

技術調査＋自己紹介

# 目次

---

- 背景
- 手法
- 分析設定
- 分析結果

# 目次

---

- 背景
- 手法
- 分析設定
- 分析結果

# 概要

---

- 論文タイトル

- 「大規模言語モデルによる投信ディスクロージャー資料の市況および見通しコメントの自動生成」

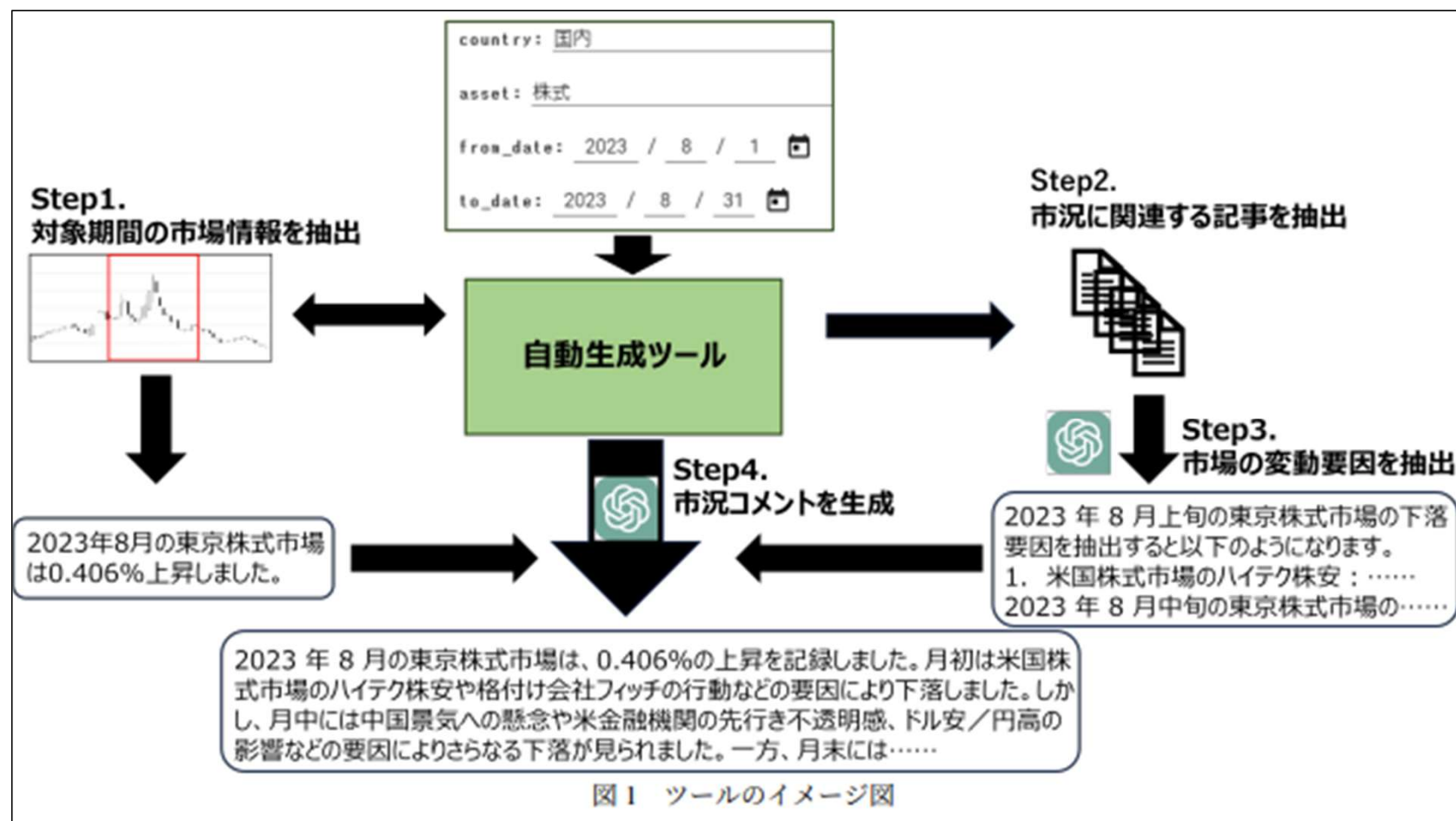
- 論文概要

- 資産運用会社においてディスクロージャー資料作成の負担大
- LLMを用いて運用報告書の市況コメント、見通しコメントを自動生成するツール作成

市況コメント　：ある期間の市場環境や経済状況についてのコメント

見通しコメント：将来の市場環境や経済状況に対する予測

# ツールのイメージ



# 目次

---

- 背景
- 手法
- 分析設定
- 分析結果

# 提案手法

---

- 分析の流れ
  - Step1. リターン/利回り変化幅の算出
  - Step2. 変動要因記事の抽出
  - Step3. 変動要因の要約
  - Step4. 市況コメントの生成
  - Step5. 見通しコメントの生成

# Step1. リターン/利回りの変化幅の算出

---

[株]

- 日次リターン

$$R_t = \frac{p_t - p_{t-1}}{p_{t-1}}$$

[債券]

- 日次リターン

$$R_t = Y_t - Y_{t-1}$$

対象資産の価格 $P_t$ , 債券の利回りを $Y_t$

対象期間を $t_0, t_1, \dots, t_n$

ただし、 $t_0$ :対象前月末、 $t_1$ :対象月初日、 $t_n$ :対象月末日



## Step2. 変動要因記事の抽出

---

- 各月の資産価格変動に影響を与えた記事を抽出
- 方法
  1. 月を上旬、中旬、下旬に分け、それぞれの期間の資産価格のリターンを計算
  2. リターンが最も下落した日と上昇した日を特定

$$t_{\min}^i = \arg \min(R_t) \quad \text{for } t \in T_i, i = 1, 2, 3 \quad (1)$$

$$t_{\max}^i = \arg \max(R_t) \quad \text{for } t \in T_i, i = 1, 2, 3 \quad (2)$$

3. 特定したの市場動向について記載されたニュース記事を抽出

## Step3. 変動要因の要約

---

- gpt-3.5は16000トークンという制限
  - ニュース記事テキストをそのまま結合したものを入力とするのは困難
  - 入力記事を要約する必要あり
- 方法
  1. 抽出したニュース記事をChatGPTに入力し、要約を実行
  2. 上旬、中旬、下旬ごとにニュース記事から変動要因を要約

変動要因の要約のための「messages」

**[system]**

あなたはエコノミストです。

**[user]**

以下のニュース記事を参考に2023年8月上旬の東京株式市場の下落要因を抽出してください。  
<3.2節で抽出したニュース記事のテキスト>

## Step4. 市況コメントの生成

---

- 月次リターンとstep3の変動要因の要約を使用し、コメント生成
- 生成方法
  - Zero-shot learning
    - ✓直接要約された情報を入力とし、コメント生成
  - Few-shot learning
    - ✓過去の市況コメントを参考にしながら新たなコメントを生成

## Step4. 市況コメントの生成

---

### • Zero-shot learning

市況コメント生成のための「messages」

[system]

あなたはファンドマネージャーです。

[user]

2023 年 8 月の東京株式市場の主な変動要因として以下の情報が得られています。

2023 年 8 月の東京株式市場は0.406%上昇しました。

<3.3 節で生成した上旬の変動要因の要約>

<3.3 節で生成した中旬の変動要因の要約>

<3.3 節で生成した下旬の変動要因の要約>

情報は以上となります。上記の情報を参考に以下のフォーマットのマーケットレポートを作成してください。

2023 年 8 月の東京株式市場は0.406%上昇しました。

月初は、、、月中は、、、月末は、、、

### • Few-shot learning

市況コメント生成のための「messages」(1-shot)

[system]

あなたはファンドマネージャーです。

[user]

2023 年 7 月の東京株式市場の主な変動要因として以下の情報が得られています。

～zero-shot と同様のため省略～

情報は以上となります。上記の情報を参考にマーケットレポートを作成してください。

[assistant]

<指定した投資信託の 7 月の市況コメント>

[user]

2023 年 8 月の東京株式市場の主な変動要因として以下の情報が得られています。

～zero-shot と同様のため省略～

情報は以上となります。上記の情報を参考にマーケットレポートを作成してください。

## Step5. 見通しコメントの生成

---

- Step4の「messages」を使用し、以下の3パターンの見通し作成
  - 楽観的な見通し
  - 中立的な見通し
  - 悲観的な見通し

見通しコメント生成のための「messages」

～市況コメント生成のための「messages」～

[assistant]

<3.4 節で生成した市況コメント>

[user]

これまでのマーケットレポートを踏まえて、東京株式市場の「楽観的な見通し」、「中立的な見通し」、「悲観的な見通し」を教えてください。

# 目次

---

- 背景
- 手法
- 分析設定
- 分析結果

# データセット

- 時系列データ（右図）

- Refinitive Datastreamから取得

- ニュース記事

- 期間：2022年1月1日～2023年9月30日

- 元データ：ロイターニュース記事

- 件数：市場サマリーor マーケットサマリーを含む記事990件

- 市況コメント、見通しコメント

- 期間：2023年4月～2023年9月

- 資料：各資産運用会社の月次投信ディスクロージャー資料

- 評価：各市場ごと3ファンド、合計12ファンドのデータを使用

表1 分析に使用する市場データ		
資産	地域	
	国内(日本)	国外(米国)
株式	TOPIX	S&P 500
債券	Nomura BPI 総合指数 最終利回り(単利)	米国債 10年物 利回り

# 精度評価指標（市況コメント）

---

- 定性評価

- 実際の市況コメントと自動生成されたコメントにトピックを付与
- 以下の評価指標を使用

$$\text{一致率} = \frac{|A \cap B|}{|A \cup B|}, \text{適合率} = \frac{|A \cap B|}{|A|}, \text{再現率} = \frac{|A \cap B|}{|B|}$$

$A$ :生成した市況コメントに含まれるトピックの集合、 $B$ :実際の市況コメントに含まれるトピックの集合

- 定量評価

- コサイン類似度
  - ✓ 2つのテキスト全体の類似性を測る指標（範囲：[-1,1]）
- BERT Score
  - ✓ テキスト間の意味的類似性を測る指標（範囲：[-1,1]）
  - ✓ コサイン類似度より詳細なトークンレベルの類似性を評価



# Hallusinationの影響評価（市況コメント）

---

- 市況コメントに対して行う
  - 各トピックを以下のように人手で分類

- (1)要因が起っていない, または, 事実と異なる
- (2)要因は起っているが, 一般的に因果関係がない
- (3)要因は起こっており, 一般的に因果関係があることが知られているが, 極性が事実と一致していない
- (4)要因は起こっており, 一般的に因果関係があり, 極性が一致している

# 精度評価指標（見通しコメント）

---

- 「自動生成コメント」と「実際の見通しコメント」の識別テスト
  - 実際の見通しコメント2問、自動生成コメント2問
- 方法
  - 「楽観的」、「中立的」、「悲観的」な見通しから1つ抽出
  - 社員(運用部社員と若手社員、各6名)が実際のものか自動生成されたものか判別

# 目次

---

- 背景
- 手法
- 分析設定
- 分析結果

# 結果：定性（市況コメント）

---

- 国内債券市場のzero-shotの一致率が低い
  - ニュース記事内で国内債券市場の変動要因に関する記述が少なく、無関係な要因を列挙したコメント生成
- few-shotの方が精度が高い
  - 過去のトピックと似たコメントを生成するため
  - 債券市場は変動要因が少ないので、過去データを与えると精度向上

表2 市況コメントの定性評価結果

市場	評価指標	zero-shot	few-shot	人手作成間
国内株式	一致率	0.108	0.124	0.176
	適合率	0.207	0.235	-
	再現率	0.171	0.197	-
国内債券	一致率	0.044	0.114	0.319
	適合率	0.052	0.172	-
	再現率	0.200	0.245	-
国外株式	一致率	0.110	0.123	0.190
	適合率	0.160	0.223	-
	再現率	0.230	0.213	-
国外債券	一致率	0.138	0.224	0.207
	適合率	0.188	0.287	-
	再現率	0.395	0.350	-

# 結果：定量（市況コメント）

---

- Zero-shotよりもfew-shotの方が精度が高い
  - 定性と同じ結果
- 人手よりもfew-shotが精度が良い市場が存在
  - 国内株式市場以外は有用
  - 株式市場は変動要因が多く一致率が低い傾向

表5 市況コメントの定量評価結果

市場	評価指標	zero-shot	few-shot	人手作成間
国内株式	LUKE Cos-Sim	0.756	0.770	0.797
	SBERT Cos-Sim	0.773	0.788	0.799
	BERT Score	0.702	0.745	-
国内債券	LUKE Cos-Sim	0.667	0.742	0.682
	SBERT Cos-Sim	0.650	0.742	0.696
	BERT Score	0.676	0.755	-
国外株式	LUKE Cos-Sim	0.751	0.750	0.739
	SBERT Cos-Sim	0.786	0.795	0.768
	BERT Score	0.694	0.735	-
国外債券	LUKE Cos-Sim	0.716	0.789	0.675
	SBERT Cos-Sim	0.730	0.793	0.704
	BERT Score	0.709	0.784	-

# 結果：Hallusination（市況コメント）

- Few-shotの方がHallucinationが多い
  - トピックが少ない市場(債券市場)において、過去の市況コメントを参考にし、生成を行うことが原因

表3 市況コメントの Hallucination 割合

市場	トピック分類	zero-shot	few-shot
国内株式	(1) 事実ではない	0.017 (=1/59)	0.114 (=20/175)
	(2) 因果関係なし	0.136 (=8/59)	0.086 (=15/175)
	(3) 極性不一致	0.000 (=0/59)	0.097 (=17/175)
	(1)+(2)+(3)	0.153 (=9/59)	0.297 (=52/175)
	(4) 正しい記述	0.847 (=50/59)	0.703 (=123/175)
国内債券	(1) 事実ではない	0.020 (=1/51)	0.022 (=2/91)
	(2) 因果関係なし	0.098 (=5/51)	0.066 (=6/91)
	(3) 極性不一致	0.020 (=1/51)	0.055 (=5/91)
	(1)+(2)+(3)	0.137 (=7/51)	0.143 (=13/91)
	(4) 正しい記述	0.863 (=44/51)	0.857 (=78/91)
国外株式	(1) 事実ではない	0.017 (=1/58)	0.029 (=3/105)
	(2) 因果関係なし	0.034 (=2/58)	0.048 (=5/105)
	(3) 極性不一致	0.069 (=4/58)	0.038 (=4/105)
	(1)+(2)+(3)	0.121 (=7/58)	0.114 (=12/105)
	(4) 正しい記述	0.879 (=51/58)	0.886 (=93/105)
国外債券	(1) 事実ではない	0.023 (=1/44)	0.033 (=3/90)
	(2) 因果関係なし	0.136 (=6/44)	0.156 (=14/90)
	(3) 極性不一致	0.000 (=0/44)	0.000 (=0/90)
	(1)+(2)+(3)	0.159 (=7/44)	0.189 (=17/90)
	(4) 正しい記述	0.841 (=37/44)	0.811 (=73/90)

# 結果：見通しコメント

---

- 運用年数が長いほど、正解率が高い
- Hallucinationが生じている
  - 誤ったまま生成する可能性があるので、現段階では運用は厳しい

表4 見通しコメントの定性評価の結果			
	運用部の 正解率	運用部以外(若手) の正解率	全体の 正解率
Q1	1.000	0.833	0.917
Q2	1.000	0.833	0.917
Q3	0.667	0.500	0.583
Q4	0.667	1.000	0.833
全体	0.833	0.792	0.813



見通しコメントは作成&評価が難しい

# まとめ

---

- 市況及び見通しコメントを自動生成するツールを作成
  - 課題は盛沢山
- 市況・見通しコメントを書く際の参考にできる
  - Hallusinationのチェックは必須
- 感想
  - 適切な評価指標を考えるのが難しい
  - GPT4.0にしたらかなり有用そう
  - RAGに変更したら精度高まりそう



# 付録：コサイン類似度

---

- ベクトル同士の向きの類似度を測る指標
  - 今回は、2つのテキスト間の意味的な近さを測る

## 定義（コサイン類似度）

ベクトル  $x$  と  $y$  は、 $\mathbb{R}^n$  において、

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

とする。コサイン類似度  $\cos(x, y)$  は以下のように定義される。

$$\cos(x, y) = \frac{\langle x, y \rangle}{\|x\| \|y\|} = \frac{\sum_{k=1}^n x_k y_k}{\sqrt{\sum_{k=1}^n x_k^2} \sqrt{\sum_{k=1}^n y_k^2}}$$

# 付録：BERT Scoreの計算方法

---

- F1がBERT Score

**Step 1:** 正解となるテキスト，生成したテキストのそれぞれを，BERT に入力することで各トークンの分散表現を取得

**Step 2:** 正解となるテキストの各トークンと生成したテキストの各トークンの全てのペアのコサイン類似度を計算

**Step 3:** トークンごとに得られたコサイン類似度の最大値を用いて，適合率，再現率，F1 スコアを計算

正解となるテキストを BERT に入力することで得られた分散表現を  $x = \langle x_1, \dots, x_k \rangle$  とし，生成したテキストを BERT に入力することで得られた分散表現を  $\hat{x} = \langle \hat{x}_1, \dots, \hat{x}_l \rangle$  とすると，以下に示す式で計算される．

$$\text{再現率} = \frac{1}{|x|} \sum_{x_i \in x} \max_{\hat{x}_j \in \hat{x}} x_i^T \hat{x}_j \quad (3)$$

$$\text{適合率} = \frac{1}{|\hat{x}|} \sum_{\hat{x}_j \in \hat{x}} \max_{x_i \in x} x_i^T \hat{x}_j \quad (4)$$

$$\text{F1 スコア} = \frac{2 \times \text{再現率} \times \text{適合率}}{\text{再現率} + \text{適合率}} \quad (5)$$

-----