

機械学習を用いたユーザインタフェースにおけるユーザ意図の分析

昆野 佑樹 小林 彰人 木村 悠貴 太田 晶 高地 昭彦
日本工学院八王子専門学校

1. はじめに

近年、機械学習は我々の生活に、深く浸透しはじめ、様々な所で使われている。チャットボット（すべてではない）などを初めとした、対話型インターフェースや、携帯などで使われる多要素認証（Apple の FaceID など）に、機械学習は実装されている。しかし、これらの機械学習の技術を、ユーザにとって直感的で、使いやすいインターフェースとするまでには、多くの課題が存在する。そこで本研究では、上記のような複数の機械学習を使ったユーザインタフェース（以下 UI）構築の前段階として、ユーザが咄嗟に出す、言語情報や、非言語情報であるジェスチャーが、どのような場面で使われ、また人とコミュニケーションを取る際との違いについて、調査を行った。

2. マルチモーダルインタラクションの必要性について

マルチモーダル性についての研究は、2000 年代に入る以前から考察されている。[1-3]の中でも語られているが、画像認識や音声認識など、こういった人間の五感にあたる感覚の一部に認識が偏る場合、その場の空気や状況などを、あらゆる情報から推測、処理する事が困難になり、円滑なコミュニケーションを取る事ができなくなる。

また、これらの対話型インターフェースの難点として、複数人の会話に弱い点があげられる。特にこういった、雑音が多い場面での誤認識の可能性や雑音が多い時に反応しないなどの、音声だけでは認識できない状況が存在する。

音声のみでの認識は、とても優秀だが、こういった点で、マルチモーダル性は、誤認識を減らすために必要になる。このような場面での、視線推定や顔の向きで、誰に対して何を話しているかの認識ができる。[4]

このような問題に対応するために、音声に加えて画像認識も活用したマルチモーダル性は、円滑なコミュニケーションを取るために必要な要素である。[4]の研究では、興味のある対象に対し、人は顔や体の向きといった非言語情報から、会話に対する関心を推定できることが報告されている。この研究では、70.5%の確率で関心がある話題に対して、上記のような身体動作をする事がわかった。

また、スマートスピーカーに注視のインタラクションを加える事で、誰に対してコミュニケーションを取りたいかについての研究は、[5]で行われている。

3. 提案手法

3.1 研究手法

本研究では、人と実際にコミュニケーションを取る際、どのような言動が現れるのか、また同じ場面で、スマートスピーカーやジェスチャー認識カメラとの差に、どのような情報が現れるのかを、日本工学院八王子専門学校 IT カレッジ / AI システム科 1 年（57 名）から、アンケート形式で調査を行った。

また、この研究結果は、現段階では、途中なため、研究によって得られたデータをワードクラウド化して、提示する。

3.2 ワードクラウドとその結果について

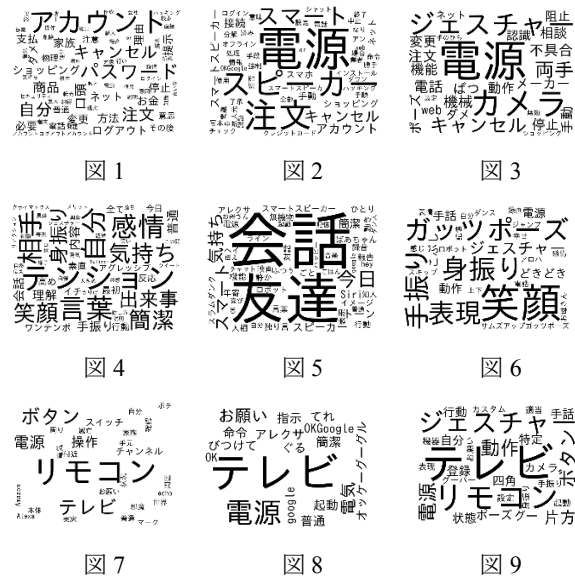


図1~3では、「あなたのアカウントを使って、対象（人、スマートスピーカー、ジェスチャー認識カメラ）がネットショッピングを勝手に行った」という想定で、アンケートを取った結果である。このワードクラウドから得られた情報として、図2,3からは、「電源」といった言葉が出てきた。

このワードは、「電源を切る」「電源を抜く」といったワードによく使われ、人以外とのコミュニケーションを、こうい

った突発的な状況では、取らないといった選択をする物が多かったと本研究では、考えている。

図4-6では、「嬉しかった事があった時に、どのようにして今の心情を伝えるか」という想定で、アンケートを取った。

図4-6のワードをみると、喜びの気持ちを伝えようといったワードは、図6が一番多く見られた。

図7-9では、「テレビをつけてください」という想定で、アンケートをとった。結果として、全てのワードクラウドの中で、最もワード数が少ない事がわかった。ここから考えられる内容は、テレビを付けるといった動作では、どのようなUIをしようとしても、ジェスチャー、音声含め差が出にくいという事が分かった。

4. おわりに

今回の研究では、ワードクラウドで、現在本研究がどのような事に取り組んでいるかについて、提示した。

今後の課題としては、このデータを元に、実際に行われたジェスチャーとその場面についての考察をしていこうと考えている。

参考文献

- [1]“汎用 AI 実現のための鍵となる自律性と マルチモーダル性についての考察” The 31st Annual Conference of the Japanese Society for Artificial Intelligence, 2017 人工知能先端研究センター
- [2]“音声・画像・画像をもつインタラクションシステム”
電子技術総合研究所
- [3]“GUIからマルチモーダル UI(MUI)に向けて”(株)東芝マルチメディア技術研究所 新田恒雄
- [4]“身体動作に基づいた対話時の話題に対する聞き手の関心の推定” 今井研究室 1231149 山倉 和樹
- [5]“スマートスピーカーにおける注視の入出力を用いたインタラクションの効果” 「いい加減な」インターフェース2 川口 一画