

能動的情報収集を目的とした対話終了判断に有効な特徴量の分析

Effective feature analysis to construct decision model

for a system to end communicating with a human

大塚 尚樹[†] 村上 真[†] 山際 基[†] 上原 稔[†]
 Naoki OTSUKA Makoto MURAKAMI Motoi YAMAGIWA Minoru UEHARA

1. はじめに

近年では情報空間の情報を再利用する研究が活発になってきている[1]。再利用する研究の例として情報検索が挙げられる。検索対象となる情報には、Webに発信されている情報、実空間のいたるところに設置されたセンサによって収集される情報、自律的に移動するロボットに搭載されたセンサによって収集される情報、人間の体に付けられた複数のセンサからデータを得るウェアラブルコンピューティングによって得られる情報などがある。一方で私たちは、感想情報が再利用する価値があると考え、ある人のプロフィール、状況が得られていて、音楽、映像作品、著作物などを聴いたり、観たり、読んだりしたときの感想がわかるとする。この情報を多く収集するとどんな人がどのような時にどのような曲などを聴くとどのような気分になるのかがわかる。そうすると同じようなプロフィール、状況下の他の人に適切な曲などを提供できる可能性があるからである。

しかし、感想情報はウェブに発信されることはあるが全員が発信するわけではない。また環境中やロボットに搭載された受動的なセンサによる情報収集では、感想情報を持った人が誰かとコミュニケーションをしていなければその情報を収集することができない。ウェアラブルコンピューティングでは、対話を必要とせず生態情報から感想情報を得られる可能性があるが、生態情報にノイズがのりやすいことや、環境の変化によっても生体情報は変化するため感情の変化によって得られた生体情報なのかの判断が難しいこと、収集した情報がどのような意味をもつのが明確でないことといった欠点がある。本研究ではロボットが能動的にユーザと対話し、感想情報を言語情報として収集することができる能動的情報収集対話システムを提案する。

従来のタスク遂行型対話システム[2]と情報収集型対話システムを比較する。まず、タスク遂行型対話では、ユーザにシステムを使用する意図があり、ユーザから対話を開始するが、情報収集型対話ではユーザにシステムを使用する意図がないため、システムが対話開始の判断をしなければならない。また、タスク遂行型対話では早く正確にタスクを達成することを目的としているが、情報収集型対話では多くの情報を得ることを目的としているため、そのような対話制御を行う必要がある。さらに、タスク遂行型対話ではタスクが達成された時に対話が終了されるのに対し、情報収集型対話では必要な情報が得られた時に対話が終了される。しかし、ユーザにはシステムを使用する意図がないため、その前に対話を終了したいと思う可能性がある。その場合は、システムが適切に判断して対話を終了しなくて

はならない。本研究では終了段階の問題を扱う。

2. 対話終了判断モデルの構築

本研究で提案するシステムは、ユーザが対話を終了したいと思った時、表情や動作、発話状態から状況を判断して終了し、そうでない時は対話を継続する。本システムを実現するためには、視覚的な情報である表情、姿勢、動作や聴覚的な情報である発話内容、韻律と終了判断との対応関係を明らかにし、対話終了判断モデルを構築する必要がある。そこで、実際に対話を終了したくなるような工夫をしたインタビュー実験を行う。その様子をカメラやマイクで収録し、そこから得た表情や動作、音声のデータを分析する。本研究では、被験者のインタビューに答えたいという状況の動作・音声とインタビューに答えたくないという状況の動作・音声をそれぞれ比較して分析し、対話終了判断モデルを構築することを目指す。

3. 実験

被験者実験によりインタビューに答えたい状況の動作・音声データと答えたくない状況の動作・音声データを収録し、動作データから対話終了判断が可能なモデルを構築する。また、音声データから終了判断に有効な韻律情報の分析を行う。

3.1 動作及び音声データの取得

被験者がインタビューに答えたくない状況を作り出す為に、長時間のインタビューを行った。提案するシステムに近い状況にする為に、パソコン1台、スピーカ1台を被験者の部屋に用意し、ボイスチャットソフト(Skype)を使い、質問者はネットワークを介して別室からインタビューを行った。さらに、被験者の部屋のパソコンにはロボットの動画を流し、質問者の音声はボイスチェンジャーソフト(MorphVox)によって変換してインタビューを行った。質問事項はごく普通のものを選択した。被験者の動作を3時間座標として得るために、頭、両肩、両肘、両手首、背中の8箇所カラーマーカを付け、8方向から8台のカメラで撮影した。また、ヘッドセットマイクにより被験者の音声を録音した。音声データ取得実験の様子を図1に示す。実験終了後に被験者にアンケートを行った。アンケートの内容は、時間が長く対話を終了したいと感じたか、どこで長いと感じたかの2つである。

動作は1人分、音声では3人分のデータを取得した。アンケート結果より、動作の被験者は30分、音声の被験者はそれぞれ20分、10分、10分経過したところで、時間が長いと感じ対話を終了したいと感じたという結果が得られた。よって、その部分で対話継続可能な状態のデータと対話継続不可能な状態のデータに分け、モデル構築と韻律の

[†]東洋大学 Toyo University

分析に使用した。



1図 実験の様子

3.2 動作データによる対話終了判断モデルの構築

対話継続可能状態の映像データと対話継続不可能状態の映像データから答える気がある状態の基本姿勢 3 種類 (300 フレーム) を対話継続可能状態、答えたくない状態の動作 (伸びをする、姿勢を後ろにずらす、顔の前で手を組む、頬杖をつく) 4 種類 (310 フレーム) を抽出した。次に色情報からカラーマーカを抽出し、視体積交差法によりマーカの 3 次元座標データを計算し、3 次元動作データを得る。得られた答える気がある状態の 3 次元動作データと答えたくない状態の 3 次元動作データから GMM により終了判断モデルを構築する。動作データの次元数は 24 次元 (マーカ 8 個×3 次元)、GMM の混合数は継続可能状態は 3、継続不可能状態は 4 とした。最後に、分割交差検定 (10 分割) によりモデルの評価を行った。表 1 にモデルの正解率を示す。

表 1. モデルの正解率

テストデータ	データ数	正解数	正解率 (%)
継続可能	300	291	97.0
継続不可能	310	296	95.5
計	610	587	96.3

継続可能状態のデータに対し 97.0%、継続不可能状態に対し 95.5%、平均して 96.3%の高い正解率を得ることができた。

3.3 終了判断に有効な韻律情報の分析

対話継続可能状態と対話継続不可能状態では声の大きさ、応答時間、話す速度に変化が現れると考え、声の大きさであるパワー、質問者が質問をしてから回答者が返答するまでの時間の長さであるポーズ長、単位時間あたりのモーラ数である発話速度について分析を行った。

対話継続可能状態と対話継続不可能状態のパワー平均値を図 2 に、ポーズ長平均値を図 3 に、発話速度平均値を図 4 にそれぞれ示す。

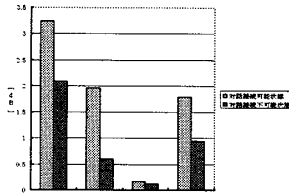


図 2 対話継続可能状態と対話継続不可能状態のパワー平均値

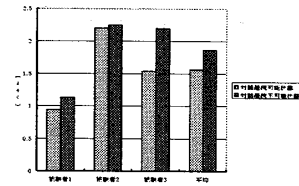


図 3 対話継続可能状態と対話継続不可能状態のポーズ長平均値

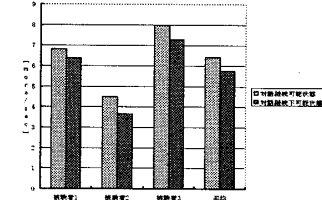


図 4 対話継続可能状態と対話継続不可能状態の発話速度平均値

実験より、対話継続不可能な状態では対話継続可能な状態に比べて、声の大きさは小さくなり、ポーズ長は長くなり、発話速度は遅くなるということが明らかになった。また、棄却域 5 パーセントで有意差検定を行った結果、全てにおいて有意差ありとなった。

4. 結論

本稿では、動作や音声から対話を終了すべきかどうかを判断できるモデルの構築を目的とし、被験者実験により対話継続可能状態と不可能状態の動作・音声データを取得し、動作データから対話終了判断が可能なモデルを構築した。その結果、平均 96.3%の正解率が得られた。また、音声データの韻律情報を分析した結果、対話継続不可能な状態では対話継続可能な状態にくらべて声の大きさは小さくなり、応答時間は長くなり、話す速度は遅くなるということが明らかになったが、その他の韻律情報にも注目し分析を行う必要がある。今後は、対話終了判断に有効な韻律情報を終了判断モデルに組み込み、動作と韻律情報から対話の終了を判断可能なモデルを構築する予定である。

参考文献

- [1] A. Singhal, "Modern information retrieval: a brief overview," *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*, vol. 24, no. 4, pp. 35- 43, 2001.
- [2] R. Cole, L. Hirschman, and L. Atlas, et al., "The challenge of spoken language systems: research directions for the nineties," *IEEE Trans. on Speech and Audio Processing*, vol. 3, no. 1, pp. 1- 21, 1995.