# 体表面筋電信号からの個人特徴の抽出と

# バイオメトリクス認証への応用

# 生体計測工学研究室 2281219 奥村紘平

# (主査:京相 雅樹 教授 副査:桐生 昭吾 教授, 和多田 雅哉 教授)

# 1. はじめに

現在,日常生活において様々な場面で本人確認のためにバイオメトリクス認証が行われている. バイオメトリクス認証とは行動的または身体的な特徴を用いた個人認証方式である.この認証方法 は登録や認証にかかる時間が短く精度も高いといった特徴がある.一方で表面的な身体的特徴 を用いるため盗難のリスクがあり,一度情報が流出すると変更できないという欠点がある.

この問題を解決するためにため、本研究では生体信号の一つである体表面筋電信号に着目した個人認証システムの開発を目的として研究を行っている.体表面筋電信号を含む生体情報を用いるメリットとして表面的な身体的特徴とは違い、生体内部の情報であるため情報が流出する可能性が低く、波形の偽造や複製が困難であることなどが挙げられる<sup>[1]</sup>.

これまでの研究では、0.1s 間隔で切り出した体表面筋電信号から特徴量を抽出・選択しサポートベクタマシンによって識別率を求めた.本研究では識別率向上を目的として切り出し秒数を 0.5s に変更しサポートベクタマシンで識別率を求めた.また、識別に寄与し普遍性のある特徴量の傾向を汲み取った.

### 2. 原理

#### 2.1 筋電位

筋電位とは,生物の筋線維が収縮活動するときに発生する活動電位である.その筋肉内で発生する微弱な電場の変化を縦軸,時間経過を横軸に取った物を筋電図という.筋電位の発生順序はα運動神経細胞で生じた活動電位が神経筋接合部に到達すると,筋小胞体から Ca<sup>2+</sup>が放出される. Ca<sup>2+</sup>濃度の上昇に伴い筋線維の活性化が促され,筋線維が収縮し筋電位が発生する. Fig.1 に筋電発生のメカニズム, Fig.2 に筋電図の記録手法を示す<sup>[2]</sup>.



Fig.1 Mechanism of myoelectric generation

Fig.2 Electromyogram Waveform

# 2.2 ランダムフォレスト<sup>[3]</sup>

ランダムフォレストは,決定木とアンサンブル学習の2種の手法を組み合わせたアルゴリズムである.決定木とは,データを分類や回帰するための木構造のモデルであり,入力データを階層的

に分割して最終的な予想を行う.しかし,単一の決定木は過学習する傾向がみられる.また,アン サンブル学習は,複数の機械学習モデルを組み合わせてより高い予測性能を得る手法をしている. ランダムフォレストでは,データセットからランダムに重複を許してサンプリングを行い,各決定木が 異なるデータセットを元に構築するブートストラップサンプリング,各決定木の分割時に,特徴量の 一部だけをランダムに選択して分割の候補とすることにより全ての特徴量を考慮するよりも多様な 決定木を構成するランダムな特徴量選択,各決定木が独立に予想を行い,最終的な予測結果を 回帰であれば平均値,分類であれば多数決を行う平均化の3つの特徴を有することによって過学 習を抑制し,安定性の高い性能を得ている.ランダムフォレストの使用用途としては,分類・回帰・ 特徴量の重要度の評価など汎化性能が高いことを利用して様々な問題に対して広く使用されて いる.本実験では特徴量の重要度の評価に使用した.

ランダムフォレストによる特徴量重要度抽出では、ある特徴量でノードを分割することでどれだけ ジニ不純度を下げることが出来るかを考え、特徴量毎にその和を算出する.ジニ不純度は式(1)で 求めることが出来る.ここで、K はクラス数,p(k)はその領域でのクラス k の割合を表している.

$$G(k) = \sum_{k=1}^{K} p(k) \left\{ 1 - p(k) \right\}$$
(1)

また,重要度は式(2)で求めることが出来る. ここで, |F(j)|はノードの分割に特徴量 j が使用された回数,  $N_{parent}(i)$ はノード i におけるデータ数,  $G_{parent}(i)$ はノード i におけるジニ不純度,  $N_{leftchild}(i)$ はノード i の左側子ノードにおけるデータ数,  $G_{leftchildt}(i)$ は左側子ノードにおけるジニ不純度,  $N_{rightchild}(i)$ はノード i の右側子ノードにおけるデータ数,  $G_{rightchildt}(i)$ は右側子ノードにおけるジニ不純度, 純度を表している.

$$I(j) = \sum_{i=1}^{|F(j)|} [N_{parent}(i)G_{parent}(i) - \{N_{leftchild}(i)G_{leftchild}(i) + N_{rightchild}(i)G_{rightchild}(i)\}]$$
(2)

*I(j)*の値が大きい特徴量jがデータ分類に大きく貢献したことを意味する. つまり, *I(j)*の値が大きい 特徴量jを重要な特徴量であると考える.

## 2.3 サポートベクタマシン

サポートベクタマシン(SVM)はデータを多次元空間にマッピングし、その空間上でクラスを分離 する境界を見つけることで、データを異なるクラスに分類する。その際に、SVM は最も近いデータ 点との間に最大限のマージンを持つ境界を選択することを重視する。このマージンを最大化するこ とにより、未知のデータをより高い信頼性で分類することができる特徴がある。

学習データの特徴ベクトルを $\mathbf{x}^T = (x_1, x_2, ..., x_n)$ , 重みパラメータを $\mathbf{w}^T = (w_1, w_2, ..., w_n)$ とする. また各グループから分離超平面に最も近いデータをそれぞれ $x_+, x_-$ とすると, マージンdは式(1)で 求めることが可能である.

$$d = \frac{|w^{T}x_{+} + b|}{||w||} = \frac{|w^{T}x_{-} + b|}{||w||}$$
(1)

ここで、分離超平面Hと, x<sub>+</sub>, x<sub>-</sub>の間にはデータが存在しないため, SVM の問題は「全てのデー タは超平面からの距離が少なくともd以上であるという条件下で, マージンdを最大とするw, bを求 める」という問題に帰着する. 即ち, 式(2)によって SVM の最適化を行うことができる.

$$\{\widehat{\boldsymbol{w}}, \widehat{\boldsymbol{b}}\} = \arg \max_{\boldsymbol{w}, \boldsymbol{b}} d(\boldsymbol{w}, \boldsymbol{b}) \ s. t. \frac{|\boldsymbol{w}^T \boldsymbol{x}^i + \boldsymbol{b}|}{||\boldsymbol{w}||} \ge d$$
(2)

# 3. 研究方法

## 3.1 計測方法

Fig.3 に計測の概略図を示す.7人の被験者に対し,手首から8cmと16cmの部分に電極を貼り表面筋電図を計測した.なお表面筋電図波形は安静区間を10秒,ハンドグリップを用いた運動区間を10秒設け被験者ごとに3セット(計60秒間)計測を行った.また計測した表面筋電図データに対し,Table1に示すディジタルフィルタを用いてノイズ除去を行った.



Fig.3 Measurement position

Table 1 Digital filter settings

6	C
Parameter	Setting
Filter type	Butterworth
Low Cut Filter [Hz]	10
High Cut Filter [Hz]	100
Hum Filter [Hz]	50

#### 3.2 特徵量抽出

計測した表面筋電図を機械学習に適したデータに変換するため,運動区間のみの筋電図 波形の抽出及び特徴量抽出を行った.

波形については,運動区間のみの波形をプログラムによって切り出した. さらに, 0.1s 間 隔の計 100 サンプルを 1 データとして被験者ごとに 300 データの 5 名分で計 1500 データ, 0.5s 間隔の計 500 サンプルを 1 データとして被験者ごとに 60 データの 5 名分で計 300 デ ータをそれぞれ抽出した.

1 データ毎に切り出した表面筋電図波形から複数の特徴量を抽出した.本研究では tsfresh<sup>[4]</sup>と呼ばれる時系列特徴量の抽出器を用い,最大値,最小値,平均値等の統計的な特 徴量や,フーリエ係数などの様々な関数や次数を使用した特徴量を抽出した.また,抽出 された特徴量の中で全て同じ値のものや,非数や無限大になっている特徴量は消去した. 特徴量抽出器を用いた目的は,先行研究で用いていた基礎的な統計量に加えて複合的かつ 膨大な量の特徴量を算出することで表面筋電図による個人認証における万人に対して普遍 的な特徴量の選定を行うためである.

### 3.3 特徴量の重要度算出

特徴量抽出器によって得られた特徴量に対してランダムフォレストを用いた特徴量の 重要度の算出及び特徴量選択を行った.重要度の算出と選択を行うことで,得られた特徴 量の有用度をあらかじめ数値化し、冗長なデータから意味のある量的データを取り出し、 得られた量的データを取捨選択することで重要な特徴量のみを得ることが出来る. これ により、識別の解釈性を向上させることや計算コストの減少による処理時間の短縮などの メリットを得ることが出来る.

本実験では、特徴量の重要度の高い順に 100%, 80%, 60%, 40%, 20%の場合での識別 率を機械学習によって求めた.

#### 3.4 識別率検証および評価手法

抽出した特徴量を説明変数として、機械学習アルゴリズムで個人のデータごとに分類を 行った.本実験ではこれまでの研究で用いていた切り出し秒数が 0.1s であるデータに加え て、0.5s で切り出したデータを使用し、非線形カーネル SVM の識別結果を算出した.カー ネルは RBF カーネルを使用し、コストパラメータとガンマパラメータはグリッドサーチを 用い、最適な数値を用いた. 0.1s の場合は学習データを 200 データ、識別データを 100 デ ータとし 0.5s の場合は学習データを 40 データ、識別データを 20 データとし正解率を求め た. Table2 に SVM のパラメータ設定を示す.なお本研究では機械学習モデルの識別率を 正解率(accuracy)で定義した.正解率は全データのうち、正しく識別した割合を示す指標で ある. Table3 に混同行列の定義及び accuracy の算出法を示す.

Table 2 Support vector machine settings

Table 3 Confusion matrix

Parameter	Setting			predicted	
Kernel	RBF			Positive	negative
C	The best of fit by grid	4	positive	ТР	FN
C	search	true	negative	FP	TN
Gamma	The best of fit by grid	TP + TN			1
	search	а	$accuracy = \frac{1}{TP + FN + FP + TN}$		

#### 3.5 特徴量の普遍性の確認

3.4 の実験で得られた識別率の整合性と体表面筋電信号から得られる特徴量から個人識別に有効性と普遍性がある特徴量の傾向をつかむためにこれまでに使用していた5名の被験者に2名新たに加え,計7名の体表面筋電信号を用い,被験者7名中5名を1組とし,組み合わせを変え,計21組それぞれに0.5s間隔で切り出した波形を用いて3.2と3.3に示す手法で特徴量の抽出と選択を行った.また,それぞれの組に対してSVMを用いた個人識別を行った.それぞれのパターンで正解率を算出し,各組で一番正解率の高かったパターンで用いられていた特徴量名をカウントした.

## 4. 実験結果

#### 4.1 計測方法

計測実験で計測した被験者 1 人の体表面筋電信号を Fig.4 に他の被験者の体表面筋電信 を Fig.5 に示す. Fig.4 と Fig.5 を見比べてどちらも安静時と運動負荷時の振幅の違いが明



瞭に分かり,振幅の最大値に違いがある体表面筋電信号波形を得ることが出来た.

#### Fig.4 EMG waveforms of a subject



## 4.2 特徴量抽出

0.1s 間隔で切り出した体表面筋電信号から 222 個の特徴量を抽出できた.また,0.5s 間 隔で切り出した体表面筋電信号からは 3.4 と 3.5 の実験条件から複数の組で抽出を行った ため 239 個から 286 個までの間で特徴量が抽出された.得られた特徴量は、時系列の特徴 量として主に使われる最大値,平均値,それぞれが観測される位置,分散値,標準偏差, エントロピー,偏自己関数,前後の差分の平均値,フーリエ係数など切り出した筋電位波 形の特徴を様々な観点から数値化したものであった.

## 4.3 特徴量の重要度

それぞれの切り出し条件で得られた特徴量を 3.4 で示す方法によって得られた特徴量の 個々の重要度を算出し,重要度が高い順にソートを行った. Fig.6 に 0.1s 間隔で切り出した 際に得られた重要度のグラフを示す.重要度が一番高いもので 0.016 程度であった.切り 出し秒数が 0.5s の場合でも同等の結果が得られた. Fig.6 から重要度の最大値は約 0.016 と なった.また,重要度が 0.004 ほどである特徴量が多数あることが確認され,重要度が 0 で あるものは存在していなかった.しかし,他のグループでは重要度が 0 であるものも複数 確認された.



Fig.6 feature importance(222types)

## 4.4 識別率検証

Table5 に SVM において切り出し秒数を変化させた際の個人識別結果を示す。0.1s 間隔 よりも 0.5s 間隔で切り出した方が約 1.83 倍の正解率を得ることが出来た。0.5s では上位 20%の特徴量数を用いた際に最大の正解率 97.97%を得た。

Data state	100%	80%	60%	40%	20%
0.1s	53.50	52.30	50.30	51.50	50.10
0.5s	89.89	92.92	89.89	91.91	97.97

Table 5 Accuracy by cutout seconds

## 4.5 特徴量の普遍性について

Table6 と Fig.7 には今回対象とした被験者 21 組に対して,これまでと同様の手順で個 人識別を行った結果を示す.また, Fig.8 には各組で正解率が最大となった際のパターン で用いていた特徴量をカウントし,多い順にしたものを示した.さらに Table7 には今回 の実験で全ての組で使用された特徴量名をまとめた.

Table6 より,特徴量の選択範囲を狭くするほど平均の個人識別結果は上昇が見られた. Fig.7 の分布をみると,使用する特徴量の数を減らしていくほどデータのばらつきが少なくなりまとまりがある結果となった.また,上位 40%と 20%の特徴量を使用した際の 個人識別結果の最大値は 100%を取ることがあった.

そして, Fig.8 から今回1組以上に使用された特徴量は全部で145種類であり,その中でも今回対象とした被験者21組に対して,全ての被験者組に当てはまる特徴量はTable7より18種類であった.その18種類の特徴量の内12種類は時系列データを量子化し,その変換されたデータに対して様々な関数を用いて計算を行うものであった.



Table 6 全 21 組の平均個人識別結果

Fig.7 21 組の各パターンにおける個人識別結果



Feature name

Fig.8 選択された被験者の組の数ごとの特徴量

change_quantiles_	f_agg	"var"	isabs_False_	qh_1.0	_ql_0.6
change_quantiles_	_f_agg_	"var"	isabs_False_	_qh_1.0	_ql_0.4
ar_coefficient_coeff_5_k_10					
change_quantiles_	_f_agg_	"mean"	isabsTrue_	_qh_1.0_	_ql_0.6
change_quantiles_	_f_agg_	"mean"	isabsTrue_	_qh_1.0_	_ql_0.4
change_quantiles_	f_agg	"var"	isabs_True_	_qh_0.8_	_ql_0.2
change_quantiles_	_f_agg_	_"var"	_isabs_False_	_qh_1.0_	_ql_0.2
change_quantiles_	_f_agg_	"mean"	isabsTrue_	_qh_1.0_	_ql_0.2
change_quantiles_	f_agg	"var"	isabs_True_	_qh_0.8_	_ql_0.4
	me	an_abs_	change		
quantile q 0.2					
change_quantiles_	_f_agg_	"mean"	isabs_True	_qh_1.0	ql_0.0
absolute_sum_of_changes					
			of_changes		
change_quantiles_	f agg	"var"	isabs_True_	qh_1.0	ql_0.4
change_quantiles	f_agg qu	"var" antile	_isabs_True_ _q_0.9	_qh_1.0_	_q1_0.4
change_quantiles_ change_quantiles_	f_agg qu f_agg	"var" antile "var"	isabs_True 0.9 isabs_False	_qh_1.0 _qh_0.8	_ql_0.4 _ql_0.2
change_quantiles change_quantiles_ change_quantiles	f_agg qu f_agg f_agg	"var" antile "var" "var"	isabs_True _q_0.9 _isabs_False	<u>qh</u> 1.0 <u>qh</u> 0.8 <u>qh</u> 1.0	<u>ql_0.4</u> <u>ql_0.2</u> <u>ql_0.2</u>

Table 7 全組に選ばれた特徴量名

# 5 考察

## 5.1 計測方法

稀に振幅が飛びぬけて大きくなる瞬間がいくつか見られた.この現象は計測中に体動が 乗ってしまったことや力を入れる際に瞬間的な筋肉の動きなどが推測される.また、4.1 で 示したように被験者によって掌握運動中の振幅の最大値に差があったが、これは個人の筋 肉量の違いや電極を貼付した箇所の固定による影響であると考えられる.

### 5.2 特徵量抽出·重要度

切り出した体表面筋電信号から 222 個から 286 個の特徴量を抽出することによって、従 来の表面筋電図の波形からは読み取れない特徴量を生み出すことができたと考えられる. また、重要度については最大でも0.016であり0から1で正規化される特徴量の重要度と しては低い結果となった.しかし、ランダムフォレストで求められる重要度は単一の特徴 量の評価であるため機械学習した際に他の特徴量と組み合わされることで重要な情報とな る可能性があるため特に問題視する事象ではないと考えている.

#### 5.3 識別率検証

4.4 で示したように、切り出し秒数を 0.5s に変更した際に約 1.83 倍の正解率の向上が見られた. また、重要度の最大値が 0.1s のときは 0.016 であったが 0.5s のときは 0.031 であり約 2 倍に向上していた. この結果からも切り出し秒数が個人識別に与える影響が大きいことが伺える.

この結果については、今回の被験者の波形が偶然個人識別において特徴的で結果に反映 されている可能性も考慮されるため、被験者数を増やし普遍性のある結果であるかを確認 するために 4.5 の実験を行った.

#### 5.4 特徴量の普遍性の確認

Table6 の 0.5s で切り出した際の正解率と Table5 の結果を見比べても個人識別結果にあ まり差がないことから偶然相性のいい被験者がそろっていた可能性は低いと考えられる. また、Fig.7 での分布の広がりや外れ値を見ると特徴量選択の重要性を再確認するとともに 3.3 の特徴量選択手法が正しく機能していることが分かる.このことから,実用化に向けた 正解率の獲得という問題の解決に大きく近づくことができたと考えている.

また、Fig.8 より少なくとも1 組以上に選ばれた特徴量は様々な関数や次数から算出され た特徴量でありその抽出部分に何か偏りがあるわけではないと考えられる.しかし、選ば れた回数に注目すると Table7 より全 21 組に選ばれた特徴量 18 種類の中で特徴量名 に"change\_quantiles"を冠した時系列データを量子化し、その変換されたデータに対して 様々な関数を用いて計算を行うものが 12 種類あり明らかな偏りが生じていることから特 にその処理を行うことで体表面筋電信号において普遍性があり個人識別に有用な特徴量に なり得る可能性がある特徴量となる可能性が示唆された.その他にも Fig.8 より,特徴量 名に"ar\_coefficient"を冠した自己回帰係数に関する特徴量も高い頻度で使用されていた.

#### 6 まとめ

本研究では、前腕屈筋群を標的とし、手首から 8cm と 16cm の部分に配置した電極から 計測した筋電データから tsfresh を用いて特徴量の抽出を行い、ランダムフォレストによ る特徴量の重要度を算出した.その特徴量を説明変数とし、SVM による個人識別を行っ た.結果として、0.1s で切り出したデータよりも 0.5s で切り出したデータの方が識別率 として約 1.83 倍の結果を得た.また、同様の手順で被験者を 2 名増やし 5 人を 1 グルー プとして計 21 組に対して識別を行った結果に再現性があり、普遍性のある特徴量の傾向 も確認された.

今後はより実用化を進めていくにあたり現在前腕屈筋群を標的として体表面筋電信号 を取得しているがこれを手の母指球筋における短母指屈筋を標的にし,個人識別を行って いくことが考えられる.

#### 参考文献

- [1] 京相雅樹, その他の生体特徴による個人認証, 生体医工学:日本エム・イー学会誌, vol.44, no.1, pp.47-53, 2006.
- [2] 增田正、佐渡山亜兵,多点表面筋電図,東京電機大学出版局, pp9-10, 2019.
- [3] A. C. Muller and S. Guido (邦訳:中田秀基), Pythonではじめる機械学習-scikit-learn で学ぶ特徴量エンジニアリングと機械学習の基礎,オイラリー・ジャパン, 2018.
- [4] Maximilian Christ, Nils Braun, Julius Neuffer, Andreas W. Kempa-Liehr, "Time Series Feature Extraction on basis of Scalable Hypothesis tests", Neurocomputing, volume 307, pp. 72-77, 2018.