

修士論文の和文要旨

研究科・専攻	大学院 情報理工学研究科 知能機械工学専攻 博士前期課程		
氏名	宮崎 齊	学籍番号	1332064
論文題目	台詞から想起される感情を表現するロボット動作の自動生成		
要旨	<p>現在、人間共存型ロボットの多くは、あらかじめ作成した動作データに基づいて対話時の動作等を実行している。しかし、個々の動作を人間が手作業で作成するという手法では、動作の種類が限られ、動きに多様性を与えることが難しい。また、作成に要する作業量も膨大である。</p> <p>そこで本論文では、与えられた台詞から想起される感情を機械学習によって推定することで、台詞に対応する感情を伴った動作を自動生成することを目的とする。</p> <p>動作生成の枠組は、学習過程と認識生成過程に分けられる。学習過程では、まず台詞と動作のセットを収集し、前処理を施した上で学習データセットを作成した。学習する台詞としては、日常会話文と感情を強く含んだ文の双方を使用した。動作としては、台詞に沿ったロボットのポーズを手動で作成したものをを用いた。次に、作成した学習データセットに対して Multimodal Latent Dirichlet Allocation (MLDA) 手法を用いて学習を行い、いくつかの特徴的なカテゴリに分類した。MLDA を選んだ理由として、複数種類のデータ（台詞、動作）を関連付けて学習できること、観測できない種類のデータ（動作）を観測したデータ（台詞）から推定できることが挙げられる。認識生成過程では、新たに入力された台詞を各カテゴリに分類し、その台詞に合った動作を類似した台詞から推定・生成する。提案手法では台詞からの動作の生成とほぼ同様の手順で、動作からの台詞の生成が可能である。</p> <p>評価のために主観評価実験を行った。実験は、被験者に1つの台詞に付けた2種類の動作を見比べてもらい、動作に関する質問に答えてもらう形式で実施した。2種類の動作のうち、一方は提案手法、他方は比較対象の手法で生成した動作とし、それらを比較することで、提案手法と比較対象の相対評価とした。比較対象には、手動生成、ランダム、動作無しの3種類を用意した。ランダムは、意味なく小さく手を動かし続ける動作であり、会話中の動作の生成によく用いられる。質問は、一組の動作につき5問で、台詞と動作の整合性、親和性、人間らしさ、動きの滑らかさ、感情表出に関するものとした。実験の結果、手動生成、提案手法、ランダム、動作無しの順で、被験者はロボットの動作に好印象を感じたことがわかった。提案手法により従来のランダムな動作をするロボットに比べて自然な動作を自動生成できるようになった。</p> <p>本研究の成果によって、ロボットの会話中の動作の自動作成が可能になるとともに、ロボットの行動の幅を広げ、より人間的な表現を実現することができるようになる。また、動作と台詞の相互変換が可能であるため、入力として人間の動作をロボットの動作に変換したものを与えることで、人間の感情や意図を簡易的に推定することができる。この成果は、人間のことを「察し」、人間的な行動をとることができるロボットの実現への第一歩となる。</p>		

平成 26 年度 修士論文

台詞から想起される感情を表現する
ロボット動作の自動生成

学籍番号 1332064

氏名 宮崎 齊

知能機械工学専攻

先端ロボティクスコース

指導教員 金子 正秀 教授

副指導教員 長井 隆行 教授

提出日 平成 27 年 2 月 27 日

概要

近年、ロボットは活躍の場を大きく広げており、工場や研究室の中だけでなく、街中の商業施設の中や博物館、病院など、我々の程近くで人間の生活を支えている。人間の生活圏で行動するロボットに強く求められるのが、人間との自然なコミュニケーション能力と人間の行動から意図、感情を理解する能力である。これらは人間同士の相互関係の形成にも大きく寄与している能力であり、ロボットが人間と信頼関係を構築する上でも非常に重要な能力であると考えられる。人間同士であれば発話内容、発話時の動作などから相手の感情や意図を読み取り、それに合わせた発話や行動を返すことで意思疎通し、友好的な関係を形成していく。しかし、現在使われているインタラクションロボットの中には一方的な会話しかできないものも少なくなく、発話中の動作もワンパターンなものが多い。人間の生活圏内でより人間的に振る舞うことのできるロボットを実現するためには、更なるコミュニケーション機能の向上が必要であると考えられる。

本論文では、機械学習を用いたロボットのための動作の自動生成システムについて提案する。事前に発話内容（台詞）とそれに合った動作を学習データとして学習することによって、学習していない新たな台詞に対して台詞の意味まで考慮したような違和感のない動作をつけることを可能にする手法を示す。また、動作の生成時にランダム性を与えるような工夫を組み込むことによって、同一の台詞に対して多様な台詞に合った動作をつけることを可能とし、ロボット動作がワンパターンになってしまうという従来のロボット動作の問題点を解消した。本論文で提案したロボット動作の自動生成システムを取り入れることによって、会話中により人間的で自然な動作をするロボットを実現することが可能となる。提案手法と他の手法を比較する主観評価実験を行い、研究の有効性を示した。

目次

第1章 序論.....	4
第2章 自動ロボット生成システムの概要.....	5
2.1. ハードウェア構成.....	5
2.2. ソフトウェア構成.....	6
2.3. 全体の処理の流れ.....	6
第3章 学習過程.....	9
3.1. 学習の収集.....	9
3.1.1. 台詞の収集.....	9
3.1.2. 動作の収集.....	9
3.2. 台詞の前処理.....	10
3.2.1. 形態素解析.....	11
3.2.2. Bag of Words.....	11
3.2.3. TF-IDF.....	12
3.3. 動作の前処理.....	13
3.3.1. Dirichlet Process Gaussian Mixture Model.....	13
3.3.2. 動作を前処理した結果.....	14
3.4. 機械学習.....	15
3.4.1. Multimodal Latent Dirichlet Allocation.....	15
3.4.2. MLDAによる学習データの分類結果.....	16
第4章 認識・生成過程.....	18
4.1. 台詞から動作の生成.....	18
4.1.1. 入力データの認識.....	18
4.1.2. 動作生成プロセス.....	19
4.2. 動作から台詞の認識・生成.....	21
4.2.1. 入力データの認識.....	21
4.2.2. 台詞（関連単語）生成.....	22
第5章 準備実験.....	24
5.1. 台詞・動作データセットの学習.....	24
5.2. 未学習の単語を含む台詞に対する動作の生成.....	24
5.3. 一つの台詞に対する多様性のある動作の生成.....	28
5.4. 動作から台詞（関連単語）の生成.....	29
5.5. 準備実験まとめ.....	32
第6章 主観評価実験.....	33
6.1. 実験の概要.....	33

6.1.1.実験の流れ	33
6.1.2.実験環境.....	33
6.1.3.事前説明.....	34
6.1.4.実験動画.....	34
6.1.5.アンケートの内容	36
6.2.実験実施前の考察.....	40
6.2.1.動作無しとの比較	40
6.2.2.ランダム手法との比較	40
6.2.3.手動との比較.....	40
6.3.実験結果・考察	41
6.3.1.実験結果.....	41
6.3.2.動作無しとの比較	42
6.3.3.ランダム手法との比較	42
6.3.4.手動生成との比較	42
6.3.5.実験総評.....	42
第7章 結論.....	43
7.1.本研究の成果.....	43
7.2.今後の展望	43

第1章 序論

近年、ロボット技術の著しい発達によって、我々の日常生活の中でも人間にサービスする機能を持ったロボットを見かけるようになった[1][2][3]。なかでもヒューマノイド型のコミュニケーションロボットの発達は目覚ましく、ステージでダンスを披露したり、博物館などのガイドとして館内を案内したり、一般家庭や商業店舗で人間の話し相手をしたりと様々な活躍をしている。これら人間の生活圏内で人間とともに生活するロボットに強く求められる能力は、人間の感情・意思を察する能力やロボット自身の感情・意図を人間に伝える能力といった人間と自然に関わるための能力である。特に、医療・介護の現場や、小さな子供と関わるような環境においては、単に高精度な自然言語認識・理解能力やわかりやすい発話機能・発話内容、直感的なインターフェースだけでなく、感情を伴ったコミュニケーションの可能なロボットが必要であると考えられる。実際、医療現場などで用いられているパロ[4]は感情・意思の伝達を目的とした様々な機能を備えているし、いやし型赤ちゃんロボット「スマイビ」[5]の様にロボット自身の意思を人間に伝達することに特化したロボットも開発されている。また、ヒューマノイド型のサービスロボット Pepper[6]は「世界初の感情認識パーソナルロボット」というコンセプトを打ち出している。ロボット技術の更なる発展に伴い、ロボットは今後更に人間の生活空間内に浸透していき、ロボットと人間が関わる場面は増加していくと考えられる。それに従って、ロボットには更なる多機能化とコミュニケーション機能の質の向上が求められている。

現在、実際に使用されているコミュニケーションロボットには、発話するたびにワンパターンな動作をするものが少なくない。また、バリエーションのある動作のできるロボットの多くは、発話の内容などによって動作の生成方法に差をつけるようなことはしていない。しかし、発話中の動作というのは人間同士のコミュニケーションにおいて非常に大きな意味を持っていることが知られている。また、ジェスチャなどの身体動作に関しては、石田らが論文[7]中の仮説として「ロボットは機能だけでなく、その機能を身体で表現することによって人間に知的な印象を与える」と言及しているように、ロボット-人間間のコミュニケーションにおいても、大きな意味を持っていると考えられる。

そこで、本研究では従来のコミュニケーションロボットよりも人間的で自然な動作のできるロボットの実現を目指し、機械学習を利用した発話内容に合ったロボット動作の自動生成手法を提案する。これによってコミュニケーション中に人間に好印象を持たせることのできるロボットを実現する。

第2章 自動ロボット生成システムの概要

2.1. ハードウェア構成

ハードウェア構成を以下に示す.

- CPU : Intel(R) Core(TM) i7-2600 3.40GHz
- メモリ : 8.0 GB
- ビデオカード : NVIDIA GeForce GTX 560 Ti
- インタラクションロボット : NAO T-14 v4

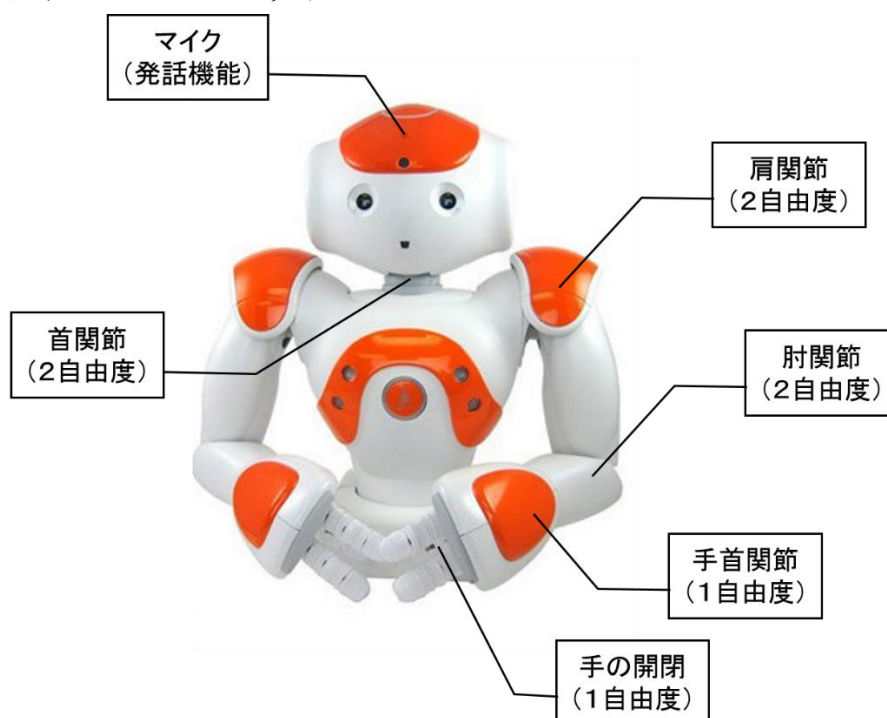


図 2.1. インタラクションロボット NAO T-14

本研究では、人間との会話時にロボットの発話する台詞に合わせた動作を自動生成するシステムを提案する。そのためロボットの外見としては、人間が話し相手と認識できる形状で、できれば親しみの持てるようなものが望ましい。そこで本研究では、インタラクションロボットとして Aldebaran Robotics 社製のヒューマノイドロボット NAO T-14 (以下、NAO と表記) を使用した。NAO は人間的な見た目をしており、発話していても違和感を覚えづらい。また、頭部に 2 自由度、左右の腕にそれぞれ 6 自由度あり、可動域は大きくないものの、ある程度人間的な仕草をすることができる。今回使用したモデルの NAO には足がないが、人間が会話中に主に注目する動作は首や手など上半身の動作であると考え、上半身のみのモデルで十分と判断した。後述する実験などは、すべて NAO を机上に設置し、被験者と顔を向かい合わせるような状態で行った。

2.2. ソフトウェア構成

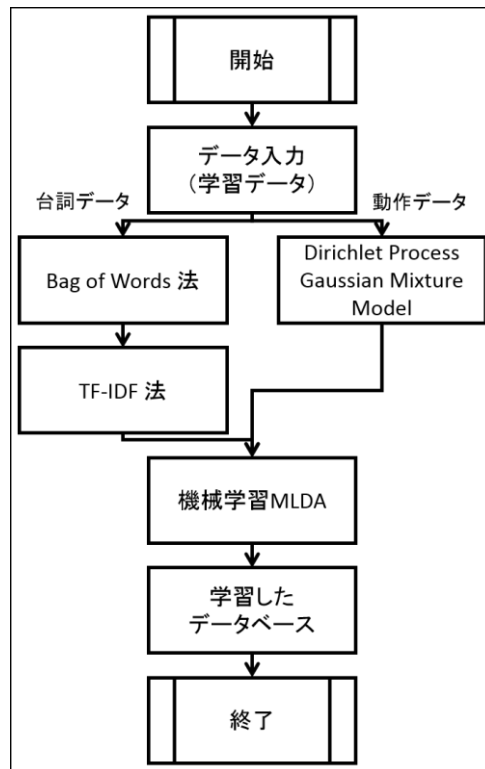
ソフトウェア構成を以下に示す.

- 計算機用 OS : Windows8.1 Enterprise
- インタラクションロボット用 OS : Nao qi ver.2.1.2.17
- VisualStudio2010
- OpenCV 2.4.9
- Mecab 0.996
- Python2.7
 - Jinja2 ver.2.7.3
 - MarkupSafe ver.0.23
 - PIL ver.1.1.7
 - Theano ver.0.6.0
 - backports.ssl-match-hostname ver.3.4.0.2
 - certifi ver.14.05.14
 - cssselect ver.0.9.1
 - decorator ver.3.4.0
 - ipython ver.2.2.0
 - lxml ver.3.4.0
 - matplotlib ver.1.3.1
 - mecab-python ver.0.996
 - networkx ver.1.9.1
 - numpy ver.1.8.1
 - pygame ver.1.9.1
 - pynaoqi-python2.7 ver.2.1.0.19
 - pyparsing ver.2.0.2
 - pyreadline ver.2.0
 - python-dateutil ver.2.2
 - pyzmq ver.14.3.1
 - qibuild ver.3.5.1
 - requests ver.2.4.1
 - scikit-learn ver.0.15.0
 - scipy ver.0.14.0
 - six ver.1.7.3
 - tornado ver.4.0

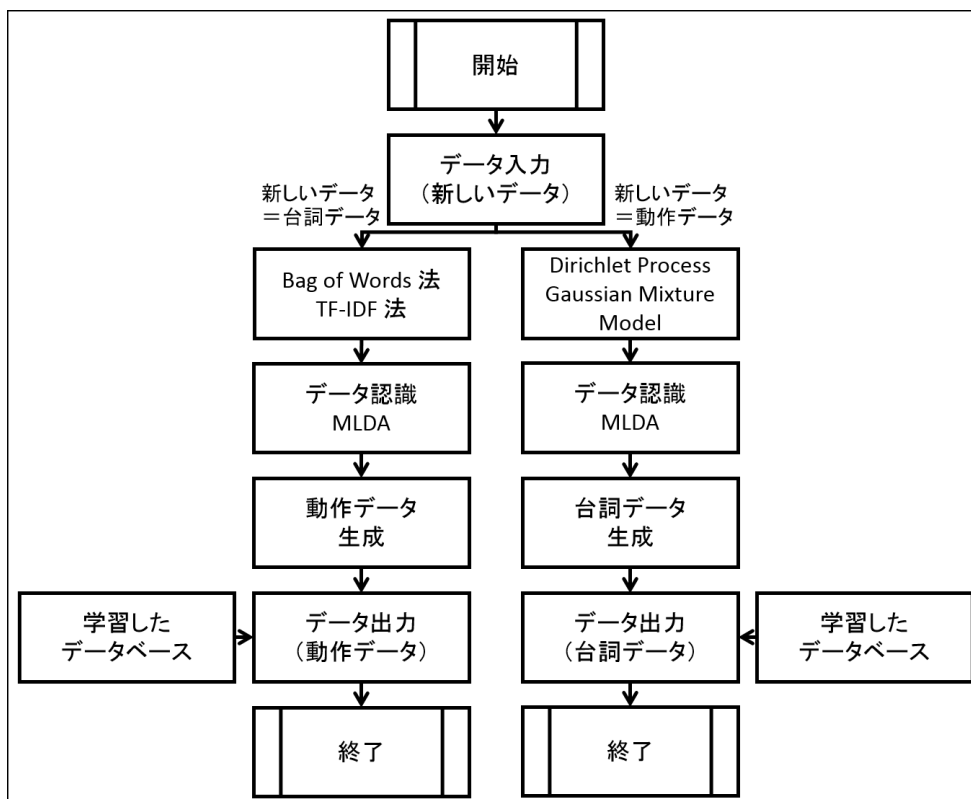
2.3. 全体の処理の流れ

本研究では、機械学習を利用してロボットの動作の自動生成を実現する。そのため、システムの全体的な流れは、学習過程と認識・生成過程の2つに分けることができる。それぞれの過程の処理の流れを図 2.2 に示す。

学習過程では、大量のデータセット（動作・台詞）を参照し、台詞と動作の関係性を学習する。学習用のデータを作成するため、まず台詞を収集し、それに対応する動作（ポーズ）を、ロボットを直接手で動かすことで作成する。台詞と動作のデータセットを学習データとする。収集した台詞、動作には、それぞれ前処理を施して機械学習で扱いやすいデータの形にする。具体的には、自然言語である台詞を Bag of Words でベクトル化し、TF-IDF 重みを付与する。動作に関しても、Dirichlet Process Gaussian Mixture Model（以下、DPGMM と表記）を利用し、類似動作に分類しておく。前処理した学習データセットに対して、



(a) 学習過程



(b) 認識・生成過程

図 2.2. システム全体の流れ ((a) 学習過程, (b) 認識・生成過程)

Multimodal Latent Diriclet Allocation 手法（以下、MLDA と表記）で学習を行い、動作と台詞の対応関係を得る。学習の結果、学習データセットは分類され、それぞれ特徴を持った複数のカテゴリを形成する。本研究では、学習の結果得られるカテゴリの内、いくつかは感情、意図を表していると考えられる。

認識・生成過程では、新たに入力された台詞をカテゴリ分類し、入力した台詞に合った動作を類似した台詞から推定・生成する。まず、新たに入力された台詞に学習時と同様の前処理を施し、機械学習で扱いやすいデータに変換する。その後、MLDA を利用し、新たに入力された台詞が、学習過程で分類したカテゴリ中のどのカテゴリに分類される確率が高いか認識する。最後に、入力した台詞が分類されたカテゴリ内の動作から、入力した台詞に合った動作を推定し、出力する。提案手法では、台詞からの動作の生成とほぼ同様の手順で、動作からの台詞の生成も可能である。

以上の 2 つの過程で、台詞に合った自然なロボット動作生成を実現する。

第3章 学習過程

本章では、ロボット動作自動生成システムのうち、学習データから台詞と動作の関係性を得る学習過程について述べる。

3.1. 学習の収集

3.1.1. 台詞の収集

学習データ作成のため、まずロボットが発話する台詞を収集した。本研究では、学習する台詞として、中学生レベルの英語教材の会話シーンを和訳したものと学習用に作成した感情を強く含んだ台詞を合わせて 896 文を使用した。英語教材の和訳を学習する台詞の素材として利用した理由としては、会話シーンが多く、効率的に台詞が集められること、セッションごとに一貫した会話のテーマがあること、内容が単純で文章を直感的に理解することができることが挙げられる。また、学習用に作成した台詞を加えたのは、英語の教科書内にあまり登場しない「喜び」、「悲しみ」などの強い感情を含んだ台詞を学習させるためである。学習した台詞の一部を表 1 に示す。

表 1. 学習した台詞の一例

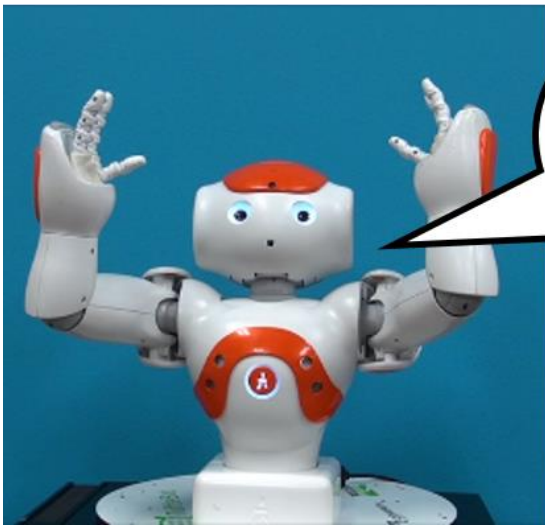
英語教材の会話文の和訳例	学習用に作成した台詞例
はじめまして。ユキです。よろしくね。	行ってらっしゃい。お気をつけて！
へえ、サキはピアノが弾けるんだ！	今日は誕生日なんだ！うれしいな！
ユキは、野球好きかい？	今日でお別れかと思うと、悲しい

3.1.2. 動作の収集

次に、各台詞に対応する動作を収集した。一つの動作（ポーズ）は、ロボットの関節角度データ 1 セット（14 次元のベクトル）として表される。動作の収集方法としては、被験者にロボットが発話する台詞を提示し、ロボットを実際に手で動かしてもらうことで、台詞に沿ったポーズを作成してもらうという方法をとった。これ以外の動作収集方法として、被験者にモーションキャプチャ用の装置を付けてもらい、台詞に沿った動作をしてもらう方法が考えられる。しかし、本研究では、人間がする動作とロボットにして欲しい動作は違うのではないかと、また、実物のロボットを見ることで、はじめてそのロボットにあった動作を作成できるのではないかとという考察の下、実際にロボットを手で動かす動作作成方法を選択した。動作作成の様子を図 3.1 に示す。また、作成した学習データセット（動作・台詞）の一例を図 3.2 に示す。



図 3.1. 動作作成の様子



やったあ！
デバッグが終わった！

図 3.2. 学習用データの一例

以上の手順で作成した台詞と動作のセットを学習データとして利用した。

3.2. 台詞の前処理

収集した台詞は自然言語の状態のため、そのままでは機械学習するためのデータとして取り扱いづらい。そこで自然言語から数値的なデータへの変換を前処理として行う必要がある。

3.2.1. 形態素解析

台詞を数値データに変換する前に、まず台詞を形態素解析する。形態素解析とは、対象言語の知識（文法のルール）や辞書（品詞等情報付きの単語リスト）を情報源として利用し、自然言語で書かれた文を、それだけで意味の分かる最小単位（形態素）に分割し、それぞれの品詞を判別する作業を指す。形態素解析には、京都大学情報学研究科-日本電信電話株式会社コミュニケーション科学基礎研究所 共同研究ユニットプロジェクトを通じて開発されたオープンソース形態素解析エンジン MeCab[8]を利用した。台詞を形態素解析した例を以下に示す。

● 形態素解析の例

- ・ 片付けと言え、明日はリサイクルの日よ。

↓形態素解析

- ・ 片付け（名詞） / と（助詞） / 言え（動詞） / ば（助詞） / 、（記号，読点） / 明日（名詞） / は（助詞） / リサイクル（名詞） / の（助詞） / 日（名詞） / よ（助詞） / 。（記号，句点） /

収集した全台詞を形態素解析し、出現した単語で辞書を作成した。辞書に登録したのは、品詞が感動詞、動詞、副詞、助動詞、名詞（固有名詞を除く）の単語で、他の品詞の単語は辞書には登録しなかった。後述の Bag of Words で台詞の数値データ化を行う際に、この作成した辞書を特徴語辞書として用いる。

3.2.2. Bag of Words

Bag of Words は文書検索システムやトピック解析などの際によく使われる表現で、特徴語の辞書を事前に用意しておき、ある文書中にどの特徴語が何回出現したかという情報でその文書を表現する手法のことである。Bag of Words による文章の変換の例を図 3.3 に示す。

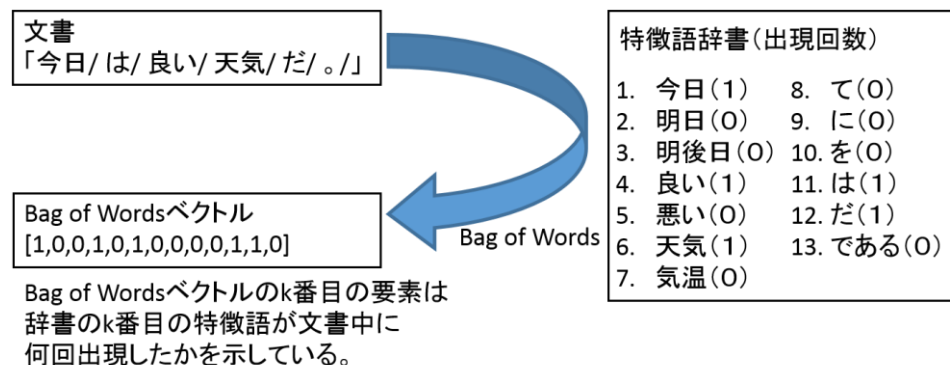


図 3.3. Bag of Words の例

3.2.1 節で作成した辞書を特徴語辞書として、全ての台詞を Bag of Words ベクトルに変換する。これによって、台詞内での単語の発生順序の情報や係り受け

の情報、辞書にない単語の情報は失われるものの、自然言語であった台詞を数値的に扱うことが可能になる。なお、Bag of Words ベクトルの次元数は辞書に登録されている単語数に等しい。

3.2.3. TF-IDF

ここまでの処理で台詞を数値のベクトルとして扱うことができるようになった。本節では、更に単語が文書内でどの程度重要であることを示す TF-IDF 重みを付与することで、学習時、重要でない単語と大袈裟なロボット動作が強く対応付けて学習されることを防ぐ。

TF-IDF 重みは、情報検索や文章要約などの分野で利用される文書中の単語に関する重みの一種である。TF-IDF は TF (Term Frequency : 単語の出現頻度) と IDF (Inverse Document Frequency : 逆文書頻度) の 2 つの指標に基づいて計算される。各指標と TF-IDF の計算方法は以下の通り。

$$TF-IDF = TF * IDF$$

$$TF_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}}$$

$$IDF_i = \log_2 \frac{|D|}{|\{d: d \ni t_i\}|}$$

ただし、 $n_{i,j}$ は単語 t_i の文書 d_j における出現回数、 $|D|$ は総文書数、 $|\{d: d \ni t_i\}|$ は単語 t_i を含む文書数である。

TF は 1 つの文書内で何度も出現する単語が重要であることを表し、IDF は複数の文書にわたって出現する一般語が重要でないことを表している。そのため、2 つの指標をかけあわせた TF-IDF は文書内である単語が持つ重要性を表す指標となる。TF-IDF 重みを学習に利用することで、一般語の重要度が下がり、一般語に大袈裟な動作が強く関連付いて学習されるのを避けることができる。

本研究では、台詞 1 つを 1 文書と考え、収集した全ての台詞に対して TF-IDF を計算し、重み付けを行った。前処理前後の台詞の一例を図 3.4 に示した。

以上の前処理は、生成・認識過程で新たに入力された台詞にも同様に行う。

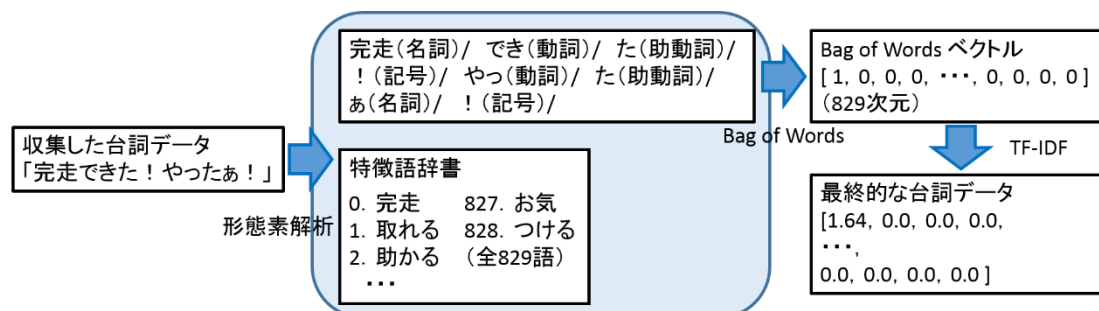


図 3.4. 台詞の前処理の一例

3.3.動作の前処理

次に動作に関する前処理を行う。収集した動作は、ロボットの関節角度データであり、今回は 14 自由度のロボットを使用しているため、14 次元のベクトルとなる。関節角度データは、そのままでも学習データとして利用可能だが、動作生成時の処理でサンプリングなどの操作を行うことを考慮すると、動作を確率的な値として扱える方が都合が良い。そこで DPGMM を用いて、一度動作を類似した動作群にカテゴリ分けし、各動作を各カテゴリの混合比の形で表現するように前処理を行った。

3.3.1. Dirichlet Process Gaussian Mixture Model

Dirichlet Process Gaussian Mixture Model (以下、DPGMM と表記) は、学習時にディクレ過程を利用し、混合比とカテゴリ数を同時に推測することによって、カテゴリ数のわからないデータ群を混合ガウスモデルで分類するノンパラメトリックベイズ法的一种である。ただし、混合比とは、あるデータに対して、そのデータが各カテゴリから生成された確率を全カテゴリ分計算し、計算の結果得られた生成確率を比として表したものである。

動作の分類に DPGMM を選択した理由は、収集したロボット動作の分布が、特徴的なポーズを中心としたある程度広がりを持ったポーズ群の集合、すなわち混合ガウスモデルになっているだろうと予測したためである。DPGMM のグラフィカルモデルを図 3.4 に示す。

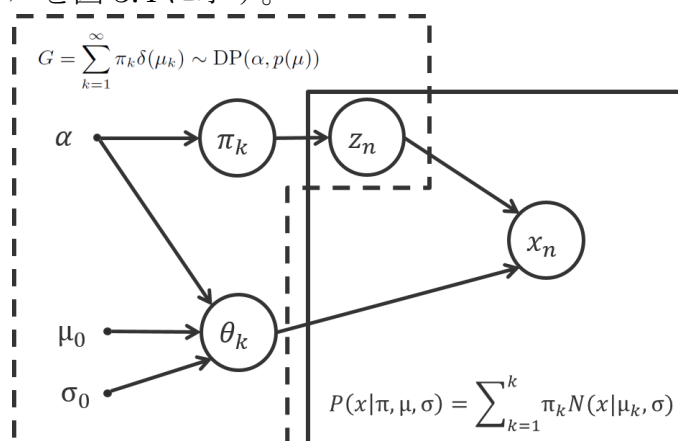


図 3.4. DPGMM のグラフィカルモデル

ただし、図中の x_n は動作、 z_n は動作 x_n のカテゴリ、 π_k はカテゴリ z_k が選ばれる確率で、このパラメータはハイパーパラメータ α により決まるディクレ過程に従う。また、 θ_k (μ_k) はカテゴリ z_k のパラメータで、 μ_0 および σ_0 をパラメータとする θ_k (μ_k) のガウス事前分布に従う。

本手法で利用した DPGMM は、ギブスサンプリングと Chinese Restaurant

Process（中華料理店過程）で実現しているため、動作 x_n が決定しているときのカテゴリ z_n が k の確率、すなわち事後確率はベイズの定理により式（3.1）で求めることができる。

$$p(z_n = k | x_1 \dots x_{n-1}, z_1 \dots z_{n-1}) \propto \begin{cases} p(x_n | k) \frac{n_k}{\alpha + n - 1} & (k = 1 \dots K) \\ p(x_n | k^{new}) \frac{\alpha}{\alpha + n - 1} & (k = K + 1) \end{cases} \dots (3.1)$$

ただし、 n_k は z_1, \dots, z_{n-1} でカテゴリ k の出現回数、 K はその時点までのカテゴリ数である。

式（3.1）から各動作がどのカテゴリから生成されたのかを表す混合比を求められる。これを用いて収集した全動作を、混合比で表現した。

3.3.2.動作を前処理した結果

動作を実際に前処理すると類似動作毎にカテゴリが形成される。前処理後の各カテゴリのイメージ図を図 3.5 に示す。

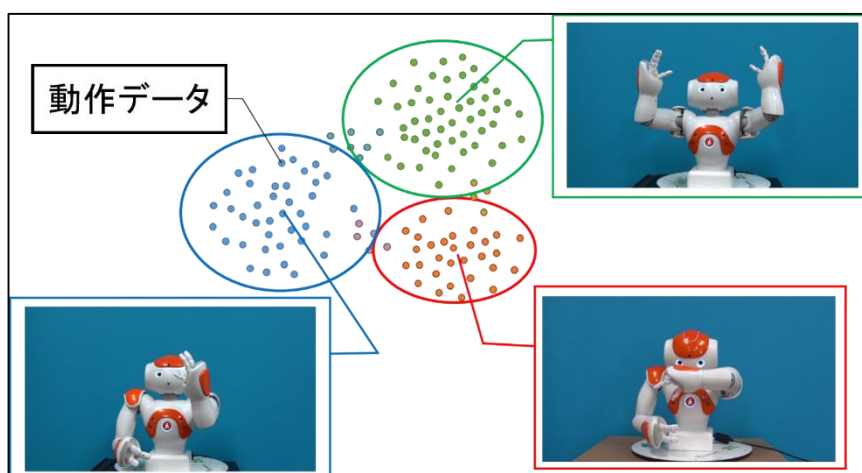


図 3.5. GMM によるカテゴリ分けのイメージ図

また、前述のとおり、動作は 14 次元の関節角度ではなく、カテゴリの混合比で表されており、その次元数はカテゴリ数に等しい。

関節角度データから混合比への変換の一例を以下に示す。

- 関節角度データから混合比への変換

- ・ 関節角度データ

[-0.0215, -0.195, -0.772, 0.661, -1.12, -1.44, -1.46, 0.688, -0.735, -0.769, 1.14, 1.43, 1.33, 0.680] (14 次元)

↓ DPGMM

- ・ 混合比

[9.25E-54, 4.46E-19, 6.30E-31, 1.00E+00, ..., 2.23E-08, 3.81E-08, 3.94E-08, 1.99E-08] (41 次元)

以上で、動作の前処理を終了した。これによって、動作生成の過程での、サンプリングなどの操作が、容易に行えるようになる。

3.4.機械学習

学習過程の最後に、前処理をした台詞と動作を学習器で処理し、台詞と動作の対応を学習する。本手法では学習器として MLDA を利用した。

3.4.1. Multimodal Latent Dirichlet Allocation

本研究では学習器として、中村らが提案した Multimodal Latent Dirichlet Allocation (以下、MLDA と表記) [9]を利用した。MLDA は、文書のトピック分析によく用いられる Latent Dirichlet Allocation (以下、LDA と表記) [10]を、マルチモーダル情報の分類へと拡張した学習器である。LDA と MLDA のグラフィカルモデルを図 3.6 に示す。

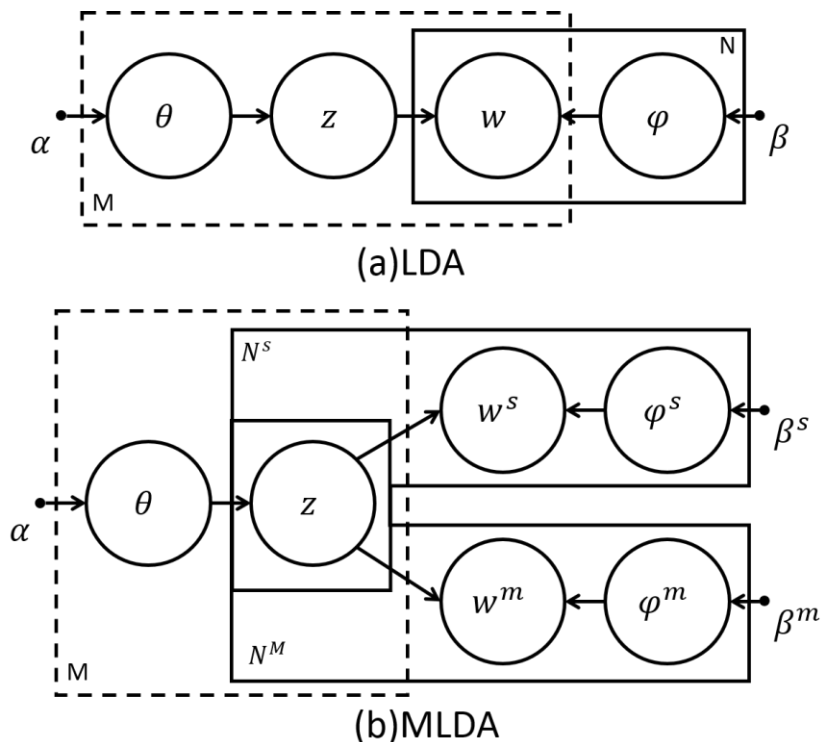


図 3.6. LDA と MLDA のグラフィカルモデル

ただし、図中の z はカテゴリ、 θ はカテゴリ z の出現確率分布を表す多項分布のパラメータで、ハイパーパラメータ α により決まるディリクレ事前分布に従う。また、 w^s は台詞、 w^m は動作、 φ^s 、 φ^m はそれぞれ w^s 、 w^m の出現確率分布を表す多項分布のパラメータで、ハイパーパラメータ β^s 、 β^m に従う。

図を見るとわかるとおり、MLDA では LDA と違い、複数種類のデータを同時に関連付けて、学習することができる。また、MLDA の利点として、複数種

類のデータのうち、一種類のデータを得ることができれば、得られたデータから未観測の他の種類データを推測することが可能であることが挙げられる。 w^s (台詞) から w^m (動作) を推定する数式を式 (3.2) に示した。

$$p(w^m|w^s) = \int \sum_z p(w^m|z)p(z|\theta)p(\theta|w^s)d\theta \dots (3.2)$$

MLDA の複数種類のデータを同時に学習できる点、また全種類のデータが得られなかったときに、未観測の他の種類データを推定することが可能である点を利用することで、本手法の特徴である台詞データと動作データを関連付けての学習、また台詞データまたは動作データしか得られなかったときの他方のデータの推定・生成が可能となると考え、本手法では MLDA を利用した。

なお、MLDA の学習、分類結果のカテゴリ数は自動決定せず、経験的に手動で決定した。HDP-MLDA に拡張することでカテゴリ数を自動的に決定することが可能ではあるが、自動決定により決定したカテゴリ数は手動で決定した最適なカテゴリ数より大きくなる傾向がある。カテゴリ数が大きくなりすぎることは、各カテゴリに感情や意図を見出すという本研究の方針からして、あまり好ましくない。そのため、本手法では MLDA のカテゴリ数を手動で決定することとし、それで動作生成に成功し、いくつかのカテゴリに適切な感情や意図を見出すことができた場合は将来的にカテゴリ数を自動化したいと考えている。

3.4.2.MLDA による学習データの分類結果

MLDA によって、学習データを分類したイメージ図を図 3.7 に示した。

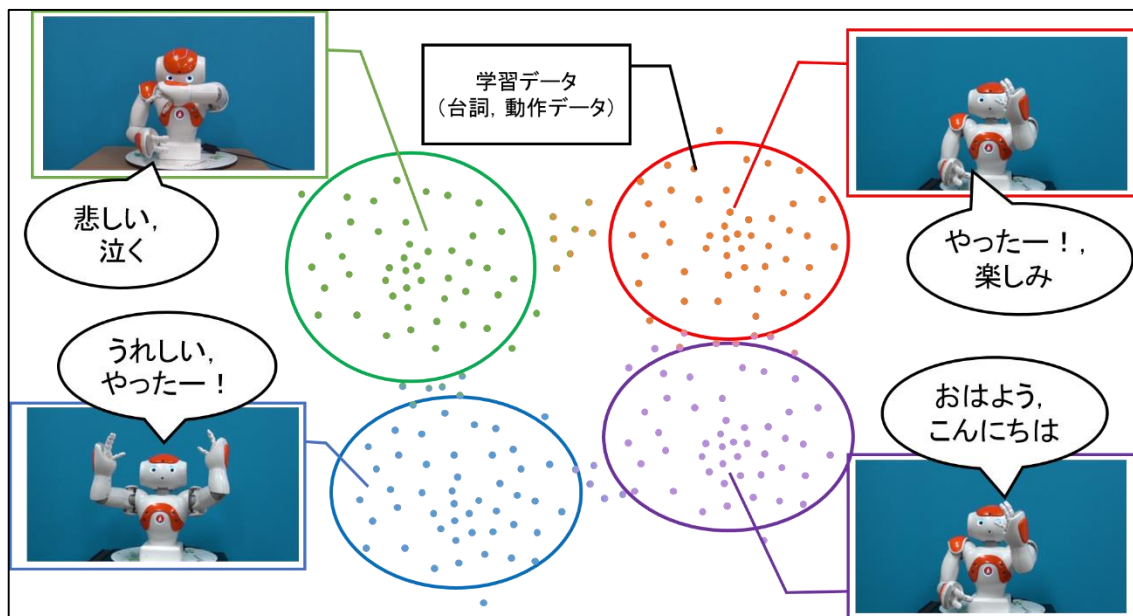


図 3.7. 学習データの学習・分類後のイメージ図

図のように台詞と動作を対応づけて学習しておき、新たなデータ (台詞・動作) が入力されたときにデータの認識と推定を行なうために利用する。

本手法では、MLDA で分類したカテゴリのうち、いくつかは感情を強く表すカテゴリになると考えている。例えば、図の左上のカテゴリは泣いているような動作と悲しい台詞が含んでいるので悲しみを表しており、左下のカテゴリはバンザイの動作と喜んでいるような台詞が含んでいるので喜びを表しているといえる。ここで付けた悲しみや喜びのラベルを、ロボット自身の感情とみなすこともできると考えている。また、図では同じ動作でも台詞が違うものや、同じ台詞でも動作が違うものを別のカテゴリに分類できていることがわかる。これによって、例えば、驚きで両手を挙げているのか、喜びでバンザイしているのか、降参の意味でお手上げしているのかを、それぞれ別のカテゴリとして分類することも可能である。

以上の工程で、ロボットは台詞と動作の学習ができた。次章以降では、本章で学習した台詞と動作の関係を利用して、新たな入力データ（台詞・動作）のカテゴリ認識や、その入力データに合った出力データ（動作・台詞）の生成を行う。

第4章 認識・生成過程

本章では、ロボット動作自動生成システムのうち、学習したデータを基に新しい入力データに合った出力データを生成する認識・生成過程について述べる。

4.1. 台詞から動作の生成

本節では、新たな入力データとして台詞が入力された時の動作の生成手法について述べる。

4.1.1. 入力データの認識

新たに台詞が入力されたとき、その入力データがどのカテゴリに属しているかを認識する。認識のために、前章で学習したデータと MLDA を用いた。

新たに入力された台詞を MLDA でカテゴリ認識するために、まず入力された台詞に対して、学習時に台詞にしたものと同様の前処理（形態素解析、Bag of Words ベクトル化、TF-IDF 重みづけ）を施して、自然言語を数値データに変換する。次に、MLDA によるカテゴリ認識を行う。MLDA を利用すると台詞、動作の両方の種類のデータがそろっていない場合でも片方の種類のデータから、データが属するカテゴリや、他方の種類のデータを推定することができる。台詞データ w^s からカテゴリ z を推定する式 (4.1) を示した。

$$p(z|w^s) = \int p(z|\theta)p(\theta|w^s)d\theta \cdots (4.1)$$

式 (4.1) に従って、新たに入力された台詞が各カテゴリに含まれる確率を求める。新たな台詞に対するカテゴリ認識のイメージ図を図 4.1 に示す。入力されたデータの学習データ空間での位置が図のとおりであり、MLDA のカテゴリ数が 4 であった場合に得られるカテゴリ認識結果の例を以下に示す。

● カテゴリ認識結果の例

- ・ 台詞データ（自然言語）

今日は休みだ！うれしいな！！

↓ 前処理

- ・ 台詞データ（数値）

[0.0, 0.0, 0.0, ..., 1.18, ..., 0.0, 0.0, 0.0] (829 次元)

↓ MLDA

- ・ カテゴリ認識結果（入力データが各カテゴリに含まれる確率）

[カテゴリ 1 に含まれる確率, ..., カテゴリ 4 に含まれる確率]

= [0.2, 0.0, 0.7, 0.1] (4 次元)

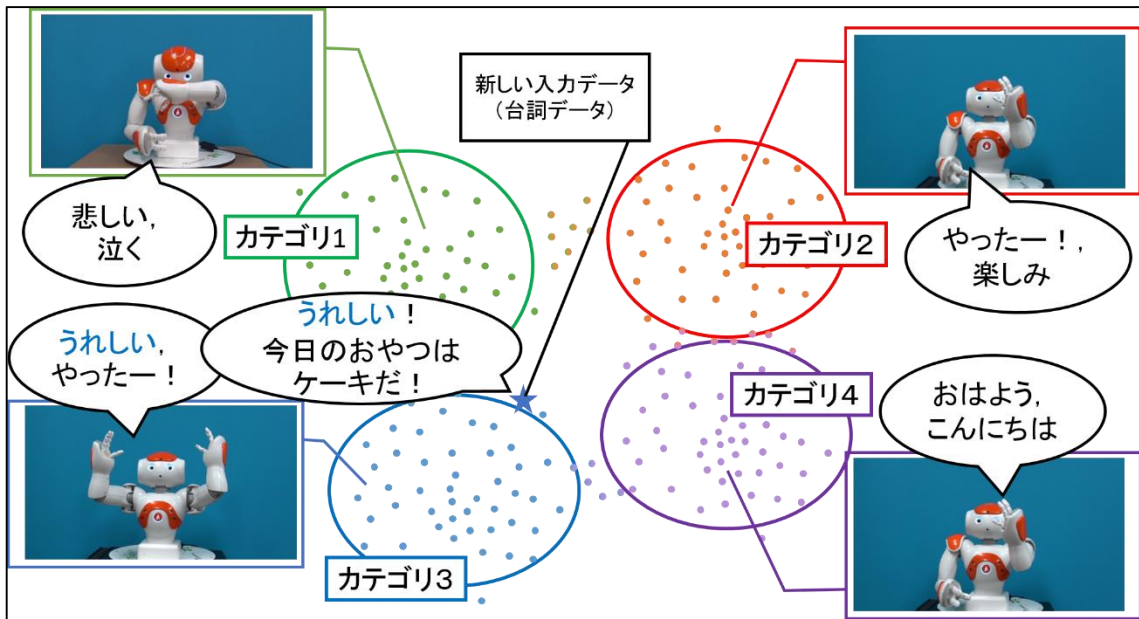


図 4.1. 入力データのカテゴリ認識のイメージ図

図 4.1 を見るとわかるように、新たに入力された台詞はカテゴリ 3 に近い特徴を持っている。そのため、カテゴリ認識結果の例で示したに、入力データはカテゴリ 3 に含まれる確率が高いと考えられる。

ここで本手法の特徴として、台詞を **Bag of Words** で処理しているため、新たに入力された台詞データに未知の単語が含まれていても、台詞中の他の単語からカテゴリの認識をすることが可能なため、結果としてある程度台詞に合った動作をつけられることが挙げられる。

例えば、新たに入力された台詞が「お父さんが出張でいなくて寂しい」というデータで、学習した台詞の中に「寂しい」という単語が全く含まれていなかったとしても、学習したデータ内に「お父さん・いない」「出張・いない」というような単語の共起とそれに合った動作が含まれていれば、新たに入力された台詞に合った動作をつけることが可能である。

4.1.2.動作生成プロセス

動作生成プロセスの処理の流れを図 4.2 に示した。

説明のため MLDA で分類した結果できたカテゴリを言動カテゴリ、GMM で分類した結果できた類似動作のカテゴリを動作カテゴリと呼ぶことにする。

まず、各言動カテゴリについて、それぞれどの動作カテゴリが生成されやすいかという確率を求める。この確率は、図 3.6 の φ^m に当たるものであり、動作について DPGMM の前処理を行ったことで DPGMM の混合比と等価な形式になっている。前処理をしないまま学習すると今後の生成の処理を行うことができない。この確率を基にサンプリングを行い、各言動カテゴリの代表となる動作

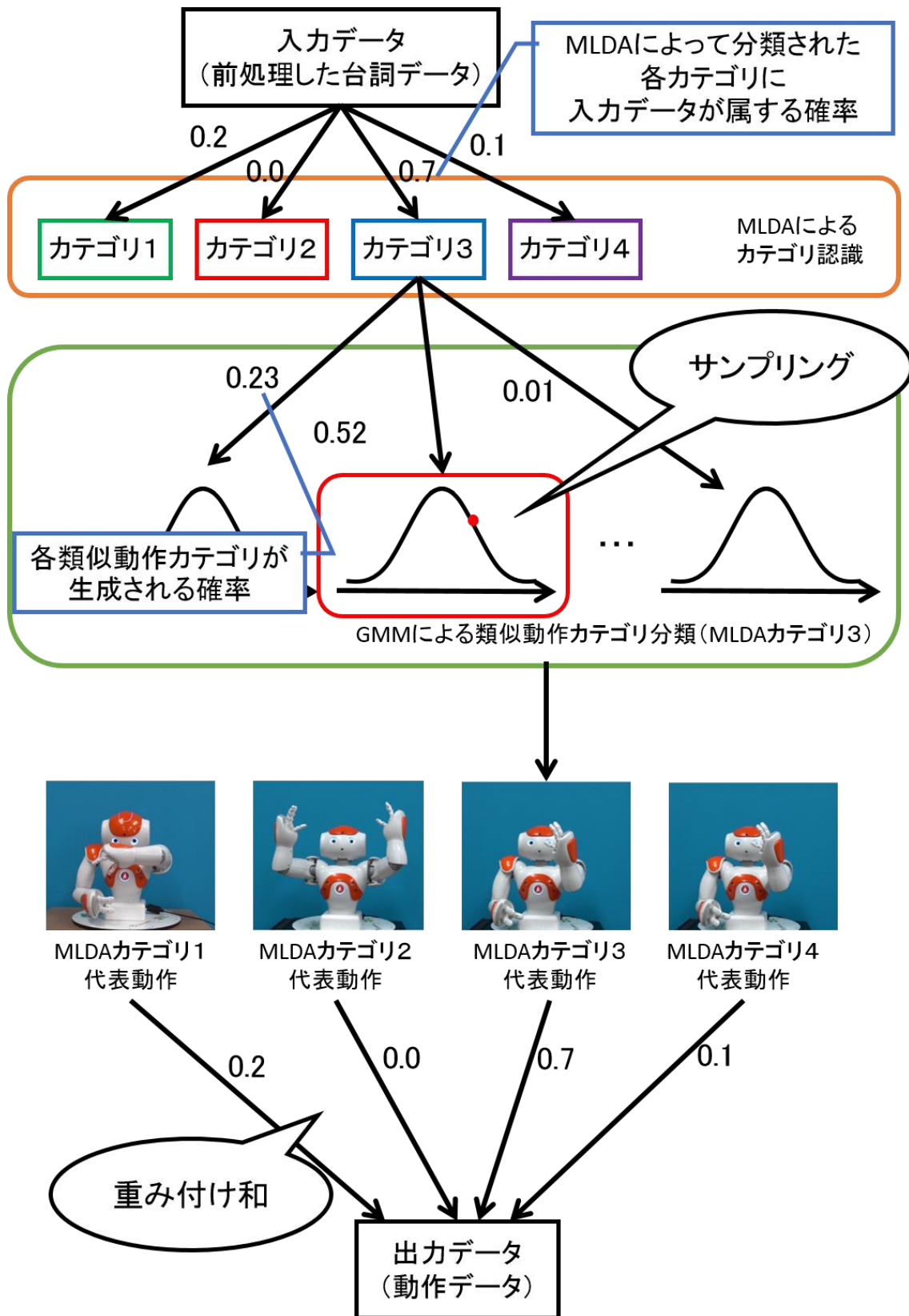


図 4.2.動作生成の流れ (一例)

カテゴリを決定する。1つの動作カテゴリは、14次元のガウス分布となっており、そのパラメータも学習時に求められている。そこで、代表となる動作カテゴリから、更にサンプリングを行って1つの代表動作を決定する。これによって、1つの言動カテゴリにつき、1つの代表動作が決定される。この時点で代表動作は、14次元のベクトルで表されるロボットの関節角度データである。

次に、4.1.1項で求めたカテゴリ分類の結果を利用して、各言動カテゴリの代表動作に重み付けをする。最後に、重み付けした各言動カテゴリの和をとったものを生成動作とする。

動作生成のプロセスの中にランダム要素であるサンプリングが含まれているため、同じ台詞を入力しても出力される動作はランダム性を持ったものとなる。

例えば、喜びの感情を含んだ台詞が入力されたときは、片手を挙げる動作カテゴリや両手を挙げる動作カテゴリが同じような確率でサンプリングされるため、両手挙げや片手挙げのポーズがランダムに生成される。このとき、うつむきの動作カテゴリがサンプリングで選択される確率も存在するが、学習したデータを基にしているため、その確率は低くなり、結果的にうつむきのポーズが生成される確率は低くなる。

つまり、本手法を利用することで、台詞に合ったポーズの中でランダムに動作を発生させることが可能となる。これによって、ロボットが発話の度に同じような動作をすることを防ぐことができる。

4.2.動作から台詞の認識・生成

本節では、新たな入力データとして動作が入力された時の台詞（関連単語）の生成手法について述べる。本手法では、主たる機能として台詞からロボット動作の自動生成を挙げている。しかし、MLDAを利用すると台詞から動作と同様に、動作から台詞の生成が可能である。

4.2.1. 入力データの認識

新たに動作が入力されたとき、その入力データがどのカテゴリに属しているかを認識する。4.1節同様、認識にはMLDAを用いた。

新たな入力データをMLDAでカテゴリ認識するために、まず入力された新しい動作に対して、学習時に動作にしたものと同様の前処理(DPGMM)をして、関節角度データをGMMの混合比で表現した。次に、MLDAによるカテゴリ認識を行う。動作データ w^m からカテゴリ z を推定する式(4.2)を示した。

$$p(z|w^m) = \int p(z|\theta)p(\theta|w^m)d\theta \cdots (4.2)$$

式(4.2)に従って、新たに入力された動作が各カテゴリに含まれる確率を求める。新たな動作データが含まれるカテゴリを推定したイメージ図を図4.1に

示す。入力された動作の学習データ空間での位置が図のとおりであり、MLDAのカテゴリ数が4であった場合に得られるカテゴリ認識結果の例を以下に示す。

- カテゴリ認識の結果の例
 - ・ 動作データ（関節角度データ）
 $[-0.0215, -0.195, -0.772, 0.661, -1.12, -1.44, -1.46, 0.688, -0.735, -0.769, 1.14, 1.43, 1.33, 0.680]$ （14次元）
 ↓ 前処理
 - ・ 動作データ（混合比）
 $[9.25e-54, 4.46e-19, 6.30e-31, \dots, 3.81e-08, 3.94e-08, 1.99e-08]$ （41次元）
 ↓ MLDA
 - ・ カテゴリ認識結果（入力データが各カテゴリに含まれる確率）
 $[\text{カテゴリ 1 に含まれる確率}, \dots, \text{カテゴリ 4 に含まれる確率}]$
 $= [0.1, 0.0, 0.8, 0.1]$ （4次元）

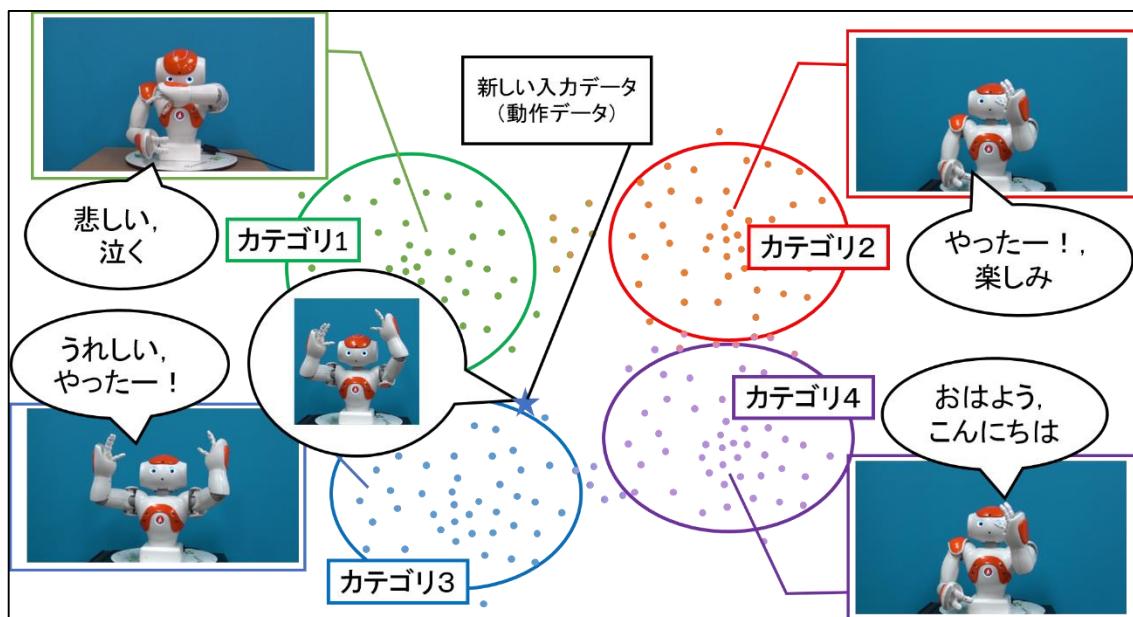


図 4.2. 入力データのカテゴリ認識のイメージ図

図を見るとわかるように、新たに入力されたデータはカテゴリ 3 に近い特徴を持っている。そのため、カテゴリ認識結果の例にあるように、入力された動作は、カテゴリ 3 に含まれる確率が高いと考えられる。

4.2.2. 台詞（関連単語）生成

関連単語生成プロセスを図 4.3 に示す。生成の手順は動作生成とほぼ同じであるが、関連単語生成をする際は、サンプリングを行わない。これは生成単語を台詞化する機能を実装していない現状、関連単語生成にランダム性は必要ないと考えたためである。

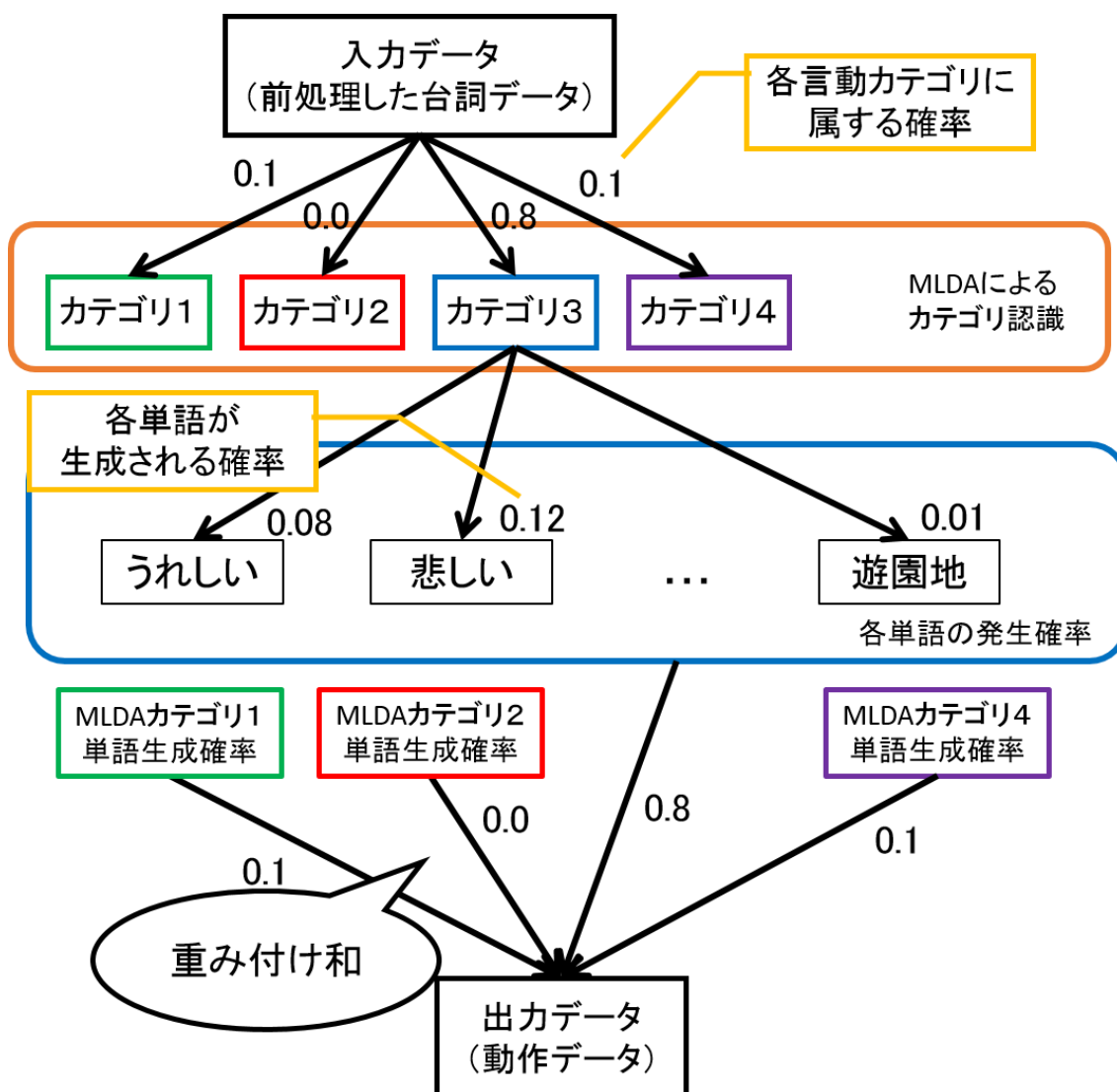


図 4.3. 関連単語生成の流れ (一例)

関連単語生成では、まず言動カテゴリごとに各単語が生成される確率を算出する。次に、各単語の生成される確率に 4.2.1 項で求めたカテゴリ認識を用いて重みづけをして、和を計算する。最終的に求められた各単語の生成確率から、確率の高いものを抽出することで、関連単語の生成が可能となる。

関連単語を台詞として利用するためには、単語の前後関係や係り受けの関係などを考える必要がある。しかし、本研究では主に台詞から動作の生成を主軸としているため、現状、生成単語の台詞化の機能は実装していない。

第5章 準備実験

本章では、本手法の動作実験を兼ねて、本手法の特徴である知らない単語の含まれた台詞への対応、同じ台詞に対する多様性のある動作の生成、台詞から動作、動作から台詞の相互推定が実現されているかを試験する。

5.1.台詞・動作データセットの学習

学習用の台詞として中学生レベルの英語教材の会話シーンを和訳したものと学習用に作成した強い感情を含んだ台詞を合わせて 896 文収集し、動作として台詞に沿ったポーズを手動で作成して、機械学習を行った。学習時 MLDA のカテゴリ数は、経験的に 8 とした。

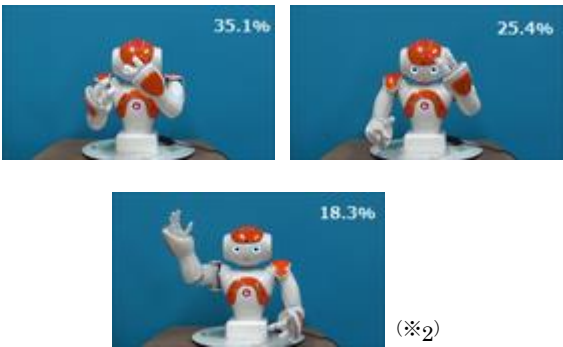

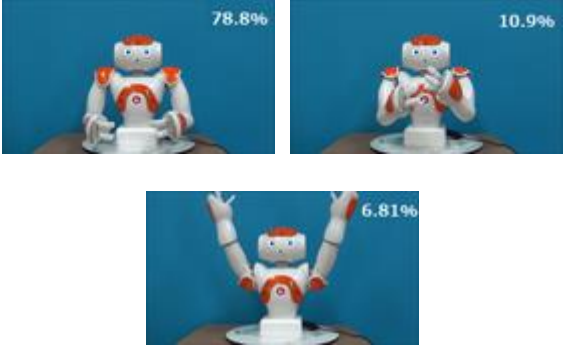
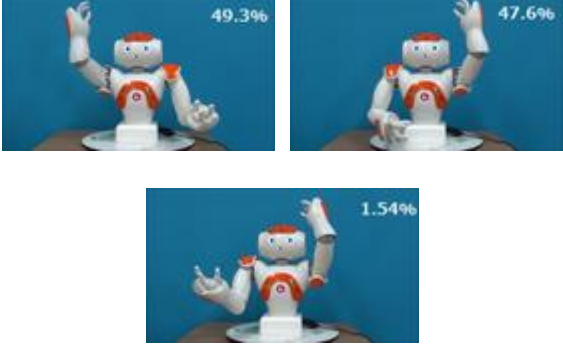
MLDA による分類・学習の結果、各言動カテゴリに分類された動作カテゴリと単語のうち、代表的なもの（動作カテゴリは生成確率上位 3 つ、単語は生成確率上位 5 単語）を表 5.1 に示した。表 5.1 を見ると、カテゴリ番号 0 は、うつむきのポーズとネガティブな単語（悲しい、泣く、泣ける）のセットで悲しいような感情を表しており、カテゴリ番号 3、4 は両手、または片手を挙げるポーズとポジティブな単語（やる＋た＝やったー、楽しい、遊園（地）、うれしい）のセットで嬉しいような感情を表していると考えられる。また、カテゴリ番号 2 では、平常状態からほぼ動かないニュートラルに近いポーズと、強い意味を持たない単語がセットになっており、曖昧な台詞に大袈裟な動作がつくことを防ぐように働いている。カテゴリ番号 6 は、片手を挙げるポーズと挨拶の台詞のセットで、相手に呼びかけながら挨拶をするような組み合わせになっている。カテゴリ番号 1、5、7 に関しては、ポーズと単語の関係に対して適当な感情などを付けづらいが、そもそも動作と台詞のセットを感情ごとに分類することが目的の研究ではないため、適当な感情ラベルを付けられないカテゴリが存在すること自体は問題ない。むしろ、今後学習データを増加させた場合、MLDA のカテゴリ自体も増加し、結果として感情ラベルを付けられるようなカテゴリの方が少なくなると考えられる。

ここで学習したデータを本章の本節以降の実験、および次章の主観評価実験用の動作生成に用いた。

5.2.未学習の単語を含む台詞に対する動作の生成

本節では、未学習の単語を含んだ台詞に対する動作生成が可能であるという本手法の特徴の一つについて、実際に未学習の単語を含んだ台詞に対する動作生成を行い、適当な動作が付けられているかを確認する。

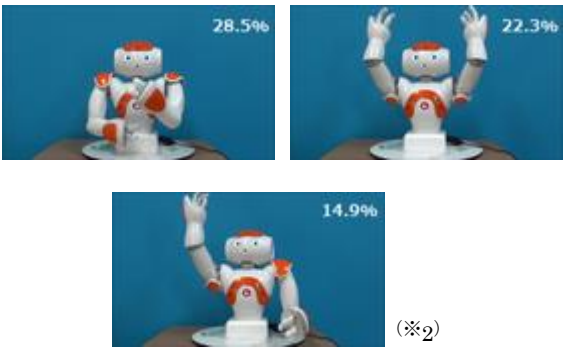
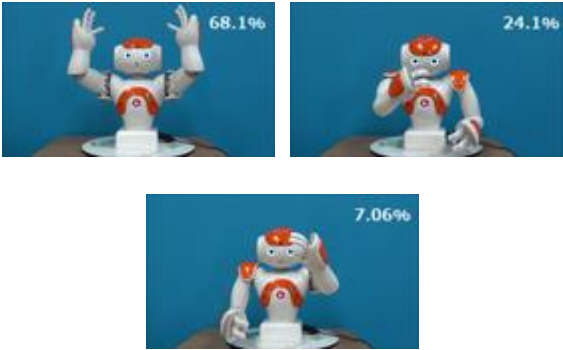
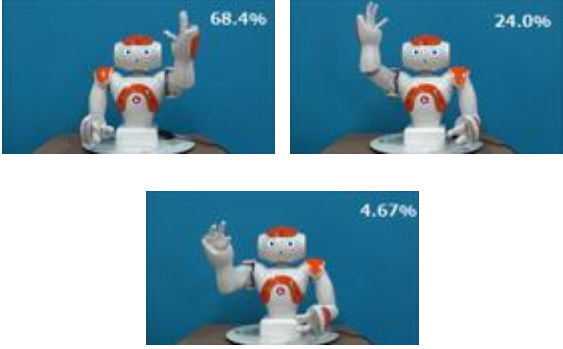
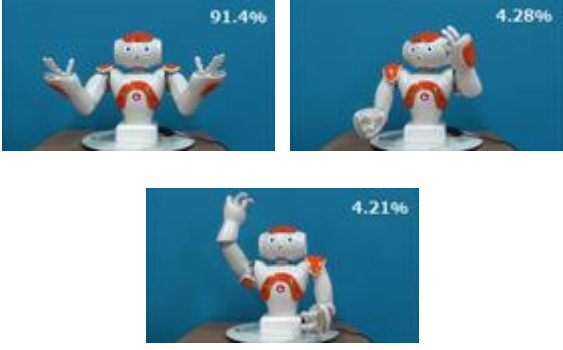
表 5.1. (a) 各カテゴリに分類された動作と台詞 (カテゴリ No.0-3)

No.	動作カテゴリの生成確率 ^(※1)	単語の生成確率
0	 <p>(※2)</p>	悲しい (6.37%) 泣く (5.16%) た (4.78%) ありがとう (3.92%) 泣ける (3.59%)
1		いい (6.85%) 行く (4.42%) 私 (4.09%) 人 (3.12%) まかせる (2.74%)
2		そう (5.18%) くらい (3.63%) だ (3.02%) いる (2.94%) ん (2.29%)
3		やる (13.1%) た (9.58%) うれしい (6.39%) だ (6.19%) もらう (2.7%)

※1 : 表中、図は動作カテゴリ中の平均動作

※2 : 図中、右肩の数字は動作カテゴリの生成確率

表 5.1. (b) 各カテゴリに分類された動作と台詞 (カテゴリ No.4-7)

No.	動作カテゴリの生成確率 ^(※1)	単語の生成確率
4	 <p>(※2)</p>	楽しい (7.6%) 遊園 (5.15%) うれしい (5.13%) 試験 (3.32%) どう (3.25%)
5		ある (5.07%) そうですね (4.36%) ない (3.35%) 悲しい (3.13%) すごい (2.92%)
6		言ってらっしゃい (7.43%) おはよう (7.33%) です (5.87%) 良い (5.53%) ます (4.49%)
7		マイク (3.62%) だ (3.54%) うん (3.1%) する (2.6%) ん (2.39%)

※1 : 表中、図は動作カテゴリ中の平均動作

※2 : 図中、右肩の数字は動作カテゴリの生成確率

実験に用いた台詞は、表 5.2 のとおりである。どの台詞も未学習の単語を含んでおり、未学習の単語が台詞中で比較的重要な意味を持っている。また、動作生成した結果を図 5.1 に示した。

表 5.2. 実験に用いた台詞

台詞	未学習の単語
お父さんが出張でいない・・・、寂しいなあ！	寂しい
単位を落として卒業できないと知ったときは、ひどく落胆したものさ・・・	落胆
ハッピーバースデー！お誕生日おめでとう！	ハッピー、バースデー
クリスマスは、おいしいケーキは食べられるし、プレゼントに新作のゲームはもらえるし、もうご機嫌だったよ！	(ご) 機嫌 ^(※3)
こんばんは、今お帰りですか？	こんばんは

※3：形態素解析の結果としては、「ご（接頭詞）/ 機嫌（名詞）」で、接頭詞はシステム上辞書に登録しないが、日本語の意味的には「ご機嫌＝上機嫌」で一つの意味になるため、表中の未学習の単語欄は（ご）機嫌とした。

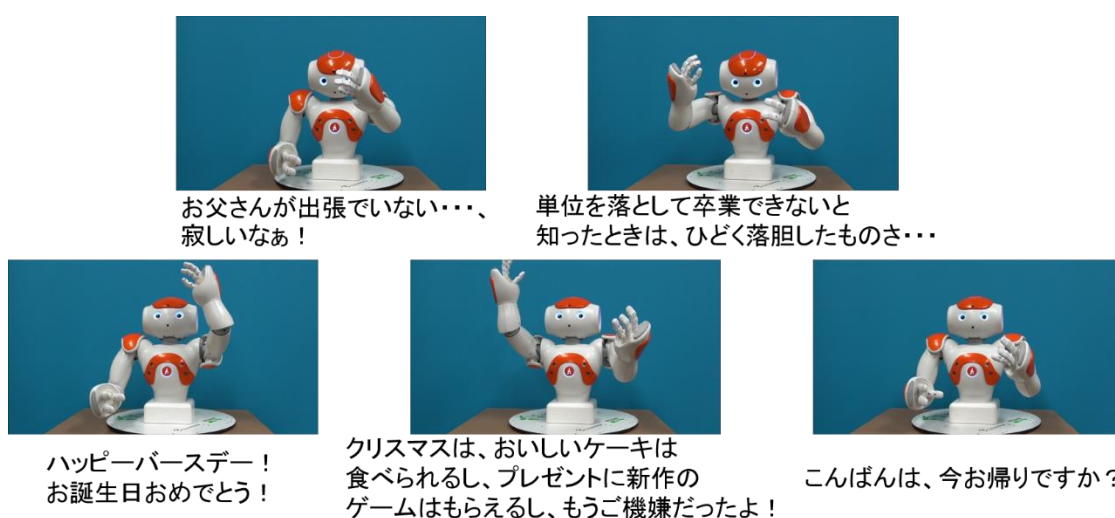


図 5.1. 未学習の単語を含んだ台詞に対する動作生成

図 5.1 を見ると未学習の単語を含む台詞に対しても適当な動作を付けられていることがわかる。「寂しい」や「落胆」に対しては、顔をうつむけたネガティブな動作が生成されているし、「ハッピー/ バースデー」や「(ご) 機嫌」には、腕を挙げたポジティブな動作がついている。これは、「お父さん・出張・いない」を含んだ台詞とネガティブなポーズのセットや、「クリスマス・ケーキ・プレゼント」「ゲーム・新作」を含んだ台詞とポジティブな動作のセットが、学習したデータの中に存在しており、台詞中の未学習の単語以外の単語から適当なポーズ

ズの推定がなされているためである。

また、図 5.1 を見ると「こんばんは、今お帰りですか？」という台詞には大きな動作がついていないことがわかる。これは台詞中に未学習の単語以外の単語が少なく、台詞を挨拶の言動カテゴリ（カテゴリ 6）に近いものだと確信できなかったためである。実際、動作生成時のカテゴリ認識の結果を詳しく見てみると、挨拶の言動カテゴリ（カテゴリ 6）に属する確率が、他の言動カテゴリの生成確率と比べて最も大きく 46.8%、大袈裟なポーズをとらない言動カテゴリ（カテゴリ 2）に属する確率が二番目に大きく 24.5%となっていた。この生成結果は、台詞に沿った動作が生成できていないと言えるが、別の側面から見ればカテゴリ分類に確信の持てなかった台詞に対して、当たり障りのない動作をすることで、違和感のある大袈裟な動作を付けてしまうことを防げたとも言える。そのため、今回は大袈裟な動作を生成しなかった結果についても、動作生成に成功しているとみなした。

5.3.一つの台詞に対する多様性のある動作の生成

本節では、一つの台詞に対して多様性のある動作の生成が可能であるという本手法の特徴について、実際に一つの台詞に対して複数回の動作生成を行い、多様な動作が付けられているかを確認する。

動作生成用の台詞は、以下のとおりである。

- 動作生成用の台詞
 - ・ 今日は、みんなで遊園地だ！楽しいな！

動作生成は同台詞に対して 15 回行った。その結果を図 5.2 に示す。

動作生成の結果を見ると、確かにほぼ全ての動作で両手、または片手を挙げて楽しいようなポジティブなポーズを生成できていることがわかる。動作生成時の認識結果を詳しく見ると、生成確率の最上位の言動カテゴリは毎回同じで、ポジティブな台詞と動作のセットであるカテゴリ 4 であった。毎回、同じ言動カテゴリが選ばれているにも関わらず、生成された動作にランダム性があるのは、動作生成の過程で確率的な処理であるサンプリングを行い、代表動作の決定をしているためである。

また、8 番目の動作について、あまり大きな動作をとることができていないが、これは提案手法の問題というよりは、学習時の経験的に調整した MLDA のカテゴリ数によるものと考えられる。うれしいことを表す言動カテゴリから生成される動作カテゴリに、大袈裟なポーズをとらないような動作カテゴリが含まれてしまっているため、このような動作が生成される。解決法としては、カテゴリ

数を調整し、最適なカテゴリ数を目指すことが考えられるが、5.2節の「こんばんは、・・・」の台詞と同様、大袈裟なポーズをとらないという結果は、動作の生成に成功しているとみなしたため、今回はこの学習結果からカテゴリ数を変更しないことにした。

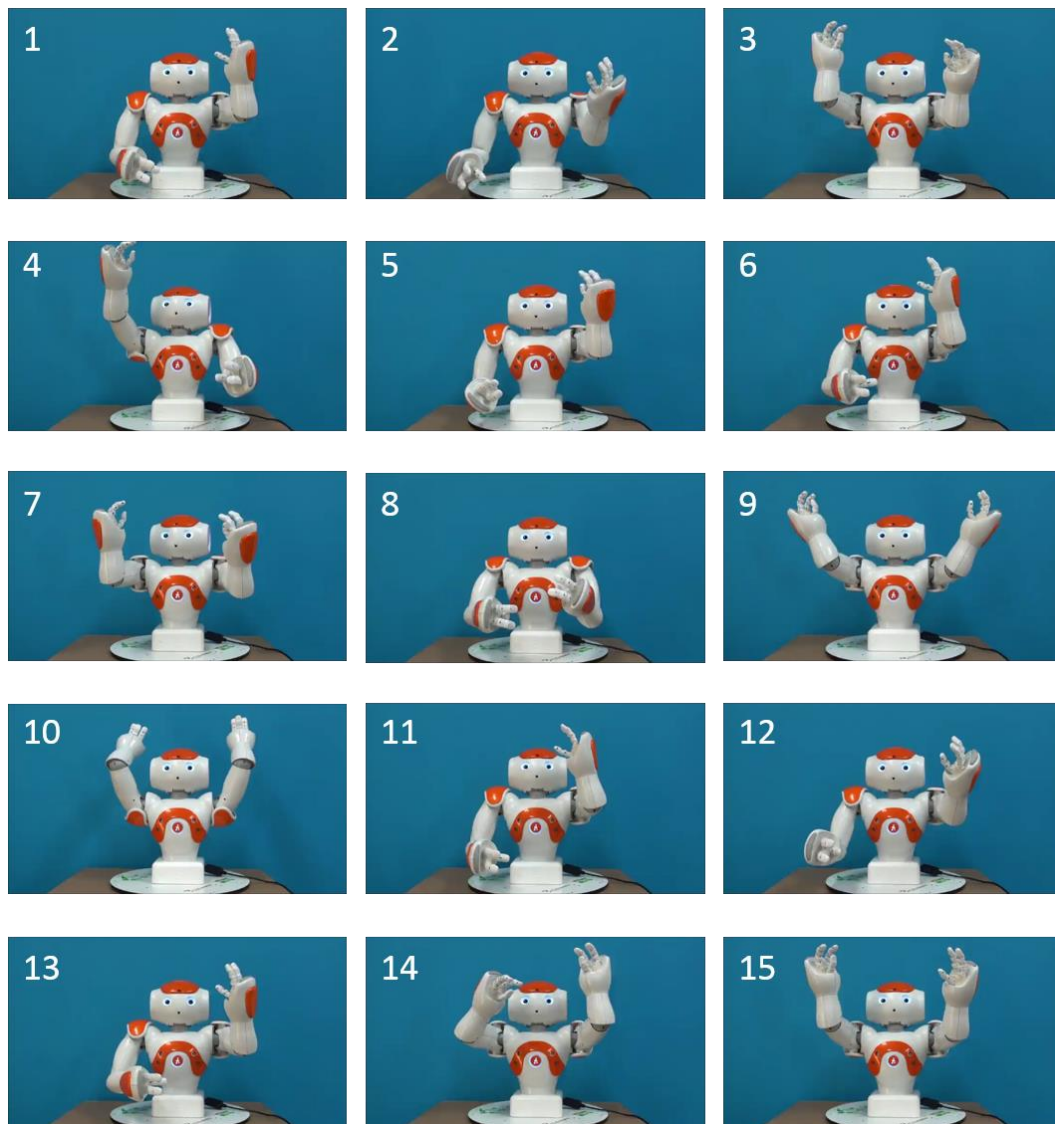







図 5.2. バリエーションのついた動作

5.4.動作から台詞（関連単語）の生成

前節までは台詞から動作の生成ばかりを取り上げてきた。本節では動作から台詞（関連単語）の生成を実際に行い、その結果について考察する。

まず、台詞生成用の動作を用意する。用意した動作は両手挙げ、片手挙げ、顔を覆ってうつむく、手をおろしてうつむく、ニュートラルポーズの 5 種類である。5 種類の動作と、生成された台詞（関連単語）を表 5.3 に示した。

表 5.3. 台詞（関連単語）生成の結果

No.	台詞生成の元になった動作	生成された単語（生成確率）
0		行ってらっしゃい (6.37%) おはよう (5.16%) です (4.78%) 良い (3.92%) ます (3.59%)
1		楽しい (6.85%) うれしい (4.42%) 遊園 (4.09%) た (3.12%) 試験 (2.74%)
2		悲しい (5.18%) た (3.63%) 泣く (3.02%) ありがとう (2.94%) 泣ける (2.29%)
3		いい (13.1%) 行く (9.58%) 私 (6.39%) だ (6.19%) 人 (2.7%)
4		いい (7.6%) 行く (5.15%) 私 (5.13%) だ (3.32%) 人 (3.25%)

動作 0 について見ると、片手挙げのポーズに対して、「いってらっしゃい」や「おはよう」などの挨拶に関する単語が生成されており、生成結果としては正しい結果になっている。入力した動作が分類された言動カテゴリも挨拶を含んだカテゴリ 6 であった。また、単語「良い」に関して、ポジティブな台詞にポジティブなポーズがついているという意味では間違っていない。とはいえ、「です」、「ます」などの一般語が含まれてしまっていることを考えると、関連単語生成の精度向上を目指すには、TF-IDF 重みの見直しなどが必要である。

次に動作 1、2 について見ると、動作 1 は両手を挙げるポジティブなポーズに対して、「楽しい」、「うれしい」などのポジティブな単語が生成されていることがわかる。台詞 2 はうつむいて顔を覆うネガティブなポーズに対して、「悲しい」、「泣く」などのネガティブな単語が生成されていることがわかる。それぞれ分類された言動カテゴリも、嬉しいようなデータセットのカテゴリ 4 と、悲しいようなデータセットのカテゴリ 0 となった。生成結果には、動作 0 同様、一般的な単語も含まれてしまっているが、大まかには単語生成していると考えられる。

動作 4 について見ると、あまり特徴的でない単語が生成されていることがわかる。これは、動作 4 がニュートラルなポーズのため、強い感情などを含まない一般語が生成されたためと考えられる。分類された言動カテゴリもカテゴリ 1 で、比較的大げさなポーズを含まないカテゴリであった。

最後に動作 3 について考える。動作 3 はうつむきの動作であるが、分類された言動カテゴリは、悲しさを表したカテゴリ 0 でなく、大袈裟なポーズをとらないカテゴリ 2 であった。これは首の関節の変化量が少なく、首以外の関節角度がニュートラルポーズとほぼ同じだったことから、動作カテゴリの混合比が、動作 3 と動作 4 でほぼ同じ値になってしまったためだと考えられる。

人間の動作を入力とした関連単語の生成も行った。人間の動作をロボットの動作に変換し、関連単語を生成した結果を図 5.3 に示した。人間のとったポーズは両手を挙げて喜んでいる状態とした。人間のポーズの取得には、Microsoft 社の Kinect センサを利用した。



図 5.3. 人間動作を入力とした関連単語生成

生成された単語の内容を見ると、「やる+た (やった)」や「うれしい」が生成されているのがわかる。また、カテゴリも喜びを表すカテゴリ 4 に分類されていた。この生成結果は、本手法によって人間のポーズから人間の感情の簡易的な推定が可能であることを示している。

5.5. 準備実験まとめ

準備実験をとおして、本手法の特徴である未学習の単語を含む台詞への対応、多様性のある動作の生成、動作と台詞（関連単語）の相互推定が可能であることを確認することができた。準備実験をとおしてシステムの動作に問題ないことがわかったので、次章では主観評価実験を行い、提案手法の有用性を評価する。

第6章 主観評価実験

本章では、提案したロボット動作自動生成手法と、他の動作生成手法を比較する主観評価実験を行い、本研究で提案した手法の有効性を評価する。

6.1.実験の概要

6.1.1.実験の流れ

本研究に関する事前知識を一切持たない本学学生 16 名を被験者として、被験者実験を行った。被験者実験の流れを以下に示す。

- 実験の流れ
 - ① 被験者に対して、実験に関する事前説明を行う。
 - ② 例題の実験用動画を見せて、アンケートに慣れてもらう。
 - ③ 実験に関する質問に答えてもらう。
 - ④ 本番の実験用動画を見せ、アンケートに回答してもらう。
 - ⑤ 実験終了を被験者に伝え、アンケートを回収する。

実験は初回 7 名、二回目 9 名で二回に分けて行った。各回で実験環境等に変化がないように注意して実験を実施した。

6.1.2.実験環境

実験は、図 6.1 のような環境で行った。

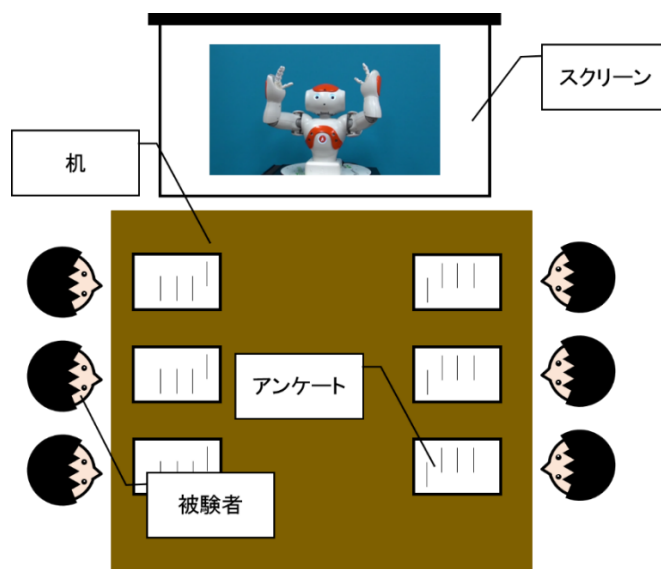


図 6.1. 実験環境

被験者には、スクリーン上に映るロボットの動作を見て机上のアンケートに答えてもらった。実験方法については、実機を直接見てもらうことも考えたが、

ロボット動作および実験環境の再現性、被験者とロボットの位置関係およびロボットを見る角度による評価の差などを考慮すると、スクリーンでロボットの動画を再生した方がより正確な実験ができると判断した。

6.1.3.事前説明

実験開始時に、被験者に事前説明を行った。被験者に行った事前説明を、以下に示す。被験者に与える情報は基本的に、今回の実験はロボット動作の自然さに関するものであること、ロボットが動作している動画を見てアンケートに答えてほしいこと、アンケートの各設問の説明のみとした。

- 事前説明（全文）

今回はロボットの動作の自然さに関する実験に協力して頂きます。

皆さんにはこれからロボットの動作を見て頂きます。ロボットはある台詞を発話しながら動作を行いません。動作は1つの台詞に対して、A、Bの2種類行いますので、皆さんはそれぞれの動作に関して、アンケートに回答して下さい。

アンケートの項目は1つの台詞につき5問あります。①はA、Bどちらの方が動作と台詞が合っているか、②、③はA、Bどちらのロボットの方が親しみやすいか、また人間らしいか、④はA、Bどちらの動作の方がスムーズか、⑤はA、Bどちらの方が感情を正しく表現できているかを聞いていますので、各台詞について判断して下さい。A、Bどちらも同じくらいに良い、ないし悪い場合はどちらも0に○を付けて下さい。

では、動作の例と回答例をお見せします。（例題の動作を見せる。）

以上で、例題は終わりです。ここから実験本番になります。

6.1.4.実験動画

実験に用いた動画は、ロボットが1つの台詞につき、動作A、B、2種類の動作を行う内容である。被験者は台詞と次の台詞の間に、動作に関するアンケートに答えることになるため、台詞間には20秒間の回答時間がもうけてある。動画は例題と本番の2種類用意し、例題では台詞1文分の動作を、本番では台詞15文分の動作を被験者に提示する。

台詞1つに与える2種類の動作A、Bのうち、どちらかは必ず提案手法で生成されたものとし、他方は比較対象の手法で生成されたものとした。2つの動作A、Bを比較することで提案手法と他の手法の相対評価が可能である。ここで、提案手法と比較するためのロボット動作の生成手法として、次の3種類の動作生成手法を用意した。

表 6.1. 比較対象の動作生成手法とそれぞれの特徴

手法の名称	特徴
動作無し	動作しない
ランダム (NAO API)	NAO T-14 用の SDK に実装されている Animated Speech(random)を利用した。腕をランダムに揺らすような動作を実行する。提案手法を含めた全動作生成手法中、唯一ポーズだけではなく、台詞中動き続ける仕様となっている。動作の詳細は後述。
手動で作成した動作	動作を各台詞に合わせて手動で作成した。最も人間の感性に合ったポーズをとる。

比較手法として、ランダムを採用した理由は、動作無しや手動生成といった人間に与える印象の良し悪しが極端な手法以外の、中間的な印象を人間に持たせる動作生成手法とも比較をしたかったためである。また、ランダム手法の生成する「特に台詞の意味とは合っていないが、とりあえず細かく動き続ける」という特徴を持った動作が既存のロボットの発話中の動作として、よく利用されるものだからというのもランダムを比較対象として、選択した理由である。

Animated Speech で生成された動作の一例を図 6.2 に示す。Animated Speech では、基本的にあまり大きな動きは生成されず、左右の手を対照的、または同方向に同時に動かすような挙動をとることが多い。挙動はランダムで、同じ台詞を入力としても、毎回違う適当な動作となる。台詞の内容は、動作に反映されない。

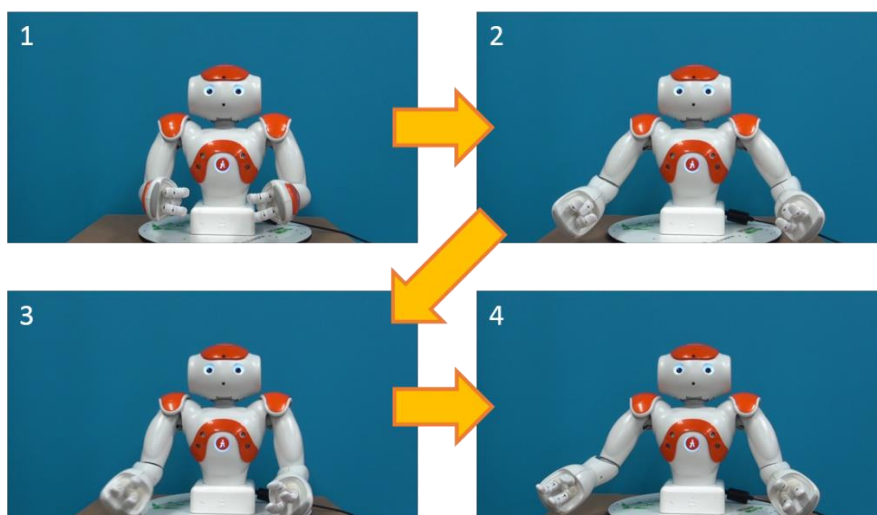


図 6.2. Animated Speech による動作生成の一例

(台詞「お父さんが出張でいない・・・、寂しいなあ！」)

図 6.2 は、台詞「お父さんが出張でいない・・・、寂しいなあ！」を入力として与えたときに出力された動作である。図を見ると、両手を一度広げ、元の状態の戻し、もう一度両手を開くという動作を行っていることがわかるが、台詞と関

係のある動作とは言えない。

実験時に提示した台詞と動作の組み合わせを表 6.2 に示す。表 6.2 中でランダム手法を実行した台詞に関しては、動作の欄を空欄にしている。これは、ランダム手法で生成される動作を一つのポーズとして示すことができないためである。動作生成用に用意した台詞は、当然、学習したデータには含まれていない。また、実験時に被験者には、どの動作がどの生成手法から生成されたものかを伝えていない。そのため、被験者は単純に動作 A、B の見た目から、ロボットの動作を比較することになる。

台詞を提示する順番としては、表す感情の同じものが連続して表れないようにした。また、動作生成手法の割り振りも、連続して同じ生成手法の組み合わせにならないように振り分けている。また、動作 A、動作 B への手法の割り振りも、常に動作 A (動作 B) が提案手法になるような事態を避けるように割り振った。これによって、見せる順番による評価の偏りを、無くすることができる。

6.1.5. アンケートの内容

実際に実験で使用したアンケートの例題部分を示す。例題以外の台詞についてもアンケートで質問する内容は同じものである。

- アンケート中の設問

例題. やぁ！こんばんは！！

① 台詞とポーズは合っていたか？

Aの方が合っている	A, Bに差はない	Bの方が合っている
-----------	-----------	-----------

② 親しみやすいロボットだったか？

Aの方が親しみやすい	A, Bに差はない	Bの方が親しみやすい
------------	-----------	------------

③ 人間味のあるロボットだったか？

Aの方が人間らしい	A, Bに差はない	Bの方が人間らしい
-----------	-----------	-----------

④ 動作はスムーズだったか？

Aの方がスムーズ	A, Bに差はない	Bの方がスムーズ
----------	-----------	----------

⑤ 感情を正しく表現していたか？

Aの方が感情を表現している	A, Bに差はない	Bの方が感情を表現している
---------------	-----------	---------------

表 6.2. (a) 実験動画内のロボットの台詞と動作の詳細 (例題)



No.	台詞	動作 A	動作 B
例題	やあ！こんばんは！！	手動 	手動 

表 6.2. (b) 実験動画内のロボットの台詞と動作の詳細 (台詞 No.1-5)

















No.	台詞	動作 A	動作 B
1	今日は休みだ！ うれしいな！！	手動 	提案手法 
2	おはよう！ 今日は早いんだね！	提案手法 	動作無し 
3	お父さんが 出張でいない・・・、 寂しいなあ！	ランダム 	提案手法 
4	やっぱり、 ホームパーティは楽しい なあ！！	動作無し 	提案手法 
5	筋肉痛で、腕が痛い	提案手法 	手動 

表 6.2. (c) 実験動画内のロボットの台詞と動作の詳細（台詞 No.6-11）

No.	台詞	動作 A	動作 B
6	ケーキ！！ 楽しみです！！	提案手法 	ランダム
7	決勝戦でシュートを 外した・・・	動作無し 	提案手法 
8	いってらっしゃい、 気を付けてね！！	提案手法 	手動 
9	プレゼント！ やったあ！！	提案手法 	ランダム
10	うう、 テストの点が 悪かった・・・、 泣いちゃう・・・。	動作無し 	提案手法 
11	遊園地に連れてって くれるの！ わあい！	手動 	提案手法 

表 6.2. (d) 実験動画内のロボットの台詞と動作の詳細 (台詞 No.12-15)

No.	台詞	動作 A	動作 B
12	今日のおやつ、 しょぼい……、 悲しいなあ！！	提案手法 	ランダム
13	今日のご飯は カレーだ！！	動作無し 	提案手法 
14	兄弟げんかは 悲しいなあ……。	手動 	提案手法 
15	うーん、そうですね、 そう思います。	提案手法 	ランダム

各質問の内容として、質問①はロボットの動作が発話する台詞と合っているかを問うもの、質問②～⑤は、ロボットの動作の変化によって、ロボットの人間に与える印象（親しみ、人間らしさ、スムーズさ、感情表出の度合い）がどう変化するものかを問うものである。なお、質問④に関してロボット動作のスムーズさを問うているが、実際は動作によって故意にスムーズさを変えるような処理はしてない。ただ、前述のとおり、ランダムな動作のみは台詞中動き続けるため、評価が他の動作と変わる可能性がある。

6.2.実験実施前の考察

実験前に立てた実験結果の予想を比較対象の動作生成手法毎に示す。

6.2.1.動作無しとの比較

動作無しとの比較に関しては、5つの質問全てにおいて、提案手法が高評価を得ると考えられる。特に質問②、③に関していえば、人間が完全に静止しながら話すことはまずありえないため、動作無しよりも何らかの動きのついた提案手法に親しみや人間らしさを感じると予想できる。また、質問⑤の感情の表出についても、動作が無くなるような感情を表出したい場合でない限りは、動作のついた本手法の方が高評価を得るはずである。

6.2.2.ランダム手法との比較

ランダム手法との比較の結果は、その時々生成された動作の出来に強く依存すると考えられる。しかし、ランダムな動作生成では大きな動作が生成されないことがわかっているため、感情を強く含んだような台詞に対する動作生成などを考えると、提案手法の方が感情を上手く表現できるのではないかと予想できる。そのため、質問⑤感情の表出に関していえば、提案手法の方が高評価を得られると予想できる。

それとは別に、ランダム手法の特徴である台詞中ずっと動作し続ける仕様は、動作無しとは反対に人間らしい状態とも言える。そのため、質問③に関しては、提案手法よりもランダム手法の方が低評価を得る可能性がある。

6.2.3.手動との比較

手動は、全動作生成手法中、最も人間の感性に合った動作になるはずである。そのため、質問①の台詞との整合性や質問⑤の感情の表出では、手動生成は提案手法よりも高評価を得ると予想できる。ただ、質問③のスムーズさに関しては、そもそも差異がないはずなので、提案手法と同程度になる可能性がある。

6.3.実験結果・考察

6.3.1.実験結果

実験結果として、アンケートを集計した結果を示す。集計方法として、提案手法以外の手法が提案手法よりも良い場合を+2、提案手法以外の手法が提案手法よりも悪い場合を-2、提案手法以外の手法と提案手法が同程度の場合を±0として、手法ごとに各質問の平均をとった。集計結果を図6.3に示す。

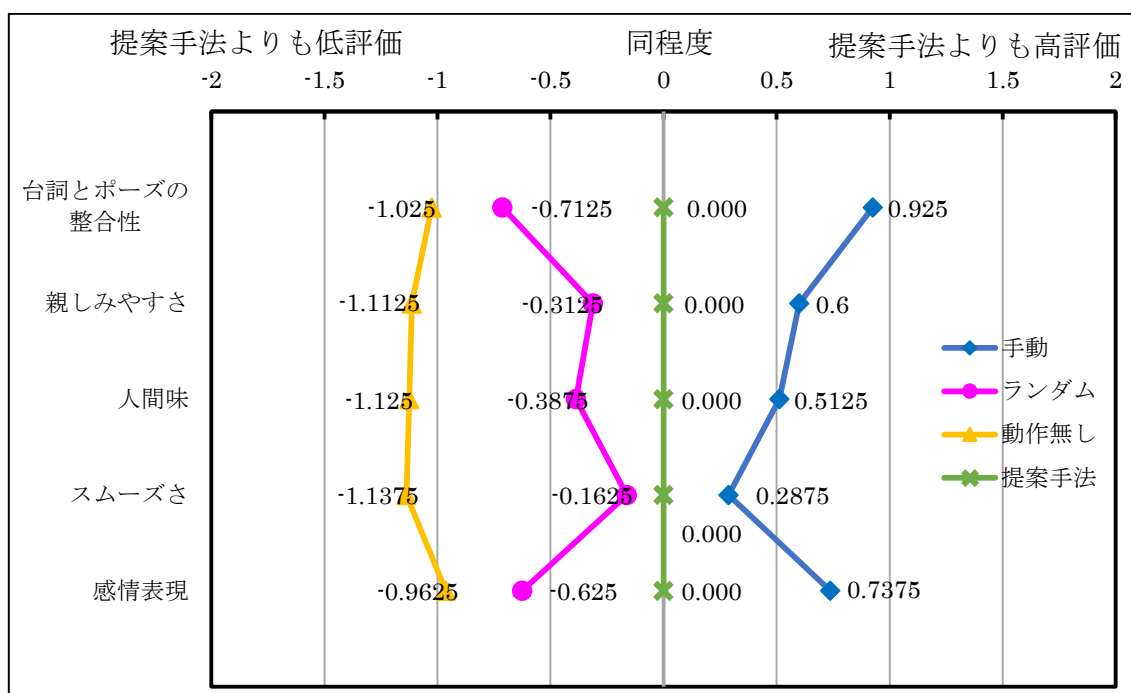


図 6.3. 提案手法と各手法の比較アンケート結果

図は各比較対象の手法と提案手法の相対評価の結果を示している。すなわち、図の左に行くほど、比較対象手法が提案手法より高評価であり、右に行くほど提案手法と比較して低評価であることを表している。また、図中央0に近いほど、比較対象手法と提案手法との差がないことを表している。

結果から全体的な評価をすると、人間から見て好ましい方から順に手動生成、提案手法、ランダム生成、動作無しとなっており、提案手法が手動生成に次いで上位になっていることがわかる。

質問ごとに見ると、台詞とポーズの整合性と感情の表出は、結果に同じような傾向がある。これは、正しい感情を示していたかという質問に対し、台詞の中の感情と動作として表出される感情を比較し、アンケートに答えていたためだと考えられる。つまり、感情の表出に関する質問に回答するときも、結果として台詞と動作の整合性を評価していたと考察できる。

また、本来どの手法もほとんど差異がないはずのスムーズさの結果に関して、各手法で結果に差が表れた。これは、台詞と動作の整合性など他の印象にスムーズさの結果が影響を受けたためと考えられる。ここから台詞と動作の整合性に違和感がなければ、人間はロボットの動きがスムーズであると感じやすくなるということがわかった。

次項以降で、比較対象手法毎に結果を詳しく考察する。

6.3.2.動作無しとの比較

動作無しに関しては、5つの質問項目全てにおいて、提案手法の方が良いという評価を得た。これは、発話時に無意識に体を動かす人間と比較して、全く動かずに発話するロボットに違和感を覚えたからだと考えられる。そのため、親しみやすさや人間味に関して、低い評価がなされている。また、そもそも動いていないため、スムーズさや台詞と動作の整合性も低い評価となっている。

6.3.3.ランダム手法との比較

ランダムな動作生成に関しては、感情の表出や台詞との整合性において、提案手法に比べて低い評価となり、ただランダムに細かく動き続ける動作より意図や感情のはっきり表れた動作の方が、台詞との整合性が高いと感じるという結果になった。また、スムーズさにおいて、動き続けるランダム手法の方が高評価を得るだろうと予想したが、台詞との整合性が高い動作の方がスムーズに感じるという結果となった。

6.3.4.手動生成との比較

手動動作生成に関しては事前の予想どおり、最も人間の感性に合った動作であるという結果となった。特に動作と台詞の整合性については、提案手法と比較して高い評価を得ている。これは、手動で台詞ごとに人間が考える最も自然な動作を付けているためで、複数の動作に重みを付けて混合し、それらしい動作を作っている提案手法と比べ、良い結果となったのは当然と言える。ただし、手動生成手法は、時間的にも作業量的にもコストが大きい。

6.3.5.実験総評

実験結果は、大まかに事前の予想どおりになったといえる。ランダムな動作生成手法を、近年のロボットの発話中の動作生成によく利用されるランダムな動作生成手法の代表と考えれば、提案手法を使用することによって、ロボットの動作をより人間的な動作にすることができた。よって、本提案手法の有用性は、主観評価実験の結果によって、十分に示すことができた。

第7章 結論

7.1.本研究の成果

ロボットに、より人間的な動作をさせるための、機械学習を用いたロボット動作自動生成手法を提案した。台詞と動作（ポーズ）を関連付けて学習することによって、台詞の内容や意味などに合致した動作を生成することが可能となった。台詞に関しては、単純なキーワードマッチングなどではないため、未学習の単語を含んだ台詞に対しても、適当な動作を付けることができる。また、動作生成の際に、学習データを基にした確率的な処理を行うことにより、同じ台詞に対して、台詞に合った範囲内で多様性のある動作の生成が可能となった。また、提案手法を利用することによって、台詞から動作の生成だけでなく、動作から台詞の生成も可能である。

本研究の成果によって、ロボットがより人間的に振る舞うことが可能になった。これによって、人間の行動を察し、人間的な対応をすることのできるロボットの実現に一步近づくことができたと考えている。

7.2.今後の展望

今後の課題として、大きく3つのことが挙げられる。

1つ目の課題は、学習データの増加である。本研究では学習データとして、896文の台詞と、それに沿った動作のデータセットを用いている。しかし、実環境でロボットを利用するには、まだ表現できる感情や意図が少ないと考えられる。現状の学習データだけでは、怒りや驚きの感情、呼びかけや提案の意図などを表現することは難しいため、様々な感情や意図を補う学習データを取り込む必要があると考えている。学習データの増加の方法としては、未知語の含まれている台詞から動作を生成する際、台詞と生成動作の関係を学習データにフィードバックし、語彙を増やすことや、周囲の人間の言葉と動作の関係を自動で取得して、学習データとすることが考えられる。

また、学習データの増加に付随して、感情をより上手く表現するために、文章に含まれる感情を推定する既存の研究[11]などを参考に、学習データの中に感情に関するデータを含め、感情空間内でデータを表現できるようにすることで、より感情豊かなロボットを実現することが可能と考えられる。

2つ目の課題は、台詞の生成の精度向上である。現状、**TF-IDF** 重みを付与したにも関わらず、一般語が生成結果に大きく表れているため、重みの計算を見直す必要があると考えられる。また、結果の出力に関して、生成された単語を文章として出力する機能の実装が必要である。これについては、文章の自動生成に関

する研究[12]が応用できると考えられる。

3つ目は、人間の感情推定の実現である。本研究の成果の一つとして、ロボットが台詞とポーズのセットを学習、分類したとき、そのうちのいくつかのカテゴリについて、感情や意図を表すものになると確認できたことが挙げられる。このカテゴリへの意図・感情ラベルの付与と、提案手法の特徴である動作と台詞の相互推定を組み合わせ用い、人間の動作や台詞を入力としてカテゴリ分類をすることによって、人間の感情や意図を簡易的に推定することが可能である。実際に、人間の動作から関連単語、およびカテゴリを推定した結果は、5.4節で既に表示したとおりである。現在は音声認識デバイスなどをシステムに組み込んでいないので、人間の台詞を動作と合わせて取得し、認識することはできないが、将来的には人間の台詞と動作を合わせて取得することで、より正確な人間の感情や意図の認識をすることができると考えている。

人間の感情推定が、提案手法によって実現されれば、一つの手法と学習データからロボット動作の自動生成と、人間の感情の簡易的な推定が可能になる。

謝辞

本研究を進めるにあたり、多大なる助言、ご指導を賜りました金子正秀教授、並びに中村友昭助教に感謝いたします。また、お世話になりました金子研究室の皆様、実験に協力して頂いた被験者の皆様に心より感謝いたします。

参考文献

- [1] 西嶋 隆, 山田 俊郎, 小川 行宏, 今井 智彦, 稲葉 昭夫, 大野 尚則: “案内ロボットの開発,” 岐阜県生産情報技術研究所研究報告第 6 号, 15, pp. 51–55 (2004)
- [2] 藤井 勝敏, 西嶋 隆, 棚橋 英樹, 山田 俊郎, 田中 泰斗, 千原 健司, 稲葉 昭夫: “案内ロボットの開発(第 3 報),” 岐阜県生産情報技術研究所研究報告第 8 号, 11, pp. 40–43 (2006)
- [3] 大塚国際美術館: “ギャラリートークロボット アート君,” 大塚国際美術館 HP < <http://www.o-museum.or.jp/> >
- [4] 柴田 崇徳: “アザラシ型ロボット・パロとの相互作用に関する研究,” 日本ロボット学会誌, Vol. 29, No. 1, pp. 31–34 (2011)
- [5] 加納, 清水: “なにもできないロボット,” 日本ロボット学会誌 Vol.29 No.3, pp.298-305, 2011
- [6] Aldebaran Robotics 社: “世界初の感情認識パーソナルロボット Pepper,” Aldebaran Robotics 社 HP < <https://www.aldebaran.com/ja/peppertoha> >
- [7] 神田 崇行, 石黒 浩, 石田 亨: “人間-ロボット間相互作用にかかわる心理学的評価,” 日本ロボット学会誌, Vol. 19, No. 3, pp. 362-371, 2001
- [8] 京都大学情報学研究科-日本電信電話株式会社コミュニケーション科学基礎研究所 共同研究ユニットプロジェクト: “MeCab: Yet Another Part-of-Speech and Morphological Analyzer,” Mecab HP < <http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html> >
- [9] 中村, 長井, 岩橋: “ロボットによる物体のマルチモーダルカテゴリゼーション,” 電子情報通信学会論文誌 D, vol.J92-D, no.10, pp.2507-2518, 2008.
- [10] Blei, Ng, Jordan: “Latent Dirichlet Allocation,” Journal of Machine Learning Research, vol.3, pp. 993-1022, 2003.
- [11] 江崎, 小町, 松本: “感情軸における感情極性制約を用いたマルチラベル感情推定,” 言語処理学会第 19 回年次大会論文集, pp.244-247, March, 2013
- [12] 富坂, 鈴木, 相澤: “自由対話実現のための自動文章作成モデルの提案,” 情報処理学会創立 50 周年記念 (第 72 回) 全国大会, vol.2, pp.627-628, 2010
- [13] 諸岡, 浜元, 長橋: “強化学習と隠れマルコフモデルの結合による自律的な動作認識,” 電子情報通信学会論文誌 D-II, vol. J88-D-II, no.7, pp. 1269-1277, 2005.
- [14] 杉山, 篠沢, 今井, 萩田: “コミュニケーションロボットのための発話とジェスチャのアサインパターンの抽出とその発展的開発手法の提案,” 電子情報通信学会論文誌 A, vol. J95-A, no.1, pp.46-59, 2012.

発表実績

- [1] 宮崎齊, 中村友昭, 金子正秀: “台詞から想起される感情を表現するロボット動作の作成,” 2014年映像情報メディア学会冬季大会, 6-5, 2014.12.18.
- [2] 宮崎齊, 中村友昭, 金子正秀: “台詞に含まれる感情表現を反映したロボット動作の自動生成,” 映像情報メディア学会メディア工学研究会, 2015.2.28.