

加速度センサを用いたジェスチャ入力における検出と分類の評価による高精度化の検討

網屋 友彰^{1,a)} 勝間 亮^{1,b)}

概要：

スマートフォンの加速度センサを利用して、スマートフォンを把持したユーザの手の動きを入力とするジェスチャ入力が新たな入力手法として期待されている。しかし、加速度値からジェスチャを認識する際の誤認識や、ノイズの多い歩行中の誤検出が問題となっている。本研究では動的時間伸縮法をもとにしたジェスチャ検出および分類手法を提案する。実験では歩行中のスマートフォンの音楽プレイヤーを想定し、静止中のジェスチャ入力の誤認識と歩行中の誤検出が同時に低くなるパラメータ設定の評価を行った。

キーワード：加速度センサ，ジェスチャ，スマートフォン，動的時間伸縮法

1. はじめに

スマートフォンには、多くのセンサが搭載されている。中でも加速度センサは、スマートフォンを把持したユーザの手の動きを検出することができる。このときの手の動きをジェスチャといい、以降では加速度センサ搭載機器を把持した状態でのジェスチャの種別を入力とする方式を単純にジェスチャ入力という。加速度センサのみを利用するジェスチャ入力は、ノイズが少ない環境で、1つあるいはごく少数のジェスチャを検出する使用場面で利用されている。ゲーム機のコントローラとして多く利用されており、スポーツゲームや格闘ゲームにおいて、実際に手を動かすという入力で、直観的な操作を可能にしている。またジェスチャ入力は、ボタンの配置を覚える必要や、画面を見る必要がないので、新たな入力手法として期待されている。期待されている。しかし、ノイズが多い環境や多くのジェスチャ候補がある場合、ジェスチャの認識精度が低くなってしまいう問題がある。ある加速度時系列データが与えられたとき、多くのジェスチャ候補から1つの正解ジェスチャを求める際のジェスチャ認識の基本的な方法は、入力として想定されるすべてのジェスチャの加速度値の時間変化を事前に測定しておき、この事前データと実際に入力したときの加速度値との類似度を計算する方法である [1] [2]。このとき入力データと最も類似している事前データのジェス

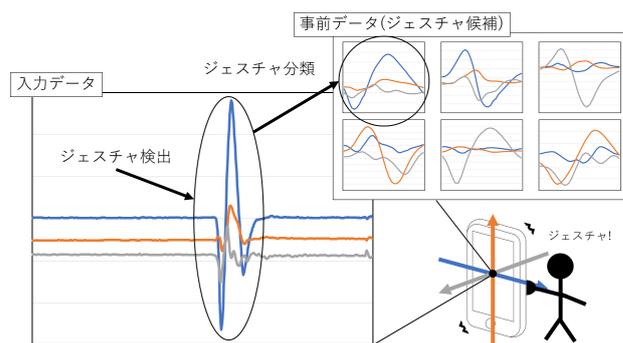


図 1 ジェスチャ認識の概要

チャが行われたと認識する。

ジェスチャ認識は、以下の2つのステップで構成されている。図1は、ジェスチャ認識の概要を示す。

- (1) ジェスチャ検出: いつジェスチャが行われたかを検出する
- (2) ジェスチャ分類: どのジェスチャが行われたかを分類する

ジェスチャ検出において、ジェスチャ入力があるときに検出できないことと、ジェスチャ入力がないときに検出してしまうことが誤認識となる。しかしこの2つの間にはトレードオフの関係があり、両方の数を減らすことは難しい。

既存研究では特徴量ベースの機械学習を用いて、ジェスチャ分類の精度を上げるものが多い [3] [4] [5]。取得した加速度時系列データには必ずジェスチャが行われたと想定する、もしくはジェスチャなしをジェスチャ分類の1候補とするため、ノイズの多い歩行中などで検出精度を保つとい

¹ 大阪府立大学
Osaka Prefecture University
a) sab01005@edu.osakafu-u.ac.jp
b) katsuma@cs.osakafu-u.ac.jp

う観点がない。本研究ではジェスチャ検出とジェスチャ分類で評価を分けることで、その両方でそれぞれ精度を算出し、パラメータ設定の評価を行った。

2. 関連研究

同じジェスチャであっても時間や状況によって動作の速さが変化する。このような場合、取得される入力データのサンプル数や加速度値にばらつきが生じるため、事前データとの類似度を求めることが難しくなる。ジェスチャの動作時間が毎回同じでないという問題点について、そのようなデータ間の類似度を計算するアルゴリズムとして動的時間伸縮法 (DTW:Dynamic Time Warping) がある [6] [7] [8]。DTW では、2つの時連続データにおいて、時間を伸縮させることで、サンプル数や加速度値の違いを解消し類似度を計算することができる。本研究では、事前データと入力データとの類似度の計算に DTW を用いる。

Izuta らの研究では、ジェスチャ検出に DTW による類似度計算を用いた [9]。このとき入力となる時連続データにおいて、事前データとの類似度を比較することでジェスチャの入力終了を待つことなくジェスチャを認識できることが示されている。

以上の研究では、取得した加速度時系列データには必ずジェスチャが行われたと想定する、もしくはジェスチャなしをジェスチャ分類の1候補とするため、ノイズの多い歩行中などで検出精度を保つという観点がない。本研究ではジェスチャ検出とジェスチャ分類で評価を分けることで、その両方でそれぞれ精度を算出し、パラメータ設定の評価を行った。

3. ジェスチャ認識アルゴリズム

入力として想定される n 個のジェスチャを G_i ($i = 1, 2, 3, \dots, n$) とし、事前に測定した G_i の加速度データ系列を事前データ B_i とする。入力データは加速度データ系列 A とする。

各時刻で A と各事前データ B_i との類似度を計算する。類似度は DTW により距離 $D(A, B_i)$ として求め、最も小さいものが最も類似しているとする。最も距離の小さい事前データ B_I との距離が、2番目に距離の小さい事前データ B_{II} との距離よりも明らかに小さくなったとき、ジェスチャを検出し、ジェスチャ G_I をジェスチャ分類結果として出力する。そのときの条件は、以下ようになる。 $\theta (> 1)$ はジェスチャ検出のための閾値である。

$$\theta D(A, B_I) < D(A, B_{II}) \quad (1)$$

θ の値が小さいときジェスチャを検出しやすくなるため、ジェスチャ入力があるときに検出できない誤認識を防げるが、ジェスチャ入力がないときに検出してしまう誤認識が

	a. 右に振る
	b. 左に振る
	c. 上に振る
	d. 下に振る
	e. 手前に振る
	f. 奥に振る
	g. 右に傾ける
	h. 左に傾ける
	i. 手前に傾ける
	j. 奥に傾ける
	k. 円を描く
	l. 三角形を描く
	m. 背面を2回タップする

表 1 候補となる基本ジェスチャ

増加すると考えられる。一方 θ の値が大きいときジェスチャを検出しにくくなるため、ジェスチャ入力がないときに検出してしまう誤認識を防げるが、ジェスチャ入力があるときに検出できない誤認識が増加すると考えられる。

4. 実験

4.1 実験目的

本実験では、ジェスチャが行われる場面はノイズの少ない静止中であり、ノイズの多い歩行中はジェスチャが行われない環境を想定する。静止中のジェスチャ分類の精度の評価と、歩行中のジェスチャ誤検出の評価を行うことで、閾値として設定した θ の適切な値を評価する。

4.2 実験環境

入力候補となるジェスチャを表 1 に示す 13 種類とし、加速度データの取得はユーザが静止した状態で行った。各ジェスチャに対して 8 個のデータを取得した。また、歩行時の加速度データは、スマートフォンを把持した状態とポケットに入れた状態で、同じ道を直進歩行したときのデータをそれぞれ 3 分程度取得した。その後、この 2 つのデータのサンプル数が同じになるよう、取得開始直後と取得終了直前のデータを取り除いた。実験機器として以下のものを用いた。

- スマートフォン iPhone8 (CPU: Apple A11 Bionic 6 コア 2.39GHz, メモリ: 2GB, 3 軸加速度センサ内蔵: サンプルング周期 100Hz)

ジェスチャデータに対して検出した	真陽性
ジェスチャデータに対して検出しなかった	偽陰性
歩行データに対して検出した	偽陽性

表 2 ジェスチャ検出の結果

あるジェスチャ入力に対して そのジェスチャを正しく出力した	真陽性
あるジェスチャ入力に対して 別のジェスチャを出力した	偽陰性
あるジェスチャを出力したとき 別のジェスチャが入力されていた	偽陽性

表 3 ジェスチャ分類の結果

4.3 実験順序

以下の順序で実験を行った。

- (1) ジェスチャ検出において、 θ の値を変化させ、F 値が最も大きくなる値を計測した。
- (2) ジェスチャ分類において、F 値によりその精度を評価した。また、ジェスチャ検出を考慮したときの分類精度も評価した。

4.4 評価方法

ジェスチャ検出において、与えたデータの種類と出力に対して結果を表 2 のように定めた。

ジェスチャ分類においても同様に、与えたデータの種類と出力に対して結果を定めた (表 3)。

そして評価指標として、再現率、適合率、F 値を次のように定めた。ジェスチャ検出において、再現率はジェスチャ入力があったときに検出できた割合、適合率はジェスチャを検出したときに実際にジェスチャ入力があった割合に相当する。ジェスチャ分類において、再現率はあるジェスチャ入力に対してそのジェスチャを正しく出力できた割合、適合率はあるジェスチャを出力したときに実際にそのジェスチャが入力されていた割合に相当する。F 値は再現率と適合率の調和平均であり、認識精度を総合的に評価できる。

$$\text{再現率} = \frac{\text{真陽性数}}{\text{真陽性数} + \text{偽陰性数}} \quad (2)$$

$$\text{適合率} = \frac{\text{真陽性数}}{\text{真陽性数} + \text{偽陽性数}} \quad (3)$$

$$\frac{1}{F \text{ 値}} = \frac{1}{2} \left(\frac{1}{\text{再現率}} + \frac{1}{\text{適合率}} \right) \quad (4)$$

$$F \text{ 値} = \frac{2(\text{再現率} \times \text{適合率})}{\text{再現率} + \text{適合率}} \quad (5)$$

$$= \frac{2 \times \text{真陽性数}}{2 \times \text{真陽性数} + \text{偽陰性数} + \text{偽陽性数}} \quad (6)$$

本研究では、取得した各 8 個のジェスチャデータから各

真陽性	713
偽陰性	15
偽陽性	6
再現率	0.979
適合率	0.992
F 値	0.985

表 4 ジェスチャ検出の精度 ($\theta=1.10$)

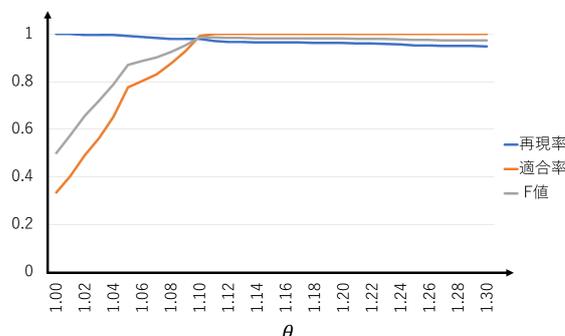


図 2 θ を変化させたときの精度の変化

1 個を事前データ、残りを入力データとし交差検証を行った。また、ジェスチャ検出においては、分類結果によらず検出の有無のみで評価した。同様にジェスチャ分類においては、必ず検出できている ($\theta=1.00$ のとき) と仮定し分類結果のみで評価した。

5. 実験結果

5.1 ジェスチャ検出

ジェスチャ検出の精度は表 4 のようになった。表に示す値は $\theta=1.10$ のときである。偽陰性数が偽陽性数より多いため、再現率が適合率より低いことが確認できた。総合的な精度の評価とする F 値は 98.5%であった。

また、 θ の値を 0.01 ずつ変化させたときの精度の変化は、図 2 のようになった。 $\theta = 1.10$ 未満のとき、再現率は非常に高い値を維持しているが、適合率と F 値が低い値である。 $\theta=1.11$ のとき F 値は最大値 98.6%となり、このときの偽陽性数は 0 であった。 θ が 1.11 を超えると再現率の低下の影響で F 値が徐々に低下することが分かった。

5.2 ジェスチャ分類

ジェスチャ分類の出力は表 5 のようになった。行は与えたデータの正解ラベル、列は出力したラベルであり、与えたデータに対して出力した個数が示されている。対角線上の個数が真陽性であり、対角線上にないもののうち、行で見たものが偽陰性、列で見たものが偽陽性である。「d. 下に振る」のジェスチャを入力したとき、そのうち 49 回は正解し、「c. 上に振る」を 3 回出力、「m. 背面を 2 回タップする」を 4 回出力し、計 7 回の偽陰性があった。また、「c. 上に振る」と出力したデータのうち、56 回は正解し、

	a	b	c	d	e	f	g	h	i	j	k	l	m	偽陰性
a. 右に振る	56	0	0	0	0	0	0	0	0	0	0	0	0	0
b. 左に振る	0	56	0	0	0	0	0	0	0	0	0	0	0	0
c. 上に振る	0	0	56	0	0	0	0	0	0	0	0	0	0	0
d. 下に振る	0	0	3(2)	49	0	0	0	0	0	0	0	0	4(1)	7(3)
e. 手前に振る	0	0	0	0	56	0	0	0	0	0	0	0	0	0
f. 奥に振る	0	0	0	0	0	56	0	0	0	0	0	0	0	0
g. 右に傾ける	0	0	0	0	0	0	56	0	0	0	0	0	0	0
h. 左に傾ける	0	0	0	0	0	0	0	56	0	0	0	0	0	0
i. 手前に傾ける	0	0	0	0	5(3)	0	0	0	51	0	0	0	0	5(3)
j. 奥に傾ける	0	0	0	0	0	4(3)	0	0	0	52	0	0	0	4(3)
k. 円を描く	0	0	0	0	0	0	0	0	0	0	56	0	0	0
l. 三角形を描く	0	0	0	0	0	0	0	0	0	0	0	56	0	0
m. 背面を2回タップする	0	0	0	0	0	0	0	0	0	0	0	0	56	0
偽陽性	0	0	3(2)	0	5(3)	4(3)	0	0	0	0	0	0	4(1)	

表 5 ジェスチャ分類の出力表

	再現率	適合率	F 値
a. 右に振る	1.000	1.000	1.000
b. 左に振る	1.000	1.000	1.000
c. 上に振る	1.000	0.949	0.974
d. 下に振る	0.875	1.000	0.933
e. 手前に振る	1.000	0.918	0.957
f. 奥に振る	1.000	0.933	0.966
g. 右に傾ける	1.000	1.000	1.000
h. 左に傾ける	1.000	1.000	1.000
i. 手前に傾ける	0.911	1.000	0.953
j. 奥に傾ける	0.929	1.000	0.963
k. 円を描く	1.000	1.000	1.000
l. 三角形を描く	1.000	1.000	1.000
m. 背面を2回タップする	1.000	0.933	0.978
平均	0.978	0.980	0.978

表 6 ジェスチャ分類の精度

	再現率	適合率	F 値
a. 右に振る	1.000	1.000	1.000
b. 左に振る	1.000	1.000	1.000
c. 上に振る	1.000	0.966	0.982
d. 下に振る	0.938	1.000	0.968
e. 手前に振る	1.000	0.949	0.974
f. 奥に振る	1.000	0.949	0.974
g. 右に傾ける	1.000	1.000	1.000
h. 左に傾ける	1.000	1.000	1.000
i. 手前に傾ける	0.942	1.000	0.970
j. 奥に傾ける	0.943	1.000	0.971
k. 円を描く	1.000	1.000	1.000
l. 三角形を描く	1.000	1.000	1.000
m. 背面を2回タップする	1.000	0.988	0.991
平均	0.986	0.988	0.987

表 7 ジェスチャ検出を考慮したときの
ジェスチャ分類の精度

3回は「d. 下に振る」が入力されていた。このとき偽陽性数は3である。「e. 手前に振る」と「i. 手前に傾ける」のように動作が類似するジェスチャ間で誤認識してしまうことが確認できた。

この結果から算出された精度は表6のようになった。平均の再現率は97.8%、適合率は98.0%であり、分類精度とするF値は97.8%であった。似たジェスチャ間で精度を低下させているため、候補となるジェスチャの選別により、精度を向上させられると考えられる。

次に、 $\theta=1.11$ のときのジェスチャ検出での結果を考慮したとき、ジェスチャ分類の出力は表5の()内の値となった。()のない数値はジェスチャ検出を考慮しないときと同じである。ジェスチャ検出で真陽性であった出力のみの分類結果である。偽陰性もしくは偽陽性であったものの出力数がそれぞれ減少していることが確認できた。

この結果から算出された精度は表7のようになった。平均の再現率は98.6%、適合率は98.8%であり、分類精度とするF値は98.7%であった。表6と比較すると、分類精度

が向上したことが確認できた。ジェスチャ分類において間違っていた出力のいくつかがジェスチャ検出の段階で出力をさせなかったからであると考えられる。

6. まとめ

本研究では、スマートフォンの加速度センサを用いたジェスチャ入力において、DTWを利用したジェスチャの認識手法での閾値設定とその評価を検出と分類の両面から行った。ジェスチャ検出の精度は最適な閾値で98.6%、ジェスチャ分類の精度は97.8%であることが確認できた。また、ジェスチャ検出を考慮したときのジェスチャ分類の精度は98.7%であることが確認できた。今後は、ジェスチャ検出とジェスチャ分類を総合的に評価する手法、より精度の高いジェスチャ認識アルゴリズムの考案、ユーザの違いによる閾値の自動設定の議論についても行う必要があると考えている。

参考文献

- [1] 坂口貴司, 金森務, 片寄晴弘, 佐藤宏介, 井口征士: 加速度センサとジャイロセンサを用いたジェスチャ認識, 計測自動制御学会論文集, Vol. 33, No. 12, pp. 1171–1177 (1997).
- [2] BAEK, J., Ik-Jin, J. and Byoung-Ju, Y.: Recognizing and Analyzing of User’s Continuous Action in Mobile Systems, *IEICE transactions on information and systems*, Vol. 89, No. 12, pp. 2957–2963 (2006).
- [3] JING, L., ZHOU, Y., CHENG, Z. and WANG, J.: A Recognition Method for One-Stroke Finger Gestures Using a MEMS 3D Accelerometer, *IEICE transactions on information and systems*, Vol. 94, No. 5, pp. 1062–1072 (2011).
- [4] XUE, Y., HU, Y. and JIN, L.: Activity Recognition Based on an Accelerometer in a Smartphone Using an FFT-Based New Feature and Fusion Methods, *IEICE Transactions on Information and Systems*, Vol. 97, No. 8, pp. 2182–2186 (2014).
- [5] HAN, X., YE, J., LUO, J. and ZHOU, H.: The Effect of Axis-Wise Triaxial Acceleration Data Fusion in CNN-Based Human Activity Recognition, *IEICE Transactions on Information and Systems*, Vol. 103, No. 4, pp. 813–824 (2020).
- [6] 村尾和哉, 寺田努: 加速度センサの定常性判定による動作認識手法, 情報処理学会論文誌, Vol. 52, No. 6, pp. 1968–1979 (2011).
- [7] Murao, K., Yoshida, G., Terada, T. and Tsukamoto, M.: Determining a Number of Training Data for Gesture Recognition Considering Decay in Gesture Movements, *Information and Media Technologies*, Vol. 10, No. 3, pp. 449–458 (2015).
- [8] Rakthanmanon, T., Campana, B., Mueen, A., Batista, G., Westover, B., Zhu, Q., Zakaria, J. and Keogh, E.: Searching and Mining Trillions of Time Series Subsequences under Dynamic Time Warping, Vol. KDD ’12, pp. 262–270 (2012).
- [9] Izuta, R., Murao, K., Terada, T. and Tsukamoto, M.: Early gesture recognition method with an accelerometer, Vol. 11, No. 3, Emerald, pp. 270–287 (2015).