

## 要素行動の含有度合いを用いたゼロショット行動認識手法に関する研究

豊増 聖実\*

藤波 香織†

\* 東京農工大学 大学院 工学府 情報工学専攻

† 東京農工大学 大学院 工学研究院 先端情報科学部門

## 1 はじめに

近年、行動認識に関する研究が盛んに行われているが、その学習手法には問題点がある。まず、広く用いられている学習手法の1つである教師あり学習では、教師データと呼ばれる行動をセンシングした際に得たデータと正解レベルの組み合わせを用いて行動の分類規則を学習する。教師あり学習では教師データをいわゆる「例題」として分類規則を学習させるため、学習時に考慮していない「未知行動」を認識することができない。この問題を解決する手法の1つにゼロショット学習がある。ゼロショット学習はクラス間に共通する補助情報を用いて未知行動の認識を行う。特に意味属性と呼ばれる行動の様子を説明する動作や姿勢状態を補助情報として、2段階の階層で分類を行う手法をDAP (Direct Attribute Prediction) モデルという[1]。近年、DAPモデルの考えを用いた行動認識に関する研究がなされ、適用可能なことが示された。しかし、従来の研究では、意味属性値がバイナリ値をとるため、認識対象にできる行動数が $2^N$  ( $N$ : 属性数) に制限されるという問題があった[2]。そのため、本研究では意味属性の値を連続値に拡張することによって、対象にできる行動数に制限のない新しい分類手法の提案を行う。

## 2 認識手法概要

本研究では連続な意味属性値を用いて、既存のゼロショット学習における問題の解決を行う。意味属性とは行動を説明するための単位動作や姿勢状態などのことで、例えば「走る」という行動には「腕が揺れている」や「移動している」という属性が含まれ、「座っている」という属性が含まれないという説明ができる。従来の研究[2]では意味属性は含まれるか否かのみを考慮した研究がされてきたが、本研究ではこれを連続値に拡張する。本研究では連続な意味属性を用いて二段階の分類を行う。提案手法の概要図を図1に示す。まず、事

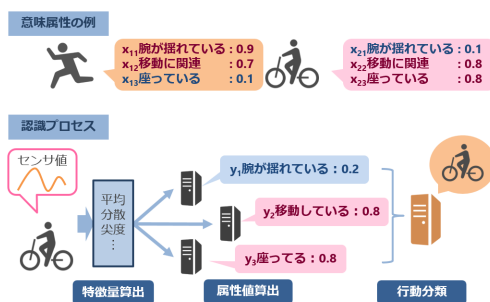


図1: 認識の流れ

前に各行動の意味属性を定義する。これは人の主観によって決められる。提案手法は特徴量算出部、属性値算

出部、行動認識部の3つに分けることができる。まず、センサデータを分類に使用できるように特徴量の計算を行う。次にそれらの特徴量を入力とし、 $N$ 個の属性値算出器により各属性値の算出を行う。属性値は連続な値を取るため回帰分析を用いて属性値算出器の学習を行う。行動分類部では、算出した属性値と事前に定義しておいた各行動の属性値から線形写像によって行動の推定を行う。写像の方法には2つの手法を提案する。1つ目は、AHP (Analytic Hierarvhy Process) [3] から着想を得た行動得点を利用する分類手法である。この手法では、属性値算出部で算出した属性値と意味属性行列を用いて行動得点と呼ばれる行動への近さを計算し、最も得点が高い行動に分類する手法である。行動得点の計算式は行動 $k$ に定義された属性 $i$ の属性値を $x_{ki}$ 、また属性 $i$ の属性値算出器から得られた予測属性値を $y_i$ として(1)式で表される。

$$\text{行動得点}_k = \sum_{i=1}^N x_{ki}y_i \quad (1)$$

2つ目は最近傍法を利用した分類手法である。この手法は二値の意味属性を用いたゼロショット学習の研究でも用いられる手法で、属性算出部から算出された属性値列と意味属性行列の各行動の属性値列のユークリッド距離を計算し、最も距離の近い行動を予測クラスとして選ぶ。

またこれらのアルゴリズムの提案に加えて、2種類のウィンドウのとり方を提案する。1つ目は、属性値算出時と行動分類時のウィンドウサイズが同一である手法である。この手法は既存手法[2]でも用いられる手法である。この手法では属性と行動が同一の長さで行われることを仮定するため、複雑な行動の分類は難しいと考えられる。2つ目は属性値算出時のウィンドウを行動分類時のウィンドウに比べて小さくする手法である。属性値の短いウィンドウを行動分類に用いる長いウィンドウにするために、属性値は連続するいくつかのウィンドウごとに平均を取る。この手法を用いることで属性値がより正確に算出でき、かつ行動を認識するのに必要なウィンドウも確保できることが期待できる。本研究では行動分類のためのアルゴリズムと、ウィンドウサイズのとり方をそれぞれ2つずつ、4通りの手法の提案を行う。

## 3 事前準備

## 3.1 データ収集

実験で用いるデータセットの要件は2つある。本研究では日常生活の複雑な行動を認識することを目的とする。そのため、多くの行動が自然な環境で取られていることが望ましい。また、意味属性を人の主観で決める性質上行動の詳細を表す動画が必要である。これら2つの要件を満たす公開データセットは筆者の知る限りない。そのため、本研究では行動の詳細がわかるサンプル動画があるデータセットの収集を行った。収集したデータは「歩行」や「食事」、「手を洗う」等の日常生活行動 (ADL; Activity of Daily Living) 23種類で、研究室内の男女14人に7つのセンサ (両手首, 両腕,

A study on a semantic zero-shot human activity recognition based on action elements content

\* Seiji TOYOMASU † Kaori FUJINAMI

\* † Department of Computer and Information Sciences, Tokyo University of Agriculture and Technology

胸, 両大腿)を取り付け, 一つの行動につき 168 分間 (12 分× 14 人) ずつ収集した。

### 3.2 意味属性行列の作成

行動に含まれる意味属性値を定義し,  $M \times N$  ( $M$ : 行動数,  $N$ : 属性数) の行列形式で表したものを意味属性行列という。本研究では, 複雑な意味属性値を, 単純な比較の繰り返しにより決定できることから, 一対比較の考えを意味属性行列作成に用いた。

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
a	0.072	0.095	0.013	0.013	0.049	0.074	0.016	0.047	0.015	0.057	0.094	0.139	0.014	0.023
b	0.124	0.107	0.018	0.013	0.141	0.133	0.016	0.033	0.014	0.051	0.091	0.153	0.015	0.029
c	0.014	0.085	0.017	0.013	0.105	0.132	0.016	0.011	0.070	0.008	0.113	0.163	0.015	0.018
d	0.041	0.021	0.058	0.017	0.034	0.022	0.018	0.029	0.016	0.053	0.059	0.015	0.134	0.075
e	0.039	0.026	0.025	0.017	0.026	0.022	0.029	0.045	0.016	0.053	0.058	0.015	0.146	0.088
f	0.044	0.028	0.061	0.017	0.030	0.025	0.043	0.044	0.016	0.057	0.019	0.015	0.139	0.095
g	0.032	0.012	0.025	0.032	0.041	0.043	0.076	0.043	0.017	0.054	0.022	0.041	0.028	0.043
h	0.026	0.010	0.025	0.038	0.026	0.026	0.079	0.047	0.016	0.056	0.017	0.018	0.024	0.050
i	0.022	0.025	0.056	0.095	0.018	0.019	0.138	0.078	0.015	0.060	0.018	0.015	0.029	0.041
j	0.021	0.026	0.054	0.099	0.016	0.018	0.140	0.010	0.123	0.007	0.019	0.015	0.032	0.042
k	0.076	0.040	0.015	0.014	0.074	0.048	0.018	0.041	0.016	0.053	0.013	0.039	0.105	0.056
l	0.075	0.102	0.012	0.014	0.106	0.133	0.018	0.050	0.017	0.057	0.092	0.103	0.015	0.022
m	0.074	0.103	0.011	0.014	0.101	0.105	0.018	0.051	0.017	0.057	0.093	0.102	0.015	0.022
n	0.021	0.028	0.053	0.103	0.016	0.017	0.109	0.009	0.138	0.007	0.021	0.016	0.029	0.030
o	0.021	0.027	0.055	0.101	0.016	0.019	0.109	0.080	0.015	0.059	0.021	0.016	0.028	0.030
p	0.036	0.025	0.022	0.019	0.030	0.024	0.045	0.050	0.016	0.055	0.027	0.016	0.120	0.069
q	0.016	0.032	0.024	0.121	0.014	0.018	0.017	0.009	0.136	0.006	0.062	0.016	0.015	0.021
r	0.014	0.031	0.037	0.105	0.012	0.019	0.015	0.097	0.015	0.059	0.021	0.016	0.015	0.015
s	0.014	0.032	0.028	0.100	0.011	0.017	0.015	0.009	0.134	0.007	0.027	0.016	0.015	0.015
t	0.014	0.038	0.027	0.103	0.011	0.017	0.015	0.009	0.130	0.007	0.068	0.016	0.015	0.016
u	0.098	0.060	0.136	0.014	0.048	0.023	0.016	0.075	0.016	0.059	0.012	0.018	0.020	0.090
v	0.039	0.036	0.137	0.014	0.032	0.024	0.017	0.078	0.016	0.059	0.013	0.018	0.015	0.053
w	0.069	0.010	0.088	0.015	0.043	0.024	0.016	0.055	0.017	0.060	0.020	0.018	0.018	0.058

a.walk, b.run, c.ride bike, d.wash face, e.wash hands, f.brush teeth, g.set table, h.make coffee, i.eat stand, j.eat sit, k.vacuum, l.upstairs, m.downstairs, n.drink sit, o.drink cup, p.wash dishes, q.use PC, r.use smartphone stand, s.use smartphone sit, t.read, u.wipe whiteboard, v.write whiteboard, w.wear jacket

1.arm-pendulum-swing, 2.cyclic-motion, 3.hand-above, 4.hands-on-table, 5.intense-motion, 6.leg-above, 7.meal-related, 8.posture-upright, 9.sitting, 10.standing, 11.symmetry, 12.translation-motion, 13.washing-related, 14.wrist-movement

図 2: 意味属性行列

## 4 評価実験と考察

### 4.1 実験概要

2章で提案した手法を実装し, 評価実験を行った。本研究は未知の行動を認識することを目的としているため一クラス抜き交差検証にて実験を行った。一クラス抜き交差検証では一クラスを抜いた行動データを用いて各属性値算出器の学習を行い, 残りのデータでテストを行う。また, 属性期算出器の学習に属性値付きの行動のデータを用いた。属性値算出器の構築に用いる特徴量は平均や分散など行動認識によく用いられる 8 種類 34 個を用いた。また, 属性値算出器の精度を上げるため, 3 つの回帰方法を試し, 行動分類時にはそれぞれ最も精度が良いものを使用した。精度の指標には二乗平均平方誤差を用いた。実験条件を表 1 に示す。また属性値算出時と行動分類時でウィンドウサイズが異なる手法では, 算出した属性値を 10 サンプルごとに平均化させることで異なるウィンドウサイズを吸収した。

表 1: 実験条件

項目	条件
サンプル数	約 5 万 (一クラスにつき)
特徴量	8 種類 32 次元 (平均, 分散, 歪度, 尖度, 各軸の相関, 平均交差回数, FFT エントロピー, エネルギー)
ウィンドウサイズ	128 サンプル (50Hz)
行動数	23 種類
属性数	14 種類
属性値算出手法	MSP; Linear, Least median of squares

### 4.2 実験結果

本実験によって得られた再現率適合率及びそれらの調和平均である F 値を表 2 に示す。属性値算出時と行

動分類時のウィンドウサイズが異なる手法は表中では WS 改善とした。ウィンドウサイズを変えることで, 適合率に関しては精度が上がったが再現率は下がった。全体的な精度を示す F 値についてはウィンドウサイズが同一の手法のほうが高かった。また算出した属性値の二乗平均平方根誤差の全属性の平均は 0.0598 だった。

表 2: 実験結果

	適合率	再現率	F 値
行動得点	0.0422	0.0374	0.0396
最近傍法	0.0345	0.0138	0.0197
行動得点 + WS 改善	0.0483	0.004	0.00739
最近傍法 + WS 改善	0.0491	0.00417	0.00769

### 4.3 考察

表 2 より, 4 つの手法のうち最も F 値が高かったのは, 行動得点を用いるウィンドウサイズが同一の手法であった。属性値算出時と行動分類時のウィンドウサイズが異なる手法では, 属性値を正確に計算しつつ, 行動の複雑さを考慮できると考えていたが, 本実験ではウィンドウサイズが同一な手法の方が精度が高かった。精度が低い原因に行動分類時に属性値を平均化させ, 時系列を考慮できなかったことが考えられる。複雑な行動には様々な属性が含まれており, それがある程度決まった順序で行われる。各属性がどの順番で行われたのかを複数, それを行動分類時に組み込めれば, より行動の複雑さを考慮できると考えられる。

分類の精度を上げるには, 属性値算出器の精度を改善することが重要であると言える。属性値算出器の精度を上げる方法の 1 つに, 学習データを属性のみを行っているデータに変えるという方法がある。行動データを用いることで属性データを追加で収集する必要がなくなるが, 行動データでは属性値を正確に学習できていない可能性がある。今後は属性のみを行っているデータを収集し, それを用いて属性値算出器の学習をし, その影響を調査する。

### 5 おわりに

本稿では連続な意味属性を用いたゼロショット行動認識手法を 4 つ提案し, 実験によりそれらの比較をした。属性値算出時と行動分類時のウィンドウサイズを変える手法では時系列を考慮できなかったため, 精度が低かった。また, 全体的に分類器の精度が低かった原因として, 属性値算出器の精度が低かったことが考えられるので, 今後は属性のみを行っているデータを用いて属性値算出器の構築を行い, 評価実験を行う。

### 謝辞

本研究は, 日本学術振興会 科学研究費補助金 (18H03228 および 19H01719) の支援を受けた。

### 参考文献

- [1] 鈴木雅大, 佐藤晴彦, 小山聡, 栗原正仁, 松尾豊. 属性ごとの観測確率を考慮したゼロショット学習. 情報処理学会論文誌, Vol. 57, No. 5, pp. 1499–1513, 2016.
- [2] Heng-Tze Cheng, Feng-Tso Sun, Martin L. Griss, Paul Davis, Jianguo Li, and Di You. Nuactiv: recognizing unseen new activities using semantic attribute-based learning. In Hao-Hua Chu, Polly Huang, Romit Roy Choudhury, and Feng Zhao, editors, *MobiSys*, pp. 361–374. ACM, 2013.
- [3] T.L. Saaty. How to make a decision: the analytic hierarchy process. *European journal of operational research*, Vol. 48, No. 1, pp. 9–26, 1990.