

対話を通じて話者の嗜好情報を学習する音声対話システム

杉原 穂 目良 和也 黒澤 義明 竹澤 寿幸

広島市立大学情報科学部

sugihara@ls.info.hiroshima-cu.ac.jp

{mera, kurosawa, takezawa}@hiroshima-cu.ac.jp

1 はじめに

近年、さまざまな分野で音声対話システムが普及しつつある。これらのシステムで扱う対話には、大きくタスク指向型対話と非タスク指向型対話がある。これまではタスク指向型対話を対象とした研究が多かったが、最近は雑談などの非タスク指向型対話についても研究が進んでいる。

しかし、従来の音声対話システムでは誰が話しかけても同じ対応しかできない。タスク指向型対話では万人に共通の対応で構わないが、雑談においては個人の嗜好によって対話戦略を変えたほうが良い場合がある。例えば、ひいきのスポーツチームが勝った時には共に喜んだりその話題を掘り下げて会話することが好まれる。逆に、負けた時には負けた話題をわざわざ出す必要は無い。

事物に対する一般的な好感度は、とにかく大量に文書を集めて統計処理することで推定可能であるが、ある特定個人の嗜好に関する文書を統計処理が可能なほど大量に集めることは難しい。さらに、一度定義した好感度もその後の学習や経験によって変化することが頻繁にある。

それでも人間は、相手との対話の中で、相手の発話から何が好きで何が嫌いかといった情報を推定することができる。例えば、「(嬉しそうに) 広島カープが優勝した!」と言う発話から、話者が広島カープを好きなことは容易に推定できる[1]。

そこで本研究では音声対話を通じて嗜好情報を学習し、学習した嗜好情報を用いて話者に合わせた対話を行うシステムを提案する。好感度推定処理では発話音声の音響的特徴から推定される話者感情と発話テキストの構文情報を用いる。そして、話者感情を考慮して返答する自然言語対話システム[2]の返答ルールを改良することで、ルール選択の条件として関連する事物の好感度を参照可能にする。評価実験では、被験者が実際に提案システムと対話することによって受ける印象を分析することで、提案手法の有用性を検証する。

2 提案手法の処理の流れ

提案手法では、音声に含まれる感情を考慮した自然言語対話システム[2]をベースとして、音声発話からの好感度推定手法[1]によって推定した好感度を考慮して返答発話を生成する。処理の流れを

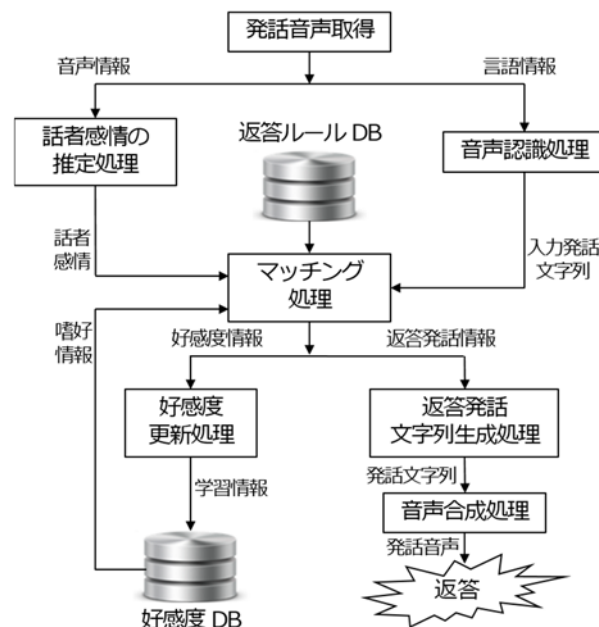


図1 提案手法の処理の流れ

図1に示す。

1. 音声認識装置を使用して入力発話音声を文字列に変換する
2. 発話音声から算出した音響的特徴を機械学習器に入力することで話者感情を推定する
3. (1)で取得した発話文字列, (2)で取得した話者感情, ユーザ発話の直前のシステム発話, 考慮する事物の好感度を条件として, 返答ルールデータベースから返答発話生成に用いるルールを1つ選択する
4. 発話文字列と話者感情から発話文字列に含まれる語句の好感度を推定する
5. (4)の推定結果にもとづいて好感度データベース内の好感度を更新する
6. 返答ルールを用いて生成した返答発話文字列を音声で出力する

3 システムの構成

3.1 音響情報からの話者感情推定

本研究で扱う感情クラスは、Plutchik[3]の8感情(受容, 怒り, 期待, 嫌悪, 恐れ, 喜び, 悲しみ, 驚き)に平静を加えた9種類の感情である。本論

文では, openSMILE[4]の IS09 特徴量セットで算出される特徴量 (音のエネルギー, 12 次元のメル周波数ケプストラム係数(MFCC), 基本周波数, ゼロ交差率, 声である確率とそれらの一次微分の音響特徴に対して 12 種類の基本統計量を算出して得られる計 384 種類) から Forward Stepwise Selection 手法を用いて絞り込んだ 55 個の特徴量を機械学習に用いる. 学習データとして, “感情評定値付きオンラインゲーム音声チャットコーパス[5]” 中の各感情の強感情演技音声 30 個を用いた.

3.2 返答ルールの構成

[2]のシステムで用いている返答ルールは“ユーザの発話文字列”, “ユーザ発話の直前のシステム発話”, “推定された話者感情”を条件として持っており, 全ての条件に適合すれば返答発話部の“発話文字列”を返答する. 本手法では話者の嗜好情報も考慮した返答生成を行うため, 条件として新たに“考慮する事物の好感度”を追加する.

“ユーザの発話文字列”とは, 直前に入力された話者発話を表す. マッチング条件緩和のためワイルドカードを使用できる. “ユーザ発話の直前のシステム発話”は, 入力された話者発話の前にシステムが出力した発話を表す. これにより対話履歴を考慮した対話を実現する. “推定された話者感情”とは, 音響分析により推定した話者感情である. “考慮する事物の好感度”とは, 入力発話に対して話者の嗜好に合った返答をするために参照する語句の好感度である.

本研究では好感度推定処理も返答ルール適用時に同時に行うため, 返答ルールの出力として返答発話文字列と好感度変化量の 2 つを設定する. 好感度推定処理については 3.3 節で述べる.

3.3 話者の感情を用いた好感度推定

好感度の推定には, 音響的特徴を用いた発話からの嗜好情報学習手法[1]を用いる. [1]の手法は, 発話事象の深層格構造と事物の好感度から快不快を計算する“情緒計算手法”を逆算させることで事物の好感度を計算する.

情緒計算手法では, 『《快/不快》は《好きな/嫌いな》対象の《接近/回避》により生起する』という Arnold の思想[6]に基づき, 主体, 対象, 動作 (接近/回避) の 3 要素の関係にのみ注目して快不快を判定している. 3 要素と生起する快不快の関係を表 1 に示す. 表 1 より, 生起している快不快が明らかになっており, かつ 3 要素中 2 要素の値が分かっている場合, 残り 1 要素の好感度を逆算することができる.

図 2 に好感度推定手法の処理の流れを示す. 例として, 話者が「(嬉しそうに) 寿司を食べた」という発話を行い, 話者にとっての寿司の好感度

表 1 情緒計算手法における好感度と生起情緒の関係

主体	対象	行動	感情
positive	positive	接近	快
positive	positive	回避	不快
positive	negative	接近	不快
positive	negative	回避	快
negative	positive	接近	不快
negative	positive	回避	快
negative	negative	接近	快
negative	negative	回避	不快

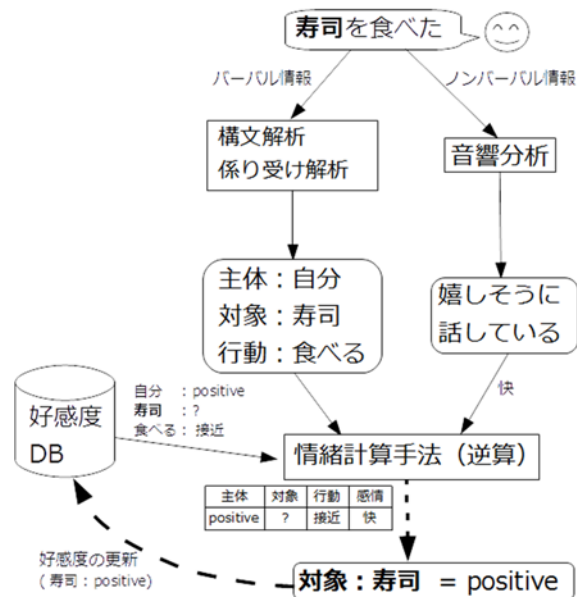


図 2 話者感情を用いた好感度推定処理の流れ

が不明だったとする. まずバーバル情報「寿司を食べた」に対して構文解析および係り受け解析を適用すると, 主体=「自分」, 対象=「寿司」, 行動=「食べる」となる. 次に好感度データベースを参照し, 各要素の好感度を得る. 今回は, 「自分」の好感度を positive, 「食べる」の値を「接近」とする. そして今回算出する「寿司」の好感度は未定である. 次に発話音声「寿司を食べた」に対して音響分析を行い, 話者の感情 (快/不快/中立) を推定する. 中立感情と判定された音声発話に対しては好感度の推定を行わない. この状態で表 1 を参照すると, 対象である「寿司」の好感度は positive であることがわかる. なお本手法により推定される好感度は positive か negative の 2 値であり, 強度は計算しない.

3.4 返答ルールへの好感度推定処理の組み込み

3.2 節で述べたように, 返答ルールでは発話文字列, 発話に現れる事物の好感度, 推定した話者感情が条件として参照されている. 返答ルール選択のための条件が適合している時点で好感度推定に必要な情報は全て揃っていることから, 本研究

では返答ルールを作成する際に事前に 3.3 節の好感度推定手法を適用して好感度の変化分を返答ルール中に設定しておく。

好感度データベースに設定できる嗜好情報は、“対象”、“好感度”、“対象のジャンル”、“学習した日付”の 4 要素からなる。対象は、好感度を与える対象となる単語である。好感度は、実数値で与えられており positive, negative, neutral を判別できる。また、同じ positive や同じ negative でも実数値がより大きいものほどより高い好感度を持つことを表す。対象のジャンルは、対象のジャンルを設定でき、同じ名前の対象が複数存在していても区別することが出来る役割を持つ。学習した日付には、嗜好情報を学習した日付を自動的に記憶することができる。

好感度は、positive や negative のようなクラスではなく、正負の符号を持つ実数値で表す。これにより、推定した好感度が positive だった場合に、強い negative が弱められたり、弱い positive が強められたりといった処理が可能となる。

4 評価実験

本論文では、以下の 3 種類の対話システムを構築し、被験者に実際に各システムと対話してもらうことで受ける印象を比較した。

A₁ (従来手法): 話者の嗜好情報を学習しないシステム

A₂ (比較手法): 対話を通じて話者の嗜好情報を学習できるが、一度学習した嗜好情報は更新できない

A₃ (提案手法): 対話を通じて話者の嗜好情報を学習できるシステム。嗜好情報の更新も可能である

本対話実験用に作成した返答ルールは 1909 個である。システムと対話する被験者は 21 才～23 才の大学生および大学院生 13 名(男性 11 名, 女性 2 名)である。対話内容は食べ物に関する話題に限定し、1 つのシステムに対して 3 回の対話を行った。3 回の対話はそれぞれ、1 回目は季節が夏、2 回目は季節が変わり冬、3 回目は同じ冬だが 2 回目より後の日、とシチュエーションを設定しており、季節によって嗜好情報が変わるような状況を設定している。A₁ (従来手法)、A₂ (比較手法)、A₃ (提案手法)それぞれのシステムと被験者の実際の対話事例を図 3 に示す。

印象評価の項目は文献[7]を参考に、“好感因子”、“性格的要因に関する項目”、“機械らしさ・生物らしさに関する項目”に関する形容詞 15 個、さらに“使用感に関する項目”に関する形容詞 3 個とした。被験者は、各形容詞について、A₁-A₂、A₁-A₃、A₂-A₃ の比較対の対比較結果を 7 段階(-3:とても〇〇の方が該当する、-2:〇〇の方が該当する、-1:やや〇〇の方が該当する、0:同じくらい、1:やや△△の方が該当する、2:△△の方が該当する、3:とても△△の方が該当する)で

<p>従来システムとの対話 1 回目(ある夏の日) システム(以下 S):「今日は暑いですね。こんな日はひんやりした冷たいアイスクリームを食べるなんてのはどうですか?」 被験者(以下 P):「かき氷が食べたいなあ」 S:「かき氷が好きなんです。私もかき氷は好きですよ」</p>
<p>従来システムとの対話 2 回目(1 回目より後のある冬の日) S:「今日は寒いですね。こんな日はひんやりした冷たいアイスクリームを食べるなんてのはどうですか?」 P:「どうしようかなあ」 S:「嫌そうな声ですね。アイスクリームは食べたくないですか?」 P:「食べたくないなあ」 S:「では、今何が食べたいですか?」 P:「しゃぶしゃぶ」 S:「しゃぶしゃぶが好きなんです。私もしゃぶしゃぶは好きですよ」</p>
<p>従来システムとの対話 3 回目(2 回目より後の冬の日) S:「今日も寒いですね。こんな日はひんやりした冷たいアイスクリームを食べるなんてのはどうですか?」 P:「んー、嫌だなあ」 S:「そうですね、わかりました」</p>
<p>比較システムとの対話 1 回目(ある夏の日) S:「今日は暑いですね。こんな日はひんやりした冷たいアイスクリームを食べるなんてのはどうですか?」 P:「かき氷が食べたいなあ」 S:「かき氷が好きなんです。私もかき氷は好きですよ」</p>
<p>比較システムとの対話 2 回目(1 回目より後のある冬の日) S:「今日は寒いですね。そういえばかき氷が好きでしたよね。こんな日はキンキンに冷えたかき氷を食べるなんてのはどうですか?」 P:「どうしようかなあ」 S:「嫌そうな声ですね。かき氷は食べたくないですか?」 P:「冬だからね、今は食べたくないかな」 S:「では、今何が食べたいですか?」 P:「おでん」 S:「おでんが好きなのですね。私もおでんは好きですよ」</p>
<p>比較システムとの対話 3 回目(2 回目より後の冬の日) S:「今日も寒いですね。そういえばかき氷が好きでしたよね。こんな日はキンキンに冷えたかき氷を食べるなんてのはどうですか?」 P:「今はかき氷はいいかな」 S:「そうですね、わかりました」</p>
<p>提案システムとの対話 1 回目(ある夏の日) S:「今日は暑いですね。こんな日はひんやりした冷たいアイスクリームを食べるなんてのはどうですか?」 P:「かき氷が食べたいなあ」 S:「かき氷が好きなんです。私もかき氷は好きですよ」</p>
<p>提案システムとの対話 2 回目(1 回目より後のある冬の日) S:「今日は寒いですね。そういえばかき氷が好きでしたよね。こんな日はキンキンに冷えたかき氷を食べるなんてのはどうですか?」 P:「どうしようかなあ」 S:「嫌そうな声ですね。かき氷は食べたくないですか?」 P:「食べたくないです」 S:「では、今何が食べたいですか?」 P:「おでんが食べたい」 S:「おでんが好きなのですね。私もおでんは好きですよ」</p>
<p>提案システムとの対話 3 回目(2 回目より後の冬の日) S:「今日も寒いですね。こんな日はキンキンに冷えたかき氷…あつ、それともあなたが好きなおでんを食べるなんてのはどうですか?」 P:「いいですね」 S:「ぜひ、今日にでも食べてみてください」</p>

図 3 ある被験者と対話システムの実際の対話事例

表 2 各項目の平均値および標準偏差

			A ₁ -A ₂	A ₁ -A ₃	A ₂ -A ₃
好感 因子	外交的	Mean	0.92**	1.69**	0.85*
		SD	1.21	1.54	1.87
	真似 したい	Mean	0.69*	1.77**	1.38**
		SD	1.20	0.97	0.84
	格好 いい	Mean	0.54**	1.15**	0.85**
		SD	0.84	1.17	0.95
	感じが いい	Mean	0.85*	1.77**	1.31**
		SD	1.46	1.53	1.64
	仲良く なれそう	Mean	1.00**	1.85**	1.46**
		SD	1.04	0.86	1.22
	また 会いたい	Mean	0.85**	1.92**	1.38**
		SD	0.95	1.07	1.00
	優しそう	Mean	0.77	1.46**	1.46**
		SD	1.48	1.69	1.50
共感が 持てる	Mean	0.85**	1.54**	1.00**	
	SD	1.03	1.08	1.18	
気が利く	Mean	0.85*	1.69**	1.46**	
	SD	1.46	1.64	1.74	
性格的 要因	真面目 そう	Mean	0.31	0.00	-0.08
		SD	1.32	1.57	1.49
	おとなし い	Mean	-0.38	-0.85**	-0.69**
		SD	1.27	1.46	1.32
	やんちゃ	Mean	0.31	0.38*	0.23
		SD	1.20	1.08	1.12
不真面 目	Mean	-0.38	-0.31**	-0.69*	
	SD	1.15	1.07	0.72	
機械らし さ・生物 らしさ	機械 らしい	Mean	-1.08**	-1.54**	-1.31**
		SD	1.73	1.60	1.07
	生物 らしい	Mean	0.31	0.85**	0.92**
		SD	1.54	1.92	1.27
使用感	楽しめ そう	Mean	0.38	1.77**	1.15**
		SD	1.27	0.89	1.17
	イライ する	Mean	-0.46*	-0.54	0.15
		SD	1.45	1.55	1.35
	規則的	Mean	-0.54*	-1.15**	-0.31
		SD	1.65	1.83	1.64

*: p<0.05, **: p<0.01

評定した。評定結果に対して、Scheffé の一対比較法 (中屋の変法) [8] を適用して分散分析を行い、有意差を計算した。各項目における回答の平均値 (Mean)、標準偏差 (SD)、および有意差の有無を表 2 に示す。

表 2 より、提案手法の評価が有意に高いことが分かった。これは、それぞれのシステムに対して似たような内容の発話をして、提案手法だけが自分の好みの変化に追従していることを感じられた結果と考えられる。次に機械らしさ・生物らしさについては、従来手法が最も機械らしく、提案手法が最も生物らしいという結果となった。これは、従来手法の「自分の好みを言ったにも関わらず同じ返答を返す」という特徴が単調な機械処理を連想させたことに比べ、比較手法および提案手法は前に話した自分の好みを記憶していたことがシステム発話から感じることでできたためだと考えられる。また、比較手法と提案手法においては、比較手法にはできなかった好みの変化に追従していることを提案手法ではできている点、より人間同士のコミュニケーションに近い印象を与えたため、評定に有意な差があ

ったと考えられる。最後に使用感についても、「楽しめそう」の項目において提案手法が良い評価を得ている。これは、話者の好みによってシステムの返答が変化するため、対話システムの反応を楽しむような被験者の気持ちによるものと思われる。一方、「規則的」の項目においては、従来手法の評定が高くなっている。従来手法ではシステムが好みを学習しないため、1, 2, 3 回目ともほぼ同じような話の切り出し方になってしまったことがシステムの反応に対し規則性を感じさせたと考えられる。

5 おわりに

本研究では、システムとの対話を通じて話者の嗜好情報を学習し、さらにその後の対話戦略に反映できる音声対話システムを提案した。好感度の推定は発話文字列と音響情報から推定した話者感情から行い、返答ルール選択条件として好感度情報を追加することで話者の好感度を考慮した返答をできるようにした。評価実験の結果、提案手法の好感因子の得点が有意に高かった。また、提案手法はより生物らしいと感じられており、使用感においても楽しめそうだと評価されていた。

今後の課題として、人手で大量の返答ルールを作成していくには限度があるため、大規模対話コーパスなどから返答ルールを自動生成する手法を取り入れることが挙げられる。また、より知的で柔軟な対話システムを作成するため、音響情報だけでなくマルチモーダルな情報からの話者感情推定も行う予定である。

参考文献

- [1] 目良, 市村ほか, 音響的特徴を用いた発話からの嗜好情報学習手法, 2012 IEEE SMC Hiroshima Chapter 若手研究会講演論文集, pp.115-118, 2012.
- [2] T.Takahashi, K.Mera *et al.*, Natural Language Dialog System Considering Speaker's Emotion Calculated from Acoustic Features, Proc. of IWSDS2016, 2016.
- [3] R.Plutchik, The emotions, New York: Random House, 1962.
- [4] F.Eyben, M.Wöllmer, *et al.*, openSMILE – The Munich Versatile and Fast Open-Source Audio Feature Extractor, ACM Multimedia Conference – MM, pp.1459-1462, 2010.
- [5] 有本, 河津, 音声チャットを利用したオンラインゲーム感情音声コーパス, 日本音響学会 2013 年秋季研究発表会講演論文集, 1-P-46a, 2013.
- [6] M.B.Arnold, Emotion and Personality, New York: Columbia University Press, 1960.
- [7] 高吉, 田中, ロボットの振る舞いと知性・性格の印象の関係, 情報処理学会研究報告(CVIM), Vol. 2007, No. 87, pp.43-48, 2007.
- [8] 中屋, Scheffé の一対比較法の一変法, 第 11 回日本科学技術連盟官能検査大会報文集, pp.1-12, 1970.