

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行われている[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ういう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体情報から感想情報を得られる可能性があるが、生体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か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。

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。

今回は、対話終了判断モデルを構築する上でまず、人がどの特徴によって対話終了を判断しているのかを明らかにしようと考えた。つまり、心的情報がどのモダリティに現れるのかを明らかにする。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。

2. 実験

2.1 実験目的

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・音声データを収録する。得られたデータの特徴があると思われる箇所を第三者に評価してもらおう。そして、人がどのような特長によって対話の終了を判断しているのかを明らかにする。

2.2 動作及び音声データの取得

インタビューに答えたい状況と答えたくない状況を作り出す為に、質問内容に興味の有無による制御をかけたインタビューを行った。質問の内容は、事前にアンケートを取り、興味があること、興味がないことが何かを取得し、質問を作成した。質問は興味があるものとないものを交互に質問した。

本システムに近い状況にする為に、質問者はパソコン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、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フラッシュ動画を流し、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。その様子をカメラで撮影し、被験者にヘッドセットマイクを取り付けて音声を録音した。撮影した時の様子を図1に示す。インタビューに答えたい状況と答えたくない状況を作り出せたかを判断するために、実験終了後、被験者にアンケートを行った。アンケートの内容は、対話を終了したくなったか、どこで終了したいと感じたか、の二つである。尚、被験者は大学生1名で実験時間は約30分である。

[†] 東洋大学 Toyo University



図 1:撮影した様子

2.3 動作及び音声データ取得実験結果

アンケートを行ったところ、被験者は対話の終了をしたくなかったと答えた。よって、このデータがインタビューに答えたい状況と答えたくない状況を作り出せたと判断した。このことより、このデータが対話終了判断の特徴分析に有効なものであると考えた。また、どこで対話の終了をしたいと感じたかという質問では時間で表すと、約6分から10分、約20分から25分と答えた。

2.4 終了特徴量の分析

得られたデータから特徴があると思われる箇所を22箇所選別した。選んだ箇所に対して質問と回答の1ターンの対話内容を書き起こしたテキスト、同じく1ターンの音声、1ターンの映像をそれぞれ作成した。

この22箇所のデータに対し、まず、評価者にすべてのテキストを提示し、最も対話する気があると感じた場合は1、最も対話する気がないと感じた場合は5の5段階で回答してもらう。さらに、なぜその評価をしたのかを具体的に答えてもらう。次に音声、映像の順で提示し同じように評価してもらう。

評価者には、この22箇所のデータを無作為な順番で評価してもらう。つまり、評価者によって22箇所のデータの順序が異なる。こうした理由は、すべての評価者が同じ順番で評価した場合、前のデータとの相対的な評価になってしまう恐れがあるためである。評価者は大学生3名である。

2.5 終了特徴量の分析結果

評価者の3名が被験者のデータに対して、対話する気があると感じた理由と対話する気がないと感じた理由をまとめた。第三者のテキストに対しての対話の継続と終了判断理由、第三者の音声に対しての対話の継続と終了判断理由、第三者の動作に対しての対話の継続と終了判断理由を、表1、表2、表3にそれぞれ示す。

表 1:第三者のテキストに対しての対話の継続と終了判断理由

対話する気ありと思った理由	対話する気なしと思った理由
具体的に述べているため	最低限の答えのみの回答だったため “なんとも思わない”という回答だったため “わからない”という回答のみのため きっぱりと否定しているため

表 2:第三者の音声に対しての対話の継続と終了判断理由

対話する気ありと思った理由	対話する気なしと思った理由
返答までに間があり考えていると感じたため	笑っていたため
声のトーンが高いため	声に抑揚がないため
語尾が伸びているため	文末に若干声が伸びているため
言い直していて伝えたいと感じたため	声のトーンが低いため 棒読みであるため だるそうであるため 即答していたため 早口であるため

表 3:第三者の動作に対しての対話の継続と終了判断理由

対話する気ありと思った理由	対話する気なしと思った理由
視線が安定していたため	目線がそれているため
質問を口ずさんでいたため	回答後に大きな動きをしていたため
視線を泳がせながらも考えていた様子だったため	質問を聞いた姿勢が崩れたため
動作しながら答えていたため	腕をいじって集中していないため
表情が明るかったため	頭がふらふらしていたため

3. 考察と結論

今回の実験より人が対話終了判断を行う主な特徴が明らかになった。テキストでは、対話する気なしと思った理由の多くが、最小限の答えのみの回答や、否定的な回答などであった。音声では、対話する気なしと思った理由の多くが、声の高さ、話す速さや語尾などに関する回答であった。動作では、対話する気なしと思った理由の多くが、目線がそれているなどの視線に関する回答や、頭がふらふらしているなどの姿勢に関する回答であった。

今後は、音声については、韻律情報などの分析、動作については、座標データを取り分析する必要がある。また、対話終了判断に有効な韻律情報と動作情報を終了判断モデルに組み込み、韻律情報と動作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.