

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

新妻巧朗¹ 田口雄哉¹ 田森秀明¹

¹ 朝日新聞社メディア研究開発センター

{niitsuma-t, taguchi-y2, tamori-h}@asahi.com

概要

本論文は、計量テキスト分析の一手法として、テキストに対する探索的なカテゴリ抽出とラベル付与に関する手法を提案する。さらに、ケーススタディとしてオープンデータに対して提案手法を適用し、その有用性を確認した。報道や市場調査の現場におけるテキスト分析では、予測ではなく洞察を得ることが目的の、仮説構築のための探索的な解析が求められている。このような場合は内容分析による手法が有効であるが、既存の手法では単語の統計情報や辞書、分類モデルの活用が中心であり、意味空間を考慮した探索的なカテゴリ分類はあまり考慮されてこなかった。そこで、文埋め込みと独立成分分析 (ICA) を用いて、意味空間を俯瞰してテキストを抽象化し、オープンドメインに対して教師なしでカテゴリを付与する手法を提案する。

1 はじめに

アンケート調査の自由記述や発言の書き起こし、SNSの投稿等のテキストデータは、人間の社会的活動の一側面が表現されており、社会調査や消費者研究等において重要なデータの一つである [1, 2, 3]。また、報道現場においても仮説構築のための探索的なデータ解析の重要性が増しており、公的統計に加えてテキストデータの活用が期待されている。

このようなテキストの内容に注目して分析する手法は、特に社会科学の領域において「内容分析¹⁾」と呼ばれてきた。内容分析は、テキストデータをカテゴリに分類することで計量可能にし、量的にテキストの特徴を把握・比較するアプローチで、主にメディア研究や社会調査において活用されてきた。なお、内容分析では、抽象化したラベルを用いてテキストを分類する作業を「コード化」と呼ぶ [4, 5] が、本論文では便宜上「カテゴリ化」と呼ぶ。また、テキストマ

1) 内容分析のアプローチには、質的手法もあるが本論文では量的手法のみを扱う。

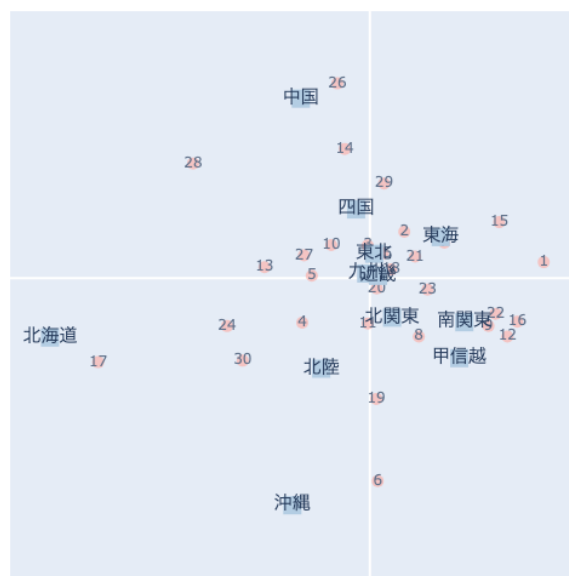


図1 提案手法によって得られたカテゴリによる対応分析

インギングを通じた内容分析を樋口は「計量テキスト分析」と定義 [4] しており、本論文もこの定義に従う。カテゴリ化による手法は、カテゴリ化されたテキストデータを、そのテキストの発信者の居住地や年齢層などの属性データと紐づけて集計することで、多角的に分析がしやすい点が有用である。図1は、景気に関するアンケート調査の結果を提案手法でカテゴリ化した上で地域ごとに集計し、それを対応分析した結果をプロットしたものである。4節にて、提案手法を用いたケーススタディとして、この対応分析の手法と結果について述べる。

カテゴリ化の手法は、単語の統計情報や辞書によるアプローチや、感情分析などの分類モデルによる手法が一般的であるが、前者は辞書の構築や単語の選定で分析者の主観が強くなりやすく、後者は既存の教師データの枠組みを越えられないという課題がある。そこで、本論文では事前学習済みモデルを用いた文埋め込みからICAを用いて独立成分を抽出し、それぞれの独立成分が表現している意味をカテゴリと考える。そして、独立成分の中でスコアが高いテキ

ストをカテゴリとして自動的に設定できる手法を提案する。これによって、埋め込み表現に内在する様々な意味空間を考慮してカテゴリを探索的に整理することが可能となり、新たな質的な観点の発見に寄与できる可能性がある。

2 関連研究

2.1 コンピュータによるカテゴリ化

カテゴリ化は、かつては人手による抽象化アプローチが中心であったが、近年はコンピュータの活用が増えており、辞書を用いた単語マッチやテキストマイニングで得られた情報をもとにおこなう事例が多い [4, 6]²⁾。また、教師あり分類モデルを使った事例では、感情分析や固有表現抽出モデルが活用されている [2, 3]。例えば、藤本らはアンケート調査に対して、探索的データ解析で得た洞察に基づいて、自由記述に対してアスペクトベース感情分析を実行し、その結果をアンケート調査に対する多重対応分析にマッピングするという手法を提案している [8]。

2.2 文埋め込みと次元圧縮

文埋め込みとは、文の意味をベクトル空間上で表現する手法の一つで、意味が類似している文章が空間上で近くに配置されるように設計されているモデルである。このような特徴から文埋め込みは文の類似性評価に対して一般的に使われてきた。近年では特に事前学習済みモデルから得られる潜在表現を使って文の類似性を学習することで、文脈を考慮できる埋め込み表現を獲得する手法が登場し、二文の類似性を予測する自然言語推論タスクにおいて大幅に性能を向上させている [9, 10]。

単語・文埋め込みを表すベクトルは、一般的にそれぞれの次元を個別に見ても意味などの特徴が解釈できないという課題がある。しかし、単語埋め込みにはそのベクトル空間の解釈に関する多くの研究があり、主成分分析 (PCA) や独立成分分析 (ICA) を使うことで、単語埋め込みを品詞や意味を解釈できる次元から構成される空間へと分解・圧縮することができるという研究がある [11, 12, 13]。特に山際らは、ICA について PCA に比べて解釈性が高く、アルゴリズムやモダリティに拘らず解釈可能な次元に分解できると実験によって示している。

2) 特に国内においてはテキスト分析ツールである「KH Coder (<https://kncoder.net/>)」の利用が目覚ましい [7]。

3 提案手法

3.1 概要

本手法は次の3つの工程からなる。

1. テキストを文埋め込みに変換する
2. 文埋め込みを ICA によって次元削減する
3. 独立成分のスコアからカテゴリを割り当てる

1と2については、先行研究である山際ら [13] の単語埋め込みによる手法を文埋め込みに用いたものである。一方で、3は本論文特有の方法であるため、次に解説をする。

3.2 独立成分スコアによるカテゴリ化

ICA の出力結果が意味的な観点からの解釈性が高い [13] という点に着目して、事前学習済みモデルから獲得された文埋め込みを ICA を通じて解釈可能な部分空間へと分解できるのではないかと考えた。

すべての文に対して、それぞれの独立成分で独立成分スコアが算出できる。そのため、本手法では ICA をそれぞれの独立成分が意味する向きへと文をソートする手法であると考え、次元圧縮後の文埋め込みの i 番目の独立成分 x_i を独立成分スコアで降順にソートする。そして、独立成分 x_i のスコアや順位に閾値を与えて、ある閾値を境に文がカテゴリに属するかどうかを振り分けることでカテゴリ化する。例えば、独立成分スコアが閾値を超えたものに1が割り当てられ、灰色の点が0となるような二値変数に変換する。この方法によって、独立成分スコアにて2.0を閾値として二値変数化した文をストリッププロットしたものが図2である。横軸は独立成分スコアで、赤色の点がカテゴリに割り当てられた点で、灰色の点が割り当てられなかった点である。

以上のような手法の性質上、どのカテゴリにも割り当てられない文が存在しうる。また、ICA は負の方向においても意味を持つ成分を解釈できるが、本論文では簡単のため正方向の成分のみを使う。

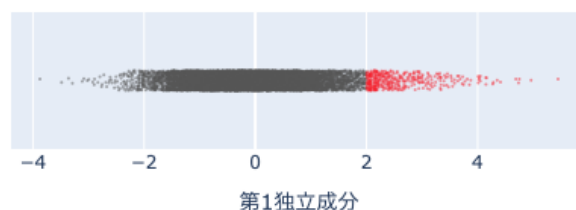


図2 第1独立成分のスコアのストリッププロット

4 ケース: 地域別不況の要因分析

4.1 分析設定

提案手法の有用性を確認するため、本項では実際のデータに対して提案手法を適用した結果を報告する。分析対象のデータとして「景気ウォッチャー調査³⁾」を用いた。景気ウォッチャー調査とは、内閣府が公表しているデータセットで、地域・業種別の景気の現状と先行きに関する五段階評価（良い、やや良い、変わらない、やや悪い、悪い）と、その判断理由を表すテキストからなる。本分析では「景気の現状判断」にて「やや悪い」と「悪い」を選択した回答者の判断理由の詳細である「追加説明及び具体的状況の説明」のテキストを2019年1月から2022年12月までの公表データから収集した。本論文では、このテキストを「不景気の判断理由」と呼ぶ。

文埋め込みモデルは、塚越ら [14] のモデル⁴⁾を用いた。ICAの実装には scikit-learn [15] の FastICA を利用した。FastICAのパラメータとしては、成分数を30、max_iter を10000、tol を1e-5 に設定した。その他はデフォルト値である。また、3.2項目における独立成分スコアの閾値は2.0で設定している。

4.2 カテゴリの集計

まず、3節の提案手法を通して独立成分とそれを閾値によってカテゴリ化する。そして、そのカテゴリに属している「不景気の判断理由」の数をデータに紐づく地域ごとに集計する。その結果として得られる集計結果の一部が表1である。

表1 地域区分別カテゴリに属する判断理由の件数(抜粋)

カテゴリ番号	4	6	15	24	30
北海道	40	15	11	50	67
東北	43	14	34	50	30
北関東	27	28	28	17	16
南関東	70	56	130	48	73
甲信越	26	21	19	21	20
東海	36	37	73	30	76
北陸	27	42	11	42	30
近畿	33	47	50	51	81
中国	38	15	46	29	40
四国	20	11	36	14	22
九州	43	33	37	34	47
沖縄	12	28	3	2	33

3) https://www5.cao.go.jp/keizai3/watcher/watcher_menu.html

4) <https://huggingface.co/cl-nagoya/sup-simcse-ja-large>

4.3 カテゴリの命名

また、カテゴリを比較・解釈しやすくするために、それぞれの独立成分でスコアの高い文を30件選択し、OpenAI社が公開しているGPT-4 (gpt-4-1106-preview) にカテゴリ名をつけさせた。次のリストは、そのカテゴリ名を抜粋したものである。

カテゴリ 4 商店街の物販低迷と飲食業界の過当競争による地域経済の冷え込み

カテゴリ 6 宿泊業界における外国人観光客減少と連休効果の限界

カテゴリ 15 製造業における受注・販売減少と構造的課題

カテゴリ 24 暖冬影響による冬物商材の売れ行き悪化と年末年始の消費低迷

カテゴリ 30 消費税増税と米中貿易摩擦による国内外の消費と観光業界の不振

使用したプロンプトおよびすべてのカテゴリ名は、付録Aに掲載する。

また、次のリストはカテゴリに属している文の例である。これはカテゴリ4に属する「不景気の判断理由」のうち、独立成分スコアの高い5件を抽出したものである。

- 11月一杯で、ほぼお遍路の仕事が終わり、12月に入り、忘年会シーズンだが前年ほど良くない。お遍路と比べ、街中では売上が上がらない。
- 商店街の中は、近年立ち飲み屋だけが多くなっている。商店街は立ち飲み屋だけでいっぱいになる可能性もある。ここ数年で60~70店舗くらい、大小様々な立ち飲み屋が広がっているようである。小売店にはブレーキが掛かって、売れていない。単価を下げてもなかなか販売の売上にはつながらない。
- 年が明けてから人が来なくなったので、毎日早じまいしている。
- 商店街一帯において、人の動きや交通量が少ない。天候は良くなっているものの、イベントを開催しても人が集まらない。
- 夜の繁華街に人はいるが乗車率が悪く、実車率も悪い。歩いてホテルに帰る県外の客が多く、地元客の飲食が少なくなっている。

カテゴリ名と対応する意味合いを持つような文が並んでおり、提案手法によるカテゴリ化が少なくとも一貫性の観点からは機能していると考えられる。

4.4 対応分析

表1を含めた地域別に集計したカテゴリごとの回答数を多重対応分析にかける。対応分析とは、クロス集計表などの項目間の関係を視覚的に表現するために使われる手法で、カテゴリカルデータのPCAとも呼ばれる [16]。また、多重対応分析とは、Blasiusら

[17] が 2 変数の対応分析を 3 変数以上に利用できるように拡張した分析手法で、実装としては `mca`⁵⁾ を利用した。これによって、集計結果に基づいてカテゴリ間の関係や地域の位置付けなどを二次元のグラフでプロットする。本分析では、ハイパーパラメータは実装のデフォルトを使用している。

図 1 は、その対応分析の結果を図示したものである。図中のカテゴリ番号に対応したカテゴリ名は、A.2 項に掲載している。

この結果から一部を抜粋して地方とカテゴリの関係性を解釈する。カテゴリ 6 は観光地における影響について述べたもので、同じ象限に沖縄が位置しているが、これは沖縄が他地域に比べて観光業の不況の影響が相対的に高いことを示していると言える。観光庁の発表している宿泊旅行統計調査 [18] によれば、各年においても関東や近畿を除くと北海道と沖縄県が最上位に位置していることからある程度は妥当な結果が得られていると考えられる。他方で、カテゴリ 15 は製造業における影響について述べたものだが、東海が非常に近く、東海の製造業の経済的な影響の度合いが他地域に比べて相対的に高いことを示していると解釈できる。中部経済産業局によれば、東海は製造業従事者数の割合が最も高い。そのため、この図は実際の統計情報に比較的近い結果が得られていると考えられる [19]。以上の情報からカテゴリ 6、15 からは産業構造による相対的な影響度が読み取れる可能性が示唆された。さらに、カテゴリ 24 は暖冬の影響によるものであるが、北海道や北陸が比較的近い位置にある。ここから、暖冬によって冬物商材が売れないことがこの 2 地域における小売店において経済的な影響が比較的大きい可能性があるという仮説を立てることができる。

5 議論

4 では、テキストデータから抽出したカテゴリデータをに紐づいている属性データの元で集計することで、さらなる洞察を得て仮説構築に活かすことができることについて述べてきた。また、本論文のケーススタディでは触れなかったが因子分析などの他の手法を利用したり、一部のカテゴリだけを用いたりすることも探索的な解析には役立つと考えられ、応用の可能性は広いと考える。

一方で、現在の提案手法には課題が 3 つある。まず、1 つ目はカテゴリ化した二値の正確な適合率と

再現率を計測することは原理的に難しいという課題がある。これは、正解の範囲がカテゴリの命名に左右されていることやカテゴリの分け方がアルゴリズムによって変わりうるためである。2 つ目は、カテゴリ化のための閾値をハイパーパラメータとして与えている点である。独立成分間でスコアの分布は異なっており、閾値を決め打ちで与えることはカテゴリの一貫性を下げってしまう恐れがある。3 つ目は、文埋め込みや次元削減アルゴリズム、それらのパラメータによって結果が大きく変わる可能性があり、カテゴリ抽出をコントロールできないという点である。

1 つ目については、カテゴリ内の文の一貫性を何らかの形で計測することで分類精度を計測することは可能であると考えている。また、この一貫性を計測することができれば、2 つ目の課題についてのアプローチも同時に考えることができるようになるだろう。例えば、一貫性の指標をなるべく保つような制約を定義し、独立成分スコアが相対的に高い文をなるべく多く選択する最適化問題として定義できるかもしれない。3 つ目については、探索的な目的においては問題にならないケースも多いが、カテゴリの粒度が粗すぎたり細かすぎたりすることがあり、適宜ハイパーパラメータの調整や場合によっては複数のカテゴリを一つにまとめる必要がある。また、このような課題にアプローチするためには、アプリケーション化においては柔軟な実装が求められるだろう。

6 おわりに

本論文では、文埋め込みと ICA を用いて文をオープンドメインでカテゴリに分類する方法を提案し、その活用方法と課題について述べてきた。特にケーススタディでは、提案手法によって得られたカテゴリから計算された属性データごとの集計値を確認し、多重対応分析にかけた。そして、その結果の一部が公表されている統計から確認できる外部情報と近いことを確認し、さらに図中のカテゴリと属性データの位置付けから仮説が立てられることを試みた。以上を通して提案手法の質的な観点での探索的解析への応用可能性が示唆された。

今後の展望としては、課題で触れたように一貫性を計測する指標を定義し、文埋め込みや次元削減アルゴリズム間での性能を比較したり、カテゴリ化の閾値を最適化問題として解くことで与えるためのアルゴリズムを検討したりしていきたい。

5) <https://github.com/esafak/mca>

謝辞

メディア研究開発センターの研究アドバイザーである MBZUAI・東北大学の乾健太郎教授および東京工業大学の岡崎直観教授に感謝を申しあげる。

また、本論文の校正は朝日新聞社の AI 校正サービス「Typoless⁶⁾」の提供を受けて実施したものである。

参考文献

- [1] Hsiu-Fang Hsieh and Sarah Shannon. Three approaches to qualitative content analysis. **Qualitative health research**, Vol. 15, pp. 1277–88, 12 2005.
- [2] Jonah Berger, Ashlee Humphreys, Stephan Ludwig, Wendy W. Moe, Oded Netzer, and David A. Schweidel. Uniting the tribes: Using text for marketing insight. **Journal of Marketing**, Vol. 84, No. 1, pp. 1–25, 2020.
- [3] Yuxin Miao. Big data text mining techniques in journalism and communication. In **Proceedings of the 7th International Conference on Humanities and Social Science Research (ICHSSR 2021)**, pp. 256–259. Atlantis Press, 2021.
- [4] 樋口耕一. 社会調査のための計量テキスト分析 内容分析の継承と発展を目指して. ナカニシヤ出版, 2023.
- [5] Steven E Stemler. Content analysis. **Emerging trends in the social and behavioral sciences: An Interdisciplinary, Searchable, and Linkable Resource**, pp. 1–14, 2015.
- [6] Lei Guo, Chris Vargo, Zixuan Pan, Weicong Ding, and Prakash Ishwar. Big social data analytics in journalism and mass communication: Comparing dictionary-based text analysis and unsupervised topic modeling. **Journalism & Mass Communication Quarterly**, Vol. 93, , 04 2016.
- [7] 樋口耕一. 計量テキスト分析および kh coder の利用状況と展望. **社会学評論**, Vol. 68, No. 3, pp. 334–350, 2017.
- [8] 大畑和也藤本一男. 多重対応分析とアスペクトベース感情分析を組み合わせた 受講者満足度調査データの分析手法の開発. 言語処理学会 第 29 回年次大会 発表論文集, pp. 255–260, 2023.
- [9] Nils Reimers and Iryna Gurevych. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan, editors, **Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)**, pp. 3982–3992, Hong Kong, China, November 2019. Association for Computational Linguistics.
- [10] Tianyu Gao, Xingcheng Yao, and Danqi Chen. SimCSE: Simple contrastive learning of sentence embeddings. In Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih, editors, **Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing**, pp. 6894–6910, Online and Punta
- [11] Tomáš Musil. Examining structure of word embeddings with pca. In Kamil Ekštejn, editor, **Text, Speech, and Dialogue**, pp. 211–223, Cham, 2019. Springer International Publishing.
- [12] Tomáš Musil and David Mareček. Independent components of word embeddings represent semantic features. **arXiv preprint arXiv:2212.09580**, 2022.
- [13] Hiroaki Yamagiwa, Momose Oyama, and Hidetoshi Shimodaira. Discovering universal geometry in embeddings with ICA. In Houda Bouamor, Juan Pino, and Kalika Bali, editors, **Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing**, pp. 4647–4675, Singapore, December 2023. Association for Computational Linguistics.
- [14] Hayato Tsukagoshi, Ryohei Sasano, and Koichi Takeda. Japanese simcse technical report. **arXiv preprint arXiv:2310.19349**, 2023.
- [15] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. **Journal of Machine Learning Research**, Vol. 12, pp. 2825–2830, 2011.
- [16] M. Greenacre. Correspondence analysis. In Penelope Peterson, Eva Baker, and Barry McGaw, editors, **International Encyclopedia of Education (Third Edition)**, pp. 103–111. Elsevier, Oxford, third edition edition, 2010.
- [17] Jörg Blasius and Michael Greenacre. Multiple correspondence analysis and related methods. **Multiple Correspondence Analysis and Related Methods**, 06 2006.
- [18] 国土交通省観光庁. 宿泊旅行統計調査報告. <https://www.mlit.go.jp/kankocho/siryoutoukei/shukuhakutoukei.html>, 2018.
- [19] 経済産業省 中部経済産業局総務企画部企画調査課. 東海経済のポイント 2 0 2 3 (3 大経済圏比較). <http://www.globus.org/toolkit/>, 2023.
- [20] Maarten Grootendorst. Bertopic: Neural topic modeling with a class-based tf-idf procedure. **arXiv preprint arXiv:2203.05794**, 2022.

6) <https://typoless.asahi.com/>

A カテゴリ名一覧とプロンプト

A.1 カテゴリ命名に使ったプロンプト

カテゴリの命名には GPT-4 の API を用い、以下のシステムプロンプトおよびユーザープロンプトを使用した。BERTopic[20] で使われているプロンプト⁷⁾を参考にして、今回の用途に合わせて修正した。

システムプロンプト

You are a great supervisor of a professional economic analyst in Japan.

Additionally, you have written great reports.

So, you can review economic texts, topics, and so on.

You MUST answer in Japanese.

ユーザープロンプト

I have a topic that contains the following sentences:

- {Sentence1}

- {Sentence2}

- {... 文の個数だけ続く }

Based on the above information, can you give a short label of the topic?

In particular, please provide a brief and short topic label that focuses on "What happened in which industry?"

Topic Label Examples

物価上昇と給与停滞による消費低迷と景気後退の懸念

新型コロナウイルス影響下における観光・消費産業の景気低迷

公共工事の受注減少と建設業界の課題

半導体不足とサプライチェーン問題による自動車・

住宅・電化製品の販売と景気の悪化

食料品・農業原材料の価格高騰と円安影響下の農業および食品産業の収益性低下

Your Answer

A.2 カテゴリ名一覧

カテゴリの番号は独立成分の番号と図 1 での赤点に記されている番号に対応している。

カテゴリ 1 消費税増税の影響と連休の影響による小売・不動産・飲食業界の売上低迷

カテゴリ 2 新聞求人広告の衰退と製造・サービス業界における採用形態の変化及び人手不足の深刻化

カテゴリ 3 天候変動及び消費税増税影響下における農産物相場の変動と小売業界の売上低迷

カテゴリ 4 商店街の物販低迷と飲食業界の過当競争による地域経済の冷え込み

カテゴリ 5 消費増税と 10 連休影響による小売・宿泊業界の売上低迷

カテゴリ 6 宿泊業界における外国人観光客減少と連休効果の限界

カテゴリ 7 消費税増税と節約志向の強化による内需低迷と各産業への影響

カテゴリ 8 農業・建設・不動産業界を中心とした市場低迷と入札競争の激化

カテゴリ 9 新型コロナウイルス影響による宿泊・飲食業界の深刻な打撃

カテゴリ 10 消費税増税と各種税制変更影響下の自動車・住宅・携帯電話市場の販売低迷

カテゴリ 11 新型コロナウイルスと自然災害が引き起こす旅行・宿泊業界の深刻な予約キャンセルと売上低下

カテゴリ 12 消費低迷による小売業界の不振と地方経済の停滞

カテゴリ 13 販売量減少と客数の低下が続く小売業界の厳しい経営状況

カテゴリ 14 新型コロナウイルス影響下の多業界にわたる売上減少と消費活動の停滞

カテゴリ 15 製造業における受注・販売減少と構造的課題

カテゴリ 16 新聞広告の市場縮小と人材採用の手法変化による出版・広告産業の売上低迷

カテゴリ 17 暖冬・消費税増税の影響によるタクシー業界の売上低下と乗務員不足

カテゴリ 18 消費税増税と各種イベント影響による飲食業界の売上低迷

カテゴリ 19 新型コロナウイルスの影響を中心とした飲食業界および小売業界の厳しい市場環境と販売減少

カテゴリ 20 新型コロナウイルス影響下における小売・サービス業界の営業停止と売上減少

カテゴリ 21 消費減少と競争激化による小売業界の売上圧迫状況

カテゴリ 22 消費行動の鈍化とイベント効果減少による飲食・小売業界の売上減少傾向

カテゴリ 23 天候変動と消費税引上げ影響による農産物価格波動と消費動向の変容

カテゴリ 24 暖冬影響による冬物商材の売れ行き悪化と年末年始の消費低迷

カテゴリ 25 東日本大震災と台風被害後の商業・観光・農産業の長期的経済低迷及び新型コロナウイルスの影響

カテゴリ 26 観光産業における来客数減少とその影響

カテゴリ 27 新型コロナウイルスの影響と消費税増税による各産業の消費行動への変化

カテゴリ 28 米中貿易摩擦と消費税増税の影響による小売・消費関連産業の景気後退

カテゴリ 29 消費税増税と節約志向の拡大による小売・サービス業界の売上低迷および来客数減少

カテゴリ 30 消費税増税と米中貿易摩擦による国内外の消費と観光業界の不振

7) <https://maartengr.github.io/BERTopic/getting-started/representation/llm.html>