環境適応的な発話機構による 文脈に厚い対話システムの提案

A Proposal of Context-Thick Dialogue System using Adaptive Utterance Mechanism in Environment

高橋諒 [1]　棟方渚 [1]　小野哲雄 [1]

Ryo Takahashi[1], Nagisa Munekata[1], and Tetsuo Ono[1]

[1] 北海道大学大学院情報科学研究科

[1]Graduate School of Information Science and Technology, Hokkaido University

Abstract: In this study, we propose the context-thick dialogue system using many context-thin dialogue systems in accordance with the subsumption architecture concept. Also, we perform the user study that participants communicate with the robot using the dialogue system we implemented based on our proposal system. As a result, we consider that interactive activity is an important element for our proposal system.

1　はじめに

近年，ソフトバンク社の Pepper[1]や，Aldebaran 社の NAO[2]を始めとしたコミュニケーションロボットが一般的に販売され，徐々に普及し始めている．特に Pepper は会話機能を有し，感情を認識できる家庭用ロボットであり，近い将来，ロボットが人間社会へ溶け込むことが予想される．その際，ロボットが長期間利用されるには，人との円滑なコミュニケーションを行えることが重要であり，中でも円滑な対話を行えることが重要な要素の一つであることは容易に想像できる．

それに関連して，これまでに様々な AI の議論が行われてきた．中でも佐藤[3]は，知的なシステムとは，文脈に厚い（多様な状況において適切な行動を行える）システムであり，それを実現するためには，多様な入力情報と多様な出力が存在するもとで，これらの間を適切に結びつける必要がある，と述べている．そして，限られた入力情報から限られた出力をする小さなモジュール（文脈に薄いシステム）を複数用意し，それらを使い分けることによる，全体としての文脈に厚いシステムの構築手法を提案している．

そこで，本研究ではこの手法に着目し，chatterbot（会話ボット)に加えて，周囲の画像情報から文章を生成するモジュールや，シソーラス（概念）辞書などの，複数の文脈に薄いモジュールを使い分けることで、周囲の環境に適応することが可能である，文脈に厚い対話システムの構築を提案し，その実現を目指す．

2　関連研究

本章では，主に本研究で使われる技術やシステムについて述べる．

人とコンピュータとのコミュニケーションの研究は古くから行われてきた．特に Joseph Weizenbaum[4]が開発した ELIZA は，有名なテキストチャットによる対話型システムの一つである．ELIZA は，相手の入力文の中から単語をキーワードとして抽出し，そのキーワードに反応してシステム内で用意されている文章の中から応答文を出力する．もしキーワードに反応できる文章がなかった場合は，相手に話の続きを促したり，話題をそらすような文章を出力する．こうすることで，ELIZA はどのような話題に対しても対話を続けることができるという頑健性を実現している．しかし，ELIZA はその反面，相手からの入力を前提としている受動的なシステムであるため，相手の入力がなければ会話が終了してしまう．また，目的のある会話（例えば，「今日の天気は何ですか」など）の場合，表層的なやり取りしか行うことができないので，対話相手が満足した答えを得ることができず，対話相手が失望してしまう恐れがある．加えて，ELIZA のような対話システムでは，環境情報を共有できないため，総じて，生活環境を共にする人とロボットとのコミュニケーションには向いてい

ない.

一方で, 近年の画像処理技術の発展は目覚ましく, 特に Deep Learning（深層学習）の登場によって, 物体認識など, 様々な画像処理技術の精度が飛躍的に上昇した．その中で，Andrej Karpathy ら[5]が開発した Neuraltalk2 は，与えられた画像から，その画像の説明文を生成するシステムである．Neuraltalk2 は, まず事前に画像とその画像の説明文が紐づけられている学習セットを用いて学習し, 学習モデルを生成する. そしてその学習モデルをもとに，新しく与えられた画像の物体認識を行い，その物体に紐づけられた文章を組み合わせることで, 画像の説明文を生成する. Neuraltalk2 を用いることで, 例えばロボットに内蔵されたカメラから画像を取得し，それを文章に変換して発話することで，環境情報を利用した能動的な発話が可能となる. しかし, Neuraltalk2 のみを用いただけでは, 環境情報の利用による能動的な発話はできても, 相手の発話内容の参照や，話題の展開ができないため，やはり人とロボットとの円滑なコミュニケーションの実現は難しい.

また，人工知能の実現を支援する目的で George A. Miller ら[6]が開発した WordNet は, オンラインデータベース上の英語のシソーラス辞書である．WordNet は，単語を synset と呼ばれる類義語のセットによってグループ化しており, また synset は上位語，下位語，反意語などの様々な語彙関係で結ばれている. WordNet を用いて文章中の単語を変化することで, 一つの文章から発展して様々な文章を生成することができ, 話題の展開や変更に役立つことが期待される．しかし，WordNet はあくまで辞書であるため, これ単体では人とロボットとのコミュニケーションは難しい.

本研究では，これらの文脈に薄いモジュールを使い分けることで， 文脈に厚い対話システムの構築を目指す.

3　提案モデル

本章では，第 1 章で述べた文脈に厚い対話システムの実現に向けて，我々が提案するモデルについて述べる.

まず，Rodney Allen Brooks[7]が提唱した，人工知能の概念である，サブサンプション・アーキテクチャについて述べる．サブサンプション・アーキテクチャとは，ロボットの複雑な行動を複数の単純な行動に分割し，それぞれを外部入力（音声など）に応じて動く小さなモジュールとして階層構造を構築する．そして，それらのモジュールを並列処理で実行

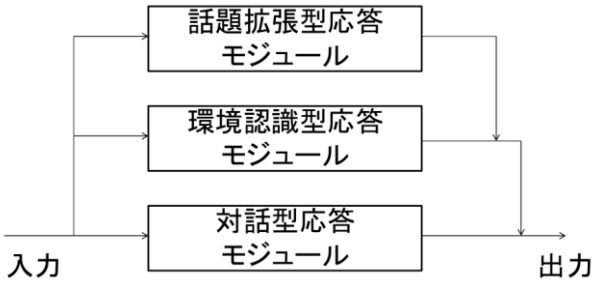

図 1:提案モデル

しつつ，競合が起きる場合は下層よりも上層の行動の方が優先的に行われることで，複雑な行動を実現するものである．本研究では，このサブサンプション・アーキテクチャの概念を用いた対話モデルを提案する.

具体的なモデルを図 1 に示す．まずロボットは人と対話するときに，最下層に位置する対話型応答モジュールにより対話を行う．しかし第 2 章で説明したように，対話型応答モジュールだけでは対話が受動的になり，相手からの入力がなければ，その時点で対話が終わってしまう．それを防ぐために，対話の止まっている時間，および対話時の発話量などから対話活性度の推定を行う．そして対話活性度が低いと判断されると，対話型応答モジュールの上層に構築されている, 環境認識型応答モジュールに移り，環境情報を文章化して能動的な発話を行う．こうすることで，対話型応答モジュールの受動的であるという欠点を克服しつつ，話題の提供を実現する．それでも，提供された話題が対話相手にとって興味のない話題だった場合，結局会話が止まったままになってしまう．その場合を想定して，対話活性度が低いままであると判断されると，環境認識型応答モジュールの上層に構築されている，話題拡張型応答モジュールに移り，辞書に設定されている語彙関係を用いて，生成された文章内の単語を変換する．こうすることで，例えば，野球→スポーツのように，野球という限定的なスポーツに興味がなくても，話題を野球からスポーツの話題に拡張することができ，対話の継続が期待できる.

現在提案するモデルは述べたとおりであるが，当然それでも会話が止まってしまうことは考えられる. だが提案モデルの大きな利点として，文脈の薄いモジュールの集合体で，かつ優先順位が決まっているため，モジュールの追加が容易である点が挙げられる．第 2 章で述べた通り，人とコンピュータの研究は数多く行われており，Web 情報を用いたものや，ユーザモデルを定義，および学習するものもある. これら過去の研究によって提案されたモデルを，新しいモジュールとして提案モデルに追加することで,

様々な応答が可能となり，最終的には人とロボットとの円滑なコミュニケーションの実現が期待できると我々は考える．

4　提案システム

本章では，第 3 章で述べたモデルをもとに，具体的に我々が実装したシステムについて述べる．

提案システムを図 2 に示す．最下層である対話型応答モジュールに関しては，頑健性があるという観点から ，JezUK[8]が提供している，ELIZA を小型化したものである eliza.py – ELIZA in Python（以下 eliza）を用いた．しかし，2 章で述べたように ELIZA には受動的である，環境情報を利用することができないなどの欠点がある．そこで，eliza の上層に，これらの欠点を補う環境認識型応答モジュールとして，Neuraltalk2 を用いた．Neuraltalk2 によって，ロボットのカメラから取得した画像を入力として，その画像の説明文を生成し発話することができるため，能動的かつ環境情報を利用した発話が可能となる．さらに，生成された文章が対話相手にとって興味のない話題であった場合を想定して，Neuraltalk2 の上層に，話題拡張型応答モジュールとして WordNet を用いた．WordNet によって，生成された文章内の単語を一つ上の上位語に変換することで，話題の拡張を狙い，結果としての対話の継続を期待する．なお，eliza，Neuraltalk2，WordNet はすべて言語が英語であるため，入出力は英語で行う．また，音声認識による誤入力を防ぐため, 入力はテキストベースとした．この場合，対話活性度の指標として発話量は利用できないため，対話が止まっている時間のみを用いて推定することとした．

提案システムの流れを説明する．システムが実行されると, モジュールの最下層である eliza が相手からの入力を待ち, 入力が行われると eliza により応答文を返す．そして，対話を続けていく中で対話が止まった場合，会話活性度によりそれを推定し，ロボットのカメラから画像を取得し，eliza の上層である Neuraltalk2 によって文章に変換され，環境情報に関する能動的な発話を行う．それでも対話活性度が低いままであると推定されると，Neuraltalk2 の上層である WordNet によって，先ほど生成した文章の単語を一つ上の上位語に変換し，話題を拡張したうえでもう一度発話する. 以下, 入力が行われるならば eliza の応答文を，対話活性度が低いのであれば Neuraltalk2 による発話，および生成された文章の単語の上位語変換を入力が行われるまで繰り返し行う．本システムによって，ELIZA による頑健性を維持したうえで，受動的であることや，環境情報を利用できないといった ELIZA の欠点を Neuraltalk2 によって克服し，かつ WordNet による話題の拡張によって対話相手に入力を促すといった，文脈に厚い対話システムを実現できているのではないかと我々は考える．

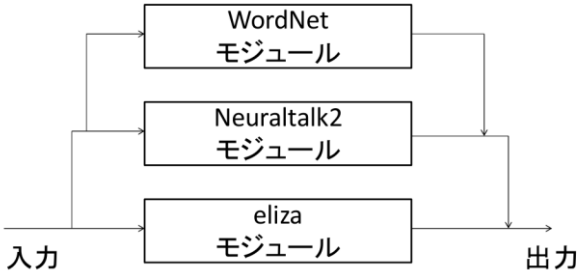

図 2:提案システム

5　ユーザスタディ

本章では，提案システムを使って実際に行ったユーザスタディについて述べる．

ユーザスタディの環境は，被験者にテーブルを挟んでロボットの向かい側に座ってもらい，被験者の向かって左側にディスプレイ，手元にキーボードを置いて行った．ロボットはソフトバンク社の Pepper を使用した．被験者は普段からパソコンを利用している 23～24 歳の男子大学生 2 名とした．

ユーザスタディでは，被験者にロボットとの簡単な対話を行ってもらった．この際，ロボットの対話システムは

① eliza のみ
② 本研究で提案しているシステム

の 2 種類のシステムがあり，それぞれのシステムについて, 各 1 人ずつユーザスタディを行った. なお，被験者の入力はテキストベース，ロボットの出力は音声で行い，ロボット側は発話した文章を胸元についているディスプレイに表示した．また②の対話システムに関して, eliza から Neuraltalk2 へ, Neuraltalk2 から WordNet へ移行する条件を, 20 秒間入力がなかった場合とした．ユーザスタディの様子を図 3 に示す．

実験の流れは，まず被験者にユーザスタディの内容を説明した．このとき，

・最低でも 4 分以上ロボットと対話を行ってもらうこと
・被験者が「Quit」と入力したとき，対話が終了すること
・②の対話システムでユーザスタディを行う人に対しては，ロボットが「Please Wait a minute」と発話し

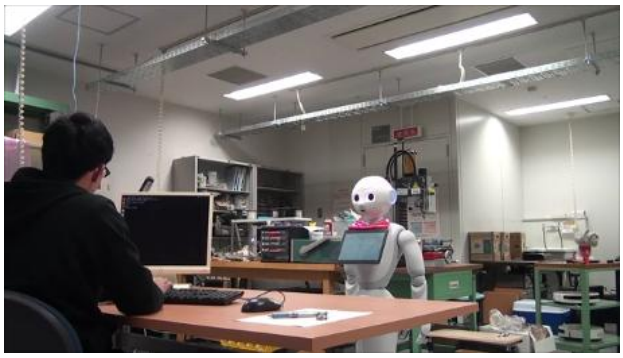

図 3:ユーザスタディの様子

たら，入力を中断して入力途中の文をすべて消してもらい，ロボットに注目してもらうこと（この際，ロボットは首を右から左へ動かしつつ，頭部についているカメラで画像を取得する）

という条件をつけた．そして，実際にロボットとの対話を行ってもらい，対話終了後に，実験アンケートに回答してもらった．アンケートに関しては，心理測定尺度集Ⅱに記載されている，対人認知の領域の特性形容詞尺度の項目[9]を参考に，パーソナリティ認知の測定に有効な特性形容詞尺度の 49 組のうち，20 組を使用して 1〜5 の 5 段階で評価してもらった. またその他に,「ロボットと対話が弾んだか」,や「ロボットとの対話が苦痛か」といったロボットとの対話に関する項目を用意し，同様に 1〜5 の 5 段階で評価してもらった．さらに加えて，

・本実験，および普段の生活の中で，会話が途切れてから気まずくなったり，間が空いたと感じるのは何秒か
・これまでにロボットと関わったことはあるか
・ロボットと被験者はお互いにどのような印象を抱いたか

などについても，自由記述欄を設けて回答してもらった，

なお，ユーザスタディ中は被験者の右斜め後方から撮影を行った．

6　結果と考察

本章では，ユーザスタディの結果，およびユーザスタディの観察からなる考察を述べる．

まず，ユーザスタディの結果を述べる．①の対話システムに関しては，対話時間は5分1秒で，対話回数は往復30回であった．また，②の対話システムに関しては，対話時間は6分51秒で，対話回数は被験者側が12回，ロボット側が20回であった．アンケート結果に関しては，被験者が少なく統計的処理が難しいため，5段階評価の部分については割愛し，自由記述の欄に関しては考察の方で触れることとする．

次にユーザスタディを観察したことによる考察について述べる．

まず①のシステムに関しては，被験者は，前半のうちは普通にロボットとの対話を行っていたが，ロボット側が似たような応答を繰り返すにつれて，後半ではロボットが人工無脳（chatterbot）ではないかと疑っている様子が見られた．実際に，被験者はロボットに対して「I feel your language seems to be some kind of chatbots.」といった入力を行っている．またアンケート内でも,「ロボットが人工無脳のような話し方に感じた」,「通常，人と人が話すときはどちらかが一方的に話すシーンも多いが，このロボットは過度に会話のキャッチボールに感じた」と述べている．総じて，被験者は対話を続ける中でロボットの対話システムが人工無脳であると気付いたと思われる．このことから，やはりelizaのようなchatterbotのみでは対話を続けることが難しいように思える．

次に②のシステムに関しては，比較的前半でロボットの発話の意味が分からずに対話が止まる場面があったが，Neuraltalk2による環境情報を利用した能動的な発話が行われることによって，対話が続いた場面が見られた．しかし後半になると，Neuraltalk2による発話が環境情報と全く関係のないものとなり，被験者が戸惑っている様子が見られた．実際，被験者はアンケート内で,「ロボットは自分のことをいろいろ知りたがって対話を広げようとしているが，自分はロボットと対話が成立していないと感じ，困った」と述べている．また，ユーザが文章を入力している途中であるにもかかわらず，ロボットは対話が止まっていると判断するなど，対話活性度の推定に問題があったように思える．これは，対話活性度の推定を，対話の止まっている時間のみによって推定していることが原因だと思われる．

7　まとめと今後の展望

本研究では，複数の文脈に薄いモジュールを組み合わせることによって，周囲の環境に適応可能である文脈に厚い対話システムの構築を提案した．

具体的には，サブサンプション・アーキテクチャの概念をもとに，下層から順に対話型応答モジュール，環境認識型応答モジュール，そして話題拡張型応答モジュールからなる階層構造を構築し，それらを対話活性度によって使い分けるこで，どのような

入力にも対応できる頑健性と，環境情報を利用することによる能動的な発話，および話題の拡張が可能な対話モデルを提案した．

さらに，対話型応答モジュールとして eliza，環境認識型応答モジュールとして Neuraltalk2，そして話題拡張型応答モジュールとして WordNet を使用し，実際に提案モデルをもとに対話システムを構築し，ロボットと対話をしてもらうユーザスタディを行った．

結果として，eliza のみの対話システムでは途中で被験者がロボットのことを人工無脳であることに気づいた様子が見られ，本研究の対話システムでは，ロボットによる能動的な発話が多くみられたが，発話の内容やタイミングに戸惑っている様子が見られ，対話の継続は見られなかった．原因として，対話活性度の推定が対話の止まっている時間のみによって推定されていたことが考えられる．

今後の展望として，まず，テキストベースの入力から音声ベースの入力に変更し，対話活性度の推定に対話中の発話量を利用するようシステムを改良する必要があると考える．またそれに加えて，モジュールを追加することによる提案システムの改良と，構築したシステムを利用したさらなる実験を行っていきたい．

参考文献

[1] Pepper: http://www.softbank.jp/robot/consumer/products/ [Accessed 9 November 2016]

[2] NAO: http://www.revast.co.jp/service/humanoid/type03.html [Accessed 9 November 2016]

[3] 中島　英之, 有馬　淳, 佐藤　理史, 諏訪　正樹, 橋田　浩一, 浅田　稔: 新しい AI 研究を目指して, 人工知能学会誌, Vol. 11, No. 5, pp. 713-724, (1996)

[4] Joseph Weizenbaum: “Computational Linguistics,” Communications of the ACM, Vol. 9, No. 1, pp. 36-45, (1966)

[5] Andrej Karpathy, Li Fei-Fei: “Deep Visual-Semantic Alignments for Generating Image Descriptions,” arXiv, 1412.2306v2 [cs.CV], (2015)

[6] Christiane Fellbaum: “WordNet: AnElectroonic Lexical Database,” The MIT Press, (1998)

[7] Rodney A. Brooks: “A Robust Layered Control System For A Mobile Robot,” IEEE Journal of Robotics and Automation, Vol. RA-2-1, pp. 14-23, (1986)

[8] eliza.py – ELIZA in Python: http://www.jezuk.co.uk/cgi-bin/view/software/eliza [Accessed 9 November 2016]

[9] 林　文俊: 対人認知構造の基本次元についての一考察, 25, pp. 233-247, (1978)