対応するクラウドソーシング処理	課題	関連セキュリティリスク
P1	データ前処理	従業者/依頼者	タスク割り当て	専門知識	分析データ漏洩
P2	データ配布	依頼者/媒体 PL	タスク指示，インスタンス提供	著作権，プライバシー保護	分析データ漏洩
P3	モデル開発	従業者	成果物提出	計算資源，モデルのプライバシー	成果物漏洩，集約シグナル品質
P4	シグナル集約	従業者/依頼者	成果物提出	品質管理	集約シグナル品質
P5	ポートフォリオ構築	従業者/依頼者	成果物提出	品質管理	集約シグナル品質
P6	運用/ステーキング	従業者/依頼者	成果物提出	品質管理，利害関係	集約シグナル品質，報酬支払い

データをそのまま共有するのではなく，難読化の処理が必要となる．ほかにも，依頼者と従業者の間の利益競合を抑止したり，従業者のバイアス混入を防ぐ手段として，データの難読化処理は有効と考えられる．

P3 モデル開発
分析データが準備できた後，分析者はデータに基づき予測モデルの作成を行なう．データ駆動の資産運用手法開発において，データに対するモデルの過学習が課題の一つである．そこで分析者は通常，訓練用データ，検証用データ，およびテスト用データを活用して，過学習に配慮したモデル構築を行なう．

P4 シグナル集約
個々のモデルを構築する際の予測パフォーマンスは，テストデータを用いて評価が行なわれる．実際に運用されるモデルのパフォーマンスは，最新のデータからの予測値で評価されるため，特定の条件下でのモデルの評価が重要となる．複数の予測モデルがある場合は，特定の条件下でのパフォーマンスのクラスタリング解析による傾向分析に基づいて，都度予測値の重みづけを更新するなどの工夫が行なわれる．

P5 ポートフォリオ構築
モデルの予測値では，必ずしも投資資産の損益予測に限定されない．例えば，ランキングやスコアの形式で相対的に投資資産を評価することで，ポートフォリオを構成する資産の重みづけを導き出す手法もある [19]．

P6 運用/ステイキング
シグナル集約時と同様，構築したポートフォリオに対しても逐次パフォーマンスを評価しながら運用される．分析者は現在の状況下での予測値を適宜提出することで，予測成果に対して報酬を受ける．分析者と依頼者の利害関係を一致させるため，報酬を延期して継続的な運用を優先させる場合もみられる．[*1]

3.2　資産運用クラウドソーシングサービス

本節では，資産運用手法開発にクラウドソーシングサービスを応用したサービスの形態を述べる．前述の通り，旧来のデータ駆動型の資産運用開発では，分析者の過学習バイアスの混入などの課題が議論されている．これらの諸課題に対応するべく，クラウドソーシングサービスの導入に加えて，依頼者と従業者の通信を媒介する (IT) 媒体プラットフォームの形態やタスクの工夫方法などが議論されている．

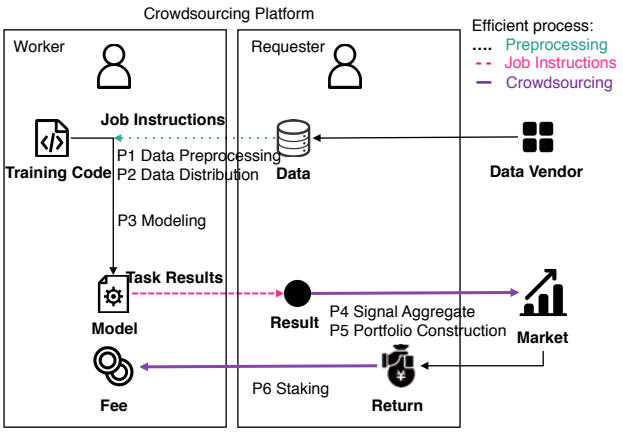

図 2　クラウドソーシングサービスを利用した資産運用手法開発のフレームワーク

具体的なクラウドソーシングサービスの形態としては，従業者への報酬の形態によって，バックテストプラットフォーム (疑似労働市場型) や投資トーナメント式 (トーナメント型) などの形態が議論されている [10]．本稿では処理フレームワークとして，データ前処理，タスク割り当ての工夫，クラウドソーシングシステムの 3 プロセスに切り分けを行なった．各プロセスの詳細は以下の通りである．

- データ前処理 (Preprocessing)
  金融関連データ，およびオルタナティブ分析で扱う非財務データは通常，データベンダーを通して入手する．旧来のデータ駆動型開発では，分析者がこれらの生データに直接アクセスし，データ前処理からモデル構築までの処理を担当していた．しかしながら，金融分析を行なうにあたり，分析者は金融の専門知識を利用して適切に生データを処理する必要がある．この解決策として，依頼者側でデータ提供前に難読化などの処理を行ったのち，従業者にデータを提供する形態が考えられる．この工夫により，従業者はデータサイエンスの知識のみでタスクを取り組むことができるようになり，分析者の金融知識レベルが集約シグナルの品質へ与える影響を依頼者側で管理できることが期待される．
- タスク割り当ての工夫 (Job Instructions)
  旧来のデータ駆動型開発では通例，分析者は取り組むべきタスクの詳細は指定されず，タスクに自由度が存

[*1] https://numer.ai

在する．このとき分析者はデータから独自の取引用のシグナルを生成して，ポートフォリオの作成とその運用に関わるリスク管理まで携わる．一方タスクの詳細を指定することにより，依頼者側で従業者のバイアス混入を抑えることが可能となる．例えば，分析で扱う訓練データやテストデータを固定したり，成果物の形態も予測スコアも可能となる．

- クラウドソーシングシステム (Crowdsourcing)
  クラウドソーシングサービスの特徴の 1 つに，分析タスクをアウトソースする形式が挙げられる．タスク依頼者は，従業者の成果物から得られるシグナルを重みづけ等の処理で集約し，運用するポートフォリオを構築する．
  クラウドソーシングシステムのもう 1 つのポイントが，従業者に対する報酬の形式である．クラウドソーシングの媒体プラットフォームの形態としては，報酬の還元方法に則して疑似労働市場型，トーナメント型，オープンコラボレーション式が提起されている [20], [21] ．資産運用手法開発の場合は通例，従業者が報酬を期待するため，疑似労働市場型とトーナメント型の 2 つの報酬方式が対象となる．

以上の 3 プロセスで構成された資産運用フレームワークを図 3 に示す．前節の資産運用手法プロセスと本フレームワークの手法を対応づけると，データ前処理の工夫がモデル構築プロセスに，タスク割り当ての工夫が従業者の成果物形態に，クラウドソーシングシステムがポートフォリオ構築および報酬プロセスに影響を与えている示唆が得られる．

3.3 クラウドソーシング評価指標による特徴づけ

本節では，図 2 で提示された資産運用手法開発システムについて，旧来の社内開発スキームやクラウドソーシングの性能評価指標，資産運用手法開発での課題指標と比較を行なった結果を表 4 に示す．クラウドソーシングの性能評価指標は [22] を参考に，コスト，匿名性，クラウドのスケール，IT 構造，実装にかかる時間，クラウドの信頼度の 6 つの指標を採用した．

はじめに資産運用手法開発の課題指標に着目してフレームワークを俯瞰したところ，クラウドソーシングベースの開発システム本体が有効な指標はデータ活用度と調査機会の数 (breadth) であった．クラウドソーシングサービスを利用する場合，プラットフォーム上で分析データや計算資源を複数の従業者に共有するため，単位データあたりの戦略数の数が増加し，生産性の向上につながることが期待されるためである．また，前処理が有効なのは，従業者の専門知識障壁，データへのアクセス障壁，データ攻撃リスクの 3 指標，タスク割り当てが有効なのは，バックテストでの過学習問題と各調査での戦略数 (depth) の 2 指標との

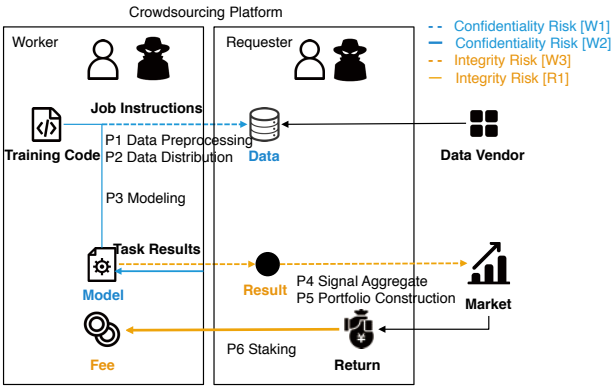

図 3 クラウドソーシングサービスを利用した資産運用手法開発のセキュリティリスク

示唆が得られた．これにより，前処理/タスク指示を活用することで，成果物の形態や従業者側の金融知識をコントロール可能であることが期待される．

次にクラウドソーシング評価指標と資産運用手法開発の課題の対応に着目したところ，クラウドのスケールと IT 構造が，資産運用手法開発の機能拡張に役立つとの示唆が得られた．また，セキュリティリスクに対しては，クラウド自体の IT 構造，匿名性，クラウドの信頼度の指標で評価できると考えられる．これらの指標は，クラウドの外部通信に関わるセキュリティリスクの評価に有効と考えられる．

4. セキュリティリスク

3 章では，データ駆動型の資産運用手法開発のプロセスの整理と，クラウドソーシングサービス導入により生じる旧来の資産運用手法との差分を議論した．本章では，前述した図 3 の資産運用手法開発のフレームワークにおけるセキュリティリスクを整理する．

以下の議論では 2 つの仮定を置く．まず第一に，依頼者と従業者の利害関係は必ずしも一致しない場合を許容する．これにより，従業者，依頼者ともに悪意を持っている可能性についても議論の対象とする．そして第二に，内部通信起因のセキュリティリスクに焦点を絞り議論を行なう．表 2 で議論したリスクのうち，Web サイトの改ざんやハッキングなど，外部通信起因のセキュリティリスクは一連のサービス機能を停止させる効果がある．本稿ではサービス稼働時のセキュリティリスクに焦点を当てるため，外部通信要因のセキュリティリスクは議論の対象外とする．

以上の前提に基づくと，従業者に悪意があるケースとして，分析データ漏洩リスク (W1)，成果物漏洩リスク (W2)，集約シグナルの品質リスク (W3) の 3 種類のセキュリティリスクが考えられる．一方，依頼者に悪意があるケースとして，成果物の改竄によるリスク (R1) が考えられる．それぞれのセキュリティリスクの状況，および関係した処理プロセスは以下の通りである．

W1 分析データ漏洩リスク (機密性)

表 4 投資手法開発における課題指標とクラウドソーシングサービスの有効性.

課題指標 [10]	クラウドソーシング評価指標 [10]	関係実体	社内開発	クラウドソーシング開発	前処理	タスク指示
分析者の経済知識	IT 構造	従業者	要	-	○	-
検証データ配布	匿名性/IT 構造	媒体 PL	生データ/チーム毎配布	-	○	-
データ管理	コスト/IT 構造	媒体 PL	チーム毎配布	-	○	-
バイアス混入	IT 構造	従業者	事後対策	-	○	○
データ活用度	IT 構造	従業者	低	○	-	-
成果物	匿名性/IT 構造	媒体 PL/依頼者	Trade/Portfolio & 予測値	-	-	予測値のみも可
IP 所有権	IT 構造	媒体 PL	企業	従業者	-	-
計算コスト負担	IT 構造	媒体 PL/依頼者	企業	依頼者 or プラットフォーム	-	-
調査機会の数 (breadth)	クラウドのスケール	媒体 PL/依頼者	低	○	-	◎
各調査での戦略数 (depth)	クラウドのスケール	媒体 PL/依頼者	低	-	-	○
単価あたりの従業者数	クラウドのスケール	媒体 PL	低	○	-	◎
セキュリティリスク	IT 構造/匿名性/クラウド信頼度	従業者/依頼者	内部通信リスク	-	○	○

分析データ漏洩リスクは，悪意のある従業者が分析で扱うデータの特徴を抽出する攻撃 (Membership inference 攻撃)[23] である．資産運用プラットフォームで扱うデータは通例データベンダーから購入したデータであり，悪意ある従業者からの攻撃の対象になりうる．本リスクの攻撃に関わるプロセスは P1，P2 である．通常当該プロセスでは，従業者は分析を行なうためにプラットフォーム上の分析データにアクセスする．一方データの特徴抽出を目的として，悪意を持った従業者がプラットフォーム上にある分析用データに対して特徴抽出攻撃を行なうリスクが考えられる．当該攻撃は機密情報の抽出にあたり，機密性に関わるセキュリティリスクに位置づけられる．

W2 成果物漏洩リスク (機密性)

クラウドサービスでは，悪意ある従業者による分析データの抽出攻撃に加えて，依頼者側から提供されるタスクや従業者の構築した成果物を抽出してモデルを推定する攻撃 (Model Inversion Attack) が懸念される [24]．本リスクの攻撃主体は従業者であり，攻撃に関わるプロセスは P3 である．従業者は当該プロセスにおいてモデル開発を行い，成果物をプラットフォームを介して提出する．プラットフォームでデータや開発アイデアを共有している場合，悪意のある従業者がプラットフォーム上で他の従業者の成果物にアクセスして情報を抽出するリスクが考えられる．当該リスクは機密情報の抽出にあたり，機密性に関わるセキュリティリスクに位置づけられる．

W3 集約シグナルの品質リスク (完全性)

集約シグナルの品質リスクとは，悪意ある従業者により提出する成果物の品質が意図的に下げられたり，従業者が提出する成果物の品質が悪意ある依頼者により意図的に変更されることで，成果物を集約して依頼者が得るシグナルの品質が低下するリスクである．悪意ある従業者が主体となって，攻撃に関わるプロセスは P3 である．具体的な攻撃例としては，モデル開発のプロセス (P3) において，従業者が意図的にランダムな値を提出するような攻撃を行なう場合 (Data poisoning)[25] である．

悪意ある依頼者が主体となる場合は，プロセス P4 のシグナル集約において，提出された成果物を意図的に改変することで，集約したシグナルの品質を低下させるように攻撃する．

当該攻撃は依頼者側が意図したシグナルを得るプロセスを阻害する，または従業者側の意図した成果物がシグナルに反映されない完全性に関わるセキュリティリスクに位置づけられる．

R1 成果物の改竄によるリスク (完全性)

クラウドソーシングサービスでは通常，従業者は提出した成果物に対して報酬を受け取る．資産運用サービスの報酬体系では，実際の市場における資産運用のパフォーマンスに応じた報酬が支払われる．そのため，パフォーマンスが悪化した場合には報酬が受け取れない場合が想定される．この状況を依頼者が悪用することで，実際にはパフォーマンスは良かったが，従業者の集約したタスクでは結果を得ることができなかったと，従業者に適切な報酬を支払わないリスクが考えられる．本リスクの攻撃主体は依頼者であり，攻撃に関わるプロセスは P6 である．当プロセスでは依頼者から従業者に報酬が支払われるが，市場の成果物から依頼者が収益を得ても，従業者に不当に低い報酬かまたは支払いがないリスクが考えられる．当該リスクは，依頼者が従業者の成果物のシグナルを意図的に改竄して正しいサービス運営を阻害するため，完全性に関わるセキュリティリスクに位置づけられる．

上述したセキュリティリスク，および資産運用プロセスとの対応をまとめを図 3 に示す．W1-W3 のセキュリティリスクはプラットフォームを介したプロセスに存在しており，データの扱い方や成果物の品質など，従業者が大きな裁量を持つフレームワークとなっていることが示唆される．各セキュリティリスクを軽減するためには，データの前処理やタスク指示のプロセスの工夫を通して，タスク依頼者側が対策を講じることが有効であると考えられる．

5. まとめと展望

本稿では，データ駆動型の資産運用手法開発のフレームワークを整理し，クラウドソーシングを用いたサービス運用に伴うセキュリティリスクを議論した．はじめにクラウドソーシングを用いたデータ駆動型の資産運用手法開発の一般的なフレームワークとして，データ前処理，タスク割り当ての工夫，クラウドソーシングシステムの 3 要素で特徴づけを行なった．次に提示したフレームワークについて，クラウドソーシングサービスの評価指標を用いて資産運用手法プロセスとの対応づけを行なったところ，クラウドのスケールと IT 構造の指標が，従来の資産運用手法開発における諸課題の解決や機能の拡張に有効であるとの示唆を得た．セキュリティリスクに対しても同指標で評価したところ，クラウド自体の IT 構造，匿名性，クラウドの信頼度の指標で評価できると考えた．これらの指標は，クラウドの外部通信に関わるセキュリティリスクの評価に有効と考えられる．さらに，本フレームワークのセキュリティリスクを機密性，完全性，可用性の観点から議論を行なった．内部通信プロセスにおけるリスクを考察した結果，分析データ漏洩リスクと成果物漏洩リスクが機密性と関わるリスク，集約シグナルの品質リスクと成果物の改竄によるリスクが完全性と関わるリスクに該当した．それぞれのセキュリティリスクと関係プロセスとの対応づけたところ，悪意のある従業者によるデータや成果物の抽出攻撃，および成果物の品質対策には，タスク指示の工夫とデータの前処理を依頼者側でコントロールすることが不可欠であることの示唆を得た．

本研究の展望としては，上記で議論した具体的なセキュリティリスクへの対策とその有効性を議論を発展させることが挙げられる．例えば，集約シグナルの品質リスクへの対策としては，依頼者側でのシグナル集約プロセスにおける Truth Inference 法などによる統計的手法を用いた品質管理の工夫が必要と考えられる．また，分析データや成果物の漏洩リスクには依頼者側での対策が不可欠であるが，データ前処理での難読化処理やタスク割り当てのテンプレート化などの工夫が必要と考えられる [26]．個々の対策の有効性の議論に向けて，各セキュリティリスクとその対策の有効性を定量的に測定し，依頼者側での適切な対策をみいだすことが有効であると考える．

参考文献

[1] Markowitz, H.: Portfolio Selection, *The Journal of Finance*, Vol. 7, No. 1, pp. 77–91 (1952).
[2] Fama, E. F.: Efficient capital markets: A review of theory and empirical work, *The journal of Finance*, Vol. 25, No. 2, pp. 383–417 (1970).
[3] Fama, E. F.: Efficient capital markets: II, *The journal of finance*, Vol. 46, No. 5, pp. 1575–1617 (1991).
[4] Fama, E. F. and MacBeth, J. D.: Risk, return, and equilibrium: Empirical tests, *Journal of political economy*, Vol. 81, No. 3, pp. 607–636 (1973).
[5] Clarke, R. G., De Silva, H. and Thorley, S.: Minimum-variance portfolios in the US equity market, *The journal of portfolio management*, Vol. 33, No. 1, pp. 10–24 (2006).
[6] Galane, L. C.: The risk parity approach to asset allocation, PhD Thesis, Stellenbosch: Stellenbosch University (2014).
[7] Bollen, J., Mao, H. and Zeng, X.: Twitter mood predicts the stock market, *Journal of computational science*, Vol. 2, No. 1, pp. 1–8 (2011).
[8] Krizhevsky, A., Sutskever, I. and Hinton, G. E.: Imagenet classification with deep convolutional neural networks, *Advances in neural information processing systems*, pp. 1097–1105 (2012).
[9] Ding, X., Zhang, Y., Liu, T. and Duan, J.: Deep learning for event-driven stock prediction, *Twenty-fourth international joint conference on artificial intelligence* (2015).
[10] de Prado, M. L. and Fabozzi, F. J.: Crowdsourced Investment Research Through Tournaments, *The Journal of Financial Data Science*, Vol. 2, No. 1, pp. 86–93 (2020).
[11] Howe, J.: The rise of crowdsourcing, *Wired magazine*, Vol. 14, No. 6, pp. 1–4 (2006).
[12] Law, E. and Ahn, L. v.: Human computation, *Synthesis lectures on artificial intelligence and machine learning*, Vol. 5, No. 3, pp. 1–121 (2011).
[13] Kajino, H., Arai, H. and Kashima, H.: Preserving worker privacy in crowdsourcing, *Data Mining and Knowledge Discovery*, Vol. 28, No. 5-6, pp. 1314–1335 (2014).
[14] Onuchowska, A. and de Vreede, G.-J.: Disruption and Deception in Crowdsourcing, *Crowdsourcing: Concepts, Methodologies, Tools, and Applications*, IGI Global, pp. 215–235 (2019).
[15] Lai, R. Y.: Structured methodology and design patterns for web services (2010). US Patent 7,831,693.
[16] Yu, B., Willis, M., Sun, P. and Wang, J.: Crowdsourcing participatory evaluation of medical pictograms using Amazon Mechanical Turk, *Journal of medical Internet research*, Vol. 15, No. 6, p. e108 (2013).
[17] Baba, Y., Kashima, H., Kinoshita, K., Yamaguchi, G. and Akiyoshi, Y.: Leveraging non-expert crowdsourcing workers for improper task detection in crowdsourcing marketplaces, *Expert Systems with Applications*, Vol. 41, No. 6, pp. 2678–2687 (2014).
[18] Yang, D., Xue, G., Fang, X. and Tang, J.: Incentive mechanisms for crowdsensing: Crowdsourcing with smartphones, *IEEE/ACM transactions on networking*, Vol. 24, No. 3, pp. 1732–1744 (2015).
[19] Nakagawa, K., Ito, T., Abe, M. and Izumi, K.: Deep recurrent factor model: interpretable non-linear and time-varying multi-factor Model, *arXiv preprint arXiv:1901.11493* (2019).
[20] Afuah, A. and Tucci, C. L.: Crowdsourcing as a solution to distant search, *Academy of Management review*, Vol. 37, No. 3, pp. 355–375 (2012).
[21] De Vreede, T., Nguyen, C., De Vreede, G.-J., Boughzala, I., Oh, O. and Reiter-Palmon, R.: A theoretical model of user engagement in crowdsourcing, *International conference on collaboration and technology*, Springer, pp. 94–109 (2013).
[22] Prpić, J., Shukla, P. P., Kietzmann, J. H. and McCarthy,

I. P.: How to work a crowd: Developing crowd capital through crowdsourcing, *Business Horizons*, Vol. 58, No. 1, pp. 77–85 (2015).

[23] Shokri, R., Stronati, M., Song, C. and Shmatikov, V.: Membership inference attacks against machine learning models, *2017 IEEE Symposium on Security and Privacy (SP)*, IEEE, pp. 3–18 (2017).

[24] Fredrikson, M., Jha, S. and Ristenpart, T.: Model inversion attacks that exploit confidence information and basic countermeasures, *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, pp. 1322–1333 (2015).

[25] Steinhardt, J., Koh, P. W. W. and Liang, P. S.: Certified defenses for data poisoning attacks, *Advances in neural information processing systems*, pp. 3517–3529 (2017).

[26] Tahmasebian, F., Xiong, L., Sotoodeh, M. and Sunderam, V.: Crowdsourcing Under Data Poisoning Attacks: A Comparative Study, *IFIP Annual Conference on Data and Applications Security and Privacy*, Springer, pp. 310–332 (2020).