F-001

教師なし学習による意図的動作と偶発的動作の判別

Discrimination between Intentional Actions and Accidental Actions by Unsupervised Learning

山田 恭士 * 春井 宏介 * 岡 夏樹 *
Yasushi Yamada Kosuke Harui Natsuki Oka

1. はじめに

情報システムやロボットと人との関わりは今後より一層密接になってくることが考えられるが，それは幅広い人々が，日常生活の中で頻繁にそれらのシステムを利用するということである．このような状況になったときに求められることは，人とそれらのものとのやり取りが人同士のコミュニケーションのようにスムーズに行えることである．例えば，ロボットが人間の動きを模倣して仕事をする場合に，人間が手を滑らせて皿を割ってしまったのを模倣してしまうのは避けたい．また，カーナビゲーションシステムにおいて，カーナビゲーションシステムが最適と思われる道を教示したとき，人間が意図して違う道を選択した場合なら，次からはそのルートを教示するように適応してほしいが，人間が誤って違う道を選択した場合は適応させたくない．これらのようにロボットやシステムが人間の動作を模倣したり，意図に適応したりする場合にスムーズなやり取りを実現するには，相手の意図を自然に読み取る能力が必要となる．

しかし，意図的かどうかを判別する意図認識能力は幼児でも持っているとされている [1] [2] にも関わらず，工学的にそれを実現する方法はまだ確立していない．

文献 [3] では，仮想空間上に音声入力とキー入力によるボール遊びゲームを作成し，システムにボールの扱い方をお手本を見せながら教える実験が行われた．この実験中に行われた意図的動作と偶発的動作を実験者が判定して，それぞれにラベルをつけ，このラベルを教師信号として教師付き学習による意図的動作と偶発的動作の判別が試みられた．

これに対して本研究では，教師なし学習による判別を試みる．これは，先に述べた幼児による意図判別においては，そのような教師信号の無い状況で判別の学習を行っているであろう事や，ロボットなどに人間の動作を模倣させたりする時などの，教師信号を用意しづらい状況も考慮に入れたからである．また，判別の特徴量については，文献 [3] で提案されたものに加え，パワーの差分やピッチも用いて判別を行う．

本稿の構成は，第2章で本研究で利用した文献 [3] の実験システムの概要について述べ，第3章では特徴量の抽出，第4章で今回用いた判別手法について述べる．第5章で今回の実験により得られたデータと，それによって行った判別の結果と考察について述べる．最後に，第6章で本研究のまとめと今後の課題などについて述べる．

2. 実験システムの概要

本研究で用いた実験データをどうやって取得したかについて，本章で説明する．本章で述べる実験はすべて櫻井ら [3] によって行われたものである．

櫻井ら [3] は，仮想空間上に音声入力とキー入力によるボール遊びゲームを作成した．ボール遊びゲームの画面構成を図1に示す．画面には，ユーザが動かすことのできる手と，システムが動かす手，ボールが配置されており，画面左上にはスコアが表示されている．区別しやすくするため，システムの手の方がユーザの手よりも少し小さめに設定してある．ユーザの手は，矢印キー (↑↓→←) を用いて上下左右に動かすことができる．またボールを上下させる，放す，つかむ，投げる，はじく動作がそれぞれ，S，Z，X，C，Vキーに対応しており，分かり易いようにそれぞれのキー上にラベルを貼り付けてある．システムの手は，上下左右への移動，放す，つかむ，投げる，はじくをランダムで行う．

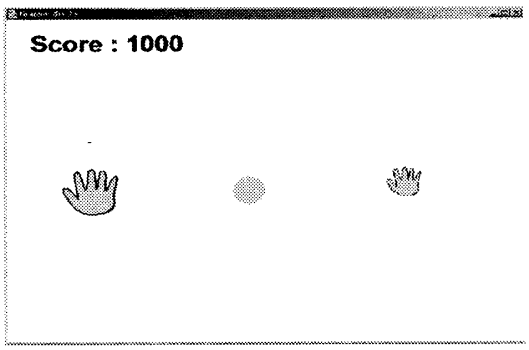

図 1: ボール遊びゲームの画面構成

この実験で行われた実験手法は次の通りである．なお，被験者は，男性9名，女性1名 (ともに20代) の計10名であった．まず，被験者に操作説明を行い十分慣れるまでボール遊びゲームを練習させた．続いて，つかむ動作やはじく動作を，キー操作と発話を用いてお手本を見せながら繰り返し教えることと，自分がお手本を見せる時は必ずボールを上下させてから行うことを指示した．この時，被験者が意図通りにつかんだり，はじいたりできると手がオレンジ色に変わるとともにスコアは上がり，意図通りの動作ができなかった場合，すなわち，つかめなかったり，はじいてしまったり，はじけなかったりするとスコアは下がり，また，一定時間ごとにもスコアが下がることを伝えた．その後，各人12分間システムにボール遊びを教える操作を行った．

全実験終了後，ビデオ録画した被験者の発話内容とキー入力の対応付けを行った．

*京都工芸繊維大学,Kyoto Institute of Technology

3. 特徴量の抽出

まず，録画したビデオからキャプチャした全被験者の動画ファイル (WMV 形式) を音声ファイル (wave 形式) に変換した．個々の被験者の wave ファイルを，つかんで投げる動作を教えるタスクの場合はつかむ動作，はじく動作を教えるタスクの場合ははじく動作の部分で切り出した．切り出す範囲は，動作完了前の発話開始時間から動作完了後の発話終了までとした．そして，音声信号処理ツールである wavesurfer を用いて特徴量の抽出を行った．

ひとりの被験者の発話のサンプルを図 2 に示す．被験者は，“ボールつかむよ、えいっああっ”と発話しており，図左側の音声波形が“ボールつかむよ”に対応しており，図右側の音声波形が“えいっああっ”に対応している．実験の映像，音声，及び音声波形から動作完了前発話の開始・終了時刻，動作完了時刻，動作完了後発話の開始・終了時刻の 5 つの時刻を取り，発話区間内でのパワー・ピッチの平均値をその発話のパワー・ピッチの大きさとし，動作完了時刻から動作完了後発話の開始時刻までの時間を発話タイミングとした．発話が無い場合の特徴量の値の取りかたは困難な問題であるが，本研究では発話が無い場合のパワー・ピッチの大きさは 0 とする．また，動作完了後発話が無い場合の発話タイミングについては，本来は∞や通常の発話タイミングと比べて十分に大きい値とするのが適当であると推測されるが，クラスタリングを行う際にそのような極端に大きな値が存在すると結果に大きな影響がでる恐れがあるので，本研究では 0 とする．

また，特に偶発的動作を行った場合に動作完了後の発話のパワーに大きな変化が表れると考えられることより，動作完了前後の発話のパワーの差分を取ったものを特徴量とする．そして動作完了後の発話の特にどの部分が判別に有効かを確認するために，動作完了後の発話を時間で 3 等分してそれぞれのパワーを求めたものも特徴量として扱う．

文献 [3] では動作完了前後の発話のパワー及び発話タイミングを離散値の特徴量として抽出している．その離散化の方法は，発話のパワーについては被験者の通常時の音声波形と比較し音声を聴き比べた上で人手によりラベル付けを行い，発話タイミングは 0.5 秒を基準にラベル付けを行っている．特にパワーに関してより客観的に離散化を行うために，本研究では ID3 アルゴリズムにより判別を行う際は，各特徴量の最大値と最小値を基準として 3 等分することで離散化をしたものを用いる．

4. 判別手法について

本章では，本研究で用いた意図的動作と偶発的動作を判別する手法について述べる．

学習時に学習パターンがその所属クラス名 (本研究の場合は意図的動作か偶発的動作かの 2 つのクラスになる) とともに与えられる学習法を教師付き学習 (supervised learning) という．本研究では教師付き学習としては決定木学習を用いる．

一方クラスのラベルが付いていないパターンを用いて行う学習法を教師なし学習 (unsupervised learning) という．本研究では教師なし学習としては K-means アルゴリズムを用いる．

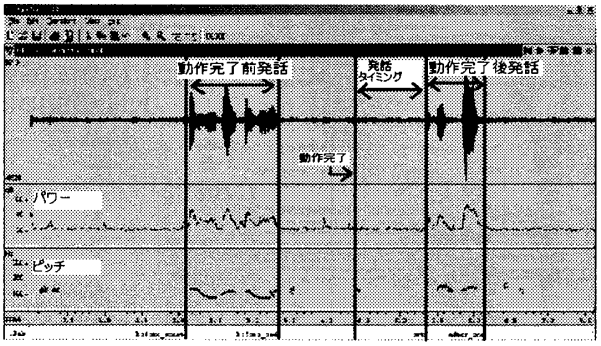

図 2: 被験者の発話サンプル

4.1 決定木学習

決定木 (decision tree) とは，入力データがどのクラスに属するかを決定する分類方法の 1 つである．入力データは属性と属性値の組の集合からなっている．入力データを決定木のルートノードから入力すると，各ノードでは入力データの属性値のテストを行い，その結果によって決定木をたどって，末端のノードでクラスが決定する．決定木の学習とは，訓練用の入力データを用いて，決定木を自動的に生成する帰納推論の 1 手法である．

本研究では，離散値にした特徴量を使って判別を行う場合は情報利得により決定木を作成する ID3 [4] アルゴリズムを用い，連続値の特徴量を使って判別を行う場合は，情報利得から得られる情報利得比によって連続値を区切る閾値を決定した上で決定木を作成する C4.5 アルゴリズム [5] を用いる．

4.2 K-means アルゴリズム

K-means アルゴリズム [6] とは，データから互いに似ているものを集めていくつかの集団 (クラスタ) としてグループ分けを行うクラスタリングの一手法である．始めにランダムに初期値を選び，それらとの距離を元にクラスタ分けと重心計算を繰り返してクラスタリングを行う．その手順は以下の通りである．

- ランダムにパターンを k 個選び，クラスタのセントロイド (重心) とする．
- 各パターンを最も近いセントロイドのクラスタへと併合する．
- 併合したパターンを用いて各クラスタのセントロイドを再計算する．
- 以上の手順を，パターンの別のクラスタへの移動や二乗誤差の最小値の減少が起こらなくなるまで繰り返す．

ここでいう二乗誤差とは，それぞれのパターンとそのパターンの含まれるクラスタのセントロイドとの距離の二乗の総和のことである．

本研究では，所属クラス名を除いた特徴量を用いてK-meansアルゴリズムを行い，2つのクラスタに分割し，片方のクラスタを意図的動作のクラスタと仮定した場合の正解率と，その逆の仮定をした場合の正解率を求め，高いほうの正解率をそのクラスタリングによる正解率として評価した．また最初の2つのセントロイドの決定には，数個ずつのパターンをランダムに取ってきた上でそれぞれの重心を求め，最初のセントロイドにするという方法を取った．

5. 判別の結果と考察

実験により得られた動画ファイルから発話を伴った意図的動作と偶発的動作を抽出した結果，表1に示す数のデータが得られた．これらのデータの様々な種類の特徴量を元に，データマイニングツールであるweka[7]を用いて意図的動作と偶発的動作の判別を行った．

その結果を表2，表3に示す．それぞれの表では，被験者ごとの正解率，正解率の平均をまとめ，さらに個人差の影響を調べるために被験者全員のデータを合わせて判別をした場合の正解率を“全体”として，一番下の行に記した．

動作完了前後のパワー・ピッチ及び発話タイミングを特徴量として用いた場合の判別結果は表2のようになった．ID3アルゴリズムによる正解率の平均が50.1％となり，データ内の意図的動作の占める割合と比較しても9％も低いが，連続値を用いたC4.5及びK-meansアルゴリズムによる正解率の平均がそれぞれ62.3％，66.2％と高いことから，本研究での特徴量の離散化の方法が適当でなかったためID3アルゴリズムでの判別の精度が低くなったと考えられる．連続値を用いた正解率はデータ内の意図的動作の占める割合を上回っており，これらの特徴量が判別に有効であると考えられる．正解率の高い決定木に注目すると，いずれも最上位のノードに発話タイミングが置かれていることからこの中でも特に発話タイミングが判別に有効な特徴量と考えられる．また，被験者により正解率に開きがあり，判別の精度は被験者による影響が大きいといえる．

表3は動作完了前後の発話のパワーの代わりにそれらのパワーの差分を取ったものを特徴量として用いた場合の判別結果である．それぞれ表2と比べて正解率が下がっているが，特に正解率の個人差がより大きくなっている．正解率の低い被験者のデータに注目すると，パワーの差分の値が非常に小さく，取る値の範囲も狭くなっていた．動作完了前発話が主にシステムに呼びかける発話であり，被験者によってその呼びかけ方は大きく異なるので，正解率の個人差が大きくなったと考えられる．

これらの判別においてK-meansアルゴリズムによる判別の正解率はどれもC4.5アルゴリズムと同程度の結果が得られており，教師なし学習による意図的動作と偶発的動作の判別は教師付き学習による判別と同程度に有効であるといえる．そして被験者全員のデータを合わせて判別をした場合の正解率はどれも低く，個人差が大きいため，正解率を上げるためには個人適応する必要性があることがわかる．

表1: ボール遊びゲームにより得られた各被験者のデータ数及び意図的動作Nと偶発的動作Pの数

被験者	意図的・偶発的動作の数		データ数
	N	P	
A	32	19	51
B	22	16	38
C	25	33	58
D	22	42	64
E	23	32	55
F	10	10	20
G	17	25	42
H	7	15	22
I	12	11	23
J	14	12	26

6. おわりに

本研究では，意図的動作と偶発的動作を判別するための特徴量として，動作完了前後の発話のパワーとピッチ，発話タイミングを用いる際に，その判別手法に教師なし学習を提案し，その有効性を検証した．

今後の課題として，以下に示す点について検討していく必要がある．

- 本研究で利用したボール遊びゲームでは偶発的動作時のダメージは少なくミスからの復帰も容易であるため，それほど特徴的な発話は得られていないので，その点を改善した実験方法の検討が必要
- 音声以外にも，表情や視線，身振りなどの特徴量も抽出することで，意図的かどうかを判別できる精度がさらに上がるものと考えられるため，発話以外の特徴量の検討
- 偶発的動作の中でさらに特徴的なパターンというのも存在する可能性があるので，クラスタリングのクラスタ数を3つ以上にするなどして検討する．またその際には，特に発話が無い場合の特徴量の値の適切なとり方についても検討
- 現段階では判別の汎用性は低いので，各個人に適応する手法の検討

参考文献

[1] Michael Tomasello(訳者：小林春美)：“The Pragmatics of Word Learning”，認知科学，4(1)，pp. 59–74，(1997).

[2] Meltzoff, A. N.: “Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children”, Developmental Psychology 31, pp. 838–850, (1995).

[3] 櫻井 晴章, 岡 夏樹: “随伴する発話の韻律情報に基づく動作意図の理解”，情報科学技術レターズ，pp. 107–109, (2004).

[4] J. Ross Quinlan: “ Induction of Decision Trees ”, Machine Learning 1(1), pp. 81–106, (1986).

[5] J. Ross Quinlan: “ C4.5: Programs for Machine Learning ”, Morgan Kaufmann, San Mateo, CA, (1993).

[6] Jain, A. K. and Dubes, R. C.: “ Algorithms for Clustering Data, Prentice Hall ”, (1988).

[7] http://www.cs.waikato.ac.nz/ml/

表 2: 動作完了前後のパワー・ピッチ及び発話タイミングを用いた意図的動作と偶発的動作の判別結果

	ID3 アルゴリズム	C4.5 アルゴリズム	K-means アルゴリズム
被験者	正解率 (%)	正解率 (%)	正解率 (%)
A	72.5	74.5	76.5
B	50.0	68.4	57.9
C	39.7	58.6	58.6
D	43.8	53.1	60.9
E	63.6	72.7	69.1
F	35.0	40.0	60.0
G	54.8	73.8	66.7
H	36.4	68.2	72.7
I	47.8	52.2	65.2
J	57.7	61.5	73.1
平均	50.1	62.3	66.2
全体	46.9	60.1	61.0

表 3: 動作完了前後のパワーの差分・ピッチ及び発話タイミングを用いた場合の判別結果

	C4.5 アルゴリズム	K-means アルゴリズム
被験者	正解率 (%)	正解率 (%)
A	70.6	72.5
B	63.2	58.0
C	55.2	54.5
D	65.6	64.1
E	81.8	63.6
F	70.0	64.7
G	66.7	59.5
H	68.2	59.1
I	34.8	52.3
J	73.1	61.6
平均	64.3	61.0
全体	59.4	53.4