

独立成分分析と Fisher の線形判別による 内閣府景気ウォッチャー調査データの分析

椎名唯圭 加藤真大 井口亮

みずほ第一フィナンシャルテクノロジー株式会社

{yuika-shiina, masahiro-kato, ryo-inokuchi}@fintec.co.jp

概要

内閣府の景気動向調査である景気ウォッチャー調査では、さまざまな職業の人々による現在および将来の経済状況に関する評価が判断根拠テキストとともに公表されている。本研究では、この景気ウォッチャー調査データをテキスト埋め込みの手法で数値データに変換し、景気判断に影響を及ぼす要因を定量的に分析する。テキスト埋め込みは解釈が難しいことで知られているが、本研究では独立成分分析と Fisher の線形判別分析を組み合わせることで、高次元のテキスト埋め込みデータを解釈可能な低次元データに変換することで、人々の景気判断に寄与しているテキストの要素について調査する。

1 導入

景気ウォッチャー調査は、内閣府が公表している市場調査であり、その目的、調査対象、および調査事項は以下のとおりである：

目的：「地域の景気に関連の深い動きを観察できる立場にある人々の協力を得て、地域ごとの景気動向を的確かつ迅速に把握し、景気動向判断の基礎資料とすることを目的とする」ことである。

調査対象：「家計動向、企業動向、雇用等、代表的な経済活動項目の動向を敏感に反映する現象を観察できる業種の適当な職種の中から選定した 2050 人を調査客体とする」ことである。

調査事項：「景気の現状に対する判断（方向性）とその理由」および「景気の先行きに対する判断とその理由（先行き：調査時点から三ヶ月後）」である。

回答者は、現状の景気および将来の景気について、以下の 5 段階で評価を行う：

現状：良くなっている・やや良くなっている・変わらない・やや悪くなっている・悪くなっている。

先行き：良くなる・やや良くなる・変わらない・や

や悪くなる・悪くなる。

本研究では、簡単化のため、この 5 段階評価を、「好景気」(4 と 3)「中立」(2)「不景気」(2 と 1)として扱う。加えて、回答者は景気判断の根拠を 1~2 文の短文で述べる。例を表 1 に示す。

1.1 貢献

本研究では、景気ウォッチャー調査データに対してテキスト埋め込みと次元削減の方法を用いることで、人々の景気判断に影響を及ぼしている要素を調査する。本研究の定式化は、判断文を特徴量とし、好景気・中立・不景気という景気判断をラベルとする 3 クラス分類問題に基づいている。

提案手法では、景気判断の根拠文を OpenAI のテキスト埋め込みモデル「text-embedding-3-small」を用いて埋め込み表現に変換する。テキスト埋め込みは、自然言語の文をベクトル表現で表すことで、定量的な分析を可能にする手法である [1]。ただし、埋め込み表現で得られるベクトルそのものを解釈することは難しいとされている。

本研究では、独立成分分析 (Independent Component Analysis, ICA) と Fisher の線形判別分析 (Linear Discriminant Analysis, LDA) を適用し、テキスト埋め込みをそれぞれ異なる二つのベクトルに変換することでテキスト埋め込みの解釈を行う。独立成分分析で得られるベクトルはテキスト埋め込み中の特徴的な要素を強調し、線形判別分析で得られるベクトルは景気判断に基づく分類に寄与する要因を表している。両者の相関を計測することで、分類に寄与する特徴的な要素を特定する。

2 問題設定

景気ウォッチャー調査データの根拠文にテキスト埋め込みを施すことで得られる行列を $X = (X_i)_{i \in [n]} = (X_{i,j})_{i \in [n], j \in [d]} \in \mathbb{R}^{n \times d}$ 。ここで、 $X_{i,j}$

表1 Caption (4: 良 - 2: 中立 - 0: 悪) (1: 現在, 0: 未来)

年月日	景気判断	職種	現在/未来	根拠
2016年03月	2	不動産業	1	・近隣の不動産屋の話では、時期的に大学生の来店が多いものの成約する賃料は低くなる傾向で、家計は良くなってないことがうかがえる
2017年1月	4	フェリー	0	・例年、雪解けと同時に輸送量が増加する傾向があるため、今後についてはやや良くなる
2021年11月	3	百貨店	0	・新型コロナウイルス次第のところはあるが、旅行に行くからと、服やバッグ等を買求める客が増えている。通常営業に戻って、催事も予定どおり開催されているので、今後良くなるのではないかと

表2 データセットに関する情報.

		2016年	2018年	2020年
現在の景気判断	サンプルサイズ	15168	14756	15564
	景気判断の平均	1.815	1.947	1.384
将来の景気判断	サンプルサイズ	17279	16788	17525
	景気判断の平均	1.906	2.030	1.440

は第 i 番目の根拠文の第 j 次元の要素を表しており、 n は分析対象のサンプルサイズ、 d は埋め込みの次元を表している。私たちはインデックス集合を $[n] := \{1, 2, \dots, n\}$ で表す。また、 X_i は \mathbb{R}^d のベクトルであるとする。私たちの埋め込み表現の次元数は $d = 1536$ 次元である。各回答 X_i には景気判断 $Y_i \in \{\text{好景気, 中立, 不景気}\}$ が付随している。

紙面の都合により、2016年の景気ウォッチャー調査データを分析に用いる。それぞれの年のデータセットは現在の景気判断と将来の景気判断に分けられる。サンプルサイズはおおよそ15000から17000程度の回答数であり、詳細は表2に記載する。景気ウォッチャー調査データは地域ごとに分けられているが、本研究では地域を区別せずに一つのデータセットにまとめて分析を行う。

私たちは text-embedding-3-small を用いて根拠文を変換しており、その次元数は1536次元 ($d = 1536$) である。

3 ICA と LDA による分析

ICA (独立成分分析) は多変量解析の手法の一つであり、観測された多変量信号を統計的に独立した成分に分解することを目的としている。すなわち、多次元の確率変数ベクトル $X = (X_1, X_2, \dots, X_d)^T$ を、正規分布に従わない互いに独立な確率変数 $Z = (Z_1, Z_2, \dots, Z_k)^T$ の線形和 $X = AZ + n$ として表し、観測された確率ベクトル x を s に分解する手法である [2]。ここで、 d と k はそれぞれ X と Z の次元を表し、 Z は独立成分と呼ばれる。 A は混合行列、 n はノイズベクトルである。

ICA は混合行列 A を推定することで独立成分 Z を得る。ICA の理論的基盤は、成分間の独立性を最

大化することに基づいている。ここでいう独立性とは、相関がゼロであることに加えて、成分の確率分布が異なることを意味する。ICA では、白色化と直交化を通じて観測行列 $X = (X_{i,j})$ を変換し、成分の非ガウス性を強調することで、独立性を最大化するように混合行列 A を推定する。この過程により、ICA は外れ値を強調する成分を得る傾向があることが知られている。

[3] は、ICA が埋め込み表現の解釈に有用であることを示した。[3] では、ICA を用いることで、単語や画像の埋め込みに含まれる「解釈可能な突起」を発見できると述べている。しかし、[3] の対象は単語埋め込みであり、文埋め込みに対する適用可能性は議論されていない。この点に関して、[4] が内閣府の景気ウォッチャー調査データを用いて文埋め込みへの適用を試みている。

LDA は、特徴量 X をクラス Y の分類に寄与する低次元空間に射影し、その空間でクラス分類を行う手法である。具体的には、観測値 $\{(X_i, Y_i)\}_{i=1}^n$ が与えられたとき、以下の手順で分析を行う。ここで、 X_i は1536次元の埋め込みベクトル、 $Y_i \in \{\text{不景気, 中立, 好景気}\}$ は景気判断のラベルである。

まず、埋め込みベクトル全体の平均を \bar{X} 、ラベル $y \in \{\text{不景気, 中立, 好景気}\}$ に対応する埋め込みベクトルの平均を \bar{X}_y と定義する。次に、クラス内共分散行列 S_W とクラス間共分散行列 S_B を以下で定義する：

$$S_W := \sum_y \sum_{i \in C_y} (X_i - \bar{X}_y)(X_i - \bar{X}_y)^T,$$

$$S_B := \sum_y N_y (\bar{X}_y - \bar{X})(\bar{X}_y - \bar{X})^T,$$

ここで、 C_y はラベル y に属するサンプルの集合、 N_y はそのサンプル数である。

LDA では、射影行列 W を以下の基準に基づいて求める：

$$J(W) := \text{tr}\left((W^T S_W W)^{-1} (W^T S_B W)\right),$$

すなわち、射影後の空間でクラス内分散を最小化し、クラス間分散を最大化するように W を定める。

ICA は、テキスト埋め込み中の特徴的な要素を発見するのに役立ち、一方、線形判別分析 (Linear Discriminant Analysis; LDA) は分類に寄与する要素を特定するのに有用である。しかし、どの独立成分が分類に寄与しているのか、また、線形判別分析の射影軸がどのような意味を持つのかは明らかではない。

[5] では、この両手法の欠点を補完するアプローチとして、線形判別分析の射影軸と相関の大きい独立成分に着目する手法を提案している。この手法では、射影軸ベクトルと高い相関を持つ独立成分ベクトルが、テキスト埋め込み中で分類に寄与する特徴的な要因を表すものと考えられる。

本研究では、[5] の手法を景気ウォッチャー調査データに適用し、人々の景気判断に影響を与える特徴的な要因を明らかにする。

4 景気判断根拠の分析

本節では、景気ウォッチャーデータに対して ICA と LDA を用いた分析を行った結果を示す。本分析では、1536 次元の文章埋め込みベクトルを ICA を用いて解釈可能な 100 次元のベクトルに変換し、その後 Fisher の線形判別分析 (LDA) を用いた分類を行った。

4.1 景気判断根拠の分析

分析結果の参考として、図 1 に現状の景気判断文 (左図) と将来の景気判断文 (右図) に対して線形判別分析を行った結果を示す。さらに図 2 と図 3 に現状と将来の景気判断分類の各射影軸と相関が高い独立成分についての調査結果を示す。本分析では、射影軸に対する相関の絶対値が大きい順に上位 10 個の独立成分を取り出し調査対象とした。また各独立成分を代表する景気判断文章を抽出し要約を行った。

4.2 現状の景気判断根拠の分析

図 1 (左図) より、射影軸 1 は「好景気」と「不景気」に関連する文章を分類する軸として機能し、正方向には好景気に関連する文章が、負方向には不景気に関連する文章が強く現れていることが確認された。さらに、図 2 (上段) より、射影軸 1 と正の相関が最も高い独立成分にはサービス業を中心に経済活

動が活発な文章が多く含まれており、特定の業種における好景気の影響が明確に現れていた。その他の正の相関を持つ独立成分も、天候や為替等の外的要因によるポジティブな影響を反映する内容であり、豪雨災害からの復興による受注増加や、マイナス金利下での住宅需要など、好景気に関する内容が表現されていた。一方、射影軸 1 と負の相関が高い独立成分では、不景気の影響が強調されている。具体的には、原材料価格の高騰や景気停滞による販売量の減少、不景気に伴う来客数や売上の減少といった要素が目立ち、製造業や小売業を中心に不景気の影響を示唆している。

これらの結果により、射影軸 1 は市場の景気状況を反映し、人々の消費傾向や産業や地域別の経済動向を反映する指標であることを示している。

また、図 1 (左図) より射影軸 2 は「中立」と「好景気・不景気」を分類する軸として機能し、軸の正方向には好景気または不景気に関連する文章が、負方向には中立に関連する文章が強く表れていることが確認された。射影軸 2 と正の相関が高い独立成分は、好景気と不景気の内容が混在し、具体的にはサービス業と製造業の収益構造違いや熊本地震による不景気など、経済の好不況が業種や地域によって大きく異なる様子が表現されている。一方、射影軸 2 と負の相関が高い独立成分では、変化の少ない経済状況が反映されており、不景気の影響が経済全体に広がり、景気の不透明感から経済活動の停滞が続く状況を示している。

これらの結果により、射影軸 2 は市場経済全体の変動性を反映する軸であり、自然災害や為替等の外的要因に対する感応度としても有効な指標であることを示している (図 2 (下段))。

4.3 将来の景気判断根拠の分析

本節では現状の景気判断分析と同様に、将来の景気判断根拠文に対しても同様の分析を行った。

図 1 (右図) は、将来の景気判断文に対して線形判別分析を行った結果を示す。また、図 3 に、線形判別分析の射影軸の調査結果を示す。

図 1 (右図) より、射影軸 1 は「好景気」と「不景気」に関連するデータを分類する軸として機能し、正方向には好景気に関連する要素が、負方向には不景気に関連する要素が強く現れていることが確認された。さらに図 3 (上段) から、射影軸 1 と正の相関が高い独立成分には期待感が表れており、政策や

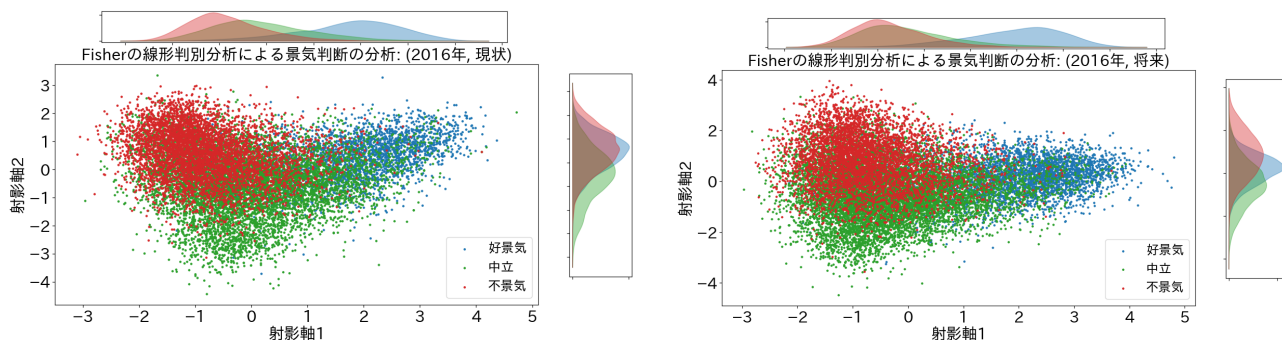


図1 現状の景気判断文（左図）と将来の景気判断文（右図）に対して線形判別分析を行った結果

イベントが経済活動を促進し、市場にポジティブな印象を与えることがわかる。一方、射影軸1と負の相関が高い独立成分には、不景気による課題や消費税増税、異常気象など経済活動への懸念が示されている。これらの結果により、射影軸1は、経済活動や政府の政策、イベント、天候等の幅広い要因が消費者心理に与える影響を表す重要な指標である。

また、図1（右図）により射影軸2は「中立」と「好景気・不景気」を分類する軸として機能し、軸の正方向には好景気や不景気に関連する文章が、負方向には中立に関連する文章が強く表れていることが確認された。射影軸2と正の相関が高い独立成分には、「売り上げ減少」や「問い合わせの増加」等、好景気や不景気が経済全体に与える影響が反映されており、射影軸2と負の相関が高い独立成分には、経済の停滞が続き変化のない状況が示されている（図3（下段））。これらの結果から、射影軸2は足下の景気状況が市場経済に与える影響の大きさを表しており、日本経済における景気動向を多角的に捉える重要な指標であることを示している。

4.4 景気判断根拠文の分析のまとめ

射影軸1は景気状況や景気期待の方向性を示す指標として機能し、好景気や不景気の業種や地域の違いを明確に捉えることができる。一方、射影軸2は景気の変動性や外的要因による景気への影響の大きさを評価する指標として機能し、現在の経済状況や為替動向、天候等の要因が経済活動に与える影響を捉えることができる。また、現状分析では、具体的な課題や外的要因の影響が強調される一方で、将来分析では、政府の政策や地域イベントに対する期待感が景気動向を左右する重要な要素として示された。本手法により、現状の景気判断分析で社会課題や具体的影響を的確に把握し、将来の景気判断分析で政策やイベントの影響を捉えることが可能である。

る。その結果、経済動向を多角的に捉え、包括的に経済を理解するための重要な視点を得ることができる。

さらに、本手法は回答者によるラベルミスの検出にも寄与する。好景気を示す点の集合内に存在する不景気を示す点について、景気判断根拠文を確認したところ、「海外からの旅行客が増加している」や「改装後に花屋として認識されるようになり、毎日客が来るようになった」等、その内容は好景気を示すものであった。これらは回答者により誤った景気判断ラベルが付けられたことを示している。本手法により、景気判断ラベルの誤りを捉え、景気ウォッチャー調査データの有効性をさらに高めることが期待される。

5 結論

本研究では、景気ウォッチャー調査データを対象に、独立成分分析（ICA）と線形判別分析（LDA）を組み合わせた手法を用いて、景気判断に寄与する特徴的な要因を定量的に分析した。ICAによりテキスト埋め込みの特徴的な成分を抽出し、LDAにより分類に寄与する要素を特定することで、景気判断の根拠となる解釈可能な要因を明らかにした。結果として、人々の景気判断に影響を与える重要な要因として、消費傾向や市場動向、政策の影響などが抽出され、これらが景気の現状や将来の評価にどのように結びつくかを定量的に示すことができた。さらに、回答者によるラベルミスの検出にも寄与し、データセットの信頼性や有効性を向上させる手段としての活用可能性も示された。本手法は、テキストデータの解釈可能性を高めつつ、政策立案や経済予測における意思決定支援ツールとしての可能性を示唆している。今後は、より広範なデータセットや他の埋め込み手法を用いた分析を通じて、さらなる応用可能性の検討が期待される。

参考文献

- [1] John Wieting, Mohit Bansal, Kevin Gimpel, and Karen Livescu. Towards universal paraphrastic sentence embeddings. In **International Conference on Learning Representations (ICLR)**, 2016.
- [2] Christian Jutten and Jeanny Hérault. Blind separation of sources, part i: An adaptive algorithm based on neuromimetic architecture. **Signal Processing**, Vol. 24, No. 1, pp. 1–10, 1991.
- [3] Hiroaki Yamagiwa, Momose Oyama, and Hidetoshi Shimodaira. Discovering universal geometry in embeddings with ICA. In **Conference on Empirical Methods in Natural Language Processing (EMNLP)**, 2023.
- [4] 新妻巧朗, 田口雄哉, 田森秀明. 計量テキスト分析のための文埋め込みによる探索的カテゴリ化. 言語処理学会第 30 回年次大会, pp. 494–499, March 2024.
- [5] 加藤真大, 持橋大地, 浦川通, 新妻巧朗, 田森秀明. 文埋め込みに基づく朝日歌壇短歌の分析. Technical report, 2024.

A 付録

A.1 景気判断分類における各射影軸の調査結果

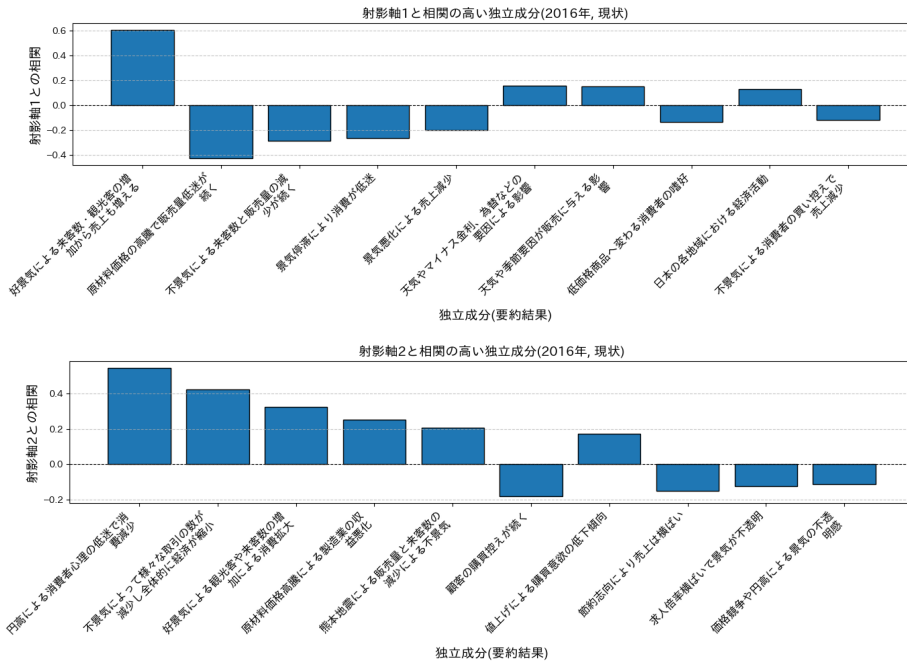


図2 現状の景気判断分類における各射影軸の調査結果 (上段：射影軸1, 下段：射影軸2)

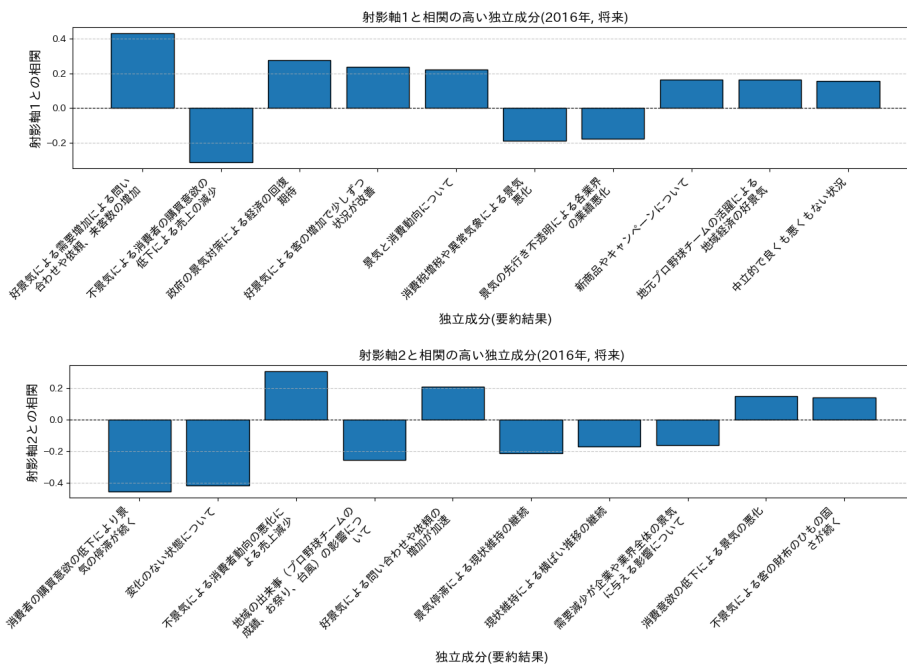


図3 将来の景気判断分類における各射影軸の調査結果 (上段：射影軸1, 下段：射影軸2)