

深層学習による多次元時系列データを用いた指文字認識

白石優旗¹⁾, 土屋智彦²⁾, 加藤伸子¹⁾, 米山文雄¹⁾, 設楽明寿³⁾

筑波技術大学 産業技術学部 産業情報学科¹⁾

筑波技術大学大学院 技術科学研究科 産業技術学専攻²⁾

筑波大学大学院 図書館情報メディア研究科 図書館情報メディア専攻³⁾

キーワード: 手話言語, データグローブ, 機械学習, 畳み込みニューラルネットワーク

1. はじめに

昨今のダイバーシティ推進により, ろう・難聴者と聴者が共に生活する機会の増加がより一層期待され, 両者の円滑なコミュニケーションの実現は急務である。その際, ろう・難聴者は音声をクリアに聞き取ることができないため, 代替手段を用いて情報獲得の支援を行う「情報保障」は必須である。これまでの情報保障は, 聴者の発生する音声を文字に変換してろう・難聴者に提示することが主流であった。しかしながら, コミュニケーションは双方向的であることが本質のため, ろう・難聴者の発話(手話・指文字)を聴者に文字または音声で提示することも必要となる。

1対1の局面では, 双方が文字を使う(筆談)等の工夫でコミュニケーション自体は可能であるものの, そのような対応がそもそも困難な状況がある。例えば, 講演・会議・グループでの話し合い等, ろう・難聴者が多数の聴者に対し手話を用いて発話するケースである。本研究ではこれらの状況を第一義的に想定する。

本研究では, 高精度の手話認識の実現により, 本課題の解決を目指す。本稿では, 動的指文字(手指を動かしながら提示する指文字)を含む全指文字を対象とする。その際, 認識手法には深層学習を用いる。また, センサとしてデータグローブを採用することにより, カメラと異なり環境光やオクルージョン(手と手の重なり)の影響を受けずに, またカメラの存在を意識せずに, 聴講者の方を向いて発話することができる。

2. 関連研究

現在の手話認識研究は, カメラ(RGBカメラ・深度センサを含む)を用いたものが主流である。また, 手話のサブセットである「指文字」や「フレーズ」を認識対象としたものがほとんど[1]であり, 実用的なレベルで「文」を認識するまでには至っていない。データグローブとカメラを共

に利用したものについてであれば, Ongらの報告[2]があるものの, 隠れマルコフモデルを活用した認識により, 英数字の指文字37文字と手話30単語の識別率80%程度に留まっている。データグローブのみを用いた手法については, Mumjadiらの報告[3]において, 様々な機械学習による手法(ただし, 深層学習は含まれていない)を比較・検討することで識別率92%を達成したと報告されているものの, 識別対象はフランス手話の内, 手指を動かさずに提示する静的指文字22文字に限定されている。

3. 開発システム

データグローブには, 長時間装着しても負担が少なく, かつ安価であり, 比較的軽量である導電繊維編み込み手法[4]を採用する。システム構成を図1に示す。

手の動作方向は慣性センサ MPU6050(3軸加速度センサ, 3軸ジャイロセンサ)を用いて計測する。5本指の動作量(曲げ量)はデータグローブの導電繊維の抵抗値の変化を, 分圧回路により電圧に変換することで計測する。これらの値は, Arduinoボードにより計測される。計測データは, USB接続されたPCにリアルタイムで転送され, 保存される。深層学習による学習・評価はPC上でTensorFlowにより行われる。

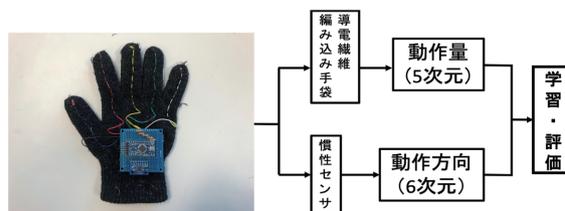


図1 システム構成

4. データ採取実験

指文字データの採取のため, 日常的に手話を使用してい

るろう・難聴の実験協力者 20 名 (20 ~ 27 歳) に対し、日本の指文字 76 文字 (濁音, 半濁音, 長音, 拗音を含む) を対象としたデータ採取実験を行った [5]。具体的には、データグローブを着用し、指文字の動作を一文字ずつ 1 秒間行ってもらった。なお、サンプリングレートは 200 個 / 秒で、指の動作量 (5 次元) と手の動作方向 (加速度 3 次元, 角速度 3 次元) について計測データ (多次元時系列データ) を取得した。また、データの収集と同時に手動でラベリングを行った。この一連の動作を 5 回繰り返した。なお、本実験は筑波技術大学倫理審査委員会の承認を得て実施した。

5. 識別アルゴリズム

識別には、畳み込みニューラルネットワーク (Convolutional Neural Network, CNN) を用いた。活性化関数には ReLU 関数を用いた。誤差関数には交差エントロピー関数 (式 1) を、学習アルゴリズムには RMSProp を用いた。

$$E = - \sum_k t_k \log y_k \quad (式 1)$$

ここで t_k は正解ラベル (one-hot 表現), y_k はネットワークの出力である。

本稿では、先行研究 [5] の課題を解決するため、(1) データスクリーニングと特徴量の改良、(2) 識別ネットワークの改良、(3) 動的指文字を含む 76 文字に対しての評価実験に取り組んだ。

(1) については、ジャイロドリフトを防ぐため、リアルタイムで加速度と角速度の各値から角度に算出する Madgwick フィルタ [6] を用いて角度 (\sin と \cos の 6 次元) を算出し、識別に用いた。また、汎化能力向上のため、200 個 / 秒のサンプリングデータに対し、移動平均によるデータ減少を行い 4 個 / 秒とした。

(2) については、図 2 に示すとおり、入力の際に、動作量、加速度、角速度、角度のそれぞれに分岐させた。その後 CNN を通し、最後に指文字の数に合わせて Dense を通して出力するネットワークを構築した。

(3) については、20 分割交差検証法で評価した。結果、平均識別率 70.0% の結果を得た。詳細については、文献 [7] を参照されたし。

6. まとめと今後の課題

本研究では、深層学習を用いて、動的指文字を含む全指文字を認識するシステムを開発した。評価実験により、20 名の実験協力者に対し、平均識別率 70.0% の結果を得た。今後は、データグローブの改良、前処理の改良を含む識別アルゴリズムの改良、並びに、学習データの追加収

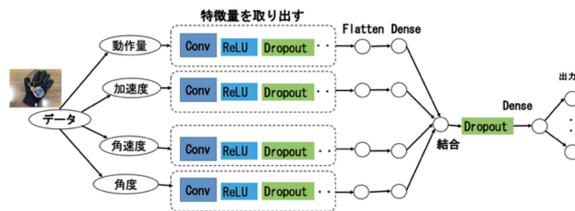


図2 識別ニューラルネットワーク

集を行うことで、識別率の向上と連続的に提示された指文字の認識に取り組む予定である。

謝辞

本研究は筑波技術大学教育研究等高度化推進事業並びに JSPS 科研費 19K11411 の助成を受けたものです。

参考文献

- [1] 松田啓佑, 山本雅人, 飯塚博幸, 手話動作分類における RCNN モデルの性能評価と内部状態解析, 2018 年度人工知能学会全国大会, 2018.
- [2] C. Ong, et. al., Sign-Language Recognition Through Gesture & Movement Analysis (SIGMA), In: Billingsley J., Brett P. (eds) Mechatronics and Machine Vision in Practice 3. Springer, Cham, 2018.
- [3] C. K. Mummadi, et. al., Real-time Embedded Recognition of Sign Language Alphabet Fingerspelling in an IMU-Based Glove, Proc. iWOAR 2017, Rostock, Germany, Sep. 2017.
- [4] 高田峻介, 志築文太郎, 高橋伸: 導電繊維編み込み手袋を用いた指の曲げ計測手法, 情報処理学 2017, Vol.2017-HCI-171, No.25., 2017.
- [5] 土屋智彦, 白石優旗, 深層学習を用いたセンサグローブによる指文字認識の改良. 情報処理学会アクセシビリティ研究会 (IPSJ SIG AAC) 第 9 回研究会; 2019-3.
- [6] S. O. H. Madgwick, A. J. L. Harrison and R. Vaidyanathan, Estimation of IMU and MARG orientation using a gradient descent algorithm, Proc. of the 2011 IEEE Int. Conf. on Rehabilitation Robotics, pp.1-7, Zurich, Switzerland. Jun. 2011.
- [7] T. Tsuchiya, A. Shitara, F. Yoneyama, N. Kato, Y. Shiraishi, Sensor Glove Approach for Japanese Fingerspelling Recognition System Using Convolutional Neural Networks, Proc. of the 13th Int. Conf. on Advances in Computer-Human Interactions (ACHI 2020), pp.152-157, Valencia, Spain, Mar. 2020.