2X-08

センサを用いた人間行動認識における Zero-shot 学習法の検討

松木 萌† 井上 創造‡

九州工業大学大学院工学府† 九州工業大学大学院生命体工学研究科‡

1. はじめに

人間行動認識はコンテキストアウェアネス，(例えば見守りシステムなど)に必要な要素の一つである[1]．ほとんどの手法は教師あり機械学習を用いるが，この既存の手法は学習データを収集する必要がある．このデータ収集は時間や手間，コストがかかると言う問題がある．Zero-shot 学習は，学習データが存在しない行動(未知行動と呼ぶ)を推定することを目的とした学習方法で，データ収集の効率化に貢献する．

一般的な行動認識はセンサデータ X から行動クラス Yを直接推定する関数 $y=f(x)$を構築するように学習を行う．しかし，基本的な Zero-shot 学習法は，センサデータ xからテキストドメインに投影する関数 $z=g(x)$を構築するために学習を行う(図 1)．テキストドメインとはテキストデータを元にして構築される単語ベクトル zが存在し，行動クラス yに対応する単語ベクトルが存在する．この時，未知行動に対応する単語ベクトルも存在する．この投影関数を用いて，新たなセンサデータを単語ベクトル空間に投影し，単語ベクトル空間 Z上で最近傍法を用いたクラス分類を行う．この時，未知の単語ベクトルに最も近く投影されたサンプルが存在すると，そのサンプルは未知行動と認識される．先行研究では，テキストドメインに word2vec を用いて構築した単語ベクトルを用いる手法を提案し[2]，投影関数の向上がこの手法の未知行動の精度を向上させることができると考察した．

本稿では，投影関数の精度向上のために 4 つの投影モデルを構築し，分析する．4 つの学習モデルの違いは，投影させる方向で，センサ空間を X 単語ベクトル空間を Z とした時，$X\rightarrow Z$, $Z\rightarrow X$, $X\rightarrow Z\rightarrow X$, $Z\rightarrow X\rightarrow Z$ の 4 種類の投影モデルを構築する．その結果，(1)投影は 2 度行うことで精度向上につながること．(2)未知行動同士が既知行動より類似すると判別が難しい，ということがわかった．

2. 関連研究

Zero-shot 学習法は，センサを用いた行動認識において多くの研究はない．しかし，画像認識の分野では多くの研究論文が存在する．それらの手法は大きく分けて 3 つの種類がある．1 つ目は特徴ドメインからテキストドメインに投影する手法 $X\rightarrow Z$[3]． 2 つ目はテキストドメインから特徴ドメインに投影する手法 $Z\rightarrow X$[4]．3 つ目は両方向の投影を学習する方法 [5]．3 つ目の手法は，特徴量ドメインからテキストドメインに投影を先に行い，投影時に用いる重み行列 Wを転置させた W^Tを用いて逆投影を行う．3 つ目の手法に関して，2 通りのモデル $X\rightarrow Z\rightarrow X$, と $Z\rightarrow X\rightarrow Z$ を構築する．

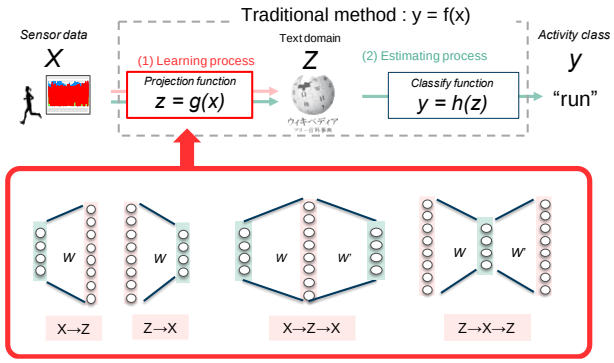

図 1. 上段は Zero-shot 学習法の概要を示す． 本稿では，$z=g(x)$に焦点を起き，4 つの投影モデル(図中下段)を構築し精度向上のための評価，分析を行う．

3. 評価方法

これらの 4 つのモデルを評価し分析するために，2 つの評価を行う．(評価 1) 投影結果を評価するために，未知行動を考えない場合で評価する．(評価 2)2 つの未知行動を設定し，推定精度を評価する．

本稿では，本研究室で収集したセンサデータセットと英語 Wikipedia から生成した単語ベクトルを用いて評価する．

1) センサデータセット:

我々は被験者の左腕にスマートフォンを装着し，“stay”, “walk”, “skip”, “jog”, “stair up” and “stair down”の 6 つの行動を行ってもらい，加速度データを収集した．被験者は 1 つの行動に対して 20 秒間行動し，これを 5 回繰り返し行った．また，このデータから特徴量を抽出するために，Slide-time-window 法を用いた．時間窓サイズは 0．5 秒，スライドサイズは 0．2 秒と設定した．

2) 単語ベクトル:

単語ベクトルを構築するために，まず，英語 Wikipedia を word2vec[6]で学習させる．学習さ

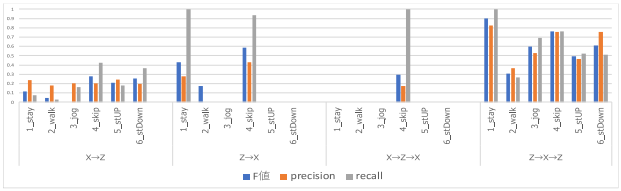

図 2. 評価 1 における，各モデルの各行動に対する推定精度．評価指標には F 値，precision， recall を用いる．

せた時点で，Wikipedia に存在する単語全てのベクトルが得られる．そこから，今回評価に必要な単語ベクトルのみを抽出した．単語ベクトルのサイズは 1000 次元である．

4. 結果

図 2 は評価 1 の各モデルの各行動に対する推定精度である．この結果から，投影モデル Z→X→Z が最もいい精度であることがわかる．なぜなら，各行動において，ある程度の精度を保っており，また，X→Z のように低い精度ではないからである．

表 1 は評価 2 の結果を表している．この時，評価 1 で最も精度の高かった *Z→X→Z* の推定精度のみを示している．精度の指標は全体サンプル数に対する正解サンプル数の割合である．推定可能であった未知行動の組み合わせは“stay”と“stair up”，“stay”と“stair down”，“stay”と“jog”である．

5. 考察

図 2 からモデル毎の精度の違いについて考察する．X → Z の学習モデルは推定精度が同じであることから，一様分布に従った推定が行われている(つまり 1/6 の確率で正解している)と考えられる．Z → X と X → Z → X の学習モデルは 0%の行動種があることから，1 つの行動の単語ベクトル付近に投影されてしまっていると考えられる．

次に表 1 と図 4 から未知行動が推定される傾向について考察する．“stay”の行動(図 4 の点)に焦点を置く．そして推定できた組み合わせ“stair up”，“stair down”，“jog”と推定精度の低い”skip”，“walk”との距離関係について考える．単語ベクトルの

表 1． 評価 2 における，Z→X→Z モデルを用いた各行動毎の推定精度を示す．行動ラベルは 2 つの未知行動を表していて，その未知行動の推定精度が表内に数字で表される．

	stay	skip	jog	walk	stair up
skip	50				
jog	**100**	54. 10			
walk	49. 66	50. 03	55. 55		
stair up	**93. 51**	59. 09	55. 59	50. 03	
stair down	**98. 21**	40. 41	55. 55	50	50. 26

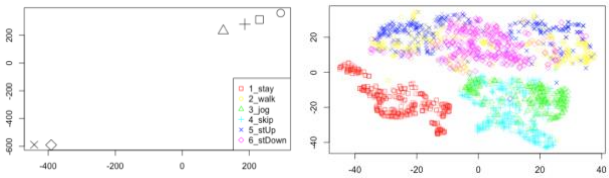

図 4. t-SNE を用いて各データセットを 2 次元に圧縮し可視化したもの．左がテキストドメイン，右がセンサードメインである．点の形が行動種を表している．

方を見てみると，”stay”と”skip”，“walk”は近い関係にある．つまり”stay”と”skip”もしくは”walk”との判別ができなかった理由は，未知行動同士が既知行動よりも類似度が近いと判別ができないと考えられる．ここから，(2)未知行動同士が既知行動より類似すると判別が難しいことがわかった．つまり 1 つの既知行動に対して 1 つの類似する未知行動がある状態がこの手法の精度向上につながると考えられる．

6. まとめ

本稿では，センサ行動認識における Zero-shot 学習法の投影関数の精度向上のために，4 つの投影モデルを構築し，評価分析を行った．その結果，(1)投影は 2 度行うことで精度向上につながること，(2) 未知行動同士が既知行動より類似すると判別が難しい，ということがわかった．

参考文献

[1] Bulling, Andreas, Ulf Blanke, and Bernt Schiele. A tutorial on human activity recognition using body-worn inertial sensors. ACM Computing Surveys (CSUR) 46. 3 (2014): 33.

[2]Matsuki, Moe, and Sozo Inoue. Recognizing unknown activities using semantic word vectors and twitter timestamps. Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing

[3] Palatucci, Mark, et al. Zero-shot learning with semantic output codes. ” Advances in neural information processing systems. 2009.

[4] 重藤優太郎, et al. Zero-shot learning における線形回帰の影響. 研究報告自然言語処理 (NL) 2015. 4 (2015): 1-8.

[5] Kodirov, Elyor, Tao Xiang, and Shaogang Gong. Semantic autoencoder for zero-shot learning. arXiv preprint arXiv:1704. 08345, 2017.

[6] MIKOLOV, Tomas, et al. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301. 3781, 2013.