

修士論文

**継続的な対話意欲を高める
テレビ視聴型雑談ロボットのための発話文評価法**

向井田 一平

2018 年 3 月 15 日

奈良先端科学技術大学院大学
情報科学研究科 情報科学専攻

本論文は奈良先端科学技術大学院大学情報科学研究科に
修士（工学）授与の要件として提出した修士論文である。

向井田 一平

審査委員：

萩田 紀博 教授 （主指導教員）

加藤 博一 教授 （副指導教官）

神原 誠之 准教授 （副指導教官）

継続的な対話意欲を高める テレビ視聴型雑談ロボットのための発話文評価法*

向井田 一平

内容梗概

本研究では、テレビ視聴型雑談ロボットを用いてユーザの対話意欲を高めるロボットの発話文の評価法を提案する。ロボットの発話がユーザの対話意欲に与える影響が明らかになっていないため、まずロボットの発話を受けてユーザが感じた対話意欲の変化を主観評価する。収集した対話意欲の評価データを用いてユーザ対話意欲モデルを作成し、発話文がユーザの対話意欲に与える影響を推定する。このユーザ対話意欲モデルの推定精度を明らかにするために、スポーツとニュースの2種類のテレビ番組に対して発話文と9～38名の被験者からの対話意欲評価を収集し、K-fold クロスバリデーションに基づいて、モデルの推定精度を評価する実験を実施する。実験の結果、スポーツでは平均二乗誤差 0.101、ニュースでは平均二乗誤差 0.107 の精度で対話意欲評価を推定できることを示す。推定結果から、ユーザのテレビ番組への興味度合いの違いが対話意欲へ与える影響について考察する。被験者が回答したアンケートから対話意欲に影響する発話文の傾向についても述べる。これらの実験結果から、ユーザ対話意欲モデルを用いてロボットの発話文を評価し、ロボットがユーザの対話意欲を高める発話文を選択可能であると明らかになった。

キーワード

雑談ロボット, テレビ視聴, 継続対話, 対話意欲

*奈良先端科学技術大学院大学 情報科学研究科 情報科学専攻 修士論文, NAIST-IS-MT1651107, 2018 年 3 月 15 日.

Robot Utterance Evaluation for TV Chat Robots Allowing Daily Continuous Conversation*

Ippei Mukaida

Abstract

This paper proposes an evaluation method of robot utterance for TV chat robots allowing daily continuous conversation. First, we make subjective evaluations on interaction motivation as when users are uttered by the robot, because there is no relationship between user's interaction motivation and robot utterances. Second, a model of the interaction motivation is induced from the collected interaction motivation datasets. The model predicts user motivation created by robot utterance. The model was tested by 2 dataset of sport and news genre using 9-38 subjects. The k-fold cross validation is used for the datasets. As a result, mean squared errors between training and test data are 0.101 in sport dataset and 0.107 in news dataset. The experimental results show that the robot is able to choose utterances to promote user's interaction motivation using the model.

Keywords:

Chat robot, Watching TV, Continuous Conversation, Interaction motivation

*Master's Thesis, Department of Information Science, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-MT1651107, March 15, 2018.

目次

図目次	v
第 1 章 はじめに	1
第 2 章 ロボットとの日常的対話に関連する研究	3
2.1 人とロボットとの関係構築	3
2.2 ロボットとの対話におけるユーザの対話意欲	4
2.3 テレビ視聴時における雑談ロボット	5
2.4 対話時の発話文の評価基準	7
2.4.1 タスク指向型対話ロボットの評価	8
2.4.2 非タスク指向型対話ロボットの評価	8
2.5 本研究の位置づけ	9
第 3 章 テレビ視聴型雑談アプリケーション Dabelive	10
3.1 Dabelive のシステム構成と機能	11
3.1.1 放送局の選択	11
3.1.2 ソーシャルコメントを利用したエージェントの発言	11
3.1.3 エージェントの発言に対する評価	13
3.1.4 エージェントの発言内容のフィルタリング	14
3.2 ソーシャルコメントの収集	14
3.3 エージェントによるソーシャルコメントの発話	16
3.4 ソーシャルコメントに対するユーザの対話意欲評価の収集	17
第 4 章 テレビ番組に関する発話文の評価	18
4.1 ユーザ対話意欲モデルに用いる特徴量	19
4.1.1 発話文とテレビ番組トピックの関連付け	19
4.1.2 ユーザ属性	19
4.1.3 発話文の対話意欲評価値	19
4.2 ユーザ対話意欲モデルの作成処理	20

4.2.1	テキストの前処理	20
4.2.2	テレビ番組のトピック抽出	21
4.2.3	テキストのベクトル化	22
第 5 章	対話意欲評価の推定実験	24
5.1	実験方法	24
5.2	発話文に対するユーザの対話意欲評価の収集	25
5.3	実験結果: スポーツ	28
5.4	考察: スポーツ	30
5.4.1	推定した対話意欲評価値の分布	30
5.4.2	テレビ番組への興味度合いの違いによる対話意欲への影響	31
5.4.3	アンケートから判明した対話意欲に影響する発話文の傾向	32
5.5	実験結果: ニュース	33
5.6	考察: ニュース	34
5.6.1	スポーツと同等の対話意欲評価値の推定精度であった理由	34
5.6.2	Bag of words モデルの問題点	35
5.6.3	各ユーザが抱いた Dabelive への印象	36
5.6.4	推定した対話意欲評価値に基づく発話文の選択	36
第 6 章	おわりに	43
	謝辞	45
	参考文献	46
	発表リスト	50

図目次

2.1	テレビ視聴型雑談ロボットとの対話風景 [11]	6
2.2	テレビ視聴型雑談ロボットとの対話例 [11]	7
3.1	Dabelive の使用例	11
3.2	Dabelive のシステム構成概要	12
3.3	放送局の選択画面	13
3.4	設定画面	14
3.5	エージェント対話画面	15
4.1	ユーザ対話意欲モデルの作成フロー	18
4.2	ユーザ対話意欲モデルによる対話意欲対話意欲の推定	18
4.3	トピック抽出の模式図	21
4.4	Bag of words モデルによるテキストのベクトル化例	22
5.1	ユーザ対話意欲モデル (実験)	24
5.2	実験用に改良した Dabelive のエージェント対話画面	26
5.3	Russell の円環モデルにおける理由ボタンの配置	26
5.4	推定した対話意欲評価値の平均二乗誤差 (スポーツ), 1182 次元	29
5.5	興味度合い毎に推定した対話意欲評価値の平均二乗誤差 (スポーツ)	30
5.6	対話意欲評価値のヒストグラム データセット:スポーツ, 回帰手法:SVR(RBF カーネル)	31
5.7	各興味度合いの対話意欲評価値ヒストグラム データセット:スポーツ, 回帰手法:SVR(RBF カーネル)	32
5.8	推定した対話意欲評価値の平均二乗誤差におけるスポーツとニュースの比較	34

第 1 章 はじめに

インターネットの普及や独居高齢者の増加に伴い、年齢問わず日常的な会話機会が減少している [1][2]。日常会話が不足すると、幸福度の低下や、生活環境の悪化により身体機能が低下する。特に高齢者の場合は、脳機能の低下により認知症に発展する可能性もあり、大きな社会問題となっている。また、平成 28 年度の高齢社会白書 [2] では、日常生活において友人や知人との会合・雑談に充実感を感じる高齢者 (女性) が半数以上存在していることがわかっている。以上より、日常生活における会話は重要な意味を持っていると考えられる。

日常生活における会話機会の不足に対する対策の一つとして、日常的に対話ロボットを利用してコミュニケーションを活性化させる試みは従来から数多く研究されてきた。神田ら [3] は、人とロボットとの長期間のインタラクション実験を実施したところ、ロボットが人間社会に受け入れられるためには、人とロボットに友好的な関係を築く必要があると結論づけている。そこで、人とロボットが友好的な関係を築くためには、ロボットの振る舞いをより人間に近づける必要がある [4][5]。しかし、近年の Pepper[6] や、AI スピーカーの Google Home[7]・Amazon Echo[8] 等のタスク指向型ロボットは特定のシナリオに沿った対話の進行を前提としている。そのため、シナリオ外の内容の対話には対応できず、タスク指向のみでは日常的な対話の継続は困難である。docomo の雑談対話 API[9] や、Microsoft の女子高生 AI「りんな」[10] のように、目的のない対話である雑談型の対話ロボットも存在する。しかしながら、これら雑談型の対話ロボットは応答文の意味的破綻こそ改善されてきたものの、ユーザの興味を引くような魅力的な応答文は少ないため、日常会話のような長期にわたる対話には対応が困難である。よって、ロボットとの対話を日常的に継続するためには、ユーザが自発的に会話したいと感じるように、ユーザがロボットと話すモチベーションである対話意欲を高める人-ロボット間の対話方法を考案する必要がある。

この課題に対し、Minami ら [11] はソーシャルメディア上のコメント (以下、ソーシャルコメント) を使ったテレビ視聴型雑談ロボットを提案した。ソーシャルコメントは人が書き込んでいるため、従来では機械的生成が困難であったユーモアを含む魅力的な内容が含まれている。Nishimura ら [12] は Minami らのロボットとテ

テレビ番組の盛り上がり場면을共有することで、ユーザに共感の印象を与えて対話意欲を向上させた。一方、これらの手法ではソーシャルコメントを無作為に選択して発話しており、ユーザの対話意欲を高めるソーシャルコメントを選択して発話しているわけではない。そのため、場合によってはユーザの意思にそぐわない内容となり、ユーザの対話意欲を低下させてしまう問題がある。ユーザの対話意欲をより向上させるためには、対話意欲を反映した選択基準が必要となる。しかし、従来研究における発話文の選択基準は、ユーザの発言に対するロボットの発話の整合性 [13][14] にとどまっており、ユーザの対話意欲を基準とした発話文選択は実現されていない。

そこで、本研究ではテレビ視聴時というシチュエーションにおいて、ユーザの対話意欲の向上する発話文を選択して発話し、より対話意欲を向上させる雑談ロボットを実現する。テレビ視聴型雑談ロボットの対話において、各発話文がユーザの対話意欲に与える影響は不明であるため、まず発話文が与えるユーザの対話意欲の影響を主観評価として収集する。次に、収集した発話文と対話意欲の主観評価を用いて、未知の発話文がユーザに与える対話意欲への影響を評価する。

以下、本論文は2章でロボットとの日常的対話を目指した関連研究について述べる。3章では、発話文を対話意欲の観点で主観評価するためのテレビ視聴型雑談アプリケーション「Dabelive」について詳述する。4章では、発話文がユーザに与える対話意欲への影響を評価する手法を説明し、5章の実験にて評価の精度を明らかにする。そして6章で本研究のまとめと今後の展望について述べる。

第 2 章 ロボットとの日常的対話に関連する研究

本章では、本研究と関連する従来研究について紹介する。まず、古くから社会心理学の分野で研究されてきた人と人の関係構築を参考した人とロボットとの関係構築について述べる。次に、人がロボットと関係を構築するために必要とされる対話が続けるモチベーション (対話意欲) に関して、対話意欲を高めるためにロボットに必要とされる要素を 3 つ紹介する。その中で、特に本研究が対象とするロボットの発話内容について、テレビ視聴型雑談ロボットを紹介する。次に、テレビ視聴型雑談ロボットが抱える発話文選択の問題について、従来の発話文評価法を挙げて議論する。最後にヒューマンロボットインタラクションの分野における本研究の位置づけを述べる。

2.1 人とロボットとの関係構築

人間社会において、人は同僚・友人・恋人・家族のように人同士で関係を持って生活している。社会心理学においては、人は自身にメリットがある関係 (交換関係) や、信頼や相互扶助のルールが重視される関係 (共同的关系) を築くと言われている [15]。これらの関係性が深いほど、長期に渡って関係が継続し、コミュニケーションも多くなることが多い。逆に、関係が浅いとコミュニケーションが少ない傾向にある。

人とロボット間のコミュニケーションにおいても、コミュニケーションを増やすためにはロボットと良好な関係性を築く必要がある。そこで、ロボットと長期間触れ合った場合、人同士と同様にロボットとも関係性が構築できる可能性を検証した実験を神田ら [3] が実施した。彼らは小学校に対話ロボットを 9 週間に渡って設置し、子供たちにロボットと触れ合わせた。子供たちは実験初期では目新しいロボットに興味を示し、ロボットの周りには常に誰かが滞在していたが、実験が進むにつれて興味が薄れ、ロボットと話す人数が減少していった。最終的に、被験者 37 人中 10 人が、半分以上の日数の間ロボットと対話し、友好的な関係を築いたと結論づけている。関係を築いた子供たちの中には、自身の秘密をロボットに打ち明けた者もあり、心理学の自己開示の観点からみても深い関係性であるといえる。しかし、関

係を築けた子供たちは一部であり、他の大多数の子供たちは1週目でロボットに飽き、対話をやめてしまっている。このことから、将来的にロボットが人間社会に受け入れられるためには、より多くの人と関係が築けるよう、より効果的な人-ロボット間の対話手法を確立する必要があるといえる。

2.2 ロボットとの対話におけるユーザの対話意欲

人がロボットと関係を構築し、より長期間、日常的にコミュニケーションする社会を実現するためには、人が自主的にロボットと対話をしたいと感じる必要がある。本研究では、このロボットと対話を続けるモチベーションを対話意欲と定義する。

対話意欲の向上を目的とした従来研究はいくつか報告されてきた。宮澤ら [4] は、ロボットとの対話において、ユーザの対話意欲を向上させる要素として、以下を主張している。

1. ロボットのデザインやふるまいを人に近付ける
2. 発話の音響的特性・言語的特性を人に近付ける
3. 発話文の新奇性を高める

1 に関しては、ロボットのデザインは単純に人に近ければ良いわけではなく、対話の環境・シチュエーションや、ロボットの機能・動作とあったデザインにする必要がある。心理学では、人は自身に似た人を好む性質があると明らかになっており [16]、これはロボットとの対話でも同様である [17]。2 に関しては、適切な発話の速度・ピッチ・イントネーションや発話量が対話意欲の向上につながる。Nishimura ら [18] は、心理療法で信頼関係を築く技術とされているペーシングをロボットに適用し、ユーザと信頼関係を構築する手法を提案している。ペーシングとは、聞き手がミラーリング・マッチング・バックトラッキングの3つからなるスキルを用いて話し手に合わせ、安心感や親しみやすさを生む技術である。ミラーリングで体の動作、マッチングで声に関する音響特性、バックトラッキングで発話内容を話し手と揃える。ペーシングは心理療法に限らず、ビジネスや教育など様々な分野で用いられている技術であり、人とロボットの対話にも一定の効果があつた。

3 の発話内容については、無意識下で作用する 1 や 2 よりも話し相手が意識する

要素であるため、対話意欲に大きな影響を及ぼすと予想される。以上の理由から、本研究では特に 3 に該当する発話文の内容と対話意欲の関係に焦点を当てている。宮澤らは 3 について、聞き手の予想を上回る内容や、期待を裏切る内容が、聞き手を飽きさせることなく対話意欲を向上させると述べている。発話文の新規性を高める手法の一つに、発話文にユーモアのある内容を含める手法がある [19][20]。ユーモアのある発話文は、聞き手がより積極的にロボットと対話を続ける効果が確認されているが、自動生成したユーモア文は前後の発話文と意味的整合性が無い場合があるという課題が残されている。ユーモアの内容は話題や会話のシチュエーションに依存するため、日常対話における様々な状況に対応するには困難である。

2.3 テレビ視聴時における雑談ロボット

前述した発話文にユーモアを含める手法の問題点を解消した手法として、Minami ら [11] のテレビ視聴時における雑談ロボットがある。彼らは人とロボットがテレビを共同で視聴する環境において、テレビの内容に関する話題について話す雑談ロボット（以下、テレビ視聴型雑談ロボット）を提案した。図 2.1 にテレビ視聴型雑談ロボットとの対話風景を示す。このテレビ視聴型雑談ロボットは、リアルタイムで放送しているテレビ番組の内容に対してコメントをする形で発話する。この発話にはソーシャルメディア上のユーザがテレビを見て書き込んだ実況コメント（ソーシャルコメント）が使用されているため、テレビの内容に合致したユーモアのある魅力的な発話が実現できている。ソーシャルコメントのみではユーザの発話に返答できないため、この研究ではユーザの発話を認識し、機械的に自動生成した応答文と、相槌・復唱を組み合わせ、人同士の対話に近い自然な対話を実現している。図 2.2 にテレビ視聴型雑談ロボットとの対話例を示す。

テレビ視聴型雑談ロボットの処理時間によって発話内容と番組の内容が食い違う問題や、番組の音声とロボットの発話が重複して聞き取り辛い問題に対しては、Nishimura ら [21] がテレビ番組を録画してタイムシフト再生することで、事前に番組内容を解析して適切なタイミングでロボットを発話させる手法を提案している。適切なタイミングでの発話はテレビ視聴型雑談ロボットをより人間らしく感じさせ、ユーザの対話意欲の向上に貢献した。



図 2.1 テレビ視聴型雑談ロボットとの対話風景 [11]

以上のように，テレビ視聴型雑談ロボットはテレビ視聴時という限られた環境下ではあるが，ユーザの対話意欲を向上させる可能性を示した．しかし，発話するソーシャルコメントがユーザの対話意欲に与える影響は考慮されておらず，新しいコメントが無作為に選ばれる．ソーシャルコメントには，ユーザの対話意欲を向上させるコメントだけでなく，対話意欲に影響のないコメントや，逆に対話意欲を低下させるコメントも含まれる．例えば，「あー」「まあ」「えー」などの感嘆詞のみの文や，テレビ番組の出演者への誹謗中傷といったコメントである．このように様々な種類の文が混在しているソーシャルコメントの中から無作為に選択して発話すると，ユーザの対話意欲を意図せず低下させてしまう問題が発生する．したがって，よりユーザの対話意欲を向上させるためには，対話意欲を基準とした発話文の選択が必要になる．

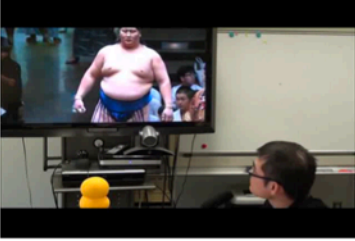
被験者とテレビ番組の状況	被験者とロボットの対話内容	コメントNo	タイムスタンプ
ロボットが美人な観客を見つけたと 発話した場面 ユーザは相撲の取り組みでは無く観客に 注目していたロボットに驚いている	(ソーシャルメディアコメント) 今日は美人が多いなあ	No.1	0:00
	(ユーザ) 観客の方観てるのかよ	No.2	0:02
	(相槌) うん	No.3	0:06
	(復唱) 一般観客か	No.4	0:09
	(ソーシャルメディアコメント) ひいつ、でけえ！	No.5	0:12
画面に映った関取の大きさにロボットが 驚いた場面 ユーザもこの関取は大きいと感じ ロボットに同意している	(ユーザ) たしかにデカイよね	No.6	0:14
	(相槌) うん	No.7	0:17
	(自動応答文生成) スポーツは わかりやすいですね	No.8	0:22
	(ユーザ) 相撲はわかりやすいね	No.9	0:25
ソーシャルコメントを利用できなかった為 自動応答文生成でロボットの発話文が 生成された場面 ソーシャルコメントでは無いことに 気付かずユーザは対話を継続させた			

図 2.2 テレビ視聴型雑談ロボットとの対話例 [11]

2.4 対話時の発話文の評価基準

ヒューマンロボットインタラクションの分野において、ロボットとの対話に関する研究はタスク指向型・非タスク指向型の対話ロボットの2つに分類される。タスク指向型の対話は、ロボットが請け負ったタスクを完遂するために行う対話である。例として Pepper を用いた店頭販売のロボットや、Google Home や Amazon Echo などの AI スピーカーを用いた情報検索が挙げられる。非タスク指向型の対

話は、目的のない対話のことであり、主に雑談が当てはまる。本節では、これら 2 種類のロボットの対話を評価する手法について、従来研究を説明する。

2.4.1 タスク指向型対話ロボットの評価

タスク指向型対話ロボットは、タスクが決まっているため対話内容が限定される。そのため、非タスク指向型に比べて発話文の生成難易度は低い。よって、タスク指向型対話ロボットの分野では、対話の質を評価する尺度が提案されている。その一つに PARADISE[22] というロボット評価のためのフレームワークがある。この PARADISE では、ロボットとの対話の満足度をタスクの成功度とタスクのコストから推定する。この手法は様々なタスク指向型の対話ロボットの研究で適用されている [23][24][25]。

長期的なロボットの使用を想定した場合、ロボットとの各対話における満足度はユーザの対話意欲に大きく影響すると考えられる。しかし、PARADISE を雑談で用いるには問題がある。PARADISE はタスクを表す属性値が必要であるため、事前にタスク (話題) を指定しなければならないが、雑談において事前に網羅的に話題をリストアップすることは困難である。したがって、非タスク指向型対話の対話意欲の評価に PARADISE は適用困難である。

2.4.2 非タスク指向型対話ロボットの評価

非タスク指向型対話ロボットは話題が限定されないため、話題に適した発話文の生成や、ユーザの発話に対して意味的に破綻しない返答の生成が課題となっている。そのため、非タスク指向型対話ロボットの分野においては、ロボットの発話文の評価尺度は、前後の文脈との整合性を評価したものが多い。Liu ら [14] は、ロボットの対話を自動評価する評価法において、生成された発話文の整合性や合理性を評価する手法についてまとめている。彼らによると、発話文の整合性や合理性の評価手法は従来からいくつも提案されているが、人間の主観的な評価と相関が弱いという課題があり、現在も研究が続けられている。

Higashinaka ら [13] は、人とロボットの対話において、対話の破綻を防ぐために

ロボットの不適切な発話を検出する課題に取り組んでいる。具体的には、不適切な発話の検出手法を確立するために、対話データセットの構築や発話文の評価法、対話破綻検出のワークショップの結果報告を述べている。

このように、現在までに研究されている非タスク指向型対話ロボットの主な評価手法は、発話の意味的整合・対話崩壊の防止に焦点が当てられており、前節で述べたようなユーザの対話意欲を基準とする評価尺度の研究は少ない。

2.5 本研究の位置づけ

本研究では、2.3 節で述べたテレビ視聴型雑談ロボットの発話文がユーザの対話意欲を意図せず低下させてしまう問題に取り組む。人同士のコミュニケーションでは、友好的な関係を築きたい相手に対して、相手の好む内容や共感する内容を選択して話す。同時に、相手の機嫌を損ねたり、相手を傷つける内容は、関係を築きたい相手に対しては選択しない場合が多い。このような発話内容の取捨選択は、人とロボットの間にも同様に効果があると予想される。よって、テレビ視聴型雑談ロボットでも、よりユーザにとって魅力的で対話意欲を向上させる内容の発話文選択が有効であると考えられる。そこで、対話意欲を基準とした発話文選択を実現するために、発話文を対話意欲の尺度で評価することを考える。しかし、2.4 節で述べた通り、従来提案された発話文の評価尺度では非タスク指向型対話におけるユーザの対話意欲を評価することは困難である。そのため、ユーザの対話意欲を基準とした新たな発話文の評価法が必要である。よって、本研究ではまず、テレビ視聴型雑談ロボットが発話した内容がユーザの対話意欲に与える影響を、データセットとして収集する。そして、収集した対話意欲に関するデータセットから、ユーザの対話意欲を推定して発話文を評価する手法を提案する。

第 3 章 テレビ視聴型雑談アプリケーション Dabelive

ユーザの対話意欲を基準とした発話文の評価をするためには，ユーザにとって魅力的だった発話内容を特定する必要がある．そのために，本研究ではテレビ視聴型雑談ロボットの発話した内容がユーザの対話意欲に与える影響を，ユーザの対話意欲データとして収集する．本章では，対話意欲のデータセットを収集するシステムとして，テレビ視聴型雑談ロボットをスマートフォンアプリケーションという形で実装したテレビ視聴型アプリケーション「Dabelive (だべらいぶ)」について述べる．

ロボットの発話した内容がユーザの対話意欲に与えた影響を収集するには，以下の要件が必要となる．

- ロボットの発話についてユーザ自身が感じた対話意欲の変化を入力できるインターフェース
- より多くのユーザに対話してもらうために対話システムを一般公開する

そこで，本研究ではスマートフォン上で動作するテレビ視聴型雑談アプリケーション「Dabelive (だべらいぶ)」を利用する．Dabelive とは，スマートフォン上で動作する対話エージェントが，リアルタイムに放送されているテレビ番組についての発言をするアプリケーションである．ユーザは図 3.1 に示すように，テレビ視聴時に Dabelive を起動し，アプリ内のエージェントと共にテレビを視聴する．エージェントはユーザが視聴しているテレビ番組についての発言をし，ユーザに発言を促す．Dabelive で扱う発話文はテレビ番組の内容に対応するソーシャルコメントとした．

Dabelive は，iOS と Android 上で動作するスマートフォンアプリケーションとして株式会社 amirobo tech が開発し，App Store[26] (iOS) と Google Play Store[27] (Android) に無料で一般公開されている．よって，より多くのユーザからソーシャルコメントに対する対話意欲の評価を収集できる．



図 3.1 Dabelive の使用例

3.1 Dabelive のシステム構成と機能

Dabelive のシステム構成を図 3.2 に示す。Dabelive はクライアントであるスマートフォンアプリケーションと、ソーシャルコメントを収集・配信するコメントサーバからなる。以下、Dabelive に実装した機能について説明する。

3.1.1 放送局の選択

ユーザが視聴しているテレビ番組を特定するために、図 3.3 に示すようにユーザに視聴している放送局を選択させる。表示される放送局名は、図 3.4 に示す設定画面の「お住いの地域」にて選択した地域で視聴可能な放送局名が表示される。

3.1.2 ソーシャルコメントを利用したエージェントの発言

図 3.5 に示す通り、エージェントはリアルタイムに放送されているテレビ番組の内容についてのソーシャルコメントを一定間隔で発話する。発話した内容はディスプレイ上に表示される。

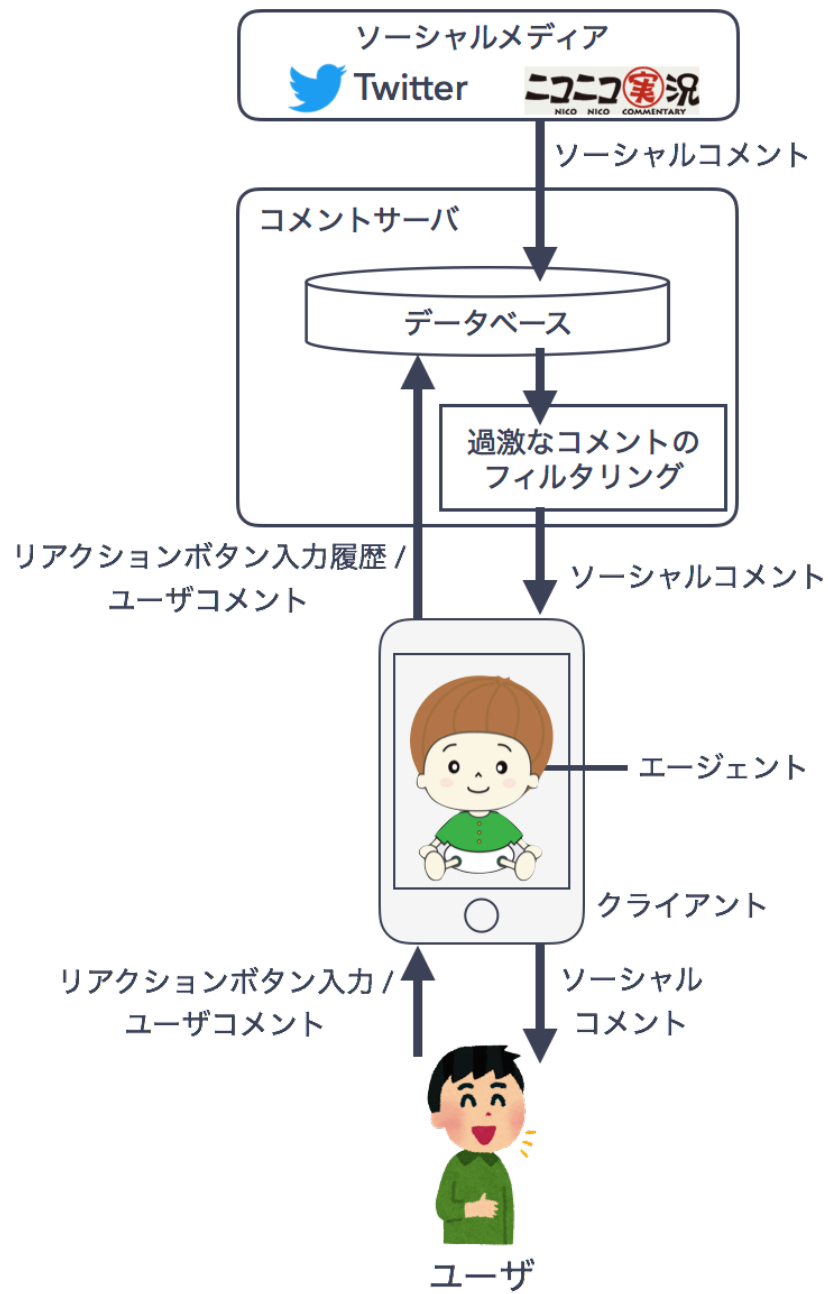


図 3.2 Dabelive のシステム構成概要

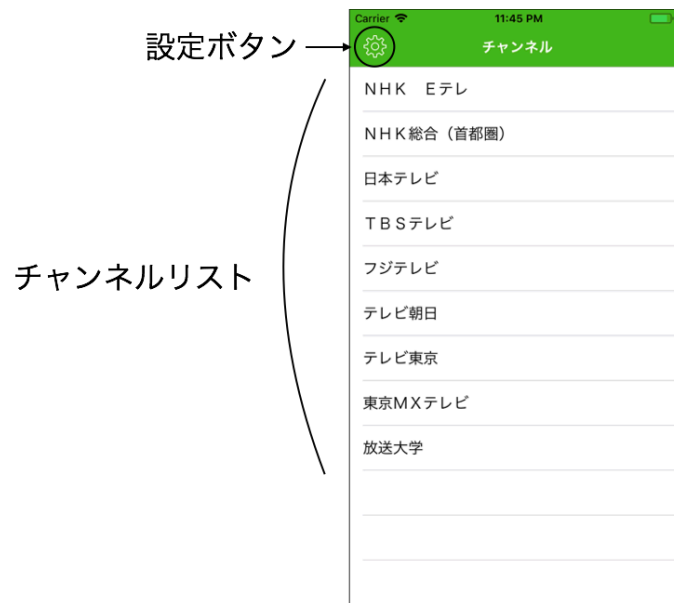


図 3.3 放送局の選択画面

3.1.3 エージェントの発言に対する評価

図 3.5 に示すエージェント対話画面には，エージェントが発話したソーシャルコメントの内容に対して，ユーザの評価を反映できるインターフェースを用意している．ユーザは，エージェントの発言に同意や共感を感じたときには「いいね」、否定や反論のときは「うーん」の評価ボタンを押す．評価ボタンを押した際には，エージェントが表 3.1 に示す返答文を発話する．

表 3.1 評価ボタンの返答文

評価ボタン	エージェントの返答文
いいね	いいね，だよね，うんうん
うーん	そうかな？，えー，そっかー



図 3.4 設定画面

3.1.4 エージェントの発言内容のフィルタリング

ソーシャルコメントには公序良俗に反する語句や、他者を誹謗中傷する内容のものが含まれている過激なコメントが存在する。そのため、過激なコメントをエージェントが発話すると、ユーザが不快に感じる可能性がある。よって、Dabelive はエージェントが過激なコメントを発話しないようにするソーシャルコメントのフィルタリング機能を持つ。フィルタリング機能は、図 3.4 の設定画面から有効にできる。

3.2 ソーシャルコメントの収集

ソーシャルコメントの収集処理は Minami らの手法 [11] を用いる。ソーシャルコメントはニコニコ実況 [28] と Twitter[29] から最新のコメントをリアルタイムで

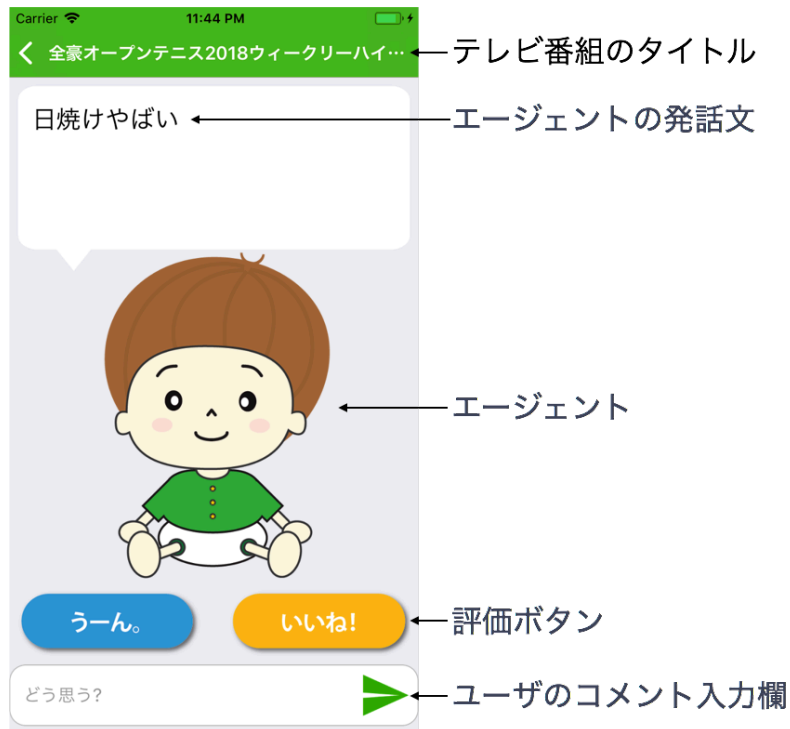


図 3.5 エージェント対話画面

表 3.2 ソーシャルコメントの属性情報

属性名	説明
station	コメントを書き込んだソーシャルメディアのユーザが視聴していた放送局
timestamp	ソーシャルコメントが書き込まれた時刻
user	ソーシャルコメントを書き込んだユーザを特定する文字列
comment	ソーシャルコメントの本文
source	取得元のソーシャルメディア名

収集する。一つのソーシャルコメントに対して表 3.2 に示す 5 つの属性情報をデータベースに記録する。

テレビ番組には全国放送のものとローカル放送のものが存在するが、ソーシャルコメントは放送局単位で収集しているため、全国放送とローカル放送の区別がつか

表 3.3 テレビ番組の iepg データ詳細

属性名	説明
pid	テレビ番組の識別番号
station	テレビ番組が放送される放送局名
start	開始時刻
end	終了時刻
title	テレビ番組のタイトル
genre	テレビ番組のジャンル
subgenre	

かない。そのため、ローカル放送の時間帯にソーシャルコメントを配信してしまうと、視聴地域によってはテレビ番組の内容と対応していないソーシャルコメントがクライアントへ送信されてしまう。これに対処するため、Dabelive のコメントサーバは全国放送の時間帯にのみソーシャルコメントを配信する仕様にした。全国放送の時間帯は、地方局を含めた全国の放送局の番組表を G ガイド テレビ王国が提供する iepg データ [30] から特定する。各番組の iepg データからは表 3.3 に示す情報が取得できる。各地域の iepg データを比較し、pid または title が同一であれば全国放送とみなす。

3.3 エージェントによるソーシャルコメントの発話

クライアントはコメントサーバへユーザが選択した放送局のソーシャルコメント要求を送信する。コメントサーバはクライアントからソーシャルコメント要求を受信すると、クライアントへソーシャルコメントを一定間隔で送信する。送信するソーシャルコメントは送信時点で最新のコメントを選択する。ソーシャルコメント要求時に過激なコメントのフィルタリング要請が含まれていた場合、コメントサーバの禁句データベース内に記録されている禁句を含むソーシャルコメントは送信し

表 3.4 使用する音声合成 API

OS	API
iOS	AVSpeechSynthesizer[31]
Android	TextToSpeech[32]

ない。

クライアントでは受信したソーシャルコメントのテキストを元に音声合成処理を施して発話する。使用する音声合成 API は表 3.4 に示すスマートフォン OS 標準のものとする。

3.4 ソーシャルコメントに対するユーザの対話意欲評価の収集

前述の通り，Dabelive ではエージェントの発言に対して，ユーザが評価ボタンとコメント入力により返答できる。これらのインターフェースによって入力された情報を，ソーシャルコメントに対するユーザの対話意欲評価とする。具体的には，全ユーザの評価ボタンの入力履歴をソーシャルコメント毎にコメントサーバ内のデータベースに記録する。

ユーザが書き込んだコメントは，Dabelive 由来のソーシャルコメントとしてコメントサーバのデータベースに記録し，他のユーザが利用するクライアントへ配信する。Dabelive 由来のソーシャルコメントを他者に配信することで，より多くのユーザがコメントを評価できる。

第 4 章 テレビ番組に関する発話文の評価

本章では，ユーザの対話意欲を基準としたロボットの発話文の評価法について述べる．この評価には，3 章で述べた Dabelive を利用して収集した発話文とユーザの対話意欲のデータセットを用いる．Dabelive は，アプリ内のロボット (エージェント) がソーシャルコメントを発話文として発話する．

より対話意欲を向上させる発話文を選択するためには，発話文がユーザに与える対話意欲の影響をモデル化し，発話文を自動的に評価する必要がある．本研究では，このモデルをユーザ対話意欲モデルと定義し，以下の特徴量を用いた重回帰分析で解く．ユーザ対話意欲モデルの作成フローを図 4.1 に示す．図 4.2 のように，作成したユーザ対話意欲モデルに，新規の発話文の特徴量を入力すると，入力した発話文の対話意欲評価値を推定して出力する．

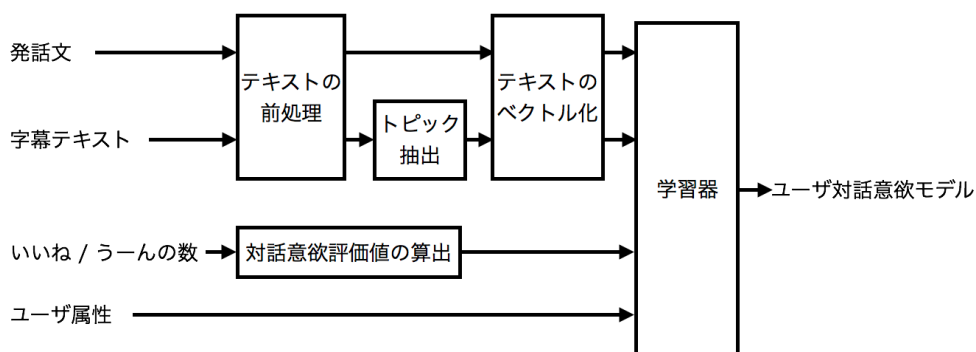


図 4.1 ユーザ対話意欲モデルの作成フロー

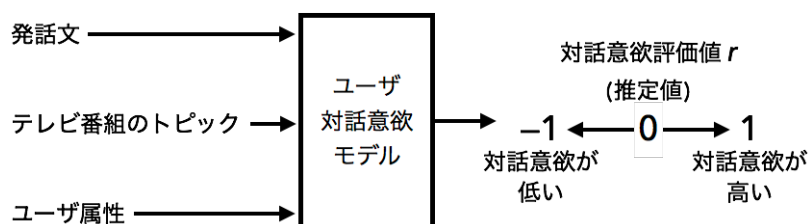


図 4.2 ユーザ対話意欲モデルによる対話意欲の推定

4.1 ユーザ対話意欲モデルに用いる特徴量

モデル作成には、以下の特徴量を用いる。

- 説明変数 1: 発話文 (4.1.1 項)
- 説明変数 2: 発話文を発話したときのテレビ番組のトピック (4.1.1 項)
- 説明変数 3: ユーザ属性 (4.1.2 項)
- 目的変数: 発話文の対話意欲評価値 (4.1.3 項)

以下、各特徴量の詳細を述べる。

4.1.1 発話文とテレビ番組トピックの関連付け

発話文は、テレビ番組のトピックに関連した内容である。よって、似たテレビ番組のトピックのときには、似た発話文が存在しやすいと予想される。テレビ番組の内容に深く関連している発話文が多いため、発話文単体では意味の分からない内容が多い。したがって、発話文とテレビ番組トピックには、共起関係が成り立つと仮定し、発話があったときに放送していたテレビ番組のトピックと、発話文のペアを一つの特徴としてモデルに入力する。なお、テレビ番組の内容はテレビ番組の字幕テキストから後述するトピック抽出処理を経て抽出する。

4.1.2 ユーザ属性

発話文が与えるユーザの対話意欲への影響はユーザの属性によって異なると考えられる。例えば、野球はユーザによって応援するチームが異なる。そのため、ユーザが応援しているチームを肯定するようなコメントは対話意欲が高まり、逆にユーザが応援するチームを否定するコメントは対話意欲が低くなる可能性がある。

4.1.3 発話文の対話意欲評価値

発話文の対話意欲評価値とは、発話文がロボットに発話されたときにユーザの対話意欲へ与える影響を定量化したものである。対話意欲評価値は、「対話意欲が向

上した (good 評価) / 低下した (bad 評価)」の 2 値をもとに算出する。ある発話文の good 評価数を g , bad 評価数を b , 発話文の配信数を c とした時の、発話文の対話意欲評価値 r は式 4.1.1 と定義する。対話意欲評価値 r は 1 に近づくほど、より対話意欲を向上させる発話文である。逆に、 r が -1 に近づくほど、より対話意欲を低下させる発話文であることを示す。この対話意欲評価値 r は発話文が持つ対話意欲への影響力を示すため、対話意欲評価値 r をユーザ対話意欲モデルの目的変数とする。

$$r = \frac{g - b}{c} \quad (-1 \leq r \leq 1) \quad (4.1.1)$$

4.2 ユーザ対話意欲モデルの作成処理

図 4.1 に示したモデル作成フローについて、処理の詳細を説明する。

4.2.1 テキストの前処理

発話文や字幕テキストに含まれる語は、同じ単語でも半角/全角が揃っていないものや、不要なスペース、記号等含まれている。よって、前処理ではテキストに含まれるカタカナ文字はすべて全角、アルファベットや数字は半角に変換する。アルファベットに関しては、すべて小文字で統一する。スペースと記号はすべて削除する。

本研究で発話文として用いているソーシャルコメント特有の特徴として、笑いの度合いを表すネット用語の「w」が文章中に含まれることが多く、「w」の連結数が発話文毎に異なる。表記の揺れを統一するために、本手法では 2 文字以上の「w」の連続をすべて 1 文字の「w」に変換する。この「w」変換の副作用として、「www (World Wide Web)」を表現できなくなる。しかし、ソーシャルコメントにおいて「www (World Wide Web)」の使用頻度は、「w (笑いの度合い)」に比べてごくわずかである。したがって、より使用頻度の高い「w (笑いの度合い)」を優先する。

4.2.2 テレビ番組のトピック抽出

テレビ番組の字幕テキストから，ある時刻に放送されていたテレビ番組の内容が取得できる．一つの字幕テキストは 5～10 秒ほどの，ごく局所的な内容しか表現しておらず，人間の目からもテレビ番組のトピックを推測し難いものである．そこで，本手法では有安ら [33] のトピック抽出手法を参考に，分単位の粒度でテレビ番組のトピックを取得し，発話文と関係付ける．

トピック抽出の模式図を図 4.3 に示す．ある発話文に対応するトピックは，発話文が発話された時刻 t から $t - T$ までの期間 T の字幕テキスト群に含まれる名詞から決定する．ここで，テキスト群に含まれる出現頻度の高い名詞が，そのテキスト群のトピックであると仮定する．したがって，期間 T の字幕テキスト群に含まれる名詞のうち，出現頻度が高いものから N 個をテレビ番組のトピックとする．

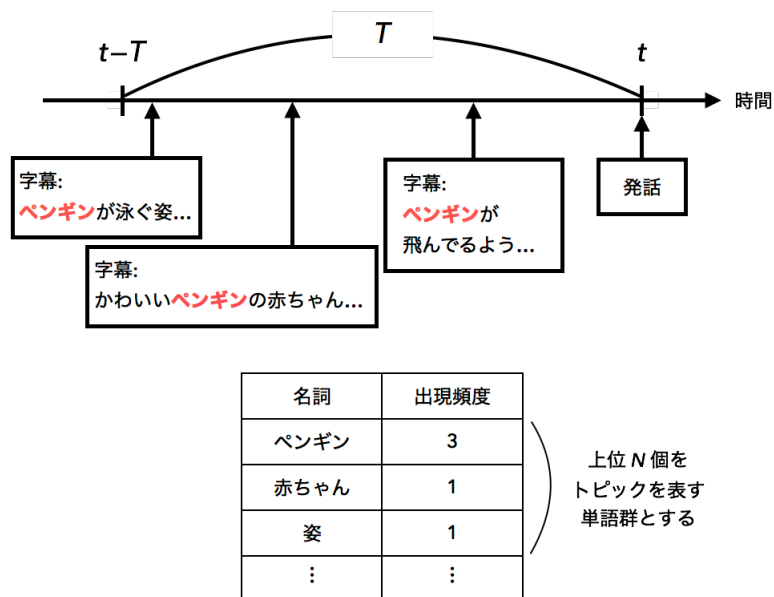


図 4.3 トピック抽出の模式図

4.2.3 テキストのベクトル化

発話文とテレビ番組のトピックである名詞群は重回帰分析を適用する際にベクトルに変換する必要がある。これらのテキストのベクトル化には Bag of words モデルを用いる。Bag of words とは、テキスト中に含まれる単語の頻度をベクトルにしたものである。ベクトル化の例を図 4.4 に示す。Bag of words を採用した理由としては、発話文とテレビ番組トピック内で共起する単語を特徴として捉え、ベクトル表現できるからである。Bag of words の処理手順は以下のとおりである。

1. 使用する全テキストを単語単位に分割
2. 使用する全テキストに出現する単語について、識別 ID を割り振った辞書を作成
3. ベクトル化するテキストに含まれる単語の出現頻度をカウント

Bag of words を用いた際の次元数は辞書のサイズと等しくなる。日本語の場合、助詞や接続詞など単体では意味を成さない単語が存在するため、本手法では Bag of words でテキストをベクトル化する際は、名詞・動詞・形容詞のみを用いる。名詞に含まれるもののうち、代名詞や疑問詞などもテキストを特徴づける用途において役に立たないとされている。このようなストップワードと呼ばれる単語は Bag of words から除外する。ストップワードのリストは SlothLib[34] の公開しているものを用いる。表 4.1 に用いたストップワード一覧を示す。

text1: 人民 の 人民 に よる 人民 の た め の 政治

辞書		
ID	単語	text1のベクトル
0	人民	3
1	政治	1
2	の	3
⋮	⋮	⋮

図 4.4 Bag of words モデルによるテキストのベクトル化例

表 4.1 ストップワード一覧

あそこ	ごっちゃ	とおり	みたい	都	誰	品	列	右	自分	以上
あたり	これ	とき	みつ	道	用	力	事	次	ヶ所	以下
あちら	これら	どこ	みなさん	府	界	法	士	先	ヵ所	幾つ
あっち	ごろ	どこか	みんな	県	会	感	台	春	カ所	毎日
あと	さまざま	ところ	もと	市	首	作	集	夏	箇所	自体
あな	さらい	どちら	もの	区	男	元	様	秋	ヶ月	向こう
あなた	さん	どっか	もん	町	女	手	所	冬	ヵ月	何人
あれ	しかた	どっち	やつ	村	別	数	歴	一	カ月	手段
いくつ	しょう	どれ	よう	各	話	彼	器	二	箇月	同じ
いつ	すか	なか	よそ	第	私	彼女	名	三	名前	感じ
いま	ずつ	なかば	わけ	方	屋	子	情	四	本当	
いや	すね	なに	わたし	何	店	内	連	五	確か	
いろいろ	すべて	など	ハイ	的	家	楽	毎	六	時点	
うち	ぜんぶ	なん	上	度	場	喜	式	七	全部	
おおまか	そう	はじめ	中	文	等	怒	簿	八	関係	
おまえ	そこ	はず	下	者	見	哀	回	九	近く	
おれ	そちら	はるか	字	性	際	輪	匹	十	方法	
がい	そっち	ひと	年	体	観	頃	個	百	我々	
かく	そで	ひとつ	月	人	段	化	席	千	違い	
かたち	それ	ふく	日	他	略	境	束	万	多く	
かやの	それぞれ	ぶり	時	今	例	俺	歳	億	扱い	
から	それなり	べつ	分	部	系	奴	目	兆	新た	
がら	たくさん	へん	秒	課	論	高	通	下記	その後	
きた	たち	ぺん	週	係	形	校	面	上記	半ば	
くせ	たび	ほう	火	外	間	婦	円	時間	結局	
ここ	ため	ほか	水	類	地	伸	玉	今回	様々	
こっち	だめ	まさ	木	達	員	紀	枚	前回	以前	
こと	ちゃ	まし	金	気	線	誌	前	場合	以後	
ごと	ちゃん	まとも	土	室	点	レ	後	一つ	以降	
こちら	てん	まま	国	口	書	行	左	年生	未満	

第 5 章 対話意欲評価の推定実験

4 章で述べた発話文の評価法を用いて，ユーザの対話意欲評価値を推定する実験を実施した．この実験より，テレビ視聴型雑談ロボットの実際の発話文に適用したときの，対話意欲評価値の推定精度を明らかにする．本章では，実験の詳細と得られた結果，結果に対する考察について述べる．

5.1 実験方法

ユーザ対話意欲モデルを作成するためには，テレビ番組の内容に関する発話文と，各発話文に対してユーザの対話意欲評価 (good/bad 評価) が必要である．よって，まず Dabelive を用いて発話文に対するユーザの対話意欲評価を収集し，ユーザ対話意欲モデルを作成するためのデータセットとする．次に，収集したデータセットを用いてユーザ対話意欲モデルを作成する．本実験では図 5.1 に示すように，ユーザ対話意欲モデルに用いるユーザ属性は省略する．

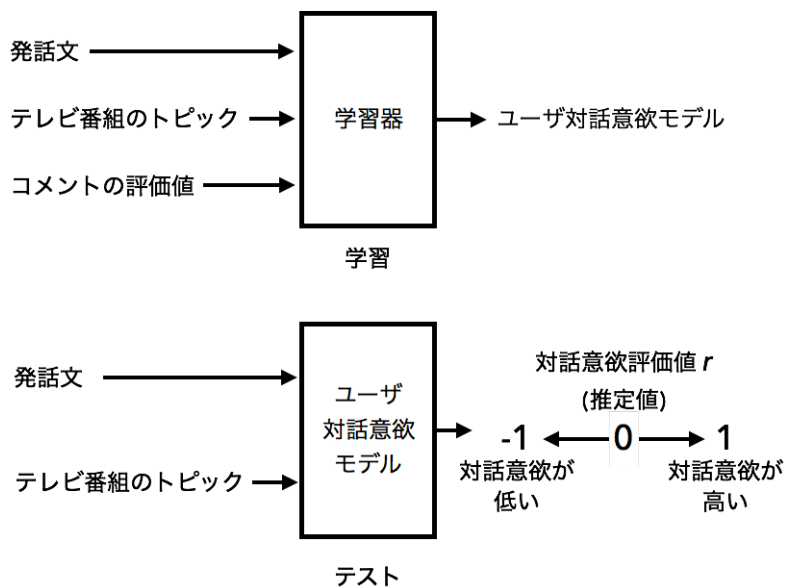


図 5.1 ユーザ対話意欲モデル (実験)

ユーザ対話意欲モデルの作成に用いるパラメータについては，トピック抽出の期間を $T = 5[\text{min}]$ ，トピックに含める名詞は出現頻度の上位 $N = 10$ とした．日本語の形態素解析には，形態素解析エンジン MeCab[35] を用い，MeCab 用のシステム辞書として Web の単語に強い mecab-ipadic-NEologd[36] を使用する．モデルが出力した推定結果は K-fold 法を用いたクロスバリデーションで評価する．本実験では $k = 5$ とする．

5.2 発話文に対するユーザの対話意欲評価の収集

ユーザの対話意欲評価を収集するために，被験者に対して Dabelive 内のエージェントと共にテレビ番組を視聴してもらった．テレビ視聴時，エージェントは放送しているテレビ番組の内容に関するコメントを一定間隔で発話する．被験者はエージェントの発話を受けて，自身の対話意欲の変化を評価ボタンで評価する．評価の仕方については，被験者に以下のように指示した．

- キャラクターともっと話したい場合は「いいね」を押してください
- キャラクターと話したくない場合は「うーん」を押してください

いいねを good 評価，うーんを bad 評価として扱う．被験者は iOS10 以上の iOS 端末に Dabelive をインストールして実験に参加した．

発話文が対話意欲に与える影響には個人差があると考えられ，被験者間で評価ボタンを押す基準も異なると予想される．よって，図 5.2 に示すように，Dabelive に評価ボタンを押した理由を表す理由ボタンを配置した．理由ボタンの選定は Russell の円環モデル [37] 上に図 5.3 のように配置されるように選択した．Russell の円環モデルとは，人間の感情を 2 次元のグラフで表したものである．縦軸が脳の活性度，横軸が快・不快を示す．

テレビ番組はスポーツとニュースの 2 種類のジャンルを選択し，それぞれのジャンルでデータセットを構築する．発話文の内容はテレビのトピックに強く依存するため，まずは似たトピックの番組として日毎に番組のトピックが大きく変化しない大相撲をスポーツの番組として選択した．スポーツのデータセットは，ユーザ対話意欲モデルの分析に用いる．一方，放送されているテレビ番組の多くは，日々ト

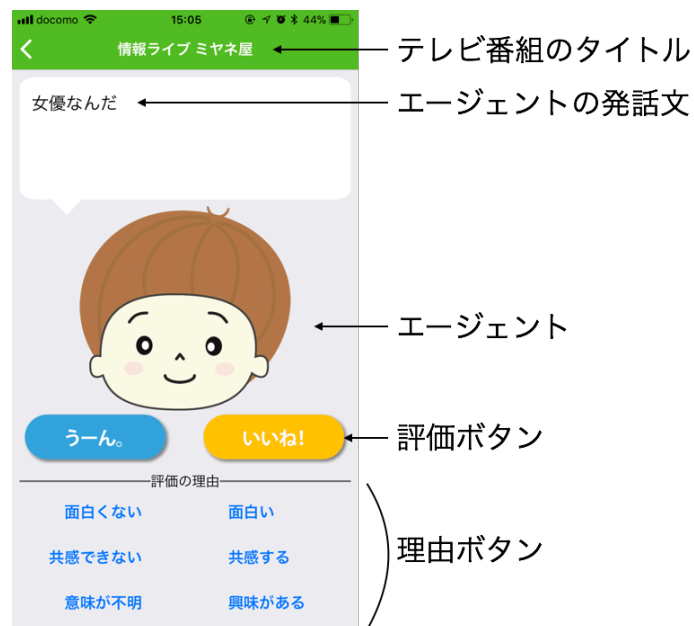


図 5.2 実験用に改良した Dabelive のエージェント対話画面

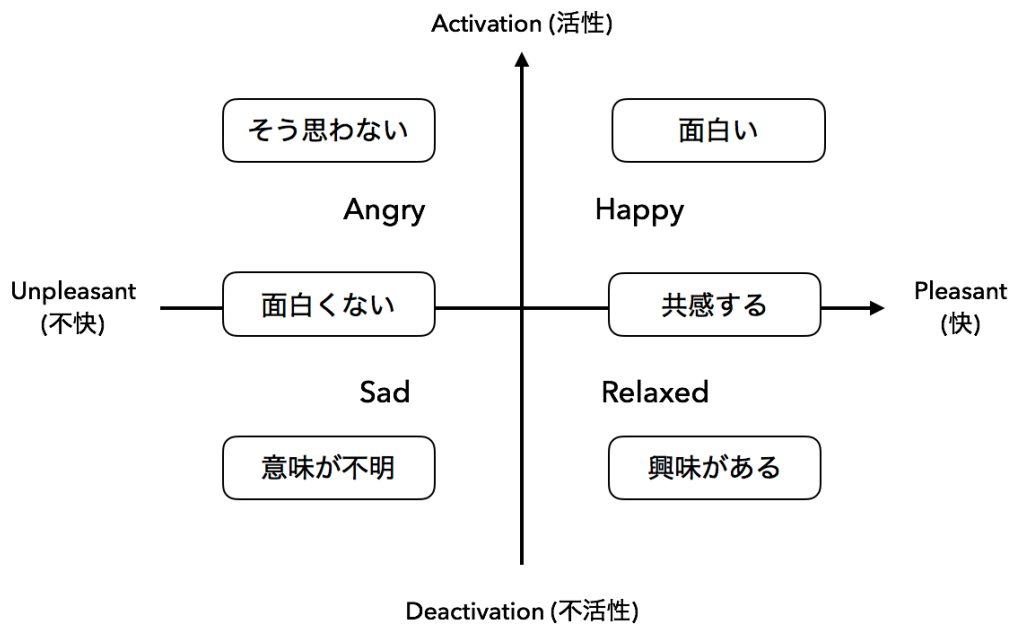


図 5.3 Russell の円環モデルにおける理由ボタンの配置

表 5.1 収集したユーザの対話意欲評価が付加された発話文

	スポーツ	ニュース
テレビ番組	大相撲十一月場所	ミヤネ屋 報道ステーション NEWS ZERO 新・情報 7days ニュースキャスター サンデーモーニング 真相報道バンキシャ！ Mr. サンデー
期間	2017/11/12 ～ 11/14 2017/11/16 ～ 11/17	2018/01/08 ～ 01/14 2018/01/22 ～ 01/28
平均視聴時間	10 時間/人	13 時間/人
被験者	20～40 代の男女 9 人	10～60 代の男女 38 人
発話文数	2,683 件	13,846 件
発話回数	57,810 回	143,533 回
評価件数	23,064 件	51,749 件

ピックが移り変わるものも多い。そのため、もう一つのデータセットとして、実際に想定されるテレビ雑談環境に近いニュース番組を選択した。ニュースのトピックは時事に影響されるため、日々トピックが移り変わる番組である。また、スポーツとニュースの 2 種類は、Nishimura ら [21] の実験でも用いられており、テレビ番組の内容を理解する情報源の違いがユーザの対話意欲に影響を与えると考察されている。彼らの実験では、スポーツは視覚から得られる情報で内容を理解するテレビ番組、ニュースは主に音声から得られる情報で内容を理解する番組として選択していた。

被験者には Dabelive に初めて触れる人が含まれるため、Dabelive に慣れさせる必要がある。よって、視聴を開始してから最初の 1 時間は練習時間としてデータセットには含めないものとする。収集したデータセットの詳細を表 5.1 に示す。表 5.1 の発話文数は、練習時間に発話されたものを差し引いた数である。

収集したデータセットを用いて Bag of words モデルの辞書を作成する際、出現頻度が極端に少ない単語・極端に多い単語は文章を表す単語として適さない。よって、本実験においては、スポーツのデータセットでは出現頻度が 3 回以下の単語、全体の 3 割以上の単語は除外、ニュースのデータセットでは出現頻度が 10 回以下の単語、全体の 3 割以上の単語は除外して辞書を作成した。数値は表 5.1 に示す 2

表 5.2 視聴後アンケートの項目

項目	回答方式
年齢	数値
性別	選択
番組自体に興味がありましたか？	1～3 の 3 段階から選択
どのような基準でコメントを評価しましたか？	自由記述

種類のデータセットのサンプル数を考慮し、経験的に決定した。

テレビ番組視聴後には番組ごとに視聴後アンケートを実施した。アンケートでは、テレビ番組への興味の度合いと、具体的な評価した理由の収集を目的とする。アンケート項目を表 5.2 に示す。

5.3 実験結果: スポーツ

テレビ番組がスポーツのときの発話をユーザ対話意欲モデルに入力して、対話意欲評価値を推定する。解析手法は、サポートベクタ回帰 (SVR) と Random Forest を採用した。SVR のカーネルは linear・RBF・多項式を用いた。Random Forest の決定木の本数は 1000 とした。 $k = 5$ の k -fold クロスバリデーションにて算出した 5 つの平均二乗誤差の平均を図 5.4 に示す。図 5.4 のグラフの縦軸は平均二乗誤差を示しており、この値が小さいほど誤差が少なく推定した対話意欲評価値が正確であることを表す。グラフは、学習データに含まれない未知のデータをテストデータとして入力し、対話意欲評価値を推定した結果である。次元数は 1182 次元である。境界値の処理として、推定値 r が $r < -1$ の場合は $r = -1$ 、 $1 < r$ の場合は $r = 1$ に丸める。また、入力するデータは正規化済みである。

図 5.4 より、テスト時に最も推定値の平均二乗誤差が少ないものは、非線形回帰である SVR の RBF カーネルであるとわかる。一方、同じく非線形回帰である Random Forest の誤差は線形回帰である SVR の多項式カーネルより大きくなっており、多項式カーネルのほうが推定精度が高い結果となっている。しかし、図 5.4 の結果はハイパーパラメータのチューニングをしていないため、本実験での

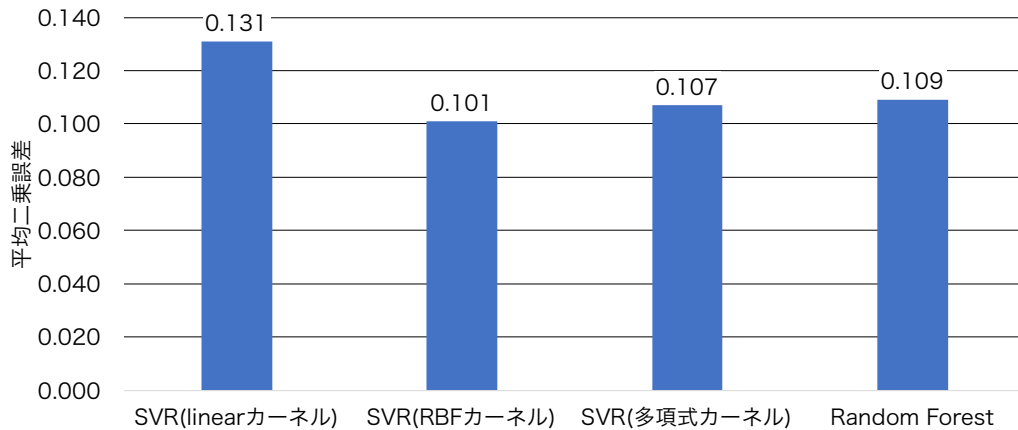


図 5.4 推定した対話意欲評価値の平均二乗誤差 (スポーツ), 1182 次元

Random Forrest の決定木本数ではモデルのバリエーションを十分に低下できていない可能性がある。SVR にも同様のことがいえ、 C, ϵ, γ をグリッドサーチにて調整すれば、ある程度の推定精度の向上が予測される。

Nishimura ら [21] の実験で、ユーザのテレビ番組への興味度合いは対話意欲に影響を与える可能性が示唆されている。そのため、視聴後アンケートで回答された 3 段階の興味度合いでデータセットを分割し、興味度合い毎にユーザ対話意欲モデルで対話意欲評価値を推定した。各興味度合いの詳細を表 5.3 に示す。テレビ番組への興味度合いは 1 時間毎に収集しているため、同一被験者が異なる興味度合いを回答している場合がある。図 5.5 に興味度合い毎の平均二乗誤差を示す。

すべての解析手法において、図 5.5 に示す興味度合い毎の平均二乗誤差は全体のものより大きく、平均で 0.119 増加している。特に、興味ありの平均二乗誤差が全体の 2 倍以上になっており、推定精度が低下した。この推定精度低下の原因は、表 5.3 に示すように、興味度合い毎のグループ分けにより、1 グループあたりの被験者数が少なくなったためだと考えられる。最も被験者数が少ない興味ありグループでは、4 章で述べた対話意欲評価値の計算式に当てはめると、1 人の評価が異なることで値は 0.3 変化する。そのため、1 サンプルの正解値の変化が大きくなり、推定の難易度が高まったのだと考えられる。

表 5.3 各興味度合いでのデータセット詳細 (スポーツ)

	興味なし	どちらでもない	興味あり
1人当たりの総視聴時間	10 時間	10 時間	9 時間
被験者数	5 人	7 人	3 人
発話文数	2683 件	2683 件	2337 件

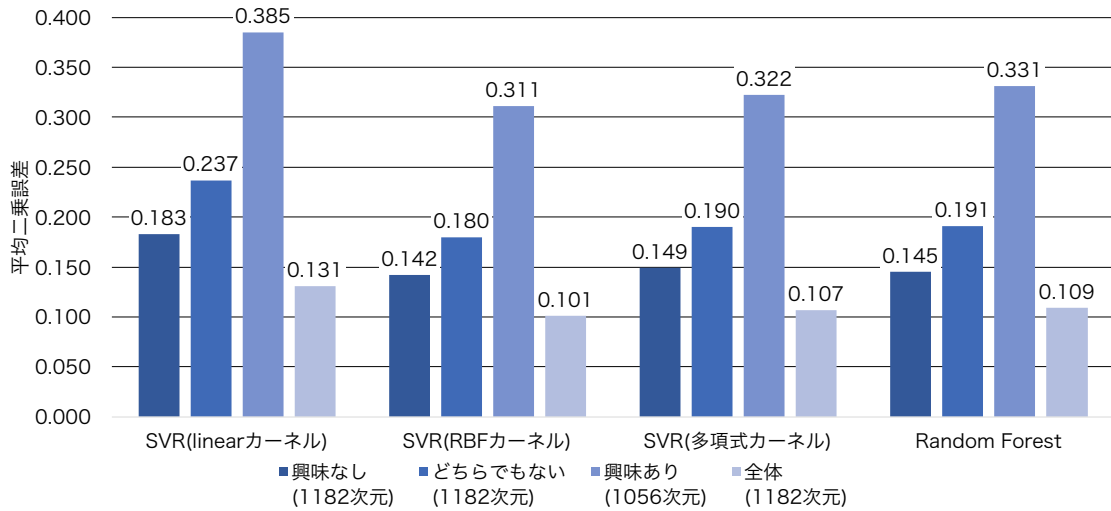


図 5.5 興味度合い毎に推定した対話意欲評価値の平均二乗誤差 (スポーツ)

5.4 考察: スポーツ

5.4.1 推定した対話意欲評価値の分布

本実験で作成したユーザ対話意欲モデルが出力した対話意欲評価値は、図 5.6 のように分布している。図 5.6 はスポーツのデータセットにおける対話意欲評価値の正解値と、SVR(RBF カーネル) を用いたときの推定値のヒストグラムである。縦軸が頻度、横軸が対話意欲評価値となっており、横軸の階級は 0.1 区切りである。ヒストグラムには、図 5.4 に示した k-fold クロスバリデーション ($k = 5$) の結果から、全 5 試行のテスト時データを用いた。

図 5.6 からわかるように、ユーザ対話意欲モデルで推定した対話意欲評価値は、0 付近のものが多く、境界値 (-1 や 1) 付近のものは非常に少ない。そのため、正

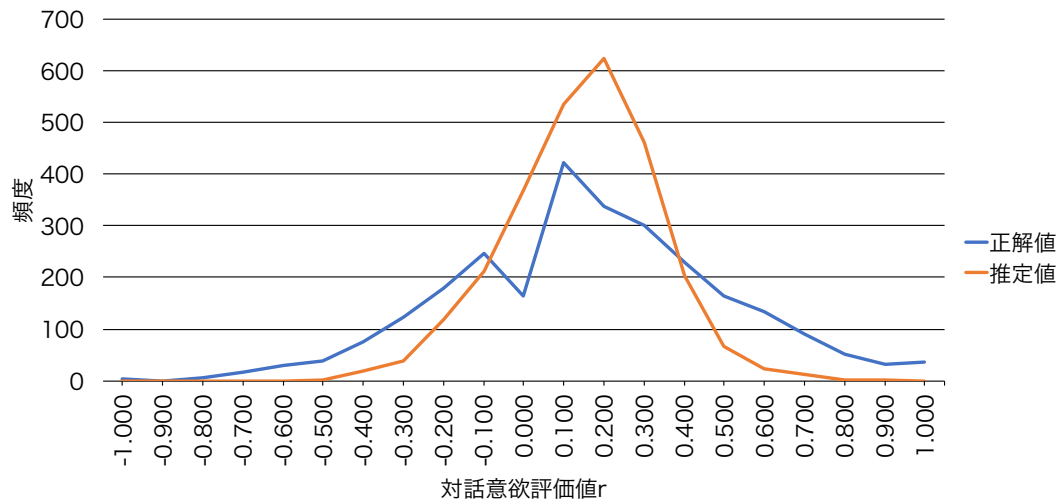


図 5.6 対話意欲評価値のヒストグラム データセット:スポーツ, 回帰手法:SVR(RBF カーネル)

解値が境界値付近の発話文に関しては，推定値の誤差が大きい結果となった．この原因としては，境界値付近のサンプル数の少なさが考えられる．図 5.6 における対話意欲評価値 r の正解値は，全 2683 サンプル中 $r \leq -0.8$ のものが 4 サンプル， $0.8 \leq r$ のものは 120 サンプルしか存在しなかった．対話意欲評価値の分布は図 5.6 のグラフから正規分布に従っているため，境界値付近のサンプルが統計的に希少であることは正しい．以上のことから，対話意欲評価値が境界値付近になる発話文はサンプル数が少ない傾向にあるため，推定難易度が高いといえる．そして，対話意欲評価値が境界値付近になる発話文は，ほとんどのユーザの対話意欲を大きく向上させる，あるいは大きく低下させる発話文である．つまり，万人の対話意欲を向上あるいは低下させる発話文は，非常に希少であり推定難易度も高いといえる．

5.4.2 テレビ番組への興味度合いの違いによる対話意欲への影響

5.3 節では，スポーツのデータセットに対してアンケート回答から発話文をテレビ番組への興味度合いで 3 分割し，対話意欲評価値を推定した．その結果，全体に比べて各興味度合いのグループでは推定値の誤差が大きくなった．本項では，その

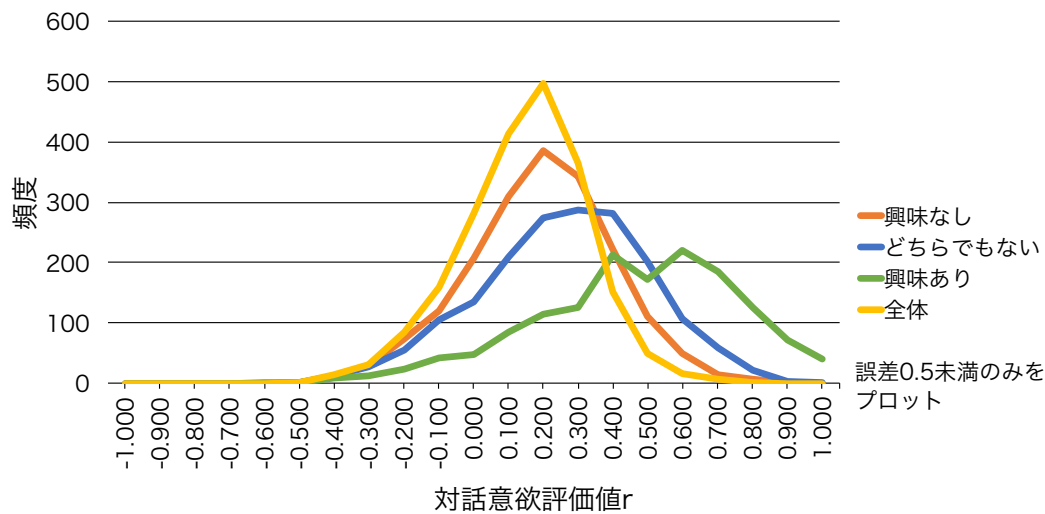


図 5.7 各興味度合いの対話意欲評価値ヒストグラム データセット:スポーツ, 回帰手法:SVR(RBF カーネル)

他に興味度合いのグループ分けが被験者間の対話意欲評価に与えた影響を考察する。各興味度合いの対話意欲評価値のヒストグラムを図 5.7 に示す。図 5.7 には、SVR (RBF カーネル) で推定した対話意欲評価値のうち、正解値との誤差が 0.5 未満のもののみをプロットしている。図 5.7 より、全体と比較して各興味度合いにおける対話意欲評価値の分布は、1 の境界値側に遷移している。特に興味度合いが大きくなるほど、より対話意欲の向上を示す 1 の境界値側に遷移している。これは、テレビ番組への興味度合いが高いほど、推定した対話意欲評価値が高いことを示している。このことから、興味度合い毎の推定結果は被験者数の減少により誤差が増加したものの、本来推定すべき対話意欲を大きく向上させる発話文をより多く推定できたといえる。

5.4.3 アンケートから判明した対話意欲に影響する発話文の傾向

被験者には視聴後アンケートにて、評価基準について質問した。回答の結果、被験者が好む発話文と嫌う発話文の傾向が窺えた。

被験者が好む発話文は、スポーツの場合、共に応援してくれる内容、共感、面白

い、番組の登場人物に関する補足情報などであった。これは宮澤ら [4] の主張する「発話文の新奇性」に該当している。逆に被験者が嫌う発話文は、アスキーアートや「w (笑いの度合い)」のみの文、テレビトピックと関係のない発話、意味の分からない内容であると分かった。特に意味の分からない内容に関しては、発話文がソーシャルコメントであるため、ネット用語が多く含まれることが影響していた。そのため、ソーシャルコメントを発話文として用いる場合は、発話前に文を加工して、口語で用いる言葉遣いや聞き手に伝わる単語に変換する必要があると考えられる。

被験者の中には、登場人物を弄る発話が実験当初は不快であったが、次第にエージェントがその人物に好意を抱いているからこそ弄っていると判明したため、対話意欲が向上した者も存在した。また、実験の中でネット用語を次第に理解し、エージェントの発話の意味がわかるようになってきたため、good 評価が増えた被験者もいた。これらの事例から、テレビ視聴型雑談ロボットとの対話には第 1 段階としてロボットとの対話に慣れるための期間が必要であると推察される。慣れ期間にユーザの対話意欲を低下させ、ユーザが対話を放棄してしまうとロボットが継続利用されなくなるため、慣れ期間はロボットが超えなければならない一つの壁であるといえる。

5.5 実験結果: ニュース

テレビ番組がニュースのときの発話文をユーザ対話意欲モデルに入力して、対話意欲評価値を推定する。解析手法は、スポーツと同様にサポートベクタ回帰 (SVR) と Random Forest を採用した。 $k = 5$ の k -fold クロスバリデーションにて算出した 5 つの平均二乗誤差の平均を図 5.8 に示す。次元数は 1224 次元である。境界値の処理も同様に、推定値 r が $r < -1$ の場合は $r = -1$ 、 $1 < r$ の場合は $r = 1$ に丸める。また、入力するデータは正規化済みである。

図 5.8 より、ニュースのデータセットにおいて最も平均二乗誤差が小さい手法は、SVR の linear カーネルと RBF カーネルであるとわかる。スポーツとニュースの誤差を比較すると、linear カーネルの場合のみニュースのほうが誤差が 0.023 少ない。その他の手法では、スポーツとニュースの間に大きな差は見られない。このことから、トピックが大きく変化するニュースのジャンルにおいても、トピックの変

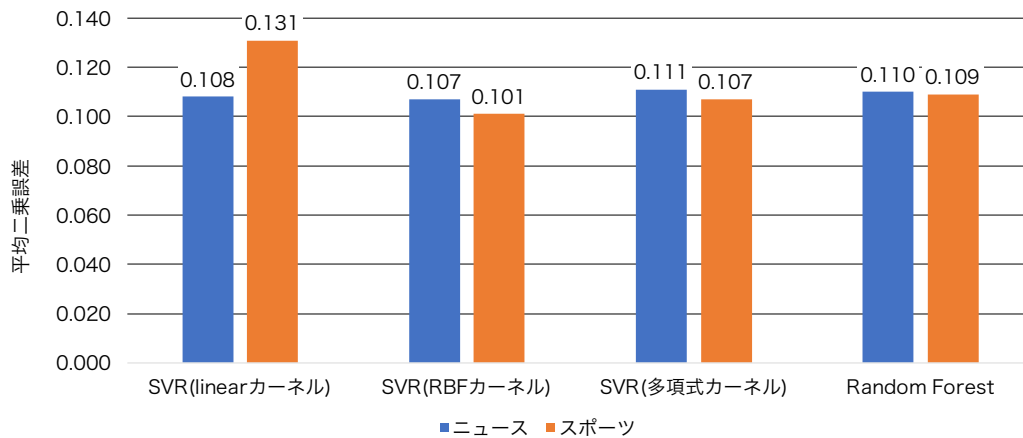


図 5.8 推定した対話意欲評価値の平均二乗誤差におけるスポーツとニュースの比較

化が少ないスポーツと同等程度の対話意欲評価値の推定精度を得られたといえる。

5.6 考察: ニュース

5.6.1 スポーツと同等の対話意欲評価値の推定精度であった理由

ニュースのデータセットを用いて推定した対話意欲評価値の精度がスポーツの場合に比べて低くなった理由について考察する。実験設計として、スポーツはニュースに比べてテレビトピックの移り変わりが小さいものとして選択した。特に本実験で採用したスポーツ番組である相撲は、ひとつの場所で2週間続けて試合があり、登場する力士も変わらない。よって、データセット内で似たテレビトピックの似た発話文が出現しやすいのだと考えられる。それに対して、ニュースのトピックは時事に依存し、1日で複数のニュースが報道される。そのため、スポーツよりはデータセット内で似たトピックの似た発話文が出現しにくい。したがって、スポーツよりニュースのほうが対話意欲評価値の推定難易度が高いと思われる。

一方、ニュースの特徴として、ニュースで扱われるトピックは1番組だけでなく複数のニュース番組で報道される。大きな事件のトピックであれば数日に渡って報道される。よって、実験設計の段階では複数のニュース番組に関する発話文を収集すれば、トピックの変化が激しいニュースでも類似する特徴を持った発話文を十分

に収集できると予想していた。そのため、ニュースのデータセットでは複数の番組を採用し、被験者の数もスポーツより多く設定した。結果的に、ニュースにおけるトピックの変化による発話文の変化は、データ数の多さで補った形となり、スポーツと同程度の推定精度を得られたと考えられる。

5.6.2 Bag of words モデルの問題点

ユーザ対話意欲モデルに用いた Bag of words モデルの問題点を議論する。実験で用いた 1 件あたりの発話文に含まれる単語数は、2~10 個ほどと字幕テキストに比べて極端に少ない傾向にある。そのため、テキスト中に含まれる単語の頻度を特微量とする Bag of words モデルでは、一つの発話文に現れる単語の頻度はたかだか 1 つとなってしまう。更に、類似したテレビトピックの発話文であっても、データセット全体で数回しか出現しない単語も多い。本実験では Bag of words の辞書を作成する際に、出現頻度が極端に少ない単語を辞書に含まない処理をしているため、稀な単語のみで構成される発話文をベクトルで表現できていない場合がある。一般的に Bag of words が用いられる文書分類等の場面では、統計的に多く出現するテキストをベクトル表現できていればよい。しかし、本研究で対象とするユーザの対話意欲を向上させる発話文は奇抜な内容や珍しい文章が含まれるため、稀な単語を正確にベクトル表現する必要がある。

Bag of words には単語の持つ意味をベクトル表現に反映できない特性がある。そのため、同じ意味だがテキスト表現上は違う単語は、別の単語として辞書に含まれてしまう。収集したデータセットでは、「稀勢の里」「キセ」など、人物の名前で多く見られた。別の単語として扱われると、「稀勢の里勝った」と「キセ勝った」のベクトルが大きく異なってしまう。

これらの問題点是对話意欲評価値の推定精度低下に起因していると考えられる。解決策としては、Word2Vec[38] という単語の意味を表現できるベクトル表現手法の採用が挙げられる。Word2Vec は単語の意味の違いを距離としてベクトル空間で表現するため、似た使われ方をする単語がベクトル空間上で近い距離に配置される。よって、稀な単語であっても似た使われ方の単語からベクトル表現可能である。

5.6.3 各ユーザが抱いた Dabelive への印象

ニュースのデータセットは、10～60 代の幅広い年齢層の男女の被験者に参加していただいた。彼らのアンケート回答から推察されるユーザの Dabelive への印象について考察する。

Dabelive はソーシャルコメントを発話文とするため、ネット用語や過激な発言が多い傾向にある。そのため、高齢になるほど Dabelive のエージェントが発話した内容は受容されにくいと予想していた。ところが、実験後のアンケートの結果からは、「レポーターを軽く野次ったコメントは笑えました (50 代女性)」, 「しばらく笑いが止まりませんでした (50 代女性)」など、好意的なコメントが多く見られた。一方、やはり発話に含まれる用語の意味が分からないという理由で bad 評価をつけたと回答した被験者は高齢になるほど増えていく傾向も見られた。ある 60 代の男性は「おちくったコメントが多すぎるのに驚いた」とコメントしており、テレビ番組の登場人物を弄るような発話が受け入れられない人も存在した。ソーシャルコメントの利用はユーザの奇をてらう発話を可能にするが、ソーシャルメディアに慣れ親しんでいない年齢層の人ほど、上記のような感想を抱きやすいと考えられる。

全体的にテレビの内容と一致していない発話や、下品な内容は bad 評価となる傾向にあり、これは全年齢に当てはまる傾向であった。漢字の読み間違いも対話意欲を低下させることがアンケートより明らかになった。上記で述べたネットの受容度合いによる影響はあるものの、対話意欲の評価の仕方は概ね全年齢で一致していた。

5.6.4 推定した対話意欲評価値に基づく発話文の選択

本研究で提案したユーザ対話意欲モデルを用いて、実際の発話文を選択した例を示し、選択された発話文の傾向を考察する。表 5.1 に示したニュースのデータセットのうち、2018/01/08～2018/01/27 までを学習データとしてモデルを作成した。そして、2018/01/28 のデータについて対話意欲評価値を推定し、1 分間で最も対話意欲評価値が高い発話文を選択した。従来手法としては、1 分間の発話文のうちランダムに一つ選択した。各手法で選択された発話文を表 5.4 に示す。

表 5.4 より、ランダムに選択した場合は下品な発話文や、攻撃的な発話文が多い傾向にある。一方、対話意欲評価値で選択した発話文は、ランダムな場合に比べて

下品・攻撃的な発話文が少なくなった。また、「・・ふさjk」、「あし」、「ふm」など文章として成立していない発話文も対話意欲評価値に基づく場合では選択されていない。対話意欲評価値の高い発話文は、テレビ番組の内容に合致しており、的を得た発話文がより多く選択されている。よって、本研究で提案するユーザ対話意欲モデルを用いて選択した発話文は、より対話意欲が高まる発話文が選択される傾向にあるといえる。しかし、1分間の選択期間に発話文が少ない場合は、ランダムに選択するときと同じ発話文が選択され、対話意欲が低下するような発話文になってしまう。そのため、適切な選択期間の設定が必要となる。今後は、表 5.4 で示したような発話文をユーザに実際に発話した際の影響について、改めて調査する必要がある。

表 5.4: 発話文の選択例

時刻	ランダムな選択 (従来手法)	対話意欲評価値による選択 (提案手法)
18:00:00	ハゲの非	資産隠しを頑張ったか
18:02:00	ちゃんと前見て謝れ	読んでる
18:04:00	大事にならなけりゃそのまま 逃げてたんだろ	大事にならなけりゃそのまま 逃げてたんだろ
18:05:00	やっぱ着物保管してるは嘘な のかな	資産を海外へ移動するのに一 生懸命
18:06:00	絶対金持ってるわな	絶対金持ってるわな
18:07:00	タイミングが悪い	結局経営者の器では無かった
18:09:00	詐欺師が詐欺認めるかよ	アカン w
18:10:00	無能神奈川県警で追い込める か	言い訳を考えるのに一生懸命
18:12:00	だれかあたしの xxxx 気持ちよ くしてくれへん w	隠したから出てきたんだろ

次ページに続く

前ページからの続き

時刻	ランダムな選択 (従来手法)	対話意欲評価値による選択 (提案手法)
18:13:00	仮想通貨のニュース自粛だっ てよ wcm 打ち切るって cc 以 外がいいだしてるらしい	いい写真やな
18:15:00	ビットコイン終了	テッちゃんかわいそう
18:16:00	思ってもないことを言うな	兄さんはこうなることを知っ ていた
18:17:00	ありえない	昭恵「私も行きます」
18:18:00	1 週間雪溶けないんだが	1 週間雪溶けないんだが
18:18:00	マスコミ w	降らなくても寒いから
18:20:00	うん謝罪なんか要らないよね	実体化しない仮想通貨のまま あ盗まれたから大ごとになら ん
18:22:00	かけえな	外国人力士もいいもんだ
18:22:00	全員首で	全員首で
18:24:00	コレは偶然ですか	こいつすぐ態度豹変して信用 できん w
18:25:00	春日野部屋のリークしたの誰 だろうね	被害者本人が声上げたのかと 思ったがちがうのか
18:26:00	cm 明け、また最初から見せら れるのかな	さっきから cm しか見てない
18:28:00	八角 4 ん w	保身に必死
18:30:00	くる・・・ぞ・・・	協会から出ていけよ
18:30:00	いいお	これは寒いだろうな
18:31:00	怖い怖い w	僕たちマスコミ
18:33:00	クスリやってんだろ w	怖い w
18:34:00	長い	勝ったな

次ページに続く

前ページからの続き

時刻	ランダムな選択 (従来手法)	対話意欲評価値による選択 (提案手法)
18:36:00	ちょっとお酒入ってるから明るい	11 人いれば余裕
18:37:00	他の離反票を確保してるなら出すだろうけど	他の離反票を確保してるなら出すだろうけど
18:38:00	流石に他人の出る出ないの言及はその部屋の力士に迷惑かかるし	バンキシヤおめでと w
18:41:00	女性をかばって	こええ
18:42:00	芸人みたい	ひええ
18:43:00	論文かけるでー	がんばれ
18:44:00	ノーマーク言ってもそばだろ w	これは無理だ w
18:46:00	3 千年って何でわかるの	落石こわすぎ
18:48:00	日本に安全地帯などないわ	ほんと信用できんな地震学者の類は
18:49:00	地球のコアを停止させたら地震なくなるよ	3 千年前とかって言い訳してるだけやろ
18:50:00	亜門は違いが分かるから	ファミチキ派
18:52:00	言えよ	量子力学の時代に未来予知とか w
18:52:00	斎藤って増毛したのに禿げだよな	唐揚げはカロリー高い
18:53:00	今年はすごいな・・・	今年はすごいな・・・
18:54:00	海岸は海岸を開拓しろよ	海岸は海岸を開拓しろよ
22:00:00	パワーおめ	アッーいいっすね

次ページに続く

前ページからの続き

時刻	ランダムな選択 (従来手法)	対話意欲評価値による選択 (提案手法)
22:01:00	春日野は理事選出るんだって な w	もはや協会いらないな
22:03:00	親方がなんで謝るの	親方がなんで謝るの
22:04:00	暴行が日常	なんで運転したんだ
22:05:00	なんや序ノ口やろこいつ	貴乃花はこれがあったから、相 撲協会に報告しなかったのか
22:06:00	マスゴミ	マスゴミ
22:08:00	ホテルがあるんだここに	くだらんコメントすんな
22:09:00	・・・おっさんはよ答えんかい	主な視聴者のじじばば引きつ けようと思ったら相撲しかね えんだろ
22:11:00	親戚って言われたら信じる	貴ノ岩って誰もう忘れたわ
22:11:00	でた、でくの棒 w	でた、でくの棒 w
22:12:00	今立行司って不在なのか	フジテレビが不祥事だよ
22:14:00	見送りじゃないか反貴乃花の 願望	寺尾も出馬しないかな
22:16:00	bba	これが協会のダメなところだ な
22:17:00	尾木は反貴乃花派	マスコミにいるいるのか
22:19:00	日本人ダメだな	邪魔
22:20:00	・・・君もおっちゃんやろ	またテレビに出まくってダメ になります
22:21:00	・・・ふさ jk	八代「相撲うぜえ
22:23:00	みえ	励みになったんやな
22:25:00	でれ	国別はじくソフトあるだろ
22:26:00	ベンツ	もうこいつが横綱でいいよ

次ページに続く

前ページからの続き

時刻	ランダムな選択 (従来手法)	対話意欲評価値による選択 (提案手法)
22:27:00	すごいニダねえ	真央ちゃんいなくなって bs し かいねーなあ・・
22:29:00	ええな	凍らせて削るしか
22:30:00	こけし界のヒロイン	大化けするか
22:31:00	あし	無理
22:31:00	スピードスケートの選手みた い	スピードスケートの選手みた い
22:33:00	その心はって聞けよ	その心はって聞けよ
22:33:00	噛み下ろした方がいい	髪型大事
22:35:00	よいやさ	成長期ってすごい
22:37:00	ベッキーは会見で嘘ついたか らな	小室はかわいそう
22:38:00	こいつを張り込もう	元から勝ってないだろ w
22:40:00	公人だけは許すはあとはどう でもいい	週刊誌として間違っていない、テ レビが叩きすぎ
22:40:00	そもそもしゅうかんしかわな い	文春がネタにして勝ってる人 がいるんだから何か問題でも あるの
22:41:00	テレビ局がずっとやってるせ いだろ	これよりコインチェックのほ うがやばいのに
22:43:00	資産隠してるだろ	読んでるだけ
22:44:00	籠池に代わるおもちゃになる かなこの社長は	わろてるで
22:44:00	詐欺ではない	コインチェックは nem 流出だ から対して話題にならない。
22:46:00	ダイエット w	失礼な

次ページに続く

前ページからの続き

時刻	ランダムな選択 (従来手法)	対話意欲評価値による選択 (提案手法)
22:48:00	nhk が自慢するのは 8k	出川コインやった
22:49:00	あいからずのいフジだな w	は一すげえな
22:50:00	ウジ「やったぜ」	木の根っこが見えた、はいはい すごいね
22:52:00	止まってるのが	こわ 1
22:53:00	だれだよ	テレビ局の映像よりきれいじ ゃん
22:54:00	やったぜ	やばいよやばいよ
22:56:00	俺も gopro 買おうかな	さっき見たけど
22:58:00	こわ	動くのかよすごいな
22:59:00	こんなん当たったら	こんなん当たったら
22:59:00	運パラ上げとこ	いるはずの無い女性の声・・・
23:00:00	ゴンドラから落ちた人	多分振り返ったら連れて逝か れるタイプの奴だわ
23:01:00	4k って地上波放送ないんでし ょう	無理だべ
23:03:00	ふ m	すごい映像
23:04:00	九死に一生を得たら笑っちゃ うよ	ひるおび感がすごい
23:06:00	あほかもうなれんわ w	富士山噴火したらまじやばい
23:06:00	1 文無しになる可能性のほうが 高い	寒いからいや
23:07:00	もう戻ってこないよ	返金するとは言ったがいつ とは言っていない・・・つま り・・・10 年後 20 年後もあり うるということ

以上

第 6 章 おわりに

本研究では、人とロボットとの日常的対話の実現を目的として、ユーザのロボットに対する対話意欲をより高める発話文の評価手法を提案した。提案した手法は、まずテレビをロボットと共同で視聴している際の雑談というシチュエーションにおいて、ロボットが発話した内容がユーザの対話意欲に与える影響をデータセットとして収集する。そして、収集した対話意欲のデータセットを用いてユーザ対話意欲モデルを構築し、新たな発話文が生成されたときにその発話文がユーザに与える対話意欲への影響を推定する。実験では、スポーツとニュースの 2 種類のテレビ番組についてデータセットを作成した。データセットの収集には、テレビ視聴型雑談アプリケーション「Dabelive」を用いてソーシャルメディアに書き込まれた実況コメントを発話させ、被験者には発話文を対話意欲の基準で評価させた。収集したデータセットを用いて対話意欲評価値を推定した結果、スポーツでは最高で平均二乗誤差 0.101、ニュースでは平均二乗誤差 0.107 の精度で推定できた。

実験からは、提案したユーザ対話意欲モデルによって、ロボットの発話文がユーザの対話意欲へ与える影響の推定が可能であると示したが、この手法で評価した発話文をロボットが実際に発話した場合については未だ明らかになっていない。発話文が及ぼしうる対話意欲への影響は評価できたが、対話意欲を向上させるタイミングや、長期でみたユーザの対話意欲の制御についても議論できていない。よって、今後の展望としては、対話意欲で評価された発話文をロボットが実際に発話した場合の調査や、ユーザの対話意欲をロボットの発話によって制御する仕組みの構築が考えられる。

本研究で利用した Dabelive は、ユーザの対話意欲を主観評価で収集したが、現状ではユーザに評価ボタンを押すモチベーションが生まれない可能性が高い。今後、より多くのユーザから発話文に対する対話意欲の評価を収集するためには、ユーザが能動的に評価するのではなく、ユーザを客観的に観察して対話意欲を自動で収集する仕組みが必要であると考えられる。最後に、Dabelive ではテレビ番組に関する発話文としてソーシャルコメントを用いたが、提案した評価法ではテレビ番組に関する発話文であればソーシャルコメントである必要はない。そして、複数人でメディアコンテンツを共有するシチュエーションにて、コンテンツに関する発話文を

用意できれば，本研究の発話文評価手法は適用できる．そのため，本研究はテレビの共同視聴に限らず，様々な非タスク指向型の対話ロボットに応用が可能であると考えられる．

謝辞

研究活動全般にわたり，指導教官の萩田 紀博教授には幅広い視野で多くの助言をいただきました．ここに感謝の意を表します．丁寧かつ熱心なご指導，ご鞭撻をいただいた神原 誠之准教授に深く感謝致します．副指導教官としての的確なご指摘をしてくださった加藤 博一教授に熱く御礼申し上げます．本研究の実験を実施するにあたり，多くの事務手続きを一手に引き受けてくださった南 あずさ秘書に深くお礼申し上げます．また，本研究で使用したテレビ視聴型雑談アプリケーション Dabelive を開発してくださった株式会社 amirobo tech 様，実験に参加してくださった被験者の方々に感謝申し上げます．そして，本研究だけでなく大学院生活で多岐にわたりお世話になった西村 祥吾先輩や，澤邊 太志先輩に心から感謝致します．その他，共に研究活動を歩み，日々切磋琢磨した環境知能学研究室の皆様にも心から感謝申し上げます．最後に，今までお世話になった両親に深く感謝致します．

参考文献

- [1] 岡田努, “現代青年の友人関係と自尊感情の関連について,” パーソナリティ研究, vol. 20, no. 1, pp. 11–20, 2011.
- [2] 内閣府, “高齢者の生活環境,” in 平成 28 年版高齢社会白書, 2016.
- [3] 神田崇行, 佐藤留美, 才脇直樹, and 石黒浩, “対話型ロボットによる小学校での長期相互作用の試み,” ヒューマンインタフェース学会論文誌, vol. 7, no. 1, pp. 27–37, 2005.
- [4] 宮澤幸希, 常世徹, 榊井祐介, 松尾智信, and 菊池英明, “音声対話システムにおける継続欲求の高いインタラクションの要因,” 電子情報通信学会論文誌. A, 基礎・境界, vol. 95, no. 1, pp. 27–36, 2012.
- [5] R. Gockley, A. Bruce, J. Forlizzi, M. Michalowski, A. Mundell, S. Rosenthal, B. Sellner, R. Simmons, K. Snipes, A. C. Schultz, et al., “Designing robots for long-term social interaction,” in Intelligent Robots and Systems, 2005.(IROS 2005). 2005 IEEE/RSJ International Conference on, IEEE, 2005, pp. 1338–1343.
- [6] ソフトバンクロボティクス株式会社, 製品情報|pepper (一般販売モデル)|ロボット|ソフトバンク, <https://www.softbank.jp/robot/consumer/products/> (2018 年 3 月 15 日 閲覧).
- [7] Google, Google home-スマートスピーカーホーム アシスタント-google ストア, https://store.google.com/product/google_home?hl=ja (2018 年 3 月 15 日 閲覧).
- [8] Amazon, Amazon | echo - スマートスピーカー, <http://amzn.asia/8Xp41Ga> (2018 年 3 月 15 日 閲覧).
- [9] 大西可奈子 and 吉村健, “コンピュータとの自然な会話を実現する雑談対話技術,” NTT DoCoMo テクニカル・ジャーナル, vol. 21, no. 4, pp. 17–21, 2014.
- [10] X. Wu, K. Ito, K. Iida, K. Tsuboi, and M. Klyen, “りんな: 女子高生人工知能,” 言語処理学会第 22 回年次大会 発表論文集, 2016.

- [11] M. Hidekazu, K. Hiromichi, K. Masayuki, and H. Norihiro, "Chat robot coupling machine responses and social media comments for continuous conversation," in *Multimedia & Expo Workshops (ICMEW)*, 2016 IEEE International Conference on, IEEE, 2016, pp. 1–6.
- [12] N. Shogo, K. Hiromichi, K. Masayuki, and H. Norihiro, "TV chat robots allowing daily-use chat with a user by synchronizing TV enthusiasm," in *Groups in Human-Robot Interaction Workshop at RO-MAN 2017*, IEEE, 2017.
- [13] H. Ryuichiro, F. Kotaro, K. Yuka, and I. Michimasa, "The dialogue breakdown detection challenge: Task description, datasets, and evaluation metrics.," in *LREC*, 2016.
- [14] C.-W. Liu, R. Lowe, I. V. Serban, M. Noseworthy, L. Charlin, and J. Pineau, "How not to evaluate your dialogue system: An empirical study of unsupervised evaluation metrics for dialogue response generation," *arXiv preprint arXiv:1603.08023*, 2016.
- [15] M. S. Clark and J. Mills, "Interpersonal attraction in exchange and communal relationships," *Journal of Personality and Social Psychology*, vol. 37, no. 1, p. 12, 1979.
- [16] D. Byrne and D. Nelson, "Attraction as a linear function of proportion of positive reinforcements," *Journal of Personality and Social Psychology*, vol. 1, no. 6, p. 659, 1965.
- [17] A. L. Baylor, "The design of motivational agents and avatars," *Educational Technology Research and Development*, vol. 59, no. 2, pp. 291–300, 2011.
- [18] N. Shogo, K. Hiromichi, K. Masayuki, and H. Norihiro, Empathetic speech synthesis applied to a chat robot to obtain the users confidence, Presented by poster in *Late Breaking Results Poster Sessions* , 2017 IEEE International Conference on Robotics and Automation, Singapore, 2017.
- [19] P. Dybala, M. Ptaszynski, R. Rzepka, and K. Araki, "Activating humans with humor—a dialogue system that users want to interact with," *IEICE*

- TRANSACTIONS on Information and Systems, vol. 92, no. 12, pp. 2394–2401, 2009.
- [20] L. Devillers, S. Rosset, G. D. Duplessis, M. A. Sehili, L. B é chade, A. Delaborde, C. Gossart, V. Letard, F. Yang, Y. Yemez, et al., “Multi-modal data collection of human-robot humorous interactions in the joker project,” in *Affective Computing and Intelligent Interaction (ACII)*, 2015 International Conference on, IEEE, 2015, pp. 348–354.
 - [21] N. Shogo, K. Hiromichi, K. Masayuki, and H. Norihiro, “A TV chat robot with time-shifting function for daily-use communication,” in *International Conference on Social Robotics*, Springer, 2017, pp. 516–525.
 - [22] M. A. Walker, D. J. Litman, C. A. Kamm, and A. Abella, “Paradise: A framework for evaluating spoken dialogue agents,” in *Proceedings of the 8th Conference on European Chapter of the Association for Computational Linguistics*, Association for Computational Linguistics, 1997, pp. 271–280.
 - [23] M. E. Foster, M. Giuliani, and A. Knoll, “Comparing objective and subjective measures of usability in a human-robot dialogue system,” in *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2*, Association for Computational Linguistics, 2009, pp. 879–887.
 - [24] M. A. Walker, R. Passonneau, and J. E. Boland, “Quantitative and qualitative evaluation of darpa communicator spoken dialogue systems,” in *Proceedings of the 39th Annual Meeting on Association for Computational Linguistics*, Association for Computational Linguistics, 2001, pp. 515–522.
 - [25] V. Rieser, O. Lemon, and X. Liu, “Optimising information presentation for spoken dialogue systems,” in *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, 2010, pp. 1009–1018.

- [26] Apple, App store, <https://www.apple.com/jp/ios/app-store/> (2018 年 3 月 15 日 閲覧).
- [27] Google, Google play store, <https://play.google.com/store> (2018 年 3 月 15 日 閲覧).
- [28] ドワンゴ, ニコニコ実況, <http://jk.nicovideo.jp/> (2018 年 3 月 15 日 閲覧).
- [29] Twitter, Twitter, <https://twitter.com/> (2018 年 3 月 15 日 閲覧).
- [30] ソニーネットワークコミュニケーションズ株式会社, テレビ番組表【g ガイド. テレビ王国】, <http://tv.so-net.ne.jp/> (2018 年 3 月 15 日 閲覧).
- [31] Apple, Avspeechsynthesizer - avfoundation | apple developer documentation, <https://developer.apple.com/documentation/avfoundation/avspeechsynthesizer> (2018 年 3 月 15 日 閲覧).
- [32] Google, Texttospeech | android developers, <https://developer.android.com/reference/android/speech/tts/TextToSpeech.html> (2018 年 3 月 15 日 閲覧).
- [33] 有安香子, 藤沢寛, and 砂崎俊二, “テレビ番組に関するメッセージ解析手法 (hybridcast 特集号),” NHK 技研 R&D, no. 133, pp. 28–34, 2012.
- [34] slothlib, Slothlib ストップワードリスト - osdn, <http://svn.sourceforge.jp/svnroot/slothlib/CSharp/Version1/SlothLib/NLP/Filter/StopWord/word/Japanese.txt> (2018 年 3 月 15 日 閲覧).
- [35] K. Taku, Y. Kaoru, and M. Yuji, “Applying conditional random fields to japanese morphological analysis,” in EMNLP, 2004, pp. 230–237.
- [36] 佐藤敏紀, 橋本泰一, and 奥村学, “単語分かち書き辞書 mecab-ipadic-neologd の実装と情報検索における効果的な使用方法の検討,” in 言語処理学会第 23 回年次大会 (NLP2017), 2017, NLP2017-B6-1.
- [37] J. A. Russell, “A circumplex model of affect,” Journal of Personality and Social Psychology, vol. 39, no. 6, pp. 1161–1178, 1980.
- [38] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient estimation of word representations in vector space,” arXiv preprint arXiv:1301.3781, 2013.

発表リスト

[1] 向井田 一平, 西村 祥吾, 澤邊 太志, 川波 弘道, 神原 誠之, 萩田 紀博. (2017). Dabelive -だべらいぶ- 魅力的な対話に向けたテレビ視聴型雑談スマホアプリ (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 IEICE technical report: 信学技報.

[2] 向井田 一平, 西村 祥吾, 川波 弘道, 神原 誠之, 萩田 紀博. (2017). テレビ視聴型雑談ロボットのためのソーシャルメディア実況コメントの評価 (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 IEICE technical report: 信学技報, 117(198), pp1-6.