

複数ロボットによる大喜利システムに関する検討

小林明美^{†1} 伊勢崎隆司^{†1} 望月崇由^{†1} 布引純史^{†1} 山田智広^{†1}

概要：本研究では、人とロボットとの関係性をより向上させるシステムとして、大喜利というシーンを設定し、複数ロボットによる掛け合いを行うロボット大喜利システムに着目する。本研究の目的は、ロボットが笑いを誘発する言葉や動作を与えるために、どのようにシステムが学習をしていくべきかを検討することである。今回、「R-env：連舞TM」を用いた複数ロボットによる大喜利システム「ロボット大喜利」の取り組みの詳細と、ユーザからの評価結果について述べる。表情やボタンといったセンサデータを学習データに利用できる可能性があることが示唆されたが、年代によって、行動に差があることが明らかになった。学習システムを作るためには、ユーザ属性に応じてセンサデータの重み付けを変更する必要があることが分かった。

キーワード：ロボット、大喜利、表情

Pilot study about Oogiri System, Japanese Comedy Show, with Some Robots

AKEMI KOBAYASHI^{†1} TAKASHI ISEZAKI^{†1}
TAKAYOSHI MOCHIZUKI^{†1} TAKASHI NUNOBIKI^{†1}
TOMOHIRO YAMADA^{†1}

1. はじめに

近年のロボット制御技術の向上により、ロボットを利用した様々なサービスが提供され始めている。受付業務にロボットが活用されている例や、福祉施設でロボットをリハビリテーションやセラピーで利用する例などがある。また、ロボットや各種デバイス、センサの高性能化、小型化、低コスト化により、色々なデバイスとの連携がより実現しやすくなり、提供可能なサービスの幅が広がることが考えられる。

ロボットを活用するサービスを進めるにあたり、ロボットやデバイスを連携させる難しさと、ロボット自体が受け入れられやすさの課題が挙げられる。

実際に現場で利用してもらうためには、簡単にカスタマイズ可能であることが望まれるが、従来様々なデバイスを連携させたサービスを開発しようとする際には、プロトコルや開発言語が異なるため、実装に時間がかかっていた。今回我々は、完全クラウド対応型デバイス連携制御技術「R-env：連舞TM」を用いて開発を行なった[1]。R-envは、NTTが開発した、デバイス間の違いを吸収した、Webブラウザを用いたブロックベースのビジュアルプログラミング環境であり、比較的簡単に実装可能であるため、ラピッドプロトタイピングツールとして利用可能である。

一方、複数のロボットが会話しているところを観察した後は、ロボットに対して会話の理解度が向上することや、自然に振る舞えるようになるなど、人とロボットとの関係

性



図 1 ロボット大喜利システム

Figure 1 Illustration of Oogiri system with some robots.

が向上することが知られている[2]。

我々は、笑いというシーンに着目し、人とロボットとの関係性をより向上させるシステムについて検討する。従来研究として、受動的メディアとしてのロボットという観点から、ロボット同士で行われる漫才を視聴すると、人間の漫才ビデオと遜色がないだけでなく、漫才の全体的な評価が高くなることが示されている[3]。我々は、複数ロボットによるお笑いの中でも、受動的メディアという観点から、大喜利というシーンを設定し、複数ロボットによる掛け合いを行うロボット大喜利システムについて検討する。

ロボットによる大喜利システムだけでなく、システムの求められる姿としては、毎回決まったシナリオを実施するのではなく、周りの状況を見て、適宜ロボットからの反応を変えていくことが課題となる。つまり、周りの状況を見て、学習していくことが重要となる。

本研究の目的は、ロボットが笑いを誘発する言葉や動作

^{†1} 日本電信電話(株) NTT サービスエボリューション研究所
NTT Service Evolution Lab.

を与えるために、どのようにシステムが学習をしていくべきかを検討することである。本研究原稿では、ロボット大喜利システムを考えた際に、そもそも学習のためにどういったデータが利用できそうか、実際に大喜利システムを色々な人に見てもらい事前検討を行った。

2. 従来研究

ロボットによる笑い感情を誘発させる研究として、与えられたお題に対してなぞかけを行ったり[4]、ニュースや物語を手掛かりとして少し言葉を変化させることにより笑いを誘発させようとする取り組みがある[5], [6]。これらの研究は、笑いが起こるメカニズムとして、違和感が生じることに着目し、違和感を生じさせるような言葉を選択するという手法が取られている。しかし、ユーザや状況に応じてルールを切り替える必要性や、手動でデータを更新していく必要がある。

3. 提案システム

以上より、ユーザから得られるフィードバックを計測し、それを評価データとして用いて学習データとモデル生成を行うことを検討する。これにより、システム利用者には負担が生じず、かつ利用者数に応じて属性ごとにデータを大量に観測することが可能となる。本研究論文では、センサデータに基づく学習データ生成に関する事前検討結果について説明する。

3.1 システム構成

今回のシステムは、図[2]の構成をとっている。ロボットとして、コミュニケーションロボットの Sota 6 体、PALRO を用いた。一番右の Sota 1 台は、ドコモの自然対話 PF (QA, 雑談, しりとり) と、RNN で機械学習したモデルを用いて、お題に対する回答を生成する。PC はそれぞれの Sota の音声合成や R-env サーバとして用い、画像表示や音声再生用のブラウザとしても利用した。また、ユーザからの評価を計測するために、面白い面白くないかを判定してもらう Raspberry Pi スイッチ (赤ボタン: 面白くない, 緑ボタン: 面白い) と、オムロン社の画像センシング技術「Human Vision Components」により表情推定可能な OkaoVision を利用した。

今回、アプリケーションを実装するために、R-env を用いた。R-env は、ロボットをはじめとする複数のデバイスを組み合わせたデバイス連携サービス作りを非エンジニア層にも可能にすることを目的に開発されており、高機能なロボットや IoT デバイスを「作る」のではなく、それらを複数組み合わせ「使う」ための技術である。特徴として、メーカーや業種問わず、多様なデバイスを連携・制御可能であることや、デバイス間の振り舞い制御を図[4]のような簡単なブラウザ上の GUI ツールで設計可能であることが挙げられる。

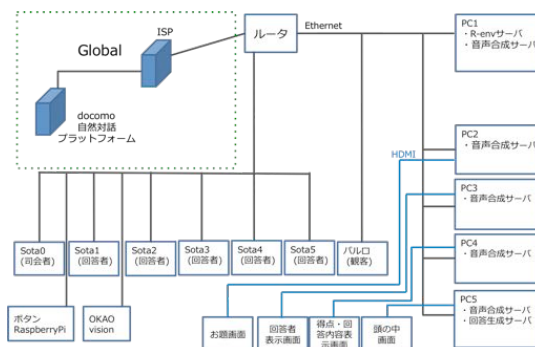


図 2 システム構成図

Figure 2 System Configuration.



図 3 大喜利シナリオでの役割

Figure 3 Role of each robots and devices.



図 4 R-env でのプログラム画面

Figure 4 Programming by "R-env".

3.2 シナリオ

今回の大喜利シナリオでは、Sota 1 台が司会、Sota5 台が回答者、Palro が観客ポジションとなり、10 分程度の大喜利形式を模したシナリオを演じる (図 3)。

シナリオは、導入部・メイン部・終了部からなる。導入部は各ロボットが自己紹介を行う。メイン部では、3 題お題が提示され、それぞれのお題に対して 3 体の Sota が回答を行う。お題出題から回答までの流れとしては、回答した Sota に対して、他の 1 体の Sota が掛け合いでツッコミを行い、司会者 Sota がコメントを行う。その後、司会者 Sota から観客に対して、面白い面白くないかをボタンにより審査するように促す指示がある。ボタンは前方に 5 セット設置され、どのボタンを押しても良い。審査は 10 秒間行われ、ボタンは何回でも押すことが可能であり、逐次得点が特典表示画面に扇子内の数字として表示される。審査後、

どちらのボタンが多く押されたかによって、回答 Sota が喜んだり悲しんだりする。回答もツッコミも行わない残り 3 体の Sota は、回答 Sota が回答するタイミングで回答 Sota の方を向き、ツッコミのタイミングで傾聴モーション（うなずきや顔を振る）を行い、審査タイミングで前に向き直る。お題間のつながりは考慮しない。終了部は、審査ボタンにより最多得点をとった Sota が優勝者となり、表彰が行われて、サヨナラをする。

4. 実験

実験では、展示会にて約 1000 人程度のお客さんに、ロボット大喜利システムを展示し、自由に観覧してもらった形式により行なった。

4.1 評価方法

アンケートによる評価は、系統的な評価が可能であるが、一つの事象全体としての評価となるため、その事象の中に、小さなイベントが連続的に起こり、そのイベントの逐次評価を行いたい場合には適切でない。特に今回はロボットが発話や動作を連続して行うため、途中で中断して評価することは、場の雰囲気を壊して正確な評価ができなくなる等の恐れがあり適切ではない。そのため、正確な評価を行い、その後の改善を行うためには、連続的な評価を取得することが重要となる。今回、連続的データとして、表情を利用した。表情はそれぞれ、笑顔・驚き・怒り・悲しみ・無表情の割合が、5つを足して 100 となるように算出される。その場の空気として、計測された表情の 5 秒間の平均と定義した

4.2 結果

図 4 は、展示全体を通しての笑顔の割合について、指数移動をとって示したものである。横軸は時間である。今回、インタビューなどより、ロボットによる大喜利はスケジュールであると回答が多かったため、大きな時間で区切った場合には、笑いのフェーズをとりきれていないのではないかと考えられる。

図 5 は、ある一回のシナリオを流した時の笑顔割合の推移を表したグラフである。横軸は時間で、それぞれ青・赤・紫の三角マークがお題提示、回答、審査時を示している。赤と青の点は、それぞれ面白いボタンとつまらないボタンが押された時間を示している。ここから、笑顔割合が上がってから面白いボタンが押される傾向があることが示唆される。

実際のフィールド実験や展示では、幅広い年代や性別など属性にばらつきが生じるため、属性が変わっても学習方法が同じ場合、適切な学習が行えない可能性がある。特に今回の展示では親子連れが多く訪れた。ボタンなどを設けて評価を行う場合、評価者が子どもの場合、ボタンを押す行為自体に楽しみを見出し、ボタンのカウント数と実際の評価が結びつかない場合が多々見られた。

4.3 考察

従来、子どもを対象とした実験の多くで言われているように、子どもは大人と違って自分自身の考えをきちんと整理できていなかったり、表現できていないので、子どもが直接評価判定を下すことは難しい。その代わりに、表情や視線などといった動作に関わる情報などを取得し、それらを評価値として利用する方法が知られている[7]。そのため、今回のようにボタンなどを設けて評価を行う場合、評価者が子どもの場合、ボタンのカウント数と実際の評価が結び

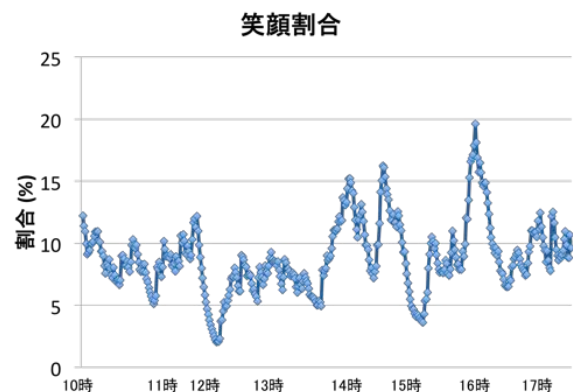


図 5 展示全体を通じての笑顔割合
Figure 5 ratio of smiling in a day

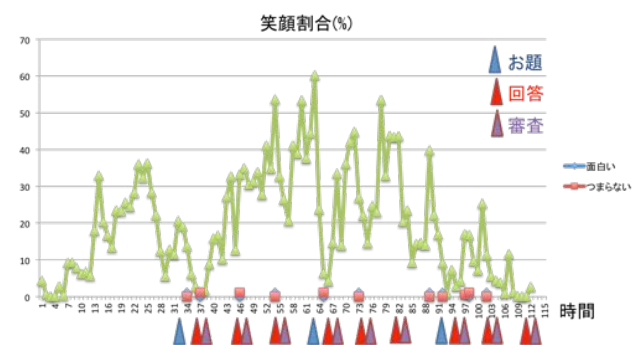


図 5 展示全体を通じての笑顔割合
Figure 5 ratio of smiling in a day

つかない一方、表情や笑い声、視線などは感情や着目度を直接表していることが多く、評価の信頼性が高くなる。逆に大人の場合、公共の場では自身の感情表現を抑える傾向にあるので、表情などのデータよりも、ボタンを押すという行為自体の評価の信頼性が高くなることが考えられる。

5. まとめ

我々は、複数ロボットによるお笑いの中でも、大喜利というシーンを設定し、複数ロボットによる掛け合いを行うロボット大喜利システムについて検討を行なった。本研究の目的は、ロボットが笑いを誘発する言葉や動作を与えるために、どのようにシステムが学習をしていくべきかを検討することである。実験により、表情やボタンといったセ

ンサデータを学習データに利用できる可能性があることが示唆されたが、年代によって、行動に差があることが明らかになった。そのため、学習システムを作るためには、ユーザ属性に応じてセンサデータの重み付けを変更する必要があることが分かった。

一方、年代が同じであっても個人間の違いもあるため、個人間で正規化した学習を行う必要があると考えられる。今後は、センサデータからお笑いの関係性の評価を行い、どういった学習アルゴリズムを利用すれば良いのかを検討したい。

参考文献

- [1] 松元崇裕, 松村成宗, 細淵貴司. インタラクシオンロボットサービスのためのクラウド型統合開発環境 (クラウドネットワークロボット). 電子情報通信学会技術研究報告. 2015, vol. 115, no. 283, p. 33-36.
- [2] 神田崇行, 石黒浩, 小野哲雄, 今井倫太, 中津良平. 人-ロボットの対話におけるロボット同士の対話観察の効果. 電子情報通信学会論文誌 D. 2002, vol. J85-DI, no. 7, p. 691-700.
- [3] 林宏太郎, 神田崇行, 宮下敬宏, 石黒浩, 萩田紀博. ロボット漫才—社会的受動メディアとしての二体のロボットの利用—. 日本ロボット学会誌. 2007, vol. 25, no. 3, p. 381-389.
- [4] 前田実香, 鬼沢武久. 単語の関連性とおもしろさを取り入れたなぞかけ生成. 感性工学研究論文集. 2005, vol. 5, no. 3, p. 17-22.
- [5] 吉田裕介, 萩原将文. 漫才形式の対話文自動生成システム. 日本感性工学会論文誌. 1012, vol. 11, no. 2, p. 265-272.
- [6] 真下遼, 灘本明代. 対立語抽出に基づく Web ニュースからの漫才ロボット台本自動生成手法の提案. 第 6 回データ構造と情報マネジメントに関するフォーラム(DEIM Forum 2014). 2014, vol. 8
- [7] 澤島康仁, 小峯一晃, 比留間伸行, 伊藤崇之, 渡辺誓司, 鈴木祐司, 原由美子, 一色伸夫. 番組視聴時の視線分布と番組内容理解度の関係. 映像情報メディア学会誌. 2008, vol. 62, no. 4, p. 587-594.