

# 行動分類のためのコーパス構築と行動分析への応用

中岡明義<sup>1</sup> 若宮翔子<sup>1</sup> 荒牧英治<sup>1</sup>

<sup>1</sup> 奈良先端科学技術大学院大学

{nakaoka.akiyoshi.nc0, wakamiya, aramaki}@is.naist.jp

## 概要

日常生活における行動の理解は、社会学、経済学や疫学などで重要な課題である。ソーシャルメディアの普及に伴い、人々の日々の行動に関わるテキストデータが蓄積されるようになり、行動分析のための材料として注目されている。本研究では、総務省統計局の社会生活基本調査で用いられている20種類の行動を基に日常生活における行動を定義した。これらの行動を「LIFE STORY」データセットにアノテーションし、公開可能なコーパスを構築した。このコーパスを用いて分類モデルを構築し、分類性能を評価した。さらに、構築したモデルを用いてCovid-19前後のエピソードから抽出した行動を定量的に分析し、コーパスの有用性を検証した。

## 1 はじめに

現代社会において、人々の日常生活の行動を把握し、分析することは、個人の幸福や健康の向上だけでなく、社会的および政策的課題の対処にも重要である。例えば、高齢者の生活習慣の分析は、高齢化社会における効果的な介護政策の設計に役立つ[1]。

近年、ソーシャルメディアやオンラインプラットフォームの普及に伴い、膨大なテキストデータが生成され、これらは日常生活の行動を把握するための有望な情報源となり得る。しかしながら、既存研究は、特定の行動に焦点を当てたものが多く、日常生活全般の行動を包括的に分析するための基盤は十分に整っていない。

本研究では、テキストから日常生活における行動を抽出する基盤の構築を目的とする。具体的には、総務省統計局の社会生活基本調査で用いられている20種類の行動を基に日常生活における行動を定義し、クラウドソーシングによって収集された「LIFE STORY」データセットにアノテーションし、公開可能なコーパスを構築した。このコーパスを用いて行動分類モデルを構築し、分類性能を基にコーパスの

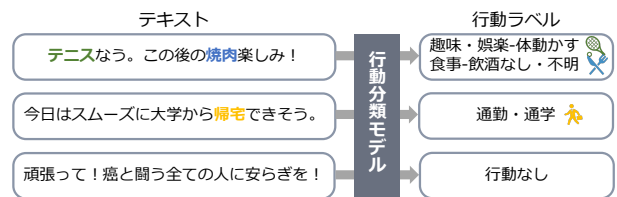


図1 テキストからの行動抽出

評価を行った。加えて、構築したモデルを用いた、Covid-19前後の行動変化を分析するケーススタディを実施し、日常行動の変化を定量的に捉えた。

図1に示すように、本研究で構築した行動分類モデルは、テキストから書き手の行動を抽出する。文脈の理解に優れた Bidirectional Encoder Representations from Transformers (BERT) [2] を基にモデルを構築することにより、より性能の高い行動分類が可能になると考えられる。

本研究で構築したコーパスは、個人や集団の生活習慣分析、地域特性の把握、時間変化に関する影響評価など、多様な応用研究に活用できる。これにより、社会的な課題の解決や政策立案に向けた新たな知見を提供することが期待される。

## 2 関連研究

テキストから行動を抽出するタスクに関する研究はこれまでも行われている。Banerjeeらは、英語のツイートを対象に形態素解析を行い、動作とカテゴリ、時間を表す単語の共起頻度に基づき行動を抽出した[3]。Nguyenらは、日本語のツイートを形態素解析を適用し、文章構造に着目することで、動作主、動作、対象、時間、位置といった動作の属性を抽出した[4]。これらの研究は、形態素解析を基に行われているため、文脈を深く考慮した抽出には限界があり、複雑な文構造に対応しきれない。Yanoらはn-gramとナイーブベイズ分類器を用いて、行動が含まれているツイートかを分類し、結果として、

表1 活動カテゴリと行動

活動カテゴリ	行動
一次活動 (I)	睡眠, 身の回りの用事, 食事-飲酒あり, 食事-飲酒なし・不明 (計4行動)
二次活動 (II)	通勤・通学, その他の移動, 仕事, 学業, 家事, 介護・看護, 育児 (計7行動)
三次活動 (III)	メディア, 買い物, 趣味・娯楽-体動かす, 趣味・娯楽-体動かさない, 交際-物理的接触, 交際-オンライン的接触, 受診・療養, 喫煙, その他の行動 (計9行動)
活動なし (IV)	行動なし

従来手法よりも Precision と Recall において高い性能を示した [5]. しかしながら, この手法も複雑な文脈や多様な表現に対応するには不十分であった.

以上の研究は, 形態素解析や統計的手法を中心に行われており, 文脈を十分に考慮した行動抽出には限界がある. 本研究では, 自然言語処理技術を用いて, 文脈を考慮した高性能なモデルを構築する. さらに, モデルの学習と評価に適したデータを提供するため, コーパスの構築も行う.

## 3 コーパス

### 3.1 材料

本研究では, LIFE STORY データセット<sup>1)</sup>[6]を使用した. LIFE STORY は, 著者らが 2017 年から開発・拡張しているオープンソース (CC-BY ライセンス) のコーパスである. 年に 4 回クラウドソーシングを行い, 7 つの感情 (悲しい, 不安, 怒り, 嫌悪感, 信頼感, 驚き, 楽しい) から想起される自由記述のエピソードを各回 1,000 人から収集して構築している.

本研究で構築するコーパスには, LIFE STORY の 1,500 件のテキストを使用した. これらのテキストは, 2023 年 2 月に収集した 1,000 件と 2023 年 5 月に収集した 500 件で構成されている. いずれも, 20 文字以上 100 文字以下のテキストを対象に, 行動を含むかどうかを分類するモデル [7] を適用し, 高い確率で行動が含まれていると判定されたテキストを抽出した.

### 3.2 行動の種類定義

本研究では, 総務省統計局が定める「一次活動」「二次活動」「三次活動」の枠組みに基づき, 行動を 4 つのカテゴリに分類する [8]. **一次活動**とは, 睡眠や食事などの生理的に必要な行動を指す. **二次活**

**動**とは, 仕事や家事などの社会生活を営む上で必要な行動を指す. **三次活動**とは, 趣味や娯楽などの自由に使える時間の行動を指す. 本研究では上記のカテゴリに加え, **活動なし**を設けた. これは, 感情や抽象的な思考といった行動に直接結びつかないテキストなどを対象とする. これにより, 網羅的なラベル付けが可能となる.

行動の詳細を表 1 に示す. さらに, 行動の特性をより詳細に分析するために, 「食事」を「飲酒あり」と「飲酒なし・不明」に分けるなどの行動の細分化も行った. これにより, 生活習慣や健康状態の関連性を分析する際の基盤として応用できる.

### 3.3 ガイドライン

本研究では, 以下のルールを設計した. 行動ラベルを付与したテキスト例を表 2 に示す.

**1 つの行動に 1 つのラベルを付与** テキスト内の各行動に 1 つのラベルを付与する. 具体例は表 2 の (1), (2) である. (1) では, 「ドラマを見る」「食べる」の記述に基づき, 「メディア」と「食事-飲酒なし・不明」のラベルを付与した. (2) では, 直接的な行動が含まれていないため, 「行動なし」のラベルを付与した.

**記述されている行動にのみラベルを付与** 明示的に記述されている行動のみを対象とし, 推測や背景情報に基づく行動にはラベルを付与しない. ただし, 「その他の行動」に関しては様々な行動が含まれるため, 主体的に行った行動に限定する. 具体例は表 2 の (3) である. (3) では「下山中」「娘と一緒に」の記述に基づき, 「趣味・娯楽-体動かす」と「交際・物理的接触」のラベルを付与した. 一方, 「道に迷った」は主体的な行動では無いため, ラベルを付与していない.

### 3.4 アノテーション

#### 3.4.1 人手でのラベル付け

ガイドラインに基づき, 2 人のアノテータがそれぞれ 100 件と 1000 件のデータに行動ラベルを付与した. このうち 2 名がアノテーションした 100 件を用いて, アノテーション一致度を算出する. アノテーション一致率は 0.95 (=95/100) であり, Cohen の  $\kappa$  係数で評価したところ, 0.93 と高い数値を得た.

アノテーション結果を図 2 に示す. 活動カテゴリ

1) <https://sociocom.naist.jp/life-story-data/>

表2 テキストと行動ラベルの例. 太字部分は行動ラベルの根拠となる行動を表す.

テキスト	行動ラベル
(1) <b>ドラマ</b> を見ながら好きなお菓子を <b>食べる</b> のが楽しい.	III: メディア I: 食事・飲酒なし・不明
(2) 気に入っていた腕時計が壊れて動かなくなったこと.	IV: 行動なし
(3) 八ヶ岳登山からの下山中に <b>娘と一緒</b> に道に迷った.	III: 趣味・娯楽・体動かす III: 交際・物理的接触

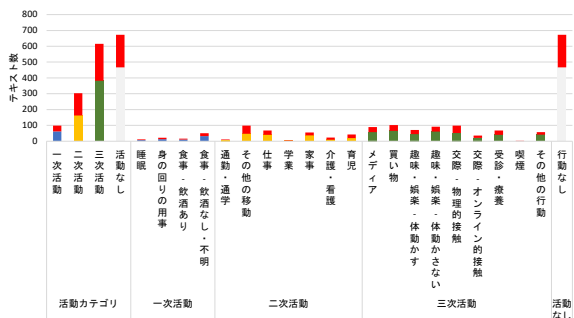


図2 活動カテゴリ, 行動ごとのテキスト数. 青色が「一次活動」, 黄色が「二次活動」, 緑色が「三次活動」, 灰色が「活動なし」を表す. 赤色は GPT-4o を用いて作成した拡張データを表す.

に注目すると、「活動なし」の次に「三次活動」のテキスト数が多いことが分かる. これは, 日常生活を記述する際に, 感情や思考などの「活動なし」や娯楽に関連する「三次活動」が特に多く記録される傾向を示している.

### 3.4.2 LLM でのラベル付け

人手によるアノテーションは高コストであるため, 本研究では LLM でラベル付けを行い, データの拡張を行った. その際, GPT-4o [9] を用いて 500 件のテキストにラベル付けを行った. GPT-4o には, アノテーションガイドラインと 21 種類の行動ラベルに対応する具体例を含むプロンプトを与え, 3-shot 形式で多様な文脈における行動を正確に分類するよう設計した (詳細は付録 A 参照).

LLM によるラベル付けの有用性を評価するために, 100 件のデータにおいて人手と LLM によるアノテーション一致度を算出した. 結果として, 一致率は 0.48 (=48/100) であり, Cohen の  $\kappa$  係数で評価したところ, 0.45 と中程度の一致度が得られた. これは, LLM によるラベル付けが完全に人手の性能に匹敵するわけではないものの, 一定の一貫性を持つことを示唆している.

結果として, 図 2 に示すように, 4 つの活動カテゴリ全てで, テキスト数が増加した. 特に, 三次活動の行動を含むテキストが顕著に増えていることが確認された.

表3 ファインチューニング時のハイパーパラメータ.

設定	パラメータ	値
共通	max length	128
	batch size	16
	optimizer	AdamW [10]
	学習率	$2.0 \times 10^{-5}$
	エポック数	100
2 クラス分類	early stopping	3 patience
	損失関数	Weighted Cross Entropy Loss
マルチラベル分類	early stopping	10 patience
	損失関数	WBCE With Logits Loss

## 4 実験

### 4.1 実験設定

次の 3 つのタスクに対応する分類モデルを構築し, その性能を基に 3 章で構築したコーパスの質を評価した. テキストを「活動あり」「活動なし」に分類する 2 クラス分類, 活動カテゴリに分類する 4 クラスマルチラベル分類, 日常生活における 21 種類の行動に分類する 21 クラスマルチラベル分類である. 分類モデルには, 東北大学の事前学習済み BERT モデルである cl-tohoku/bert-base-japanese-v3<sup>2)</sup> を使用し, 表 3 に示すハイパーパラメータで学習を行った. コーパスを 9:1 の割合で学習・テストデータに分割し, 学習データを 5 分割して交差検証を実施した. ラベル分布を維持しつつ分割を行い, 5 通りのモデルの性能平均を最終評価とした. 不均衡データへの対応として, Weighted Binary Cross Entropy (WBCE) [11] を誤差関数に採用し, ラベル毎の重みを以下の式で計算した.

$$\text{classweight} = \frac{n_{\text{samples}}}{n_{\text{samples}_{\text{class}}} \cdot n_{\text{classes}}} \quad (1)$$

ここで,  $n_{\text{samples}}$  はデータの総数,  $n_{\text{samples}_{\text{class}}}$  は各ラベルのデータの総数,  $n_{\text{classes}}$  はラベルの総数を表す. 評価指標には, Accuracy と Macro-F1 を用いた.

2) <https://huggingface.co/tohoku-nlp/bert-base-japanese-v3>

**表 4** 3つのタスクにおける各モデルの分類性能. 太字は各タスクでの Accuracy, Macro-F1 の最大値を示す. w/ DataAug は GPT-4o を用いてデータ拡張を行なったコーパスで学習したモデルを指す.

タスク	モデル	Accuracy	Macro-F1
2 クラス分類	Majority Baseline	0.53	0.346
	BERT	0.762	0.761
	BERT w/ DataAug	<b>0.786</b>	<b>0.785</b>
4 クラスマルチラベル分類	Majority Baseline	0.463	0.158
	BERT	0.627	0.504
	BERT w/ DataAug	<b>0.67</b>	<b>0.557</b>
21 クラスマルチラベル分類	Majority Baseline	0.46	0.032
	BERT	0.744	0.421
	BERT w/ DataAug	<b>0.756</b>	<b>0.469</b>

## 4.2 結果と考察

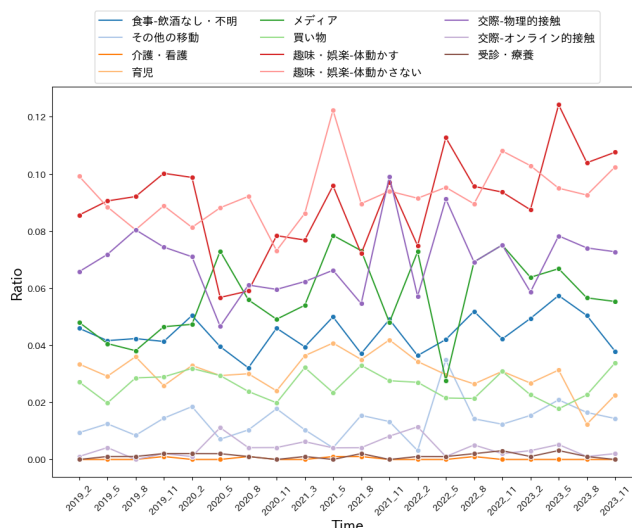
分類モデルの性能を表 4 に示す. 本実験ではベースラインモデルとして Majority Baseline を使用している. 全てのタスクにおいて構築した分類モデルが Majority Baseline より高い性能を示した. さらに, GPT-4o を用いたデータ拡張 (w/ DataAug) により, 全タスクで性能が向上した. 特に, ラベル数が多いタスクほど性能向上が顕著であり, データ拡張の有効性が確認された.

## 5 ケーススタディ

コーパスの有用性を示すために, Covid-19 前後における人々の行動を定量的に比較し, どのような行動変化をもたらしたのかを分析する.

### 5.1 設定

構築したモデルを用いてテキストに行動ラベルを付与し, 取得時期ごとの行動割合を算出して時系列変化を分析する. 2019 年 2 月~2023 年 11 月に収集された LIFE STORY の 2 万件の Joy データ (年 4 回, 各回 1,000 件) を用いた. 行動を含む記述が豊富であると考え, Joy データを選定した. モデルには, 4.2 節で 21 クラスマルチラベル分類タスクにおいて Macro-F1 が最も高かった BERT w/ DataAug を採用した. 分析する行動は, 分類結果の F1 値が 0.5 を超えた 11 種類とした. これらの行動は, 食事-飲酒なし・不明, その他の移動, 介護・看護, 育児, メディア, 買い物, 趣味・娯楽-体動かす, 趣味・娯楽-体動かさない, 交際-物理的接触, 交際-オンライン的接触, 受信・療養である (表 5). これらの行動ラベルを付与したデータを用い, 取得時期ごとの行動割合



**図 3** 行動割合の時系列変化 (2019 年 2 月~2023 年 11 月). 母数は欠損を除いた各取得時期のテキストの総数.

を算出し, 行動変化の傾向を定量的に分析した.

### 5.2 結果と考察

図 3 に 2019 年 2 月~2023 年 11 月までの 11 種類の行動割合の変化を示す. 結果として, Covid-19 の影響が顕著であった 2020 年 5 月を中心に, 行動に大きな変化が見られた. 「趣味・娯楽-体動かす」と「交際-物理的接触」がそれぞれ約 50%と約 40%減少した一方で, 「メディア」は約 45%増加した. この結果は, 行動制限や外出自粛が日常生活に与えた影響を反映しており, 特にメディア使用の増加はオンライン活動への移行を示唆している. また, 趣味や交際の減少は身体的活動や対面接触の制限を反映し, パンデミックによる行動変化を表している.

## 6 おわりに

これまでの多くの行動抽出に関する研究は, タスク固有の分類を解いていたが, 一次活動から三次活動という人間の基本的な行動レベルで, 汎用性のある分類を行った研究は少ない. 本研究では, 一次活動から三次活動に至る行動分類を目的に, コーパスと分類モデルを構築し, ケーススタディにより応用可能性を確認した.

今後の発展としては, ソーシャルメディアテキストなど異なるドメインへの適用可能性の検証や, 行動分類性能のさらなる向上が挙げられる. さらに, 構築したコーパスとモデルを活用し, 生活習慣分析, 地域特性の把握, 時間変化の影響評価など, 多様な応用研究に発展させることが期待される.

## 謝辞

本研究は、「戦略的イノベーション創造プログラム (SIP)」「統合型ヘルスケアシステムの構築」JPJ012425, JST SICORP, JPMJSC2107 および JPMJSC23C6, JSPS 科研費 JP22K12041 の支援を受けたものである。

## 参考文献

- [1] Shahla Ayoubi-Mahani, Maryam Eghbali-Babadi, Ziba Farajzadegan, Mahrokh Keshvari, and Jamileh Farokhzadian. Active aging needs from the perspectives of older adults and geriatric experts: a qualitative study. **Frontiers in Public Health**, 2023.
- [2] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. **arXiv preprint arXiv:1810.04805**, 2018.
- [3] Nilanjan Banerjee, Dipanjan Chakraborty, Koustuv Dasgupta, Sumit Mittal, Anupam Joshi, Seema Nagar, Angshu Rai, and Sameer Madan. User interests in social media sites: an exploration with micro-blogs. In **Proceedings of the 18th ACM Conference on Information and Knowledge Management**, 2009.
- [4] The-Minh Nguyen, Takahiro Kawamura, Yasuyuki Tahara, and Akihiko Ohsuga. Self-supervised capturing of users' activities from weblogs. **International Journal of Intelligent Information and Database Systems**, 2012.
- [5] Yuji Yano, Tomonori Hashiyama, Junko Ichino, and Shun'ichi Tano. Behavior extraction from tweets using character n-gram models. In **2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)**. IEEE, 2014.
- [6] 柴田大作, 若宮翔子, 伊藤薫, 荒牧英治. JIWC: クラウドソーシングによる日本語感情表現辞書の構築. 言語処理学会 第 23 回年次大会 発表論文集, 2017.
- [7] 中岡明義, 若宮翔子, 荒牧英治. ソーシャルメディアテキストからのユーザ行動の分類. 第 16 回 DEIM フォーラム論文集, 2024.
- [8] 統計局. 社会生活基本調査における 1 日の行動の種類とその説明 (行動調査), 2001. アクセス日: 2024 年 1 月 9 日.
- [9] OpenAI. Hello gpt-4o, 2024.
- [10] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. **arXiv preprint arXiv:1412.6980**, 2014.
- [11] Sophie Henning, William Beluch, Alexander Fraser, and Annemarie Friedrich. A survey of methods for addressing class imbalance in deep-learning based natural language processing. In **Proc. of the 17th Conference of the European Chapter of the Association for Computational Linguistics**, pp. 523–540, Dubrovnik, Croatia, May 2023. Association for Computational Linguistics.

## A プロンプトの詳細

GPT-4o を用いたラベル付けを行なった。GPT-4o には、アノテーションガイドラインと 21 種類の行動ラベルに対応する具体例を含むプロンプトを与え、3-shot 形式で多様な文脈における行動を正確に分類するよう設計した。Prompt 例は次のようになっている。

Prompt 例

質問: テキストを次のルールに従ってマルチラベルに分類する。

ルール

1. テキスト中に含まれるテキスト作成者が行った行動に以下の行動を割り当てる。
2. テキスト中に複数のテキスト作成者が行った行動がある場合は複数の行動を割り当てる。
3. どの行動もテキストに含まれない場合に行動なしを割り当てる。

行動の種類: 以下に行動の種類と各行動の例を示す。行動の種類は 21 種類です。

睡眠

- ・例: 夜間の睡眠, 昼寝, 仮眠
- ...

以下にマルチラベル分類の例をのせる。

要求

次の「」中のテキストに対してルールに基づいてマルチラベル分類せよ。「テニスなう。この後の焼肉楽しみ。」

応答

- 趣味・娯楽-体動かす, 食事-飲酒なし・不明
- ...

要求

次の「」中のテキストに対してルールに基づいてマルチラベル分類せよ。

表 5 2つの分類モデルによる性能比較。  
太字は行動ラベル毎の F1 値の最大値を表している。

モデル	BERT w/ DataAug	GPT-4o
睡眠	0.0	<b>0.714</b>
身の回りの用事	0.133	<b>0.591</b>
食事-飲酒あり	0.0	<b>0.706</b>
食事-飲酒なし・不明	0.583	<b>0.659</b>
通勤・通学	0.0	<b>0.762</b>
その他の移動	<b>0.857</b>	0.602
仕事	0.495	<b>0.561</b>
学業	0.0	0.0
家事	0.13	<b>0.348</b>
介護・看護	<b>0.66</b>	0.311
育児	<b>0.683</b>	0.554
メディア	<b>0.799</b>	0.774
買い物	<b>0.861</b>	0.731
趣味・娯楽-体動かす	<b>0.762</b>	0.629
趣味・娯楽-体動かさない	0.635	<b>0.636</b>
交際-物理的接触	<b>0.519</b>	0.471
交際-オンライン的接触	<b>0.653</b>	0.468
受診・療養	<b>0.861</b>	0.615
喫煙	0.0	0.0
その他の行動	0.36	<b>0.374</b>
行動なし	<b>0.853</b>	0.591

## B 行動ラベルごとの分類性能

21 クラスマルチラベル分類タスクにおいて Macro-F1 が最も高かった BERT w/ DataAug と、GPT-4o の分類性能を比較した。表 5 に各モデルのラベルごとの F1 値を示す。三次活動においては、BERT w/ DataAug が GPT-4o よりも高い分類性能を示した。これは、三次活動に関連するデータ数が他の活動に比べて多いことが影響していると考えられる。一方、データ数が比較的少ない一次活動および二次活動では、GPT-4o の方が優れた分類性能を示した。これらの結果から、データ分布の偏りがモデル性能に与える影響が示唆される。特に、一次活動および二次活動に関連するデータを増加させることで、BERT w/ DataAug の性能向上が期待される。