

初対面対話における好感の生成と発話構成要素の予測のモデル

田中 滉己[†] 井上 昂治[‡] 中村 静[‡] 高梨 克也[‡] 河原 達也[‡]

[†] 京都大学 工学部情報学科

[‡] 京都大学 大学院情報学研究所 知能情報学専攻

1. はじめに

近年、様々な音声対話システムの研究が進められている。対話をより自然で人間らしいものにするためには、感情のような内部状態をシステムが持ち、それを表現することが効果的であると考えられる。従来の音声対話システムでは、ユーザの発話に対してシステムの応答は常に1つに決まる固定的なものだった。しかし、内部状態を考慮することで、システムとユーザとのやりとりを通して互いの内部状態がゆるやかに変化していく相互作用が生まれ、ユーザの同じ発話に対してもシステムが異なる応答を返すような柔軟な対話が期待できる。システムに感情を組み込んだ対話システムの先行研究として、PicardのAffective Computing [1]がある。また、ユーザの感情やエンゲージメントを引き出すようなシステムとしては Devalut らの SimSensei [2] や Swartout らの Ada and Grace [3] がある。

本稿では、内部状態として相手への好感を持つ対話システムを提案する。好感は、対話への積極性や態度に大きく影響する重要な要素である。システムが持つ好感を表現するために、発話を構成する「反応」「エピソード」「質問」といった要素の含め方を変化させる。また、システム内部の好感を生成するために、相槌や笑いといった「ユーザの話の聞き方」と、発話量、流暢さ、発話間隔、笑いや発話構成要素といった「ユーザの話し方」をもとに、ユーザが話題に対して持っている興味やシステムに対して持っている好感を推定する。これらは各モジュールとして独立にそのモデルが学習されるが、最終的にはモデル全体での最適化を図る。

2. コーパス

本稿で用いた対話データは、オペレータによって遠隔操作されたアンドロイド ERICA [4] と被験者（ユーザ）1名との10分程度の対話18セッションである。ERICAと被験者は初対面であり、趣味・休日の過ごし方や職業などについてお互いに質問をしたり、それに答えたり、というような会話を通じて仲を深めることが目的である。オペレータへの説明として、発話は「反応（質問への回答や、相手の話へのリアクション）」、「エピソード（回答への追加情報、自分の知識や経験など）」、「質問（話を掘り下げる質問、話題転換する質問）」という3種類の要素の組み合わせで構成し、それ以外の発話はできるだけ避けること、相手への好感や話題への興味に応じて要素の組み合わせ方を変化させることを指示した。本稿ではこの3種類の要素を「発話構成要素」と呼ぶ。

対話終了後に被験者・オペレータの両者にアンケートをとり、対話中に出た話題ごとに、以下の項目について7段階で評価してもらった。

- 自分がその話題に興味を持っていたか
- 相手がその話題に興味を持っていたと思うか

A model to control favorability and predict the utterance constructional unit in first encounter dialogue: Koki Tanaka, Koji Inoue, Shizuka Nakamura, Katsuya Takanashi, and Tatsuya Kawahara (Kyoto Univ.)

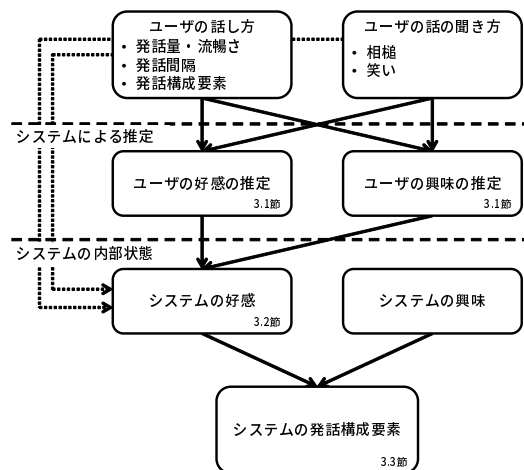


図 1: システム概要

- 自分が相手に好感を持っていたか
- 相手が自分に好感を持っていたと思うか
- 自分の相手への好感は前の話題より高くなったか
- 相手の自分への好感は前の話題より高くなったと思うか

3. 提案システム

提案するシステムの構成を図1に示す。以下、順に説明する。

3.1 ユーザの好感・興味の推定

入力されたユーザの発話音声から、ユーザの聞き方（相槌・笑い） [5]、ユーザの話し方（発話量・流暢さ・発話間隔・笑い・発話構成要素）を抽出し、それらの特徴量をもとに、ユーザが話題に対して持っている興味とシステムに対して持っている好感を推定する。

3.2 システムの好感の生成

ユーザの話題への興味とシステムへの好感の推定結果をもとに、システムがユーザに対して抱く好感を生成する。後の実験では、先述のコーパスのアンケートでのオペレータの回答を学習データとしてロジスティック回帰を行う。将来的には、ユーザの話し方・聞き方も入力として用いる。

3.3 システムの発話構成要素の選択

システムがユーザに抱く好感と、話題に対して持っている興味をもとに、発話構成要素の組み合わせを選択する。後の実験では、好感と興味は先述のアンケートでのオペレータの回答を用い、発話構成要素はオペレータの発話に対してアノテーションしたものをを用いて、ロジスティック回帰により「反応」「エピソード」「応答」それぞれの有無を分類した。

3.4 全体の最適化

3.1 から 3.3 を話題ごとに得られるアンケートデータをもとに学習したあと、ターンごとのユーザの話し方・聞き

方を入力データ、システムの発話構成要素を教師データとした誤差逆伝播法により、多くのデータで改めて全体を学習する。これにより、コーパス内でアンケートデータのない部分や、アンケートをとっていない対話データでも学習可能となる。

4. 評価

提案システムの一部について、収録した対話データを用いてその妥当性を検証した。

4.1 システムの好感の生成

オペレータが推定した被験者のシステムへの好感と話題への興味を入力としたロジスティック回帰により、オペレータが持つ被験者への好感を推定する。学習データとして、収録したコーパスのアンケートでのオペレータの回答 68 個を用い、5 分割交差検証を行った。オペレータの持つ好感は 7 段階の 4 未満を「好感なし」、4 以上 6 未満を「どちらでもない」、6 以上を「好感あり」として多クラス分類を行い、正解精度を算出したところ、0.648 となった。したがって、システムが推定したユーザが持つ好感・興味から、システム自身の好感が生成できる可能性が示唆された。

4.2 システムの発話構成要素の選択

まず、コーパスのオペレータの発言に対して発話構成要素のタグ付けを行った。そして、オペレータの持つ被験者への好感と話題への興味を入力とし、発話の構成要素「反応」「エピソード」「質問」それぞれに対して有るか無しかの 2 値分類を行うロジスティック回帰を学習した。「反応」は相手の発話が質問かどうかによって、リアクションと質問への回答に分かれるが、質問への回答は必ずあると仮定し、今回は被験者の質問以外の発話に対するリアクションの有無を対象とした。「反応」にはオペレータの発話のうち、被験者の質問以外の発話直後のもの 214 発話、「エピソード」「質問」にはオペレータの全 354 発話を用い、5 分割交差検定により適合率、再現率、F 値を計算した。結果を表 1 に示す。「質問」は高い精度で予測できたが、「エピソード」は適合率、「反応」は再現率がそれぞれ低かった。システム自身の好感・興味だけでなく、対話の文脈やユーザ発話の情報などを検討する必要がある。

また、システムが持つユーザへの好感と話題への興味、それぞれの値に対するロジスティック回帰の出力（確率値）を図 2 から図 4 に示す。実際の対話システムで発話を生成する際には、この確率に従って各要素を発生させる。「反応」の発生確率は興味が高くなるほど減少し、好感が高くなるほど増加している。これは、興味が高い話題に対してはリアクションなしで矢継ぎ早に質問をすることが多くなる一方で、好感の高い相手には丁寧にリアクションをしようとするためと考えられる。「エピソード」「質問」は好感と興味が高いほど発生確率が増加している。好感のある相手、興味のある話題ほど、自分の話を相手に多く聞いてもらいたいと思うために自分のエピソードの発話量が増え、相手の話をもっと聞きたいと思うために質問の量が増えると考えられる。

5. おわりに

本稿では、好感という内部状態を持った初対面对話システムについて述べた。初期検討として、システムの好

表 1: 発話構成要素の選択精度

要素	適合率	再現率	F 値
反応	0.763	0.345	0.443
エピソード	0.383	0.720	0.484
質問	0.793	0.695	0.737

	1 (弱い)	2	← 3	興味 4	→ 5	6 (強い)	7
1 (弱い)	0.56	0.54	0.52	0.50	0.48	0.46	0.43
2	0.57	0.55	0.52	0.50	0.48	0.46	0.44
3	0.57	0.55	0.53	0.51	0.49	0.47	0.44
4	0.58	0.56	0.53	0.51	0.49	0.47	0.45
5	0.58	0.56	0.54	0.52	0.50	0.48	0.45
6 (強い)	0.59	0.57	0.54	0.52	0.50	0.48	0.46
7	0.59	0.57	0.55	0.53	0.51	0.49	0.46

図 2: ロジスティック回帰結果 (反応)

	1 (弱い)	2	← 3	興味 4	→ 5	6 (強い)	7
1 (弱い)	0.25	0.28	0.32	0.35	0.39	0.43	0.47
2	0.27	0.31	0.34	0.38	0.42	0.46	0.50
3	0.30	0.33	0.37	0.41	0.45	0.49	0.53
4	0.32	0.36	0.40	0.44	0.48	0.52	0.56
5	0.35	0.39	0.42	0.46	0.50	0.54	0.58
6 (強い)	0.38	0.41	0.45	0.49	0.53	0.57	0.61
7	0.40	0.44	0.48	0.52	0.56	0.60	0.64

図 3: ロジスティック回帰結果 (エピソード)

	1 (弱い)	2	← 3	興味 4	→ 5	6 (強い)	7
1 (弱い)	0.41	0.43	0.45	0.48	0.50	0.52	0.55
2	0.41	0.43	0.46	0.48	0.50	0.52	0.55
3	0.41	0.43	0.46	0.48	0.50	0.53	0.55
4	0.41	0.44	0.46	0.48	0.51	0.53	0.55
5	0.42	0.44	0.46	0.48	0.51	0.53	0.55
6 (強い)	0.42	0.44	0.46	0.49	0.51	0.53	0.55
7	0.42	0.44	0.46	0.49	0.51	0.53	0.56

図 4: ロジスティック回帰結果 (質問)

感の生成と発話構成要素の選択について回帰モデルを作成し、実験と評価を行った。今後は、ユーザの好感・興味の推定、全体の最適化を行い、システム全体を完成させたのち、自律型アンドロイド ERICA に初対面对話システムを実装する。

謝辞 本稿は、JST ERATO 石黒共生ヒューマンロボットインタラクションプロジェクト JPMJER1401 の支援を受けて実施された。

参考文献

- [1] Rosalind W. Picard. *Affective computing*, volume 252. MIT press Cambridge, 1997.
- [2] David DeVault *et al.* Simsensei kiosk: A virtual human interviewer for healthcare decision support. *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-agent Systems, AAMAS '14*, 1061–1068, 2014.
- [3] William Swartout *et al.* Ada and grace: Toward realistic and engaging virtual museum guides. *International Conference on Intelligent Virtual Agents*, 286–300. Springer, 2010.
- [4] 井上昂治 他. 自律型アンドロイド Erica のための音声対話システム. *SIG-SLUD*, 75:21–24, 2015.
- [5] 井上昂治 他. 潜在キャラクタモデルによる聞き手のふるまいに基づく対話エンゲージメントの推定. *人工知能学会論文誌*, Vol. 33, No. 1, DSH-F-1–12, 2018.