

## 未学習行動推定のための Zero-shot 学習法における精度向上の試み

## Toward Improving Zero-shot Learning with Semantic Word Vectors for Activity Recognition

松木 萌<sup>\*1</sup> 井上 創造<sup>\*1</sup>

Moe Matsuki Sozo Inoue

<sup>\*1</sup>九州工業大学大学院

Kyushu Institute of Technology

Mobile Activity recognition is important for healthcare and care-life of elderly people. However, there is a problem that the method can't recognize activities which don't appear in the training data. We propose the Zero-shot learning with semantic word vectors to overcome the problem. In this paper, we evaluate the proposed method by using 5 supervised learning method to improve the accuracy. As a result, the accuracy of the unseen classes improved than naive methods, and identified several challenges.

## 1. はじめに

近年、人の行動を推定する行動認識技術が IoT (Internet of Things) の分野で注目を浴びており、高齢者の見守りシステムや健康管理システムへの応用が期待されている。行動認識の研究において、行動ラベル付きのセンサデータを教師データとする機械学習を用いることがほとんどである。しかし、ユーザーが求める全ての行動クラスを含む教師あり学習データの収集は困難なことが多い。そのため、学習データにない行動クラスを推定することはできないという問題がある (図 1)。

以降、学習データに存在しない行動クラスを未知クラス、推定できるクラスを既知クラスと呼ぶ。この未知クラスを推定する学習方法を **Zero-shot 学習法** という。この学習方法は、既知クラスと未知クラスの両方を含む別情報源 (別カテゴリと呼ぶ) のデータの特徴量を学習し、別カテゴリの特徴量からクラスを推定する方法である。このとき、別カテゴリの特徴量のクラスには未知クラスと既知クラスが存在する。従って、既知クラスのみを学習させても、別カテゴリの特徴量の類似点を推定することができれば、未知クラスの推定が可能となる。

既存研究として、行動クラスの下位行動クラスとして別カテゴリ特徴量を用意して、Zero-shot 学習法を適用するものが提案されている。下位クラスとは例えば、「走る」行動クラスの下位行動は「足を上げ下げする」「腕をふる」といったものである。しかし、この手法の別カテゴリ特徴量は手動で作成するものであり、作成の手間がかかり限界がある。

本研究では、別カテゴリ特徴量に単語ベクトルを用いた Zero-shot 行動認識手法を提案する (図 1)。このとき、別カテゴリ学習データのサンプルから別カテゴリ特徴量を学習させることを、サンプルを単語ベクトルの空間に投影学習すると表現する。我々の先行研究では、投影学習のアルゴリズムを  $k=2$  の  $k$  近傍法を用いた手法の提案を行い、リアルなデータを用いて未知行動の推定が可能であることを評価実験で検証した。本稿では、精度向上に向けて、投影学習に様々な学習法を用いた検証を行った。評価実験では、人工データと実データを用意し、それぞれの学習法の精度の比較を行った。評価に用いた機械学習は 5 つで、そのうちニューラルネットワークを用いたものが、

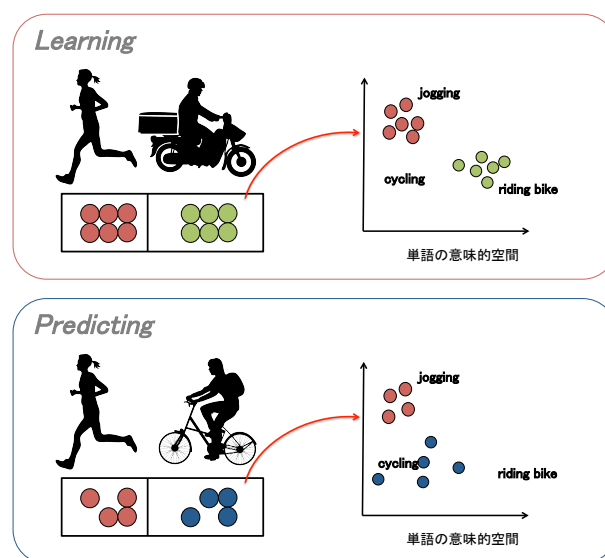


図 1 意味的な単語空間へのサンプルの投影学習。このとき、単語空間には未知行動クラスを含ませている。

先行研究に比べて、精度が向上した。また、既知行動クラスの精度を保つことが精度向上につながるということがわかった。

## 2. 関連研究

行動認識の研究の多くは、センサで集めた行動ラベル付きの学習データを学習させて、推定モデルの生成を行なう教師あり機械学習を用いる。しかし、センサデータの収集にはコストと手間がかかるため、教師なしデータを援用する半教師あり学習や、他の人のデータからサンプルを転移させる転移学習を用いた手法も提案されている。しかし、学習データにない行動が実際に起きて推定はできないという問題は依然として残る。この問題の解決策として、Zero-shot 学習法があげられる。この学習法は、画像認識 [7] や自然言語処理の分野 [8] で多く用いられる手法であり、サンプルのないクラスを推定することを目的とした手法である。

行動認識の研究でも Zero-shot 学習法が提案されている。これらの提案のアプローチとして、文献 [1] は、推定する行動クラスの下位行動を他の行動クラスのサンプルとして補えると仮定

連絡先: 九州工業大学 工学府先端機能システム工学専攻 博士課程前期 1 年井上創造研究室 松木萌 E-mail:matsuki.sousisu@gmail.com

するものである。例えば，“走る”行動の下位行動に“足を上下に動かす”や“腕をふる”といったものがある。この下位行動を“自転車をこぐ”の下位行動とすることで補う。しかし、この手法は、手で行動と下位行動の共起行列を作成することを前提とするため、多くの行動クラスや回行動に分解できない行動クラスについては限界がある。本研究は、別カテゴリ特徴量に単語ベクトルを用いた zero-shot 学習法を用いることを提案する。提案手法は、自動で単語ベクトルを生成できる word2vec ツールを用いることで、既存手法の問題を解決する。

### 3. 提案手法

本節では、未知行動クラスの推定するための単語ベクトルを用いた Zero-shot 学習法を述べる。

#### 3.1 行動認識

行動認識の目的は、センサデータから抽出された特徴量ベクトル  $x$  から、行動  $y$  を推定することである。

$X, Y$  をベクトル集合とし、 $X$  をデータの特徴量ベクトル集合とする。 $Y$  を  $X$  の特徴量データに対応する行動ラベル集合とする。また、 $X, Y$  のそれぞれのドメインを  $\mathcal{X}, \mathcal{Y}$  と示す。

一般の行動認識は、 $X$  と  $Y$  を学習データとして機械学習し、推定モデル  $f: \mathcal{X} \rightarrow \mathcal{Y}$  を生成する。推定時には、新しい入力  $x^{new}$  から

$$\tilde{y} = f(x^{new})$$

により行動  $\tilde{y}$  を推定する。

#### 3.2 Zero-shot 学習法

Zero-shot 学習法は、未知クラスを推定することを可能とする学習法である。 $f: \mathcal{X} \rightarrow \mathcal{Y}$  の学習法との違いは、間に別カテゴリの特徴量  $Z$  を挟んで、 $\mathcal{X} \rightarrow \mathcal{Z} \rightarrow \mathcal{Y}'$  の流れで推定を行う点である。ただし、 $\mathcal{Y} \subset \mathcal{Y}'$ 、つまり別カテゴリの特徴量のほうが、センサデータに対応したラベルドメイン  $\mathcal{Y}$  にはない未知行動クラスを含んでいる必要がある。

つまり、 $\mathcal{X} \rightarrow \mathcal{Y}$  の学習法ではデータセットが1つ必要であったのに対し、Zero-shot 学習法はデータセットを2つ必要とする。1つ目のデータセットは  $\mathcal{X} \rightarrow \mathcal{Z}$  で用いる  $\{X, Z\}$  がペアになったデータセットで、2つ目は、 $\{Z, Y'\}$  のペアである。

Zero-shot 学習法の流れを以下に記す。

1.  $Y'$  を説明変数、 $Z$  を目的変数とし機械学習を行い、推定モデル  $g: \mathcal{Z} \rightarrow \mathcal{Y}'$  を生成。
2.  $Z$  を説明変数、 $X$  を目的変数とし、機械学習を行い、推定モデル  $h: \mathcal{X} \rightarrow \mathcal{Z}$  を生成。
3. 推定時には、新しい入力  $x^{new}$  から  $\tilde{y} = g(h(x^{new}))$  により行動  $\tilde{y}$  を推定する。

つまり、 $X \rightarrow Z$  を学習させることで、 $X$  のサンプルを  $Z$  の空間に投影させ、投影した  $Z$  の空間にあるサンプルを判別器にかけ処理を行なう。未知行動を含む  $Z$  の空間に投影させことで、未知行動クラスである  $\mathcal{Y}' - \mathcal{Y}$  の推定を間接的に可能にしている。

#### 3.3 単語ベクトルを用いた Zero-shot 学習法

図2は本研究で提案する手法の全体構造である。

まず、3.2節で述べたように2つのデータセットを用いる。 $\{X, Y\}$  ペアのデータセットは、本稿の評価では人工データと家庭内行動センシング実験で収集された実データを用いる。また、 $\{Z, Y'\}$  ペアのデータセットは日本語 Wikipedia から word2vec で生成した単語ベクトルを用いる。

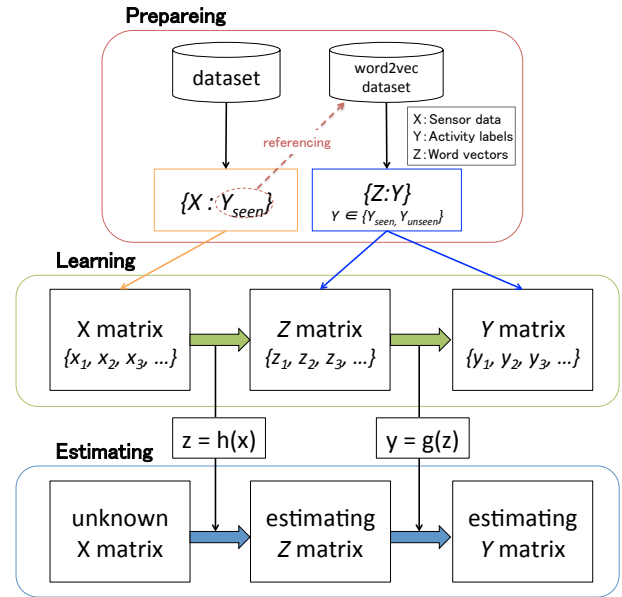


図2 提案手法の全体の流れ

本提案手法の特徴は、word2vec[6]を用いることで、2つ目のデータセットを手軽に自動生成する所である。word2vecとは、自然言語テキストから分かち書きされた単語を文中の単語を周辺の単語を情報とし、ニューラルネットワークを用いて、学習させることで、各単語にベクトルを割り当てるツールである。このword2vecを用いてWikipediaから単語ベクトルを作成し、行動ラベルに対する単語ベクトルを生成した。日本語Wikipediaは、全部で4.2億単語あり、Yはその単語の中から抽出する。つまり、多くの行動クラスが存在する大きな辞書を用いると言える。word2vecで生成した単語ベクトルの次元数は100次元であり、主成分分析で10次元に圧縮したものを本稿では用いた。

また、実験で得られた行動ラベルには、1つの単語ではないラベルも含まれており、例えば、「テレビ・ラジオ・新聞・雑誌」という行動ラベルも含まれている。これらのラベルは、行動ラベルに含まれている単語ベクトルの平均をとるといった対処をした。

次に、学習を行なう。 $Y$ に対応する投影学習を行い、関数  $h: \mathcal{X} \rightarrow \mathcal{Z}$  を生成する。学習時には、 $d$ 次元のベクトル  $x^d$  から、10次元のベクトル  $z$  の各次元を1つつ学習していく。

最後に、学習で生成された推定モデルを用いて推定を行なう。推定された  $\tilde{z}$  のもっとも近い距離にある  $z$  に対応するクラス  $y$  を推定結果とする。本稿では、サポートベクターマシンの用いる。

### 4. 評価

本節では、単語ベクトルと Zero-shot 学習を用いた行動認識において、通常の行動認識と同じく全てが既知行動クラスである状況で精度がどうなるか、未知行動クラス数が増えていった場合の精度の変化、既知行動と未知行動ごとの精度の違い、といった観点から、幾つかの学習モデルについて比較する。

#### 4.1 評価の流れ

評価方法の流れは以下である。

1. 交差検証のため、データセットから訓練データとテスト

データに分ける。

2. 訓練データから  $m$  個のクラスのデータセットを未知行動クラスとみなして、除外する。
3. 除外された訓練データを  $Z$  を用いて学習させ、モデルを生成する。
4. 生成されたモデルを用いてテストデータを推定し、精度を算出する。

以上のことを、(i)  $m = 0$  つまり通常の行動認識と同じく既知行動クラスのみの場合、(ii) 未知行動クラスを  $m = 1, 2, \dots, 12$  と増やした場合の 2 パターンで評価する。

また、投影学習モデルとしては以下のモデルを比較に用いた。

- (LM) 線形回帰
- (NN\_linear) 中間層から線形のまま出力したニューラルネットワーク
- (KNN)  $k$  近傍法
- (SVM) サポートベクターマシーン
- (NN\_log) 中間層からロジスティック回帰分析をして出力したニューラルネットワーク

## 4.2 データセット

データセットとして、提案手法の挙動を調べるための人工データと、実用にあたっての効果を調べるための実データの 2 つを用いる。ここでのデータセットとは  $X$  および  $Y$  がペアになったセンサ特徴量データセットのことである。

### 4.2.1 人工データ

人工データについては、精度向上のヒントを得るために、実データの複雑さを取り除いたデータを自作した。実データにおいては、行動クラスが曖昧であったりする **[] どれ**。例えば後述する実データにある行動クラスを見ると、“食事”と“朝食”、“夕食”という行動クラスが存在する。したがって、本稿では、推定しやすいシンプルなデータを自作した。行動クラス、特徴量の次元は実データに合わせ、12 クラス、6 次元である。作成の流れは以下のとおり。

各次元毎に、

1. 中心となる値を  $-100 \sim 100$  の間でランダムに 1 つ選択。
2. 出現確率が正規分布になるように、1000 サンプル作成。

その結果、 $1200 \times 6$  の  $X$  行列となった。

### 4.2.2 実データ

実データには、部屋の照度と消費電力から家庭内行動の認識の研究 [4] で収集されたデータセットを用いる。収集されたデータは、4 ヶ月間 35 家庭分あり、全部で 53 クラスの行動がある。本節では、そのうち 1 家庭分のデータを用いる。収集された時系列のデータから特徴量抽出処理は以下を行った。

- 一日の中の時刻 (0 - 24 時  $\times$  60 分) から、時刻  $\times$  60 + 分と計算し、値を 0 - 144 の 1 つの数値にする。日時は消す。
- 照度の平均値、標準偏差、最大値、最小値
- 一分ごとに得られる消費電力値

この家庭の行動クラスは、「睡眠」、「身の回りの用事」、「食事」、「通勤・通学」、「仕事」、「学業」、「家事」、「買い物」、「テレビ・ラジオ・新聞・雑誌」、「休養・くつろぎ」の 12 種類であった。つまり、サンプル  $\times$  パラメーターの行列  $X$  で表したとき、 $31,809 \times 6$  の  $X$  行列集合となる。

## 4.3 結果

以下では、通常の行動認識と同じく既知行動クラスのみの場合の精度、未知行動クラス数を変化させた場合の精度、既知行動クラスと未知行動クラスに分けて再現率を算出した結果を示す。適合率は再現率と反比例しており、表示する必要がないと重い、ここには記さない。

### 4.3.1 既知行動クラスのみの場合

評価実験の全体の精度を表 1 に表す。

表 1 未知行動がない場合の精度 (%)

	人工データ	実データ
KNN	100.00	82.73
SVM	100.00	83.78
LM	75.00	73.88
NN_linear	100.00	82.87
NN_log	100.00	73.45

人工データは LM 以外 100% の精度であった。また、NN\_linear が一番良い結果となった。

$k$  近傍法は 1 直線の右下がり、提案手法は一部  $k$  近傍法の精度を上回っている。実データを見ると、提案手法が他の手法を上回っている。

### 4.3.2 未知行動クラス数を変化させた場合

未知行動クラス数を 1 から 12 まで変化させた場合の正解率を図 3 に表す。

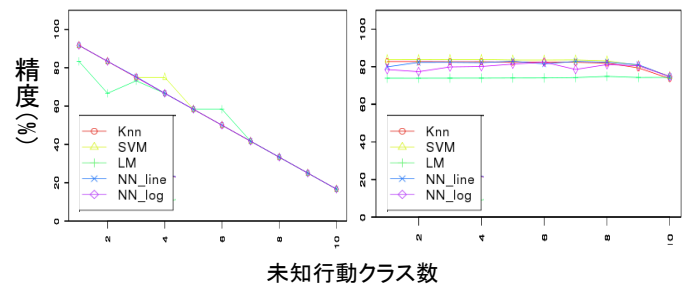


図 3 正解率。右が実データ、左が人工データ (縦軸: 精度、横軸: 未学習データ数)

人工データを見ると、全体的に右下がり、SVM と LM の結果がばらついている。また、実データについては、未知行動クラス数が増えても精度の減少はなだらかだが、 $SVM > KNN > NN\_linear > NN\_log > LM$  の順に精度が高い。

### 4.3.3 既知行動クラスと未知行動クラスごとの再現率

評価実験の未知行動クラスと既知行動クラスの再現率を図 4 に表す。

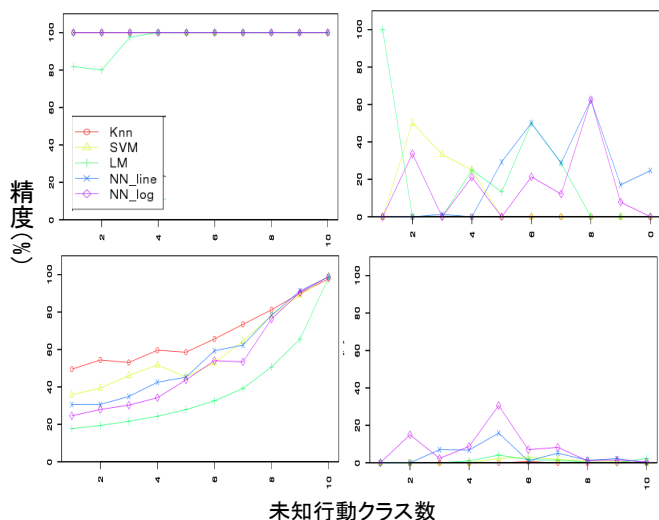


図4 未知行動クラスと既知行動クラスごとの再現率. 上が人工データ, 下が実データ, 左が既知行動クラスの精度 (横軸: 未学習データ数, 縦軸: 精度), 右が未知行動の精度

人工データを見ると, LM においては既知行動クラスについて精度が低いのにに対し, 未知行動クラスについては高くなっている. これは, 未知行動クラスばかりを推定していることを意味する. また, SVM においては未知行動クラスが増えるに連れて, 未知行動の推定を行っていない. 逆のことが NN\_linear に言える. KNN は未知行動クラスの推定はできていない. また, 全体的に未知行動を推定し, 精度がよいのは, NN\_log である.

実データを見ると, 既知行動は  $KNN > SVM > NN\_linear > NN\_log > LM$  の順に精度が高い結果となった. 未知行動は人工データと同様に NN\_log が高い精度となっている. NN\_log 以外の分類モデルは未知行動クラスの推定ができていない.

#### 4.4 考察

評価結果より, 人工データと実データの違い, 学習モデルの違いによる影響について考察する.

##### 4.4.1 人工データと実データの比較

まず, 図3から, 人工データは右下がりの形であるのに対して, 実データは横に伸びている. これは, 実データは, 行動クラスのサンプル数に大きな偏りを持っていることを意味する. 「睡眠」という行動クラスが全体のサンプル数の約5割, 「睡眠」と「家事」の行動クラスを合わせると, 約7割を占めていた. このことより, サンプルの多い既知行動の精度は保たれていることがわかる.

##### 4.4.2 学習モデルの比較

線形と分類モデルでの比較をすると, 線形モデルは未知行動の推定を積極的に行なうが, その分, 既知行動クラスの推定が下がるという問題点がある. また, 分類モデルはその逆が言える. 線形回帰モデルが全体的に精度が低いことから, 単語ベクトルとセンサデータの相関は低く, 非線形モデル学習する法が良いと考えられる. また, ニューラルネットワークに焦点を当てると, ロジスティック出力のほうが未知行動の推定を積極的に行なっている.

どのモデルでも, 未知行動の推定を積極的に行うほど既知行動の推定が下がってしまうという問題があるため, 既知行動クラスの推定を保ちつつ未知行動を精度よく推定することが今後の課題である.

## 5. まとめ

本稿では, 行動認識において未知行動クラスを推定することを目的とし, 単語ベクトルを用いた Zero-shot 学習法を提案した. 様々な学習法を投影学習に用いて評価を行い, 精度向上のための考察を行なった. その結果, (i) 行動認識の精度の向上をさせるには, 既知行動の推定を保つ必要がある, (ii) 線形もしくはニューラルネットワークを用いると未知行動クラスの推定がされやすいことがわかった. しかし, その分既知行動の推定が下がるので, これについては今後の課題である.

## 参考文献

- [1] Le T. Nguyen, Ming Zeng, Patrick Tague, and Joy Zhang. 2015. Recognizing new activities with limited training data. In Proceedings of the 2015 ACM International Symposium on Wearable Computers (ISWC '15).
- [2] Moe Matsuki and Sozo Inoue. 2016. Recognizing unknown activities using semantic word vectors and twitter timestamps. In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct (UbiComp '16).
- [3] Kim, Eunju, Sumi Helal, and Diane Cook. "Human activity recognition and pattern discovery." IEEE Pervasive Computing 9.1 (2010).
- [4] Xincheng Pan, Tomohiro Minezaki, Tatsuya Isoda, Shota Tanaka, Yuri Uchino, and Sozo Inoue. 2015. Analyzing daily-life activities and power consumptions using tablet sensors and activity annotation web system. In Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers (UbiComp/ISWC'15 Adjunct).
- [5] 井上 創造, ウェアラブルセンサを用いたヒューマンセンシング, 知能と情報, 1347-7986, 日本知能情報フェジィ学会, (2016)-12, 28 6, 170-186
- [6] Mikolov, Tomas, Wen-tau Yih, and Geoffrey Zweig. "Linguistic Regularities in Continuous Space Word Representations." Hlt-naacl. Vol. 13. (2013).
- [7] Socher, Richard, et al. "Zero-shot learning through cross-modal transfer." Advances in neural information processing systems. (2013).
- [8] Johnson, Melvin, et al. "Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation." arXiv preprint arXiv:1611.04558 (2016).