J-049

ロボットから人に話しかける判断モデルの構築

Decision Model for a Robot to Start Communicating with a Human

牛山 聡† Satoshi Ushiyama　松井 和教† Kazunori Matsui　村上 真‡ Makoto Murakami　白井 克彦† Katsuhiko Shirai

1. はじめに

近年，Web 検索システムに要求される情報は多様化している．例えば，楽曲のような膨大な集合の中から 1 つを選ぶ場合には，曲名や歌手名といった情報だけでなく，それを聞いた人の感想などを元に，そのときの気分に合う曲を探す人も多い．だが，一般的に Web はユーザが自ら加える他に情報が集積されることがないため，知識や経験，感情などのごく個人的で，全ての人が積極的に発信してくれるわけではない情報は検索に活用されにくいものになってしまっている．このような発信されにくい情報を得るためには，インタビューなどによって，収集する側が能動的に情報を引き出しに向かわなければならない．自動で情報を収集する方法としては，身体に装着したセンサにより周りの環境や人の動きといった情報を取得する，ウェアラブル・コンピュータ[1]，災害時向けに開発が進む自走ロボット[2]などがあるが，これらでは人間の頭の中にある情報を引き出すことは難しい．そこで，活躍が期待できるのがコミュニケーションロボットである．人間の生活空間で自由に動き回れるロボットが自らインタビューを実行していけば，効率良くこれらの情報を集めることができる．

コミュニケーションを図るロボットを用いた研究には，オフィス内で働いている職員について尋ねられた際，その居場所を回答するロボット[3]，会場でポスターセッションをガイドするロボット[4]など様々なコミュニケーションロボットがある．しかし，これら従来のロボットはタスク遂行型で，インタビューをするためには新たなシステムが必要になる．例えば，インタビューを開始する際には，タスク遂行型ではユーザが意図を持っているが，インタビューにおいてはユーザには意図がない．そのためインタビューの開始に当たって，人間の様子から話しかけるのに適切なタイミングを判断し，相手の進行中の作業を妨げないようにする必要がある．インタビューの継続中においてはタスク遂行型では正確で速やかなタスクの完遂が目的となるが，インタビューではより多くの情報を引き出すなどと目的が異なる．そのため，この目的のための対話戦略や有効な非言語情報を扱うシステムが求められる．インタビューを終了する場合にはタスク遂行型では明確な達成点があるが，インタビューの達成点はより多くの情報を引き出すなどと厳密でない．したがって，継続した利用に支障がないよう，人が不快を感じる前にロボット自らが対話を終了させる必要が生じる．このように，コミュニケーションの 3 つの場面で有効な判断ができることが望ましい．

本研究では話しかける場面に着目し，適切なタイミングでロボットから人間に話しかける判断モデルの構築を行う．今回我々は，オフィスで働いている人に専門分野の研究内容について聞きたいことがあるという設定で，インタビューをするために話しかける場面を想定する．人は話しかける際には相手を観察し，その動き等からその人が置かれている状況や感情を推し量っている．ロボットも同様に画像情報などを用いて話しかける必要がある．ロボットから人間に話しかけるモデル構築の先行研究として，状況判断モデル構築[5]が試みられているが，話しかけたい対象がマウスを握っている時間が長いかどうかで判断するだけに留まっていて，ロボットが人間の状況を適切に把握して，話しかけているとは言えない．そこで，本研究ではオフィスで働いている人の作業中に現れる動作に注目し，その動作や姿勢と話しかけられるかどうかの対応関係を明らかにする．その結果より，話しかけると判断できるデータとそれ以外のデータとを別々に GMM で学習させ，話しかける尤度と話しかけない尤度の優劣により話しかけて良いかどうかを判定する．

† 早稲田大学
‡ 東洋大学

2. 判断モデル用データの取得

話しかけ判断モデルを構築するために，作業を行う人にマーカを取り付け，様子を映像で記録し，動きや姿勢の 3 次元データを取得する．この作業を行う人間の映像を被験者に見せ，話しかけさせた結果により，話しかけても良い状況と判定されたデータを得る．

2.1 作業映像の収録

話しかけるかどうか判断してもらう対象となる人物の動作や姿勢の 3 次元位置計測を行うために，作業をする人物の体にマーカを装着し，作業映像を撮影する．マーカは図 1 のように頭部に 3 個，首・背中・腰・肩・肘・手首・手の甲にそれぞれ 1 個ずつ，合計 10 個配置し，作業者の右手側にステレオカメラを設置する．作業の内容は，オフィスでの仕事として代表的と考えられる，PC の画面を見てキーボードで入力，画面を見て書類への筆記，書類だけがあり書類に筆記，画面を眺めるだけの 4 種類とする．また，各々の作業中に現れる動作を，予備実験で確認された動きより，伸び，もたれ，肩回しの 3 種類とする．動作の程度によって，話しかけるかどうかの判断が異なる可能性があるので，各動作を速い・遅いや大・小で分け，さらに，画面を眺めるだけの作業には肘を机について手で顎を支える動作を加え，計 45 種類の映像(表 1)を収録する．

2.2 人体の三次元データの取得

入力デバイスとして，PointGrey 社の 2 眼ステレオカメラ Bumblebee を用いる．マーカの三次元位置情報は，最近傍識別器を用いた色ターゲット検出[6]が可能な 3D Position Measurement SDK を用いて取得する．ステレオカメラは床と水平になるように設置し，PC の画面に向かって座った被験者の右手側真横方向から撮影する．被験者が PC 画面に正対する方向が x 軸，床の法線ベクトルが y 軸，ステレオカメラからの奥行き方向が z 軸となる．

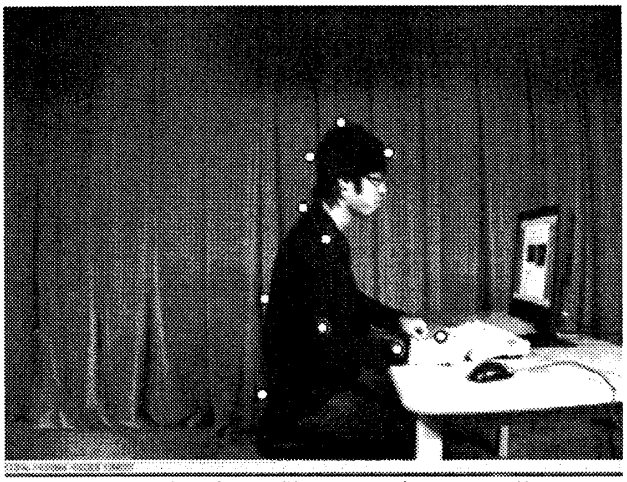

図 1. 3 次元座標取得のための各マーカの位置

2.3 話しかけ実験

被験者に作業映像を見せて，これに向かって話しかけさせる．作業映像は 1 シーン当たり 30 秒で構成されており，表 1 に示す 45 種類の映像をランダムに提示する．その中で，動作の起こる映像は「作業している途中に動作が起こり，また作業に戻る」という構成になっている．被験者は提示された映像を見て，映像中の人物に話しかけるかどうかを判断する．話しかけてよいと判断すれば，映像中の人物に話しかけてもらう．話しかけてよいと判断できない場合は何も話しかけずに 30 秒が過ぎ，次の映像が提示される．被験者が発話した音声はヘッドセットマイクにより録音する．なお，被験者は大学生の男子 4 名，女子 1 名の合計 5 名である．

2.4 話しかけ実験結果

5 人全員の結果から，話しかけた映像の中で，作業中ではなく伸び等の動作中に話しかけた確率は 98.5%であった．また，4 種類の作業内容の差での話しかける確率の違いはほとんど無かったため，動作により結果をまとめると，80%以上が話しかけた動作は伸びの速いものと遅いもの，もたれの動作が大きくで速いものと遅いもの，肩回しの動作が大きくて遅いものの 5 種類の動作であった(表 2)．これらの動作データを話しかける動作データ，それ以外を話しかけない動作データとし，話しかけ判断モデルの構築に用いる．

3. 話しかけ判断モデルの構築

話しかけても良い状況で現れる動作の特徴量を身体に取り付けたマーカの 3 次元データから抽出する．話しかけると対応付けたデータと，残りの話しかけないデータのそれぞれを GMM で学習させる．尤度の差による判断を行い，認識率を導く．

3.1 特徴量抽出

伸び，もたれ，肩回しの 3 種類の動作と速度を表現する，肩と肘と手首のなす角度・角速度，首と腰のマーカを通る直線と床平面との角度・角速度，上腕（肩から肘）の x 軸を回転軸とする回転角度・角速度，上腕の z 軸を回転軸とする回転角度・角速度，合計 8 次元の特徴量を抽出する．

表 1. 収録した作業映像

作業内容	作業中に現れる動作	動作の速さ	動作の大きさ
画面を見てキーボード入力	なし		
	伸びをする	遅い	
		速い	
	もたれる	遅い	小
			大
		速い	小
			大
	肩を回す	遅い	小
			大
		速い	小
			大
画面を見て紙に筆記	なし		
	伸びをする	遅い	
		速い	
	もたれる	遅い	小
			大
		速い	小
			大
	肩を回す	遅い	小
			大
		速い	小
			大
紙に筆記(画面無し)	なし		
	伸びをする	遅い	
		速い	
	もたれる	遅い	小
			大
		速い	小
			大
	肩を回す	遅い	小
			大
		速い	小
			大
画面を眺める	なし		
	肘をつき，手で顎を支える		
	伸びをする	遅い	
		速い	
	もたれる	遅い	小
			大
		速い	小
			大
	肩を回す	遅い	小
			大
		速い	小
			大

表 2. 各動作に対する話しかける確率

作業中に現れる動作	動作の大きさ	動作の速さ	話しかける確率(%)
なし			0
肘をつき，手で顎を支える			40
伸びをする	遅い		100
	速い		95
もたれる	遅い	小	40
		大	80
	速い	小	25
		大	95
肩を回す	遅い	小	40
		大	80
	速い	小	25
		大	60

3.2 GMM による学習

80%以上の人が話しかけると判断した動作データを話しかけるクラス，それ以外を話しかけないクラスとし，それぞれクラスを元に GMM で学習を行う．

3.3 GMM による認識実験

話しかける動作のデータ数は 12,900 フレーム，話しかけない動作のデータ数は 14,070 フレームである．それぞれのデータを均等に 10 分割し，9 割を学習データ，残りの 1 割を評価データとして，10 通りのデータセットについて GMM による認識率を算出した． GMM の混合数を 1，2，4，8，16 と変化させ，それぞれ 10 回の平均値を認識率とした(図 2).

この結果によると，話しかける動作の認識率と話しかけない動作の認識率は GMM の混合数が 4 のとき，その和が最も高くなる．このときに話しかける動作の認識率は 79.3%，話しかけない動作の認識率は 73.2%であった．

4. まとめ

情報検索のニーズの高まりとともに，人が自ら進んで発信しないような情報を収集して再利用したいという要求も高まっている．このような情報を取得するには，人間の生活空間を自由に動き回ることが可能なコミュニケーションロボットが適している．本研究では，オフィスで働いている人に専門分野の研究内容について聞きたいことがあるという設定で，それらの情報を引き出すために話しかける場面を想定した．

人はオフィスで作業している人に話しかける場合，作業している人が伸び等の何らかの動作をしたときに話しかけると仮定し，人が作業中に起こす動作に注目し，その動作や姿勢と話しかけられるかどうかの関係を実験により明らかにした．その結果として得られた話しかける動作データに基づく GMM，それ以外の話しかけない動作データに基づく GMM から，その尤度の差によって，オフィスで働いている人に話しかけて良いのかどうかを判断可能な話しかけ判断モデルの構築を行った．GMM の混合数が 4 のとき，話しかける動作の認識率は 79.3%，話しかけない動作の認識率 73.2%となり，その和が最大になった．

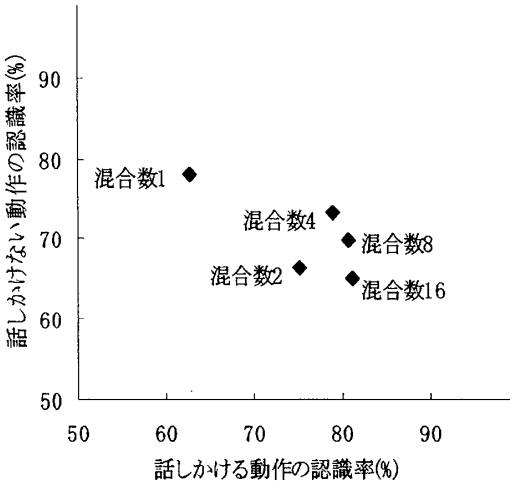

図2. GMMによる認識率の結果

本稿では，作業中に現れる動作として，伸び，もたれ，肩回しを扱ったが，他の様々な動作に対しても話しかけるかどうかの判断の要因になると考える．今後は作業中の動作の種類を増やし，それらを表現できる動作・姿勢の特徴量を抽出し，GMM を作成することによって，よりロバストな話しかけ判断モデルの構築を行う．

謝辞

本研究の一部は，早稲田大学理工学研究所の研究課題「自発的コミュニュケーション機構を有するマルチモーダルヒューマンインタフェースの研究」，平成 19 年度科学研究費基盤研究（B）課題番号 17300066「対話状況に応じた自発的コミュニケーション機構の研究」によるものである．

参考文献

[1] 大内一成，大盛善啓，松下宗一郎，土井美和子，"情報獲得機能を持つウェアラブル・オーサリングシステム"，インタラクション 2000，pp45-46，2000.

[2] 田畑克彦，稲葉昭夫，今井智彦，天野久徳，鈴木隆司，光井輝彰，"情報収集(被災者捜索)ロボットの移動機構に関する研究(第 4 報)"，岐阜県生産情報技術研究所研究報告 第 7 号.

[3] 松井俊浩，麻生英樹，J. Fry，浅野太，本村陽一，原功，栗田多喜男，速水悟，山崎信行，"オフィス移動ロボット Jijo-2 の音声対話システム"，日本ロボット学会誌，Vol. 18，No. 2，pp300-307，2000.

[4] 小出義和，神田崇行，角康之，小暮潔，石黒浩，間瀬健二，西田豊明，"強調的センサ群を用いたコミュニケーションロボット開発"，情報処理学会研究報告，Vol. 2004，No. 90，2004-HI-110，pp. 39-46，2004.

[5] 奥石欣吾，上野敦志，木戸出正継，"自発的にコミュニケーションを図るロボットのための判断モデル構築の試み"，第 16 回人工知能学会全国大会，May，2002

[6] 和田俊和，"最近傍識別器を用いた色ターゲット検出-「らしさ」に基づかない識別とコンピュータビジョンへの応用-"，情報研報 CVIM134-3，No.134，pp.17-24，2002