

深層学習を用いた保育士の行動認識における 学習データ量と認識精度の関係

藤原健之¹ 大見士² 大村廉³ 石橋尚子⁴

豊橋技術科学大学情報・知能工学課程¹

豊橋技術科学大学情報・知能工学課程²

豊橋技術科学大学情報・知能工学系³

椋山女学園大学教育学部子ども発達学科⁴

1. はじめに

現在、保育施設の急増とそれに伴う保育士不足が、保育現場の負荷を増大させ、様々な問題を生じさせている。また、保育士にとって「何がよい保育であるか」といったことは十分な研究がなされておらず、このような背景から、保育士の業務を記録・分析することが求められている。

そこで本研究では日々の保育士の行動を自動的に記録してその分析を行なうシステムの構築を目指し、ウェアラブルセンサを用いた行動認識技術による保育士の行動認識を試みる。これにより保育業務（行動）の定量的な分析が可能となると考えられる。

近年、機械学習分野では特徴量エンジニアリングを必要としない深層学習が注目を集めており、行動認識においても良い識別精度が得られることが分かってきた。このため、本研究では保育士の行動認識に深層学習の利用を試みる。

しかし、深層学習には大量の学習データが必要となる。また、業務分析などでは日々の業務で短時間しか現れない動作なども存在する。このため、継続的に学習データを取得しつつ識別器の改善を行なえるようになる事が望ましい。半教師あり学習などの手法により、このような継続的なデータの取得と識別器の学習が期待できるが、半教師あり学習の手法では初期認識モデルの精度がその後の学習効率や識別精度の上限に大きく影響することが予想される。そこで、本研究では、各種行動認識で用いられる深層学習の手法について、保育業務における行動データを用い、学習データ量と認識精度の関係の地調査を行った。

Relationship between The Amount of Training Data and Accuracy in Deep Learning Based Activity Recognition of Nursery Teachers

1 Fujiwara Kenshi, Toyohashi University of Technology

2 Omi Akira, Toyohashi University of Technology

3 Ohmura Ren, Toyohashi University of Technology

4 Ishibashi Naoko, Sugiyama Jogakuen University

2. 関連研究

Oluwaladeらは15種類の日常行動に対し、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深層学習の手法であるLSTM, BiLSTM, Convolutional LSTM, 及びCNNで行動認識を行い、それらの認識精度を比較した[1]。結果、畳み込みを行う認識モデルの方が畳み込みを行わない認識モデルより平均的に高いprecisionが得られることを示した。[1]のように一定の学習データ量に対して認識モデルを比較し、認識精度を確認した研究は存在するが、学習データ量と認識精度の関係を明示的に調べた研究は少ない。本研究では保育行動に対する各認識モデルの性質を確認するとともに、学習データ量と認識精度の関係を調査する。

3. 実験方法

実際に保育園（こども園）に勤務する保育士1名に、手首や足首など体の6箇所に3軸の加速度・角速度センサを装着してもらい（図1）、行動データを取得した。なおサンプリングレートは100Hzとした。

データの取得は実際の保育業務を対象に、9時～11時、13時～15時の合計4時間行った。ビデオによる分析から、39種類の保育行動を定義し、総実施時間が少ない（20秒以下）行動の行動を除いた27種類の行動を行動認識の対象とした。認識対象とした保育行動の一部を表1に示す。認識対象行動には、「前かがみで話している」「子供に触れている」など子供に合わせた姿勢での行動が多いなどの特徴がある。



図1. ウェアラブルセンサの装着位置

表 1. 認識対象の保育行動の一部

前かがみで話している	座って話している
足で線を引いている	見回している
連絡帳を読んでいる	紙を折っている
手をたたいている	子供に触れている

加速度・角速度データは、標準化を行った後、ウィンドウサイズ 1000ms, 重なり 75%でスライディングウィンドウを適用して各行動の部分系列を抽出した。行動のデータとして、合計 27970 インスタンスを得た。

識別器には、CNN・LSTM・CNN-LSTM を用い、それぞれ加速度データを直接入力した。CNN には既存研究において看護行動を対象とし作成された、畳み込み層を 2 層保持するモデルを用いた[2]。LSTM には日常行動を対象として作成された LSTM 層を 2 層用いたモデルを用いた[3]。CNN-LSTM には日常行動と行動間の動作を対象に畳み込み層が 3 層、LSTM 層が 1 層用いられているモデルを用いた[4]。バッチサイズはそれぞれ 128・128・256 でエポック数はそれぞれ 20・100・150 とした。バッチサイズは実行時間が長すぎず、かつ、認識精度が落ちない値、エポック数は認識精度が収束した値を用いた。

4. 実験結果

認識モデル別の、学習データのインスタンス数と認識精度 (Accuracy) を図 2 に示す。

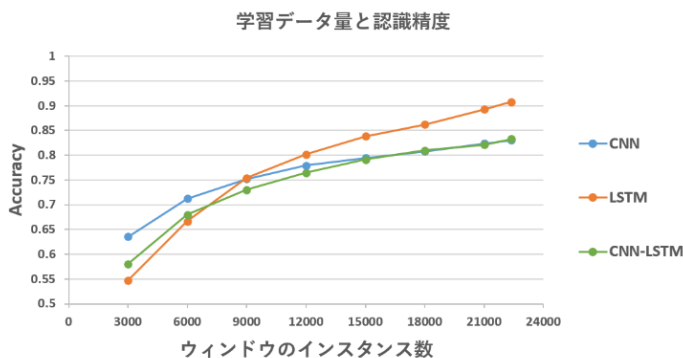


図 2. 認識モデル別の学習データ量と認識精度の関係

図 2 では、どの認識モデルにおいても学習データ量が増えると Accuracy が向上するとともに、まだデータ量に対して Accuracy の向上が収束していない。したがって、どの認識モデルにおいてもさらに学習データを増加させることで、識

別精度の向上が望めると考えられる。

LSTM では 21,000 インスタンスの学習データ量で Accuracy が約 0.90 を達成した。21,000 インスタンスは約 2 時間から 3 時間分のデータに相当する。今回、データ取得量が少なかった行動を除外しているものの、2 時間程度のデータがあればある程度高精度の認識を期待できることが分かった。今後、半教師あり学習などを導入する際にも、一つの目安となる学習データ量であると考えられる。

さらに、学習データ量に対する Accuracy の伸びが 1 番良かった認識モデルは LSTM であった。本研究で定義した保育行動の行動認識には、LSTM が適していると考えられる。

5. まとめ

本研究では保育行動における行動認識を行い、学習データ量と認識精度の関係について調べた。結果として、2 時間程度の学習データで約 0.90 の Accuracy での認識を期待できることが分かった。

謝辞

本研究は椋山女学園大学附属椋山こども園の職員の皆様、椋山女学園大学石橋ゼミ田尻萌乃様に多大なご協力を戴きました。深く御礼申し上げます。

参考文献

- [1] Bolu Oluwalade, Sunil Neela, Judy Wawira, Tobiloba Adejumo, Saptarshi Purkayastha, Human Activity Recognition using Deep Learning Models on Smartphones and Smartwatches Sensor Data, 14th International Conference on Health Informatics, 2021
- [2] Md. Golam Rasul, Mashrur Hossain Khan, Lutfun Nahar Lota, Nurse Care Activity Recognition Based on Convolution Neural Network for Accelerometer Data, UbiComp/ISWC '20, 2020, pp. 425-430
- [3] Selda Güney, Çağatay Berke Erdaş, A Deep LSTM Approach for Activity Recognition, 2019 42nd International Conference on Telecommunications and Signal Processing (TSP), 2019
- [4] Huaijun Wang, Jing Zhao, Junhuai Li, Ling Tian, Pengjia Tu, Ting Cao, Yang An, Kan Wang, Shancang Li, Wearable Sensor-Based Human Activity Recognition Using Hybrid Deep Learning Techniques, Security and Communication Networks, 2020, p. 12