

## B203

### 介護士の行動認識における学習データの選択に関する検討

井上創造<sup>†</sup> ○後藤広樹<sup>†</sup>  
<sup>†</sup>九州工業大学

#### On Selecting Training Data in Activity Recognition for Caregivers

<sup>†</sup> Sozo Inoue and <sup>†</sup> Hiroki Goto  
<sup>†</sup> Kyushu Institute of Technology

**Abstract:** *In this paper, we consider selecting training data in caregiver's activity recognition. Management of training data is important for machine learning of activity recognition algorithms. We considered training data sizes, the number of models, and ability of training at self-devices for several selection patterns of training datasets, and evaluated the accuracies of some patterns. We found that best of the patterns for caregivers is to use data of people with the same attribute for training and it outputs better results than to use data of a person be recognized.*

#### 1. はじめに

人材不足が進む介護業界において介護記録を効率化のために、我々はセンサ行動認識技術を用いて介護記録を自動的につけるシステムの研究を行っている[1]。行動認識においては機械学習を用いるため、学習データの管理が重要となるが、ある介護士の行動認識のためにどの範囲の学習データを使えば良いかは、全体の処理量と精度の点から検討し最適化する必要がある。本稿では、介護施設の行動認識において学習データをどう選択すべきかをいくつか挙げて、その計算量とモデル数、端末において処理可能かといった観点から検討する。また、そのいくつかの場合において認識精度を比較する。その結果、認識対象者と同じ属性を持つ介護人材のデータを学習に用いる手法が我々の実験[1]において有用である結果を得た。

#### 2. 関連研究

機械学習工学研究会は、機械学習システムの開発・運用にまつわるプラクティスを共有することを目的としている[2]。この中で鶴林は機械学習における学習結果のモデル管理を一つの課題として挙げている[3]が、学習データが異なればモデルも異なることから、学習データの管理も関連して重要である。本研究は介護士の行動認識のために機械学習を用いるが、ある介護士の行動認識を行うためにどの学習データを用いるのが良いかは自明ではない。例えば自身のデータを学習に使うのであれば、自身の端末のみでの学習も可能であるが、十分な学習データが集まるまでは認識できないという問題や、学習も介護士数分行う必要が発生する。一方で複数介護士のデータをまとめて学習に使うのであれば、学習の総数は少なくなるが、一回あたりの学習の時間は長くなり、また各自の端末で学習するにはデータの共有が必要になる。

このように、介護施設における介護士の行動認識のような具体的な応用における機械学習における学習データの管理についてはまだ研究がなされておらず、種々の学習データ選択手法をその精度とともに検討する必要がある。

#### 3. 学習データ選択手法の検討

本節では、介護士行動認識のための学習データの選択方法について、その計算量とモデル数、個人端末での処理（以下、これらをリソースと呼ぶ）の観点から考察する。

表1に、行動認識において、学習データの選択の仕方による各種リソースを検討した結果を示す。選択する学習データは、1)本人のデータ、2)年齢・性別・身長・体重・役職・勤務年数・業務環境といった本人と同じ属性の人のグループのデータ、または3)施設全体（または利用できる全部の）のデータの3種類を考えた。

表1では、対象者数を $n$ 人、学習対象となるセンサデータの日数 $d$ 日、1人日分のデータを学習させるのに必要な時間を $T$ 、1つの属性についての属性値の種類を $g$ 種とする。1種の属性を持つ対象の人数は等しいと仮定する。

一般的に機械学習では、データ量が増えると精度が高まるとされているが、関係のないデータ、つまり質の悪いデータばかり増えると精度の上昇よりも計算量が増える。実用化のためには計算量を抑えつつ、高い精度を出す必要がある。したがって、質の良いデータが多いことが重要である。今回の表の中では②がこの条件にもっともあてはまる。⑤に比べ、②では対象者本人に関係のない学習データが省かれることで認識精度の改善が期待される。①に比べ、②では属性の種類によっては軽微な認識精度低下で、別人への汎用性を高くできると考えられる。

表 1. 行動認識による学習データの選択と各種リソースの関係

選択された学習データ	全対象者の全学習時間	生成モデル数	個人端末で学習できるか
①本人のデータ	$ndT$	$n$	○
②本人+ 属性が同じグループのデータ	$ndT$	$g$	×
③属性が同じグループのデータ (本人以外)	$(n/g-1)gdT$	$n$	×
④属性が異なるグループのデータ	$ndT$	$g$	×
⑤全データ	$ndT$	1	×

しかし精度とそれぞれの属性の関係について評価が必要である。作成されるモデルの数は、少ないほどサーバの記憶容量を必要とせず、モデル毎の更新もしやすくなる。また計算用サーバを必要とせず、センシング用のスマートフォン内部で完結することができれば、計算の分散ができ、行動認識の実用性は高められる。ただし、学習に必要なデータが全て個人のスマートフォンで収集される必要がある。

これらの検討の結果、我々の実験[1]では②が最適である。ただし、精度による評価は必要である。

#### 4. 初期評価

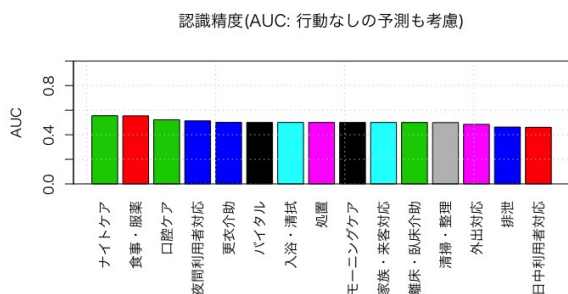


図 1. ①男性介護士自身のデータ対を学習した男性介護士の行動認識精度 (縦軸: AUC, 横軸: 行動の種類)

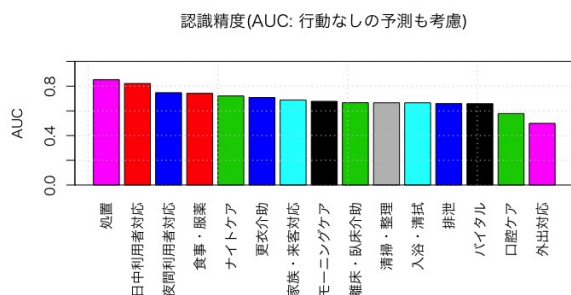


図 2. ②属性(性別:男)グループのデータ対を学習した男性介護士の行動認識精度 (縦軸: AUC, 横軸: 行動の種類)

表 1 の各場合における精度の初期評価として、①と②の場合について認識精度を比較するため、我々の実験[1]のデータセットを用いた。このデータは 31 名 2

ヶ月分の介護士のスマートフォンの 3 軸加速度データからなり、15 種の行動が記録されている。図 1, 2 にそれぞれ結果を示す。

全行動の認識精度の AUC 平均は、①が 0.503, ②が 0.689 であった。②の選択手法の認識結果は AUC が 0.5 を下回る行動がないため効果があると考えられる。

#### 5. 考察

実験前、②よりも①の方が精度のみ優れると予想したが、結果は逆転した。原因として、学習データの中に質の悪いデータが含まれており、精度の悪いサンプルがあったと予想する。またより効果の高い属性の種類として勤務年数を挙げる。熟練者は効率の良い行動ができることから、人間自身による行動の効率化があり、勤務年数が長い対象者は似た体の動きをすると予測されるためである。

#### 6. まとめ

介護施設における行動認識システムにおける学習データの選択方法について、その計算量と生成モデル数、個人端末での処理の観点から検討した。その結果、認識対象者と同じ属性を持つ介護人材のデータを学習に用いる手法が我々の実験[1]において有用である結果を得た。今後さらに複数の属性による場合分けも検討していく。

#### 7. 参考文献

- [1] 国立大学法人九州工業大学: 介護施設従事者の IoT による行動認識・予測の実施実験を実施, 九州工業大学(オンライン), 入手先 [http://www.kyutech.ac.jp/archives/025/201810/press\\_181026.pdf](http://www.kyutech.ac.jp/archives/025/201810/press_181026.pdf), (参照 2018-10-28).
- [2] 日本ソフトウェア科学会機械学習工学研究会: 機械学習工学研究会(オンライン), 入手先 <https://sites.google.com/view/sig-mlse/> (参照 2018-11-15).
- [3] 鶴林尚靖: 機械学習工学が解決すべき 8 つの研究課題, 第 1 回機械学習工学ワークショップ 2018 (MLSE2018) 論文集, No. 84, p72-74, (2018. 07).