

相手の様子を理解して対話するヒューマノイドロボット

明石 航 内田 啓太 谷津 元樹 原田 実
 青山学院大学 理工学部 情報テクノロジー学科

1 背景・研究目的

近年、医療や介護などの現場において対話ロボットが注目されている。原田研究室においても、対話システム Hermes[1]を開発し、雑談や質問応答の相手となっている。この中で、ロボットが話し相手の様子を理解した結果に基づいて話しかけることができればより会話に現実味を与えることができると考えた。

本研究では、Pepper をプラットホームに用いて、話し相手の様子を理解し、自発的に話しかけるような、人間の行動に近い動作を行わせることを目指した。

2 提案手法の概要

本研究では、画像キャプション生成システム DeepWatcher[2]と動作系システム Athena2011[3]と対話系システム Hermes の 3 つのシステムを統合し、ヒューマノイドロボットが話し相手の様子を理解して話しかけるシステムを構築した。以降、このシステムを AthenaHermes と呼ぶ。3 章で画像キャプション生成について説明する。AthenaHermes の概要を図 1 に示す。なお AthenaHermes は、Pepper と無線 LAN で接続された PC 上で動作する。

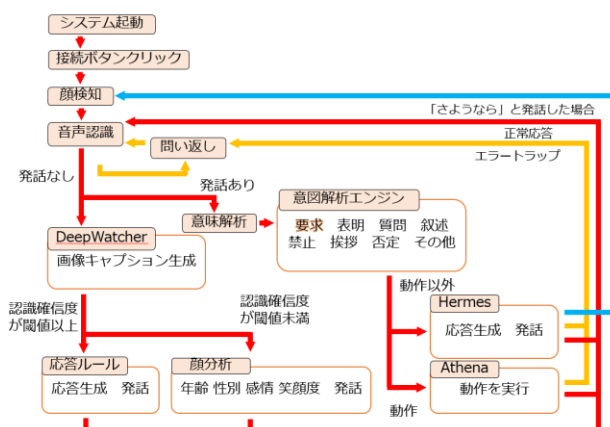


図 1. AthenaHermes における処理の流れ

3 キャプション-応答生成

画像内の状況について説明する文(キャプション)の生成からキャプションに対する応答文を生成し、ヒューマノイドロボットに発話させるまでを以下に説明する。

3.1 画像キャプション生成

Pepper の頭部で撮影した画像を、SCP ファイル転送プロトコルを用いて、コンピューター側に転送する。転送された画像は DeepWatcher に入力され、入力画像についてのキャプションとその認識確信度が出力される。

3.2 キャプション生成システム DeepWatcher

DeepWatcher は Convolutional Neural Network (CNN)によって抽出した画像特徴量および画像説明文をベクトルに変換したものから、Long Short Term Memory (LSTM)を用いて次の単語(形態素)の出現確率を計算し、最大確率の語を連結して画像のキャプションを生成する。DeepWatcher の概要を図 2 に示す。

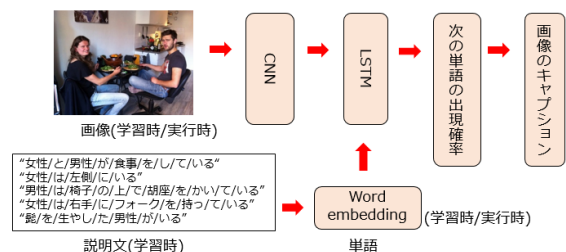


図 2. DeepWatcher の構成

3.3 キャプション-応答コーパス作成

DeepWatcher で実際に生成されたキャプションを約 1000 件集め、これらの各文に対応する応答文を手で作成し、キャプション-応答コーパスとする。

3.4 キャプション-応答コーパス検索

撮影した画像に対するキャプション G とキャプション-応答コーパスにおけるキャプション側の各文 C をグラフ照合し、意味的類似度が最大のキャプション-応答例 C-R を選択する。この流れを図 3 に示す。

Humanoid robot talking to a person based on the result of understanding its appearance

Akashi Wataru, Uchida Keita, Yatsu Motoki and Harada Minoru
 Department of Integrated Information Technology, College of Science and Engineering, Aoyama Gakuin University

生成されたキャプションG			
男性は青い服を着ている			
↑			
	キャプションC	類似度	応答R
1	男性は黒い服を着ている	75	その黒い服どこで買ったんですか？
⋮	⋮	⋮	⋮
627	男性は眼鏡をかけている	31	その眼鏡どこで買ったんですか？
⋮	⋮	⋮	⋮
1000	男性は目を瞑っている	31	眠たいんですか？

図3. キャプション-応答コーパス検索

3.5 応答編集

3.4 で選択されたコーパス内のキャプション-応答例 C-R のキャプション C はキャプション G と属性(色、数、服、乗り物、楽器など)が異なっていることが多い。属性を表す語 W が、G と C 中に類似対応して存在し、さらに C 中の語 W_C が応答 R 中にも含まれているとき、R 中の語 W_C を G 中の対応する語 W_G に置き換えて発話させる。この流れを図4に示す。

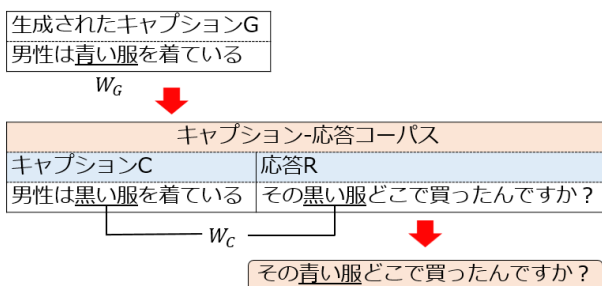


図4. 応答編集および応答発話の決定

3.6 対話事例

実際のキャプション生成に用いた画像と対話例を図5に示す。

システム: 携帯で何をしていますか?

ユーザー: ゲームを探しています

システム: ええ、ちょうどいい画面白そうゲーム探そと

ユーザー: モンスト楽しいですよ

システム: モンストらしい気配よく石配ってくれてガチャ引けるとやっける側も楽しいよね

ユーザー: モンストの運営は素晴らしいですよ

システム: モンスト運営有能だからほぼ全キャラが光る用のクエスト用意してもらえから好き

ユーザー: だからモンストは人気のかもしれないですね

システム: アプリゲームでゆっく課金したのは、白猫とモンストなんだけど、いい勝負で白猫の方が課金したかも

図5. キャプション生成用画像と対話事例

4 評価実験

Pepper の実機を用いて、Athena2011 と Hermes のみを統合したシステム(Athena2015)と、現在のシステム AthenaHermes との印象評価の比較実験を行った。Athena2015 は自発的にヒューマノイドロボットが話しかける動作が含まれていないので、話し相手の様子を理解した結果に基づいて話しかけることがどれだけ被験者がロボットに対し感じる印象に影響を与えるか評価する。

10人の男性理系学生を被験者とし、5分間の対話を行ってもらい(1.自発性 2.親しみやすさ 3.人間らしい会話 4.満足度)の4項目に関して7段階のリッカート尺度を用いて各システムを評価してもらった。Athena2015 と AthenaHermes の印象評価項目の平均値を表1に示す。

表1. 印象評価の比較実験結果

評価項目	Athena2015	AthenaHermes
自発性	2.80	5.00
親しみやすさ	3.90	5.50
人間らしい会話	3.20	4.80
満足度	3.60	5.50

全ての項目で高い評価が得られたが、特に「自発性」では従来システムと比べ、高い評価が得られた。これは、従来システムでは音声認識の際に何も話しかけないと「すみません、もう一度お願いします」などの問い返しがされるだけだったのに対し、提案手法ではキャプション生成や顔分析を行いロボットが相手の様子を理解し、話題を変えて発話するためであると考えられる。また、従来システムではシステム起動時に話しかけていたのに対し、提案手法では顔検知をして初めて話しかけるため、きちんと相手と目を合わせて対話している印象を与えたという点も好印象の要因である。

参考文献

- [1] 池田奈生, 原ひかる, 松村冬子, 原田実: “話者の意図理解に基づく対話応答システム Hermes における雑談対話”, 情報処理学会第 76 回全国大会論文集, 3P-4, (2014.3).
- [2] 小林豊, 鈴木諒, 谷津元樹, 原田実: “深層学習による日本語キャプション生成システムの開発”, 第 17 回インタラクティブ情報アクセスと可視化マイニング研究会発表予稿集(2017.11).
- [3] 田村優樹, 中野雅広, 長崎達也, 原田実: “意味解析に基づくロボット指示システム Athena2011”, 情報処理学会研究報告, 2012-NL-206, No.5, pp. 1-8, (2012.5).