

ウェアラブルセンサデータと環境設置ビデオデータを併用した 保育士行動認識

XAICHA JAIXIONG¹ 大見 士² 藤原 健之³ 大村 廉⁴

豊橋技術科学大学¹

豊橋技術科学大学²

豊橋技術科学大学³

豊橋技術科学大学⁴

1. はじめ

現在、保育施設の数や利用者の増加により、保育士一人当たりの担当児童数が増加し、その負荷の増大が現在の問題となっている。また、保育における保育士の適切な業務遂行については定量的な分析が不足している現状がある。保育士行動の分析のためのデータ獲得のため、行動認識技術を用いて保育業務の記録を行う事が有効であると考えられる。

保育行動の認識にあたり、既存研究ではウェアラブルセンサのみを用いる場合が多かった。藤原らはウェアラブルセンサから取得した加速度・角速度データのみを用いて、複数の深層学習モデルで認識精度を比較した[1]。その結果、LSTMを用いて約90%のAccuracyを得ることができると明らかにした。

一方、保育の現場では、安全性の確保などの理由から施設内にカメラを設置する機会も多くなってきており、行動認識にはウェアラブルセンサを用いると共に、環境に設置されたカメラを用いることが考えられる。そこで本研究では、ウェアラブルセンサと環境設置ビデオデータを組み合わせて保育士の行動を認識することで、認識の高精度化を目指す。

2. 関連研究

藤原らは、ウェアラブルセンサから取得した加速度・角速度データを用いて、保育士の行動認識を行った[1]。この研究では27種類の保育における行動の認識に対して適切な深層学習のモデル、および、学習データ量と認識精度の関係を調査している。結果から、LSTMを用いて約2時間程度の学習データを用いれば、90% Accuracyが得られることが分かった。

Junhyukらは、「手をたたく、ボールを投げるなど」の9種類の日常行動認識についてウェアラブルセンサから得られた加速度センサデータとともに、骨格画像と関節座標加を組み合わせで認識する手法の検討を行った[2]。それぞれのデータに対してそれぞれの深層学習モデルを用いて特徴量を抽出し、完全接続層と出力層で行動を分類した。結果より、骨格画像のみ、加速度センサデータのみ、骨格画像と加速度センサデータの組み合わせ、関節座標データと加速度センサデータの組み合わせ、骨格画像と関節座標と加速度センサデータの組み合わせで、それぞれ70.9%、84.7%、94.8%、91.8%、93.1%のAccuracyとなり、加速度センサデータと画像データを組み合わせることで高い認識精度を得られることを明らかにした。

本研究では、[2]の研究を参考に、ウェアラブルセンサデータから取得した加速度・角速度データと環境設置ビデオデータから取得した関節座標データを組み合わせて保育士の行動認識を行う。

3. 実験方法

本研究では藤原らが取得した保育士の加速度・角速度データ、環境設置ビデオカメラから取得したビデオデータを用いた[1]。

保育士1名に体の6箇所ウェアラブルセンサを装着してもらい、100Hzのサンプリング周波数で3軸の加速度・角速度データを収集した。ウェアラブルセンサの装着位置を図1に示す。

また、ビデオカメラをこども園内の部屋に3台設置し、それぞれで保育士の行動を含む部屋内の状況の様子を撮影した。ビデオカメラの設置状況は図2に表す。ビデオは30fpsで録画した。ビデオデータをOpenPoseに入力し、保育士の関節座標データ(18点)を取得した。なお、OpenPoseでは画像内に写る人々についてそれぞれ関節座標を得ることができる。つまり、保育士だけでなく子供たちの関節座標も同時に得られるが、本研究では保育士のみ関節座標データを選択して使用した。その選択方法の説明は

Nursery Activity Recognition Using Wearable Sensor and Environmental Video Data

1 XAICHA JAIXIONG, Toyohashi University of Technology

2 Kenshi Fujiwara, Toyohashi University of Technology

3 Akira Omi, Toyohashi University of Technology

4 Ren Ohmura, Toyohashi University of Technology

スペースの都合上、本稿では割愛する。

対象とする保育行動は、保育学の専門家の指導の下定義した 39 種類の行動のうち、1 台以上のビデオカメラに保育士の行動が映っており、かつその総時間が 20 秒以上であった 21 種類の行動を認識対象とした。認識対象とした保育士行動の一部を表 1 に表す。

加速度・角速度データと関節座標データを組み合わせるため、まず、関節座標データを 100Hz にアップサンプリングしてサンプリングレートを合わせた。そして、これらデータについて、スライディングウィンドウ（ウィンドウサイズ 1000ms, 重なり 75%）を用いて部分系列を切り出した。切り出した部分系列は直接深層学習モデルの入力とした。深層学習モデルには[1]の研究で高精度を得られた LSTM を用いた。LSTM の学習においてバッチサイズは 128 でエポック数は 500 とした。学習の評価には Stratified K-Fold Cross-validation(k=5)を用いて評価を行い、評価指標には Accuracy を用いた。

表 1 認識対象の保育士行動の一部

座って話している	歩いている	座っている
子供に触れている	紙を切っている	見回している
立っている	歩いて話している	物を持っている



図 1 ウェアラブルセンサの装着位置

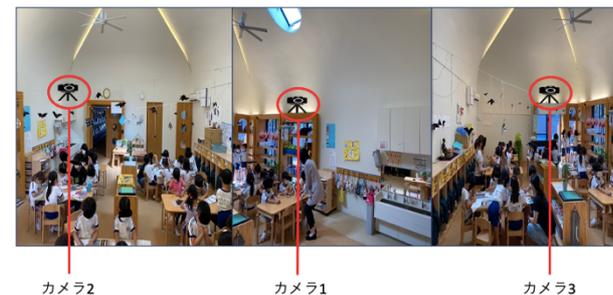


図 2 ビデオカメラの設置状況

4. 実験結果

加速度・角速度データ（ウェアラブルセンサデータ）と関節座標データ（環境設置ビデオデータ）をそれぞれ個別に使用して認識を行った場合、および、両方を組み合わせて認識した場合の平均 Accuracy の比較を図 3 に示す。

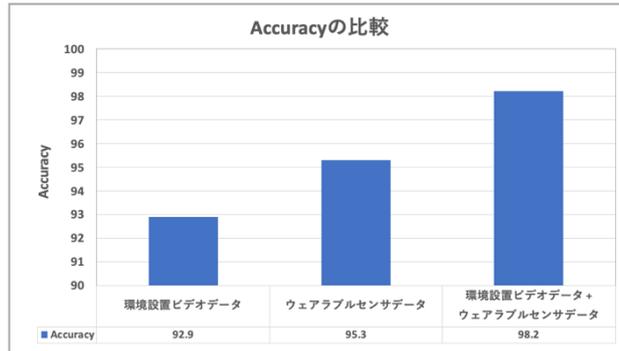


図 3 それぞれのデータに基づく認識精度

図 3 より、環境設置ビデオデータのみを用いた場合、ウェアラブルセンサデータのみを用いた場合、両データを組み合わせた場合で、Accuracy はそれぞれ 92.9%, 95.3%, 98.2% となった。ウェアラブルセンサと環境設置ビデオデータを組み合わせることで、ウェアラブルセンサデータのみで認識した場合から 2.9%, 環境設置ビデオデータのみで認識した場合から 5.3%, 認識精度が改善できることが分かった。

評価に使用したデータは保育士が画角に収まっている部分のみを抽出したものであり、保育士が画角外で行動する場合もあるため、常にウェアラブルセンサデータと環境設置ビデオデータが併用できるとは限らない。しかし、併用できる場合には併用した方がより高い認識精度を期待できることが分かった。

5. まとめ

本研究では、保育士の行動を高精度で認識するためにウェアラブルセンサから得られる加速度・角速度データ、環境設置ビデオカメラから得られる関節座標データを組み合わせて認識を行った。実験結果より、ウェアラブルセンサと環境設置ビデオデータを組み合わせた場合に 98.2% の認識精度が得られ、それぞれ単体のデータで認識を行うよりも認識精度の向上が期待できることが分かった。

参考文献

[1] 藤原健之, 大見士, 大村廉, 石橋尚子 “深層学習を用いた保育士の行動における学習データ量と認識精度の関係”, 情報処理学会第 84 回全国大会, P167-168

[2] Junhyuk Kang, Jieun Shin, Jaewon shin, Daeho Lee and Ahyoung Choi “Robust Human Activity Recognition by Integrating Image and Accelerometer Sensor Data Using Deep Fusion Network.”, Sensors . 2022; 22(1):174 , <https://doi.org/10.3390/s22010174>