子どもを対象とする雑談対話システムにおける発話理解と感情生成 Utterance understanding and emotion generation in a dialogue system for children

山本 賢人*1	大村 俊哉*2	根本 凌 ^{*2}	和田 史織*1	杉本 徹 ^{*2}
Kento Yamamoto	Syunya Omura	Ryo Nemoto	Shiori Wada	Toru Sugimoto

*1 芝浦工業大学大学院 理工学研究科 Graduate School of Engineering and Science, Shibaura Institute of Technology

*2 芝浦工業大学 工学部

College of Engineering, Shibaura Institute of Technology

We propose a dialogue system for children with emotions. In this paper, we focus on utterance understanding and emotion generation in this dialogue system. In utterance understanding, we extract utterance intention and focus. Machine learning based on utterances collected from college students gives an accuracy of about 70% for utterance intention analysis and about 90% for focus extraction for children's utterances. In emotion generation, we deal with two types of emotion: short emotion and mood. The type of response is affected by mood, and short emotion is expressed as a facial expression. Regarding expressions of emotion, college students judged about 60% of emotions generated in dialogue as valid, but we could not confirm their usefulness in dialogue with children.

1. はじめに

近年, 雑談対話システムの研究やビジネス利用が盛んである. たとえば, NTT ドコモ社の「しゃべってコンシェル」, ソフトバンク 社の「Pepper」が例として挙げられる[東中 2014]. また, 子どもを 対象とした対話システムの研究として「たけまるくん」が存在する [西村 2004].

ー方で子どもが対話システムを利用する場合,子どもが退屈 せずに楽しめるための話題や機能の充実といった課題が考えら れるが,このような課題に焦点を当てた研究は少ない.

そこで我々はシステムの対象年齢を 7~12 歳程度と仮定し, 子どものユーザとクイズや石取りゲーム,雑談を楽しむことがで きるシステムの開発を目指している. クイズや石取りゲームとい った子ども向けの対話の内容に加えて,システムに感情を持た せることや強化学習を行い子どもに適応することで,子どもが退 屈せず利用できるようにする.

本稿では、この子どもを対象とする雑談対話システムにおけ る発話理解と感情生成手法に焦点を当て評価と考察を行う.

2. システムの概要

提案するシステムの処理の流れを図1に示す.

ユーザが音声による入力を行うと,発話理解部は入力結果から発話内容を解析し,発話意図や焦点などの情報を抽出する.

そして感情更新部では,発話理解で得られた結果をもとに感 情の更新を行う.本研究では感情の概念をシステムの情緒と気 分の二つに分けて定義する.得られた情緒に基づき,キャラクタ ーの表情を出力する.また得られた気分に基づきシステムのい くつかの応答候補に対してその応答のスコアを増加させる.

一方でクイズ出題部では発話理解部で得られた現在の対話 状況に応じて次にクイズとして出題する問題文の選択や,クイズ の正誤判定を行う.

連絡先:山本賢人, 芝浦工業大学大学院 理工学研究科, 〒135-8548 東京都江東区豊洲 3-7-5 E-mail: al14097@shibaura-it.ac.jp



図 1. システムの処理の流れ

最後に応答選択部で応答文の選択を行う.本研究では現在 の対話の状態をタスクとフェーズの組として表現し,それらに加 えシステムの気分,発話行為タイプ,キーワードが含まれている か,ユーザの沈黙時間等に基づいて複数の応答文の候補のス コアリングを行い,スコアが最大の応答候補を選択する.ここで タスクとは,雑談・クイズ中・石取ゲームのいずれかである.一方, フェーズとはタスクが行われている・中断している・もう一度行う か現在確認中のいずれかである.

スコアの値は現在の状態に対しユーザの沈黙時間や相手を 褒める内容や貶す内容が発話に含まれているか等を報酬とし た強化学習により調整される.

最後に応答文を音声合成して出力する.

3. 発話理解手法

3.1 発話意図および焦点情報の抽出

(1) 発話意図解析

発話意図は発話行為タイプと属性の組で表す.発話行為タ イプはその発話がどのような意味であるかを大まかに表す.一 方属性は発話行為タイプが指し示す具体的な内容である.発 話行為タイプは[Meguro 2010]が雑談対話システム用に定めた 33 種類の発話行為タイプを参考にし,後述する発話収集の実 験での発話内容を踏まえつつ本システムに合わせて18種類定 義した.このうち4種類はシステムの発話のみに付与されるため, ユーザの発話に付与される発話行為タイプは 14 種類である. 発話行為タイプの例を表1に示す.

発話文	発話行為タイプ	焦点		
僕の誕生日は3月2日だよ.	陳述	誕生日,3月2日		
寝たことありますか.	質問	寝る, ある		
やらない.	質問返答. 否定	(無し)		

表1 発話行為タイプおよび焦点の例

発話行為タイプの付与はLIBLINEARを用いてロジスティック 回帰で行う. 属性は, 各発話行為タイプに対応した品詞の単語 を形態素解析の結果から抽出する.

(2) 焦点解析

焦点とは発話の中心となる語であり,名詞(句)と述語の組み 合わせと定義する. 述語は動詞, 形容詞, 名詞(句)のいずれか である.なお名詞または述語どちらか一方しか含まれない場合 はそれだけを抽出する. 焦点の例を表1に示す. 発話への焦点 情報の付与は, CRFsuiteを用いて形態素単位の系列ラベリング で行う.

3.2 発話データの収集および学習

(1) 学習に用いる発話の収集

発話意図解析および焦点解析の学習を行うため発話を収集 した.大学生に子どもを演じてもらい、システム役の人間と実際 の対話を想定した模擬対話を行い9人から合計563発話を収 集した.発話文は人手で書き起こしを行い,発話行為タイプと焦 点情報を付与した.

(2) 発話意図解析に用いる素性の決定

発話意図解析の学習に用いる素性の決定を行った.素性の 候補と最終的な使用の有無を表2に示す.

表 2. 発話意図解析で用いる素性の候補

素性の種類	使用の有無
単語および品詞 n-gram	1,2,3-gram
	ユーザの発話と
何発話分の n-gram を使用するか	直前のシステム
	の発話
ユーザの発話の文字数	使用
直前のシステムの発話行為タイプ	不使用
直前のシステムの発話に焦点が含まれているか	不使用
ユーザの発話の形態素解析結果に	
JUMAN の 22 種類の意味カテゴリや 4 種類の	使用
固有表現情報が含まれているか	

素性の組み合わせは,現在得られている最も正解率の良い 組み合わせに新しい素性を加えて正解率を調べ,正解率が向 上した場合その素性を加えることを繰り返した.素性は表2の上 から順に加えた.表2に示す素性の組み合わせで最高0.701 の正解率が得られた.正解率は学習データの発話のうち, LIBLINEAR が出力する確率値が最大の発話行為タイプと正 解の発話行為タイプが一致した発話の割合である.

なお LIBLINEAR のアルゴリズムは L2 正則化 primal を使用 し、パラメータはデフォルト値を使用した.また、素性のスケーリ ングは行わなかった. 正解率を調べる際には10分割交差検定 を用いた.

(3) 焦点解析に用いる素性の決定

焦点解析の学習に用いる素性の決定を行った.素性の候補 と最終的な使用の有無を表3に示す.

表 3. 焦点解析で用いる素性の候補

素性の種類	使用の有無
単語 n-gram	1,2gram
品詞 n-gram	1,2,3gram
直前のユーザとシステムの発話行為タイプ	使用
直前のシステムの発話に焦点が含まれているか	不使用
ユーザの発話の文字数	使用
発話文中での形態素の位置を正規化した値*1	使用
ユーザの発話の形態素解析結果に JUMANの22種類の意味カテゴリや4種類の固有 表現情報が含まれているか	不使用

発話文中での形態素の位置 * 10 *1

発話文の形態素数

素性の組み合わせの求め方は発話意図解析と同様である. ただし評価基準は焦点のF値および再現率を用いた.F値がほ ぼ等しい場合は再現率が高いものを選択した.表3に示す素性 の組み合わせで 0.688 の再現率が得られた.

なお内部アルゴリズムおよびパラメータはデフォルト値を使用 し、5分割交差検定で調べた.

4. 感情更新手法

4.1 感情生成モデル

[目良2010]では,感情の概念を情緒,気分,表情の3つに分 けて定義している. 情緒は瞬間的に生じる短くて強い感情であ る. 気分は生起した情緒や内的状態に起因して生じる弱いなが らも長く持続する感情である.表情は情緒や気分などが身体を 通じて表出されるものである.本稿では、感情の概念をシステム の情緒と気分の2つに分けて定義する、情緒と気分は、Russel の感情円環モデル[Russel 1980]を用いて「快-不快」を表す好 感度,「覚醒-眠気」を表す覚醒度の二次元平面上の実数値の 組として表現する.また,極座標ベクトルの方向および大きさを それぞれ情緒および気分の種類と大きさとする、本研究で扱う 情緒の種類を表4に示す.

衣 4. 生起する情緒の種類				
	覚醒	眠気		
快	喜び	平常		
不快	怒り	悲しみ		

4.2 好感度の計算

最初にユーザの発話文から名詞,動詞,形容詞,副詞を抽 出する.抽出した語の好感度データを実数値で取り出し、その 平均値をその発話文における好感度とする. 好感度データには, 単語感情極性対応表[高村 2006]を用いる. この表は, ある単語 が一般的に良い印象を持つか,悪い印象を持つかを-1から+1 の実数値で表したものである.

4.3 覚醒度の計算

覚醒度はシステムがユーザの発話に興味を持つかどうかで 判定する、システム自身の好き嫌いやユーザの発話行為タイプ、 クイズや石取ゲームを連続して続けた回数に応じてスコア付け を行い-1から+1の実数値を決定する.

システム自身の好き嫌いは、システムのプロフィールとして好 きなもの,嫌いなものを登録しておき,ユーザの発話に好きなも のの語が含まれればシステムは興味を持ち覚醒度が上がる.嫌 いなものの語が含まれれば覚醒度が下がる.

ユーザの発話行為タイプが自己開示や情報提供などを意味 する陳述である場合は、システムは興味を持ち覚醒度が上がる. クイズやミニゲームを連続して続けた場合にも覚醒度は上が るが、すぐにやめてしまった場合には覚醒度は下がる.また、長 く続けた場合にも飽きが来ると考え覚醒度は下がる.

4.4 情緒および気分の更新

情緒と気分を以下の式で更新する.まず4.2節,4.3節の手法 で好感度 Fr および覚醒度 Ar を求め初回の情緒 $E_1 = {Fr \choose Ar}$ を 決定する.次に、気分Mを決定する.初回は $M_1 = E_1$ とする.2 回目以降は1回目と同様にFrとArを求め、以下の式により情緒 E_n を更新する.

$$\mathbf{E}_{n} = (1-t) \begin{pmatrix} Fr \\ Ar \end{pmatrix} + t \mathbf{M}_{n-1} \qquad (n \ge 2)$$

tはどれだけ現在の気分を反映させるかの定数であり今回は 0.5 とした.気分の更新方法は生起した情緒の種類によって以下の二つに分けられる.

1. 生起した情緒の種類が直前の情緒の種類と異なる場合:

 $M_n = E_n$

2. 生起した情緒の種類が直前の情緒の種類と同じ場合:

 $M_{n} = \frac{\alpha}{2} \left(M_{n-1} + E_{n} \right) \left(\alpha : 氣分の減衰率 \right)$

以上により情緒および気分を更新する.

4.5 情緒に応じた表情の生成

生起した情緒はシステムの表情として出力される.[中村 1996]では人類に共通していると考えられる表情の特徴をまとめ ている.本研究ではこの特徴をもとに表情の作成を行った.

4.6 気分に応じた応答戦略

システムの現在の気分に応じ、いくつかの応答に対してその 応答の選択されやすさを変化させる.システムの気分が良いと きはシステムの方から積極的に話すようになると考え、システム の自己開示、話題に基づく応答、オープンクエスチョンの応答 をしやすくなるようにスコアを増加させる.システムの気分が悪い ときは、システムは消極的になると考え、傾聴や当たり障りのな い応答をしやすくなるようにスコアを増加させる.スコアの増加 量は現在の気分M_nの好感度と覚醒度の二乗和の平方根により 求める.

5. システムの実装および評価実験

5.1 プロトタイプシステムの実装

本研究のシステムのプロトタイプを作成した.このプロトタイプ に用いたキャラクターの喜び,怒り,悲しみ,平常時での発話状 態のグラフィックを図2に示す.



喜び 怒り 悲しみ 平常(発話時) 図 2. キャラクターの各感情のグラフィック また, 今回は音声認識を用いずユーザが話した音声をオペ

レータが入力する形で対話を行った. クイズは子どもが興味を持つと考えられる話題の問題を23問 人手で子どもに合わせた言葉遣いで作成した.

5.2 発話理解手法および感情生成手法の評価実験

前述したプロトタイプを用いて小学2年生3人,5年生2人と 対話を行った.対話では感情を表出するシステム、しないシステ ムそれぞれと各7分間対話を行い、対話するシステムの順番は ランダムに入れ替えた.合計244発話を収集した.ユーザとシス テムの対話例を表5に示す.

表 5. 対話例

番号	発話者	発話文
		はじめましてボクの名前はクリームだ
1	システム	よ,さっそくだけどクイズか石取ゲーム
		やらない, それとも僕とお話しする?
2	ユーザ	石取.
		石取りゲームは1から数を数えていって
2	ミロテレ	先に21を言ったほうが勝ちだよ. 1回で
3	V//A	3まで数をふやすことができるよ. さいし
		ょはボクからスタートだよ. じゃあ 1, 2.
4	ユーザ	21.
~	1.75)	おめでとう,君の勝ちだよもう一回や
5 システム		る?
6	ユーザ	もう一回やる.
7	システム	ボク,チョコレートが大好き.
8	ユーザ	僕はふつう.
9	ユーザ	クイズ.
		それでは問題です,あさごはんによくで
10	2751	る,ねばねばしてい大豆からできた食
10	ンステム	べ物は何でしょう,1番とうふ,2番なっ
		とう,3番ドーナツ.
11	ユーザ	2番.

最初に1のようにシステムから発話が行われる.2から6は石 取ゲームの発話であり,7から8は雑談中の発話,9から11はク イズ中の発話である.

(1) 発話理解の精度

3.2 節で述べた大学生の発話データを学習データとして, LIBLINEARとCRFsuiteの学習を行った.その後プロトタイプを 使用して子どもから得た発話データをテストデータとして与え適 合率および再現率を求めた.発話意図解析の適合率および再 現率を表6に,焦点抽出の適合率および再現率を表7に示す.

衣 0. 光間11為クイノことの週日平わよい丹妃平					
発話行為タイプ	適合率	再現率			
陳述	0.456 (21/46)	0.750 (21/28)			
質問	0.846 (33/39)	0.804 (33/41)			
質問返答.肯定	0.875 (14/16)	0.560 (14/25)			
質問返答.否定	1.000 (02/02)	0.400 (02/05)			
質問返答.その他	0.823 (14/17)	0.875 (14/16)			
話題変更		0.000 (00/02)			
相槌	0.470 (08/17)	0.615 (08/13)			
クイズ.回答	1.000 (20/20)	0.645 (20/31)			
石取	0.861 (62/72)	1.000 (62/62)			
定型	0.000 (00/02)				
要求	0.000 (00/01)				
+評価					

-評価		0.000 (00/03	3)	
その他		0.333 (03/09)		0.142 (03/21)
平	 均	0.725 (177/244)		
表 7. 焦点抽出の適合率お			およて	バ再現率
ラベル	適合率			再現率
焦点	0.717 (0160 / 0223)		0.	824 (0160 / 0194)
非焦点	0.959 (0802 / 0836)		0.	927 (0802 / 0865)
平均	0.908	(0962 / 1059)		

表 6 に関しては, 発話行為タイプの出現頻度に差がみられ, 出現頻度が高い発話行為タイプほど再現率も高い. これは学習 データにも同様の傾向がみられ, 学習データの量が十分である ため正しく判別が出来ていると考えられる. 一方で話題変更, 定型, 要求, +評価, -評価は出現数が少ないか全く出現してい ないため分類精度を調べられなかった. これらは学習データに もほとんど出現しないため, 発話行為タイプの種類自体の見直 しや収集方法を工夫する必要があると思われる.

表7に関しては、焦点の再現率が学習データの0.688に比べ 向上している.これは、子どもが話す文に複文や複雑な係り受 け関係が少ないため、文字数や形態素の位置の素性がより有 効に働いたためと考えられる.

学習データおよびテストデータの形態素数の分布をそれぞれ 図 3, 図 4 に示す.





小学生の発話は形態素数が10以上の発話が少ないことが分かる.

(2) 感情生成手法の有効性の評価

対話後に各システムについて4段階(1~4)で評価してもらった. 評価内容および各システムの平均値を表8に示す.

アンケート項目	感情あり	感情なし
上手に会話できていたか	2.8	3.0
話していて面白かったか	3.0	3.0
また話してみたいか	3.0	3.2
仲良くなれそうだったか	3.0	3.0

表 8. 各システムのアンケート評価結果

アンケートの結果から、感情を表情として表出するシステムと しないシステムで、あまり差異は見られなかった.これは実際に 実験の様子を観察して、子どもが画面を見ずに別の方を見てい たり、画面を見ていても字幕を読むことに集中していたことから、 システムの表情の変化に気づかなかったのではないかと推測さ れる.従って、システムの表情の変化を分かりやすくするために、 表情と共に背景色を感情に合わせた色に変化させるなどの工 夫をする必要があると考えられる.

(3) 感情生成手法の妥当性の評価

大学生の被験者7人に、子どものユーザとシステムの対話ログ を読み、システムの発話時に生成される情緒が妥当かを5段階 で評価してもらった.評価は、ユーザの発話に対するシステムの 発話を1対話として合計36対話に対して行った.被験者が選 択した評価のうち、最も多かったものをその対話の評価とし、各 評価に割り当てられた対話の数とその割合を表9に示す.

衣 9. 谷評価に割り目 くられに対 品数とての割合					
評価	1 妥当 である	2 まあまあ 妥当である	3どちらとも 言えない	4あまり妥 当でない	5 妥当 でない
対話数	16	6	5	5	4
割合	44%	17%	14%	14%	11%

1 および 2 の評価で約 61%の割合となった.4 および 5 の評価となった対話には、ユーザ発話が質問なのに対し怒りなどの情緒が生成されている場合があった.

6. まとめ

本研究では子どもを対象とする雑談対話システムの提案を行い、実際にプロトタイプを構築し、発話理解の精度および感情 生成の手法についての評価を行った。発話意図解析では約7 割、焦点解析については約9割の正解率で分類が行えるという 結果を得た.一方、感情生成については約6割の対話につい て生成された情緒が妥当という結果を得たが、子どものユーザ に対する感情の表出の有用性は確認できなかった。

今後は発話検出や音声認識を組み込んだ発話理解部の開発や,感情表出の仕方の見直しを行う.またシステムのクイズの 自動生成機能の実装や,強化学習による応答選択の最適化に 関する評価実験を行う.

参考文献

- [東中 2014] 東中竜一郎: 雑談対話システムに向けた取り組 み,人工知能学会第 70 回言語・音声理解と対話処理研究 会, 2014.
- [西村 2004] 西村竜一,西原洋平,鶴身玲典,李晃伸,猿 渡洋,鹿野清宏:実環境研究プラットホームとしての音声情 報案内システムの運用,電子情報通信学会論文誌 J87-D-II(3) pp. 789-798, 2004.
- [Meguro 2010] Meguro, Toyomi, Minami Yasuhiro, Higashinaka Ryuichiro, Dohsaka Kohji: Controlling listening-oriented dialogue using partially observable Markov decision processes, Proceedings of COLING, pp. 761-769, 2010.
- [目良 2010] 目良和也,市村匠,黒澤義明,竹澤寿幸:情緒 計算手法と心的状態遷移ネットワークを用いた音声対話エ ージェントの気分変化手法,日本知能情報ファジィ学会誌 22(1) pp. 10-24, 2010.
- [Russel 1980] J.A. Russel: A circumplex model of affect, Journal of Personality and Social Psychology 39(6), pp. 1161-1178, 1980.
- [高村 2006] 高村大也,乾孝司,奥村学:スピンモデルによる 単語の感情極性抽出,情報処理学会論文誌 47(2) pp. 627-637, 2006.
- [中村 1996] 中村真: 表情とコミュニケーション, 繊維製品消費 学 37(1) pp. 4-9, 1996.