

5ZD-02

短文手話動作の認識のためのセグメンテーションと識別精度の検討

栗原 泰晴[†] 黒尾 陽太[†] 川喜田 佑介[†] 西村 広光[‡] 田中 博[†] 三次 仁^{††}

神奈川県立 情報学部 情報工学科[†]
 神奈川県立 情報学部 情報メディア学科[‡]
 慶應義塾大学 SFC 研究所^{††}

1. はじめに

手話を認識するためには、複数の単語から構成される動きを認識する必要がある。本報告では、各単語の動作に分割(セグメンテーション)する方法として2つの手法を検討し、それに基づいて手話動作の認識を行った結果を述べる。

2. 使用する手話データ

本検討では、3単語から構成される日常的な5短文とそれを構成する11単語を使用した。使用した短文・単語手話動作を表1に示す。これらの手話動作は両手首と両肘の計4か所に無線加速度センサ[1]を取り付け取得した。表2に認識対象の短文データ、認識で用いる11種類の単語の識別モデルをLSTM(Long Short Term Memory)で作成するために用いた学習データ[2]、短文動作と単語動作のDTW(Dynamic Time Warping)距離を求めするために用いた単語データを示す。

表1 使用する手話短文・短文に含まれる単語

短文	1. 私/走る/汗だく	2. 私/ダイエット/中	3. 私/眼鏡/買う
	4. 私/眼鏡/忘れる	5. 息子/顔色/悪い	
単語	1. 汗だく	2. 最中(中)	3. ダイエット
	4. 走る	5. 顔色	6. 買う
	7. 眼鏡	8. 息子	9. 悪い
	10. 忘れる	11. 私	

表2 使用する手話データセット

使用用途	種類	手話者	手話動作	サンプル数(1動作)	サンプル数(合計)
認識対象	短文	D	5	1	5(1*5*1)
学習データ	単語	A,B,C	11	20	660(3*11*20)
DTW距離	単語	A	11	20	220(1*11*20)

3. 単語動作へのセグメンテーション

3.1 速度によるセグメンテーション

文章動作の各単語へのセグメンテーションにおいて速度変化に基づいて分割する方法が提案されている[3]。本検討では、動作範囲が広く常に動きのある右手首の加速度を使用し速度を求めた。ここで、積分に影響するバイアス成分は

“Segmentation and Classification Accuracy for Recognition of Short-sentence Sign Language Motions”

[†], [‡]Kanagawa Institute of Technology, Faculty of Information Technology

Yasuharu Kurihara[†], Youta Kuroo[†], Yuusuke Kawakita[†], Hiroshi Tanaka[‡], Hiromitsu Nishimura[‡]

^{††}Keio University SFC Research Institute
 Jin Mitsugi^{††}

DC カットフィルタを用いて除去した。そして、積分を行い3軸の速度を出力し、これらを合成して右手首としての速度を求めた。

本検討で使用している短文は3単語で構成されているため、速度の停留点のうち小さい点からの3点を分割点と定め、それぞれ「区間1, 2, 3」と定めた。使用した加速度と求めた速度とセグメンテーション結果の一例を図1, 2に示す。

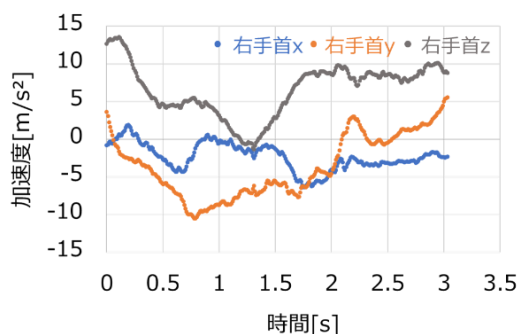


図1 “私/眼鏡/買う” 加速度(右手首)

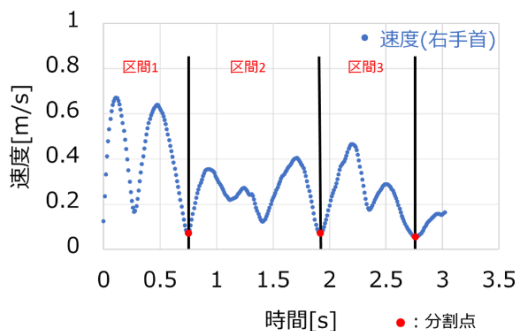


図2 “私/眼鏡/買う” 速度とセグメンテーション

3.2 DTWによるセグメンテーション

DTWによるセグメンテーションの方法を図3に示す。短文データを最初から最後の要素まで1要素ずつ増やしながら、11単語データとのDTW距離を求めた。そして、DTW距離が最小となった短文データの要素までを分割点とし、次の要素から同じ処理を行った。これを短文データの最終要素が最小のDTW距離となるまで分割点を定めた。1単語のDTW距離は20サンプルの平均である。

この提案手法から、短文5つに対して区間数が3区間の短文が2つ、2, 4, 5区間の短文が1つ

ずつあった。区間数が短文の構成単語数と一致しない原因として、短文動作に含まれる単語動作を接続する動きである「渡り」が影響している可能性がある。DTW による短文データのセグメンテーションの一例(3 区間)を図 4 に示す。同一データに対する図 2 とほぼ同じ結果が確認できる。

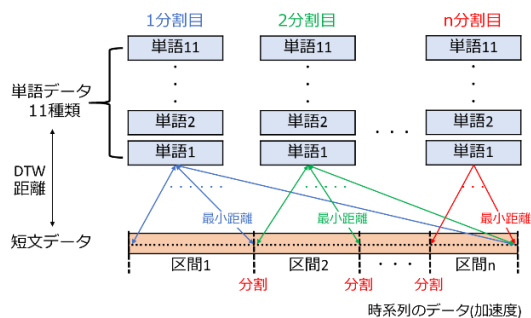


図 3 DTW によるセグメンテーションの方法

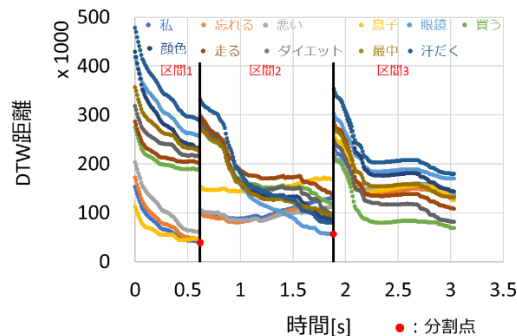


図 4 “私/眼鏡/買う” DTW のセグメンテーション

4. 尤度と DTW 距離による認識

本検討では、分割した区間内の本来の動作を“正解単語”とし 3 位内に含まれれば認識に成功したとする。提案した手法で分割した動作データに対して、LSTM で 11 単語を学習させた識別モデルでの尤度[2]の積算値と 11 単語との DTW 距離をそれぞれ算出し、上位 3 位までを認識結果とした。認識結果の一例として短文“私/眼鏡/買う”，“私/走る/汗だく”の場合を表 3, 4, 短文 5 つに対して認識に成功した区間数の結果を表 5 に示す。なお、表 3, 4 で塗られている箇所は区間内の 3 位内に含まれた正解単語である。

速度によるセグメンテーションの認識結果は、30 区間の内 25 区間は正解単語が 3 位内に含まれていた。また、DTW 距離による認識では正解単語が全て 1 位と認識される短文が 2 つあった。

DTW によるセグメンテーションの認識結果では、区間数が短文の構成単語数と一致した短文 2 つではすべての区間において認識に成功した。しかし、構成単語数と一致していない短文 3 つでは、正解単語が 3 位内に含まれない区間が多くあった。

表 3 “私/眼鏡/買う” 認識結果

分割方法	評価方法	順位	区間1	区間2	区間3
速度	尤度	1	私	買う	顔色
		2	買う	眼鏡	眼鏡
		3	息子	汗だく	最中
	DTW 距離	1	私	眼鏡	買う
		2	忘れる	汗だく	ダイエット
		3	息子	最中	走る
DTW 距離	尤度	1	私	顔色	買う
		2	顔色	眼鏡	最中
		3	息子	汗だく	顔色
	DTW 距離	1	私	眼鏡	買う
		2	息子	忘れる	ダイエット
		3	忘れる	悪い	走る

表 4 “私/走る/汗だく” 認識結果

分割方法	評価方法	順位	区間1	区間2	区間3	区間4
速度	尤度	1	買う	汗だく	汗だく	
		2	私	走る	眼鏡	
		3	悪い	眼鏡	ダイエット	
	DTW 距離	1	顔色	汗だく	眼鏡	
		2	私	眼鏡	汗だく	
		3	忘れる	最中		
DTW 距離	尤度	1	顔色	顔色	走る	汗だく
		2	私	私	顔色	走る
		3	悪い	忘れる	眼鏡	眼鏡
	DTW 距離	1	私	忘れる	ダイエット	汗だく
		2	忘れる	悪い	買う	眼鏡
		3	息子	私	最中	ダイエット

表 5 提案した方法による認識に成功した区間数

分割方法	分割数	正解単語が3位内に含まれた区間数	
		尤度	DTW 距離
速度	3	12 (15区間内)	13 (15区間内)
DTW 距離	3.4	13 (17区間内)	11 (17区間内)

5. まとめ

本検討では、短文手話動作の認識のためのセグメンテーション手法と識別精度の検討を行った。速度で分割する手法では、短文の構成単語数を把握している前提で分割し認識を行うと正解単語が 3 位内に含まれた区間が 25/30 区間あり全区間数の約 83%は認識に成功した。DTW 距離で分割する手法では、認識に成功した区間数は 24(DTW 距離)/25(速度)であり同様だが単語数の条件を与えていないので有効な手法と思われる。

謝辞

手話動作取得時にご指導いただいた株式会社ケイ・シー・シーの関係者各位に感謝いたします。本研究開発は総務省の「電波資源拡大のための研究開発(JPJ000254)」によって実施した成果を含みます。

参考文献

- [1] J. Mitsugi, et al., “Wireless Modal Analysis With Backscatter Sensors Subjected to Clock Instability,” IEEE J. RFID, vol. 5, Issue 3, pp. 269-277, Sept. 2021.
- [2] 栗原泰晴 他, “単語学習させた LSTM を用いた手話の短文動作識別に関する基礎検討”, 情報処理学会 FIT2022, M-012, 2022.
- [3] 内田佳佑 他, “手話認識のための手話動作の音素切り出し手法の検討”, 情報処理学会第 82 回全国大会, 1R-03, 2020.