

行動関連語によるマイクロ行動の分散表現の検討

A Study of Distributed Representation of Micro Activity by Action-Related Words

笹崎 海利[†] 松井 智一[†] 諏訪 博彦[†] 安本 慶一[†]
Kairi Sasazaki Tomokazu Matsui Hirohiko Suwa Keiichi Yasumoto

1. はじめに

近年, IoT 技術の発展に伴い住宅内環境のスマート化が進んでいる. とりわけ, 在宅中の居住者の行動を認識する技術については, 居住者の生活行動を認識することで, より利便性の高い宅内サービスの提供が可能になるとして注目されている. 宅内行動を認識することで, 家電の省エネ化や高齢者の見守りの役割を果たすことが可能になる. 家電制御については, 状況に応じた自律化や最適化が可能になるほか, 独居中の高齢者に対しては転倒や夜中の徘徊などといった異常な行動を検知することが可能となり, 非常時の対応を早めることが期待できる.

これらの宅内サービスの利便性を高めるには, 正確な行動認識技術が必要になり, そのためには行動をより詳細に認識することが重要になる. しかし, 従来の行動認識手法では, 「料理」「食事」等の「マクロ行動」と呼ばれる行動を推論することを目指している. そのため実際にはマクロ行動では詳細な行動を認識するには不十分でありマクロ行動を細分化した「マイクロ行動」を認識する必要がある. ただし, このようなマイクロ行動認識においては認識すべき行動が無数に定義可能であることから, 機械学習における他クラス分類での問題定義が難しいことが課題として挙げられる. これは推論すべき正解ラベルについて設定すべき粒度の決定が困難であることに起因する. 多くのマクロ行動は, マイクロ行動へ細分化が可能であり, その識別クラスの粒度はあらかじめ設定しておく必要がある.

また, これらの行動認識用の識別クラスは, 一般的には機械学習モデルの設計者が定義するものである. しかし, 行動認識結果を実サービスへ応用する場合に, サービスの提供者が求めるクラス分類が機械学習モデルに定義されているとは限らない. これまでのクラス分類では, One-hot エンコーディングのように行動データと正解ラベルを一对一で対応させるベクトル表現を用いてきた. そのため, 行動認識モデルをサービスへ実利用へ拡張するには, 新たなデータセットでの再学習かファインチューニングが必要になる.

本研究の目的は, 行動を演算可能な分散表現として扱い, サービスの提供者が行動を認識し, 居住者の行動の

意図や流れを読み解くことができるコンテキストアウェアな宅内サービスの提供を実現することである. 行動の分散表現を獲得するために, テキストの分散表現のアイデアをマイクロ行動のクラス分類へ応用し, 日常生活中に発生する会話の文脈や文章ベクトルを基にマイクロ行動の分散表現を構築する. 本稿では, 発話と行動の擬似データセットを LLM を用いて生成する手法を開発し, 擬似データセットを用いて行動の分散表現を獲得する手法について検討した内容について報告する.

本稿の構成は, 次のとおりである. 2 章では, 関連した既存研究を紹介し, 本研究の位置付けを明らかにする. 3 章では提案する LLM を利用した言語-行動エンコーダの開発手法について説明し, 4 章ではこれを実現するための実験計画について述べる. 5 章では, 本稿の結論と今後の展望について述べる.

2. 関連研究

本章では, 宅内における日常生活行動認識と教師なし学習による行動認識について関連する既存研究について調査した結果を示し, 本研究の意義について述べる.

2.1 宅内における日常生活行動認識に関する研究

Balli ら [1] はスマートウォッチに搭載された IMU, 歩数計心拍センサから得られるデータを主成分分析による次元削減を行なった上で, RandomForest を用いて「歩く」, 「座る」, 「PC 操作」, 「静止状態」など 8 種類の動作に識別した. 結果として, F 値にして 98.5% の精度で分類可能であることを示している.

Matsui ら [2] は一般家庭においても行動認識技術を利用すべくプライバシーを考慮したセンシングシステムを提案している. 非侵襲的なエナジーハーベスティングで動作するセンサとアノテーションボタンから構成されたこのシステムを用いてデータ収集を行い, 「食事」, 「調理」, 「入浴」, 「睡眠」, 「外出」の 5 種類の行動について LSTM モデルで行動認識を行った結果として再現率で 72.3% の精度で識別している.

Rai ら [3] は複数視点の映像と音声, 環境センサデータで構成されたマルチモーダルなデータセットを提案している. 行動の階層構造に着目し, 行動を撮影した映像にアノテーションを行うだけでなく映像中に見られる一つ一つの細かい行動に対してもアノテーションを行うこ

[†] 奈良先端科学技術大学院大学, Nara Institute of Science and Technology

とで、推論結果を改善できることを示している。

Chatterjee ら [4] は IMU による行動認識のために音響センサを用いた自動的なアノテーション方法を提案している。教師データを作成するアノテーションタスクを行う際に、ウェアラブルデバイスに IMU とともに搭載されている音響センサを用いてエッジ上でアノテーションを行うことで、プライバシーに配慮しつつ教師データを作成することを可能にした。

2.2 教師なし学習による行動認識に関する研究

手作業によるアクティビティ定義とデータ収集の必要性をなくすためにゼロショット学習が提案されている [5-7, 7-10]。ゼロショット学習の目的は学習データに表れない未知のデータに対して学習済みの知識を応用することである。教師あり学習では全てのクラスの正解データを集める必要があるが、ゼロショット学習ではその必要なく未知のデータに対しても推論が可能である。

センサデータを利用した行動認識に属性の埋め込み空間を利用した研究 [11, 12] がある。センサデータから抽出した特徴空間を属性空間へ射影するモデルに CNN や MLP を採用している。意味空間には属性空間を採用しているが、属性の重要度という概念を導入し異なるクラスからのインスタンスに対しては属性において異なる反応を示すものとなっている。

Tong ら [13] は生活行動を取めた動画から意味空間を取り出し、IMU ベースのゼロショット学習による行動認識手法を提案した。IMU の特徴を区別するために重要な動きについて情報を持った動画画像を利用して未知のクラスについても分類を可能にしている。

Matsuki ら [14] は単語の埋め込み空間を利用したゼロショット学習による活動認識のための意味空間を提案している。センサから得た特徴ベクトルと単語の埋め込みを意味ベクトルとして一対一に対応づけることで未知の行動クラスに対しても識別が可能であると示している。

Xia ら [15] は ChatGPT* を利用した教師なし行動認識手法を提案している。行動時に接触するオブジェクトのリストから行動を推定するタスクについて情報量の多寡で精度を比較した。単にオブジェクトのリストだけを与える場合、データセットや wikipedia からの文章を共に与える場合、そして ChatGPT に事前に生成させたオブジェクトと対応するアクティビティを学習させ知識文を共に与える場合の 3 パターンで比較を行っている。生成した知識文を用いた場合の結果が最良の結果を示したことから、ChatGPT における 2 段階のプロンプトエンジニアリングの有効性を示している。

*OpenAI: ChatGPT, <https://openai.com/blog/chatgpt> (accessed 2023-07-20)

2.3 本研究の位置付け

これまで既存研究について述べてきたが、多くの研究において、行動の粒度が「粗い」マクロ行動認識に留まっており、生活行動中のコンテキストを考慮することができていない。また、マイクロ行動認識を行う研究もあるが対象領域が限定的であることや正解ラベルのクラス数が定まらない等の課題がある。これに対して、未知の正解ラベルを推論することができるゼロショット学習による行動認識については動画像と単語の埋め込み空間を利用した手法を紹介したが、文章表現を利用するものは少ない。これらの研究は、未知の行動を推論することを目的としており、推論結果の解釈および実生活中的のサービスにおける活用を前提としているものではない。これらのことから、本研究は文章の分散表現を参考に行動についても演算が可能な分散表現を獲得する手法を開発する。従来の離散的で行動認識結果とは異なり演算により結果を解釈することが可能になるマイクロ行動認識手法であると位置付けることができる。

3. 提案手法

本章では、提案手法の概要とその構成について述べる。本研究では行動の分散表現を獲得するため宅内での発話に関する擬似データ生成と発話から行動を推論する言語-行動エンコーダを開発する。

3.1 概要

提案手法の概要を図 1 に示す。ChatGPT に代表されるような LLM (Large Language Model) を活用し、後の 3.2 節で詳細について述べる網羅的な行動環境 (マクロ行動、居住者属性、住環境等) を入力とした生活行動中の発話に関する擬似データセットを作成する。データセット構築のために利用するマクロ行動は、佐々木らの提案 [16] に基づいて決定し、各行動それぞれについて付随する行動環境のパラメータをランダムで決定し LLM に対する入力とする。行動に関連する発話の擬似データセットを作成した後、事前学習済み BERT [17] や fastText [18] などの言語処理モデルを用いて、テキストから行動や居住者属性、環境を推論する言語-行動エンコーダを構築する。最終的にこのエンコーダの中間層を取り出すことで、発話からベクトルにエンコードされた行動を抽出することができる。これを行動の分散表現とすることで認識結果を計算可能な形で解釈可能な新しいマイクロ行動認識の概念を提案する。

3.2 LLM を用いた会話の擬似データセット生成

生活行動は共通理解を得るための言語化が行われやすく、文化的な背景の影響も受けるものと捉えることができる。そのため、言語によって大規模かつ網羅的に行動

会話の擬似データ生成による行動の分散表現化

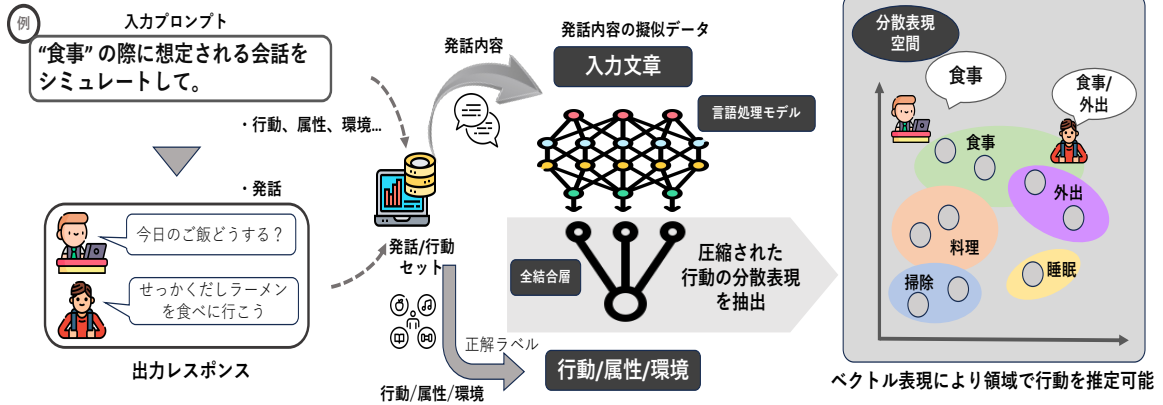


図 1: 提案手法の概要

表 1: 宅内行動一覧

行動名	行動の定義
洗顔	洗面所で顔を洗う
入浴	浴室で体を洗い湯船に浸かる
料理	キッチンで食材を調理する
食事	料理を食べる
食器洗い	食器を洗う
トイレ	トイレで用を足す
部屋掃除	部屋を掃除機で掃除する
風呂掃除	浴室を掃除する
洗濯	衣類を洗濯機に入れる
PC 作業	PC で作業する
スマートフォン操作	スマートフォンを使う
ゲーム	ビデオゲームをする
読書	本を読む
TV 視聴	TV を視聴する
睡眠	ベッドで睡眠をとる
身支度	外出の準備をする

を表現するデータを獲得することで、あらゆる生活行動に対応した行動の分散表現を構築できると考えられる。生活中におけるマイクロ行動はほぼ全て自然言語で説明可能であるという仮説のもと、言語に基づくマイクロ行動の擬似データを生成する。擬似データの生成には実際に人の会話を生活実験中から取得することが望ましいが、コストが膨大になるため現実的ではない。これを受け、人の会話をシミュレート可能で大量にデータを収集してもコストが抑えられる LLM を利用した。

具体的な LLM に対する入力として表 1 に示す佐々木

らが定義した 17 のマクロ行動のうちの 1 つと、会話をする人数、話者の年齢、居住地区、住環境、そして社会属性を与える。入力として与えるプロンプトを制御するパラメータとその例については表 2 に示す通りである。それぞれのパラメータを入力に用いる理由を述べる。マクロ行動はそれぞれに関わる発話を得るために必須である。会話の人数は 1 人の場合ホームアシスタントなどの人ではない相手であり 2 人以上であれば人である可能性が高いため会話文中における表現が変わる可能性があると考えたためである。話者の年齢を含めた理由は言葉遣いや発話内容に差異があると考えたためであり、居住地区を含めた理由は出現する単語に違いが出る可能性があると考えたためである。住環境についてはアパートでは宅内に階段がないように、行動に制約を生む可能性があり、社会属性は発話内容に持っている知識が影響を与える可能性を考慮したためである。入力するプロンプトは前述のパラメータを自動で変化させながら生成する。

図 2 に示すように、これらの入力を GPT に与えその状況下で発生する会話をシミュレートさせる。評価のために擬似データとして獲得された発話内容が人の発話内容とどれだけ異なっているかを確かめる 2 群比較の検定を実施する。

3.3 言語-行動エンコーダの構築

LLM (GPT-3.5-turbo) を利用して作成した擬似データの発話内容を言語処理モデルに与えるため次の手順を行う。

1. 前処理として入力文から不要な記号や絵文字、ストップワードを取り除く。その後、日本語形態素解析システム Juman++ を利用して形態素解析を行い N 個の単語列を取得する。

入カプロンプト

あなたは**70歳の高齢者**で、**郊外の一軒家**に住んでいます。
あなたは今自宅にいてこれから**食事**をしようとしています。そのときの会話を再現してください。
会話に登場する人数は合計**4人**で全員あなたと同居しています。
彼らの社会属性は**医療従事者**と**医療従事者**と**医療従事者**であなたとの関係は社会属性から推測してください。

出カレスポンスデータ

あなた：「今日は何作る？」
医療従事者A：「何でもいいよ。ちょっとスパイスの効いたメキシカン料理とかどう？」
あなた：「いいね、それならタコスとグアカモーレを作ろうかな」
医療従事者B：「それにチリコンカンも合うかも。私が作るよ」
あなた：「それなら私はブラウニーを焼くね」
医療従事者C：「みんなに差し入れしてくれるの？嬉しいな」
あなた：「もちろん、一緒に食べよう！」

図 2: GPT に対するプロンプトとレスポンス。プロンプト中の太字斜体部分はパラメータ

表 2: LLM に対するプロンプトを制御するパラメータ

パラメータ	入力例
マクロ行動	表 1 に示す 17 行動
人数	1 人, 2 人, 3 人, 4 人
話者の年齢	社会属性に応じた年齢
居住地区	都市部, 郊外, 田舎
住環境	アパート, 一軒家, 寮, マンション
社会属性	幼児, 児童, 学生, 会社員, 公務員, 医療従事者, 事務員, 高齢者

- 解析結果として取得した単語列を言語処理モデルに与える。ここでは入力に発話内容を与え、出力には擬似データ生成の際に入力として与えたパラメータを推論する。
- 中間層を取り出すことでベクトル空間に埋め込まれた発話として行動の分散表現を獲得する。

4. 実験概要・計画

本章では、提案手法の評価に必要な実験について述べる。本研究の手法は LLM を用いて作成した発話の擬似データセットを利用し行動の分散表現を獲得するものである。言語-行動エンコーダの品質は擬似データの品質に影響を受けるため、作成した擬似データセットの品質検証が必要となる。

- LLM で生成された発話内容と人の発話内容にどれだけの差異が生じているのかについて検証する実験を行う。定性的なアンケートを基に人が日常の生活中で発生する会話であるかどうかを検証し、擬似データセットの精度を確かめる。アンケートでは GPT

によって出力された発話内容が人による発話と思うか、そうでないかの 2 つの選択肢を選ぶ。このとき、人の発話としている選択肢は生活行動を撮影した動画などを提示し、その行動中に起こり得る発話のことであり、クラウドソーシング等を利用して収集する。アンケートによって得られた交差集計表を基に 2 群間の有意差の有無について統計的に検証を行う。

- 生成した擬似データセットの発話内容が実際に行動を表現できているかを検証する実験を行う。作成した言語-行動エンコーダで生成された行動のベクトル表現から行動が所属するクラスを可視化し、発話が行動を表現できているかを定性的に検討する。行動に関する情報が埋め込まれたベクトル表現は 300 次元以上の高次元で表現されることが多い。このベクトルに対して次元削減を行い 2 次元に落とし込んだものを 2 次元マップに散布図として射影し、行動毎に発話内容がクラス分類できるかどうかを検証する。クラスターリングは k-means 法等の教師なしアルゴリズムを用いて行う。

5. おわりに

本稿では、生活中行動認識の結果を宅内サービスへ活用することを念頭に、マイクロ行動の分散表現を開発する手法を提案した。言語処理における分散表現のアイデアをマイクロ行動のクラス分類へ応用し、日常生活中に発生する会話を基にマイクロ行動の分散表現を構築することが可能であると考えられる。

今後の展望として、本研究で構築した言語-行動エンコーダを用いた宅内行動認識手法を開発する。そのために、奈良先端科学技術大学院大学構内にあるスマート

ホームを利用して人の宅内行動をセンサ群によって認識する生活実験を実施する。発話から行動を認識するため被験者は行動中に発話により自身の行動の説明などを行う。センサデータと行動の説明を対応づけることで発話を利用してマイクロ行動にアノテーションすることが可能になる。これにより、センサデータのクラス分類の推論が可能となり、その上で推論結果をベクトル表現として解釈することが可能になるため実サービスへの利活用を実現する。

謝辞

本研究は JSPS 特別研究員奨励費 JP22KJ2287 の助成を受けて行った。

参考文献

- [1] Serkan Balli, Ensar Arif Sağbaş, and Musa Peker. Human activity recognition from smart watch sensor data using a hybrid of principal component analysis and random forest algorithm. *Measurement and Control*, Vol. 52, No. 1-2, pp. 37–45, 2019.
- [2] Tomokazu Matsui, Kosei Onishi, Shinya Misaki, Manato Fujimoto, Hirohiko Suwa, and Keiichi Yasumoto. Salon: Simplified sensing system for activity of daily living in ordinary home. *Sensors*, Vol. 20, No. 17, p. 4895, 2020.
- [3] Nishant Rai, Haofeng Chen, Jingwei Ji, Rishi Desai, Kazuki Kozuka, Shun Ishizaka, Ehsan Adeli, and Juan Carlos Niebles. Home action genome: Cooperative compositional action understanding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11184–11193, 2021.
- [4] Soumyajit Chatterjee, Avijoy Chakma, Aryya Gangopadhyay, Nirmalya Roy, Bivas Mitra, and Sandip Chakraborty. Laso: Exploiting locomotive and acoustic signatures over the edge to annotate imu data for human activity recognition. In *Proceedings of the 2020 International Conference on Multimodal Interaction*, pp. 333–342, 2020.
- [5] Ziming Zhang and Venkatesh Saligrama. Zero-shot recognition via structured prediction. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part VII 14*, pp. 533–548. Springer, 2016.
- [6] Yanwei Fu, Yongxin Yang, Tim Hospedales, Tao Xiang, and Shaogang Gong. Transductive multi-label zero-shot learning. *arXiv preprint arXiv:1503.07790*, 2015.
- [7] Mohammad Norouzi, Tomas Mikolov, Samy Bengio, Yoram Singer, Jonathon Shlens, Andrea Frome, Greg S Corrado, and Jeffrey Dean. Zero-shot learning by convex combination of semantic embeddings. *arXiv preprint arXiv:1312.5650*, 2013.
- [8] Zeynep Akata, Scott Reed, Daniel Walter, Honglak Lee, and Bernt Schiele. Evaluation of output embeddings for fine-grained image classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2927–2936, 2015.
- [9] Maxime Bucher, Stéphane Herbin, and Frédéric Jurie. Improving semantic embedding consistency by metric learning for zero-shot classification. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part V 14*, pp. 730–746. Springer, 2016.
- [10] Xun Xu, Timothy Hospedales, and Shaogang Gong. Semantic embedding space for zero-shot action recognition. In *2015 IEEE International Conference on Image Processing (ICIP)*, pp. 63–67. IEEE, 2015.
- [11] Wei Wang, Chunyan Miao, and Shuji Hao. Zero-shot human activity recognition via nonlinear compatibility based method. In *Proceedings of the International Conference on Web Intelligence*, pp. 322–330, 2017.
- [12] Hiroki Ohashi, Mohammad Al-Naser, Sheraz Ahmed, Katsuyuki Nakamura, Takuto Sato, and Andreas Dengel. Attributes’ importance for zero-shot pose-classification based on wearable sensors. *Sensors*, Vol. 18, No. 8, p. 2485, 2018.
- [13] Catherine Tong, Jinchun Ge, and Nicholas D Lane. Zero-shot learning for imu-based activity recognition using video embeddings. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, Vol. 5, No. 4, pp. 1–23, 2021.

- [14] Moe Matsuki, Paula Lago, and Sozo Inoue. Characterizing word embeddings for zero-shot sensor-based human activity recognition. *Sensors*, Vol. 19, No. 22, p. 5043, 2019.
- [15] Qingxin Xia, Takuya Maekawa, and Takahiro Hara. Unsupervised human activity recognition through two-stage prompting with chatgpt. *arXiv preprint arXiv:2306.02140*, 2023.
- [16] 佐々木渉, 大西晃正, 三崎慎也, 諏訪博彦, 藤本まなと, 水本旭洋, 荒川豊, 木村亜紀, 三木智子, 安本慶一. 生活の質の向上を目指した宅内行動・生体情報収集システムと qol アウェア家電制御の検討. 人工知能学会第二種研究会資料, Vol. 2019, No. SAI-034, p. 01, 2019.
- [17] 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.
- [18] Armand Joulin, Edouard Grave, Piotr Bojanowski, Matthijs Douze, H erve J egou, and Tomas Mikolov. Fasttext. zip: Compressing text classification models. *arXiv preprint arXiv:1612.03651*, 2016.