

同調的音声対話システムのための ユーザに特徴的な発話フレーズの検出

西田 潤^{1,a)} 大野 誠寛^{2,b)} 加藤 芳秀³ 松原 茂樹¹

概要: コミュニケーションの場において、人は言い回しが自分と似ている応答に親しみを感じる事が報告されており、対話システムにおいて、ユーザの特徴的な言い方を捉えて、それを適度に真似た応答ができれば、ユーザに親近感を与えることができると考えられる。本稿では、応答様式をユーザの個性に応じて使い分けることができる同調的対話システムの実現を目指し、その要素技術として、ユーザに特徴的な発話フレーズを検出する手法を提案する。本手法では、複数ユーザが複数トピックについて発話している状況を想定し、各フレーズが誰の何についての発話に出現したかという分布傾向を考慮して、ユーザを特徴づける度合いを推定する。CIAIR 車内音声対話コーパスのドライバ発話を用いて評価実験を行い、本手法の実現可能性を確認した。

キーワード: 用語抽出, 個性検出, ユーザ適応

Detection of User-Specific Phrases for User-Adaptive Spoken Dialogue System

Abstract: To actualize intimate communication between a user and a system, it is desired for the system to capture characteristic expressions which the user used in speaking and imitate them reasonably when it reacts. This paper proposes a technique for detecting user-specific phrases as an elemental technique to actualize a dialogue system which can react flexibly in accordance with user's personality. Our method estimates how user-specific a phrase is, in consideration of the distribution tendency of user-specific phrases in various dialogues where multiple users talked about various topics. We conducted an experiment by using a real in-car speech dialogue corpus, and confirmed the feasibility of our method.

Keywords: term extraction, personality detection, user adaptation

1. はじめに

音声・言語技術の進展を背景に、これまでに様々な音声対話システムが開発されてきた。その多くは、タスクを正確かつ効率的に達成することが重視され、システムの振る舞いはユーザに依存することなく画一的であった。しか

し、ユーザとシステムとの間で親しみのあるコミュニケーションを実現するためには、システムはユーザの個性を理解し、言い方や言い回しなど、応答様式をユーザに応じて使い分けることが望まれる。例えば、コミュニケーションの場において、人は、言い回しが自分と似ている応答に親しみを感じる事が報告されており [1]、ユーザの特徴的な言い方を捉えて、それを適度に真似た応答をすることにより、ユーザにシステムへの親近感を与えることができる。本研究では、このような対話システムを同調的音声対話システムと呼ぶ。

本稿では、応答様式をユーザの個性に応じて使い分けることができる同調的対話システムの実現を目指し、その要素技術として、ユーザに特徴的な発話フレーズを検出する手法を提案する。本手法では、各フレーズがどのユーザの

¹ 名古屋大学 大学院情報科学研究科
Graduate School of Information Science, Nagoya University,
Chikusa, Nagoya 464-8601, Japan

² 名古屋大学 情報基盤センター
Information Technology Center, Nagoya University,
Chikusa, Nagoya 464-8601, Japan

³ 名古屋大学 情報連携統括本部
Information and Communications Headquarters, Nagoya
University, Chikusa, Nagoya 464-8601, Japan

a) nishida.jun@h.mbox.nagoya-u.ac.jp

b) ohno@nagoya-u.jp

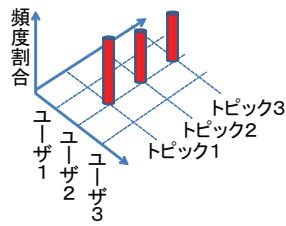


図 1 ユーザに特徴的な発話フレーズの分布

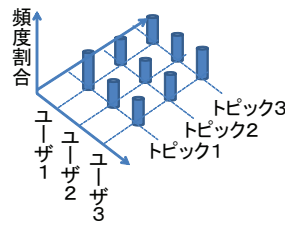


図 2 トピックワードの分布

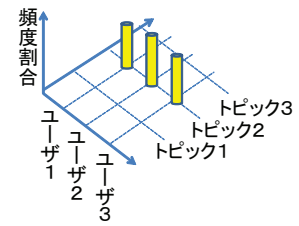


図 3 ストップワードの分布

何のトピックについての発話に出現したかという分布傾向に着目して、ユーザに特徴的な発話フレーズを検出する。CIAIR 車内音声対話コーパス [2] のユーザ発話を用いて検出実験を行い、ユーザに特徴的な発話フレーズに対する本手法の検出可能性を確認した。

2. ユーザに特徴的な発話フレーズ

ユーザに特徴的な発話フレーズとは、ユーザを特徴づける発話フレーズのことである。例えば、あるユーザが他のユーザと比べて非常に頻繁に発話したフレーズは、たとえ他のユーザの発話の中で見慣れたフレーズであっても、そのユーザを特徴づける発話フレーズといえる。逆に、そのユーザが数回しか発話していなくても、他のユーザがほとんど発話しないような独特なフレーズであれば、そのユーザを特徴づける発話フレーズとみなせる。直感的には、いわゆる口癖や文末表現などが該当する。

3. ユーザに特徴的な発話フレーズの検出

本研究では、複数ユーザが複数トピックについて対話システムと対話している状況を想定し、各発話フレーズが誰の何のトピックに関する発話に出現するかという分布傾向に着目して、ユーザに特徴的な発話フレーズを検出する。

3.1 ユーザに特徴的な発話フレーズの特徴

本研究では、図 1~3 に示すように、ユーザに特徴的な発話フレーズは、トピックワードやストップワードと異なる出現傾向を持つと仮定する。図 1~3 は、どのユーザの何のトピックに関する対話で、各フレーズがどの程度発話されるか（頻度割合）を示す模式図である。トピックワードは、ユーザに関係なく、ある特定のトピックに関する発話のみに出現し、ストップワード（助詞など）は、ユーザやトピックに関係なく、どのような発話でも出現する。一方、ユーザに特徴的な発話フレーズは、ある特定のユーザの発話のみに出現し、トピックに関係なく発話される傾向が強いと考えられる。そこで本研究では、あるユーザに特徴的な発話フレーズは以下の特徴をもつとする。

- 特徴 1 そのユーザによって頻繁に発せられる
- 特徴 2 そのフレーズを発するユーザは偏る
- 特徴 3 トピックに関係なく発せられる

表 1 分析データ

ドライバ数	497
トピックの種類数	63
発話単位数	20,736
形態素数	141,183

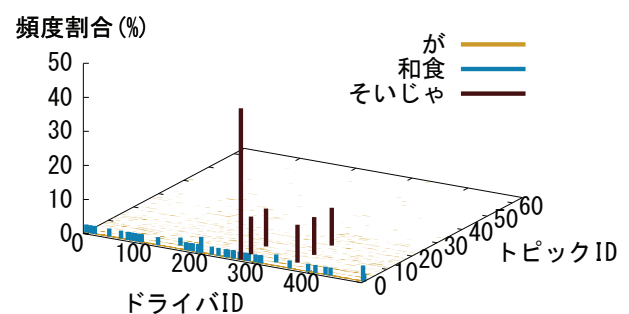


図 4 ストップワード「が」、トピックワード「和食」、口癖フレーズ「そいじゃ」の分布

実際に、CIAIR 車内音声対話コーパス [2] を用いて、ユーザに特徴的な発話フレーズ、ストップワード、トピックワードのそれぞれの代表的なフレーズに対して分布を調査した。CIAIR 車内音声対話コーパスには、ドライバとナビゲータとの間で遂行された対話の音声データの書き起こしが収録されている。各対話には、トピックとして店検索や道案内などが設定されており、発話ごとにトピックのラベルが付与されている。また、すべての発話に対して、ChaSen[3] を用いて形態素解析を施している。辞書には、店名などの固有名詞を独自に追加した Unidic[4] を用いた。分析には、ドライバ 497 人分の発話データを用いた。分析データの基礎統計を表 1 に示す。なお、発話単位とは、発話を 200ms 以上のポーズで区切った単位である。

図 4 に、ユーザに特徴的な発話フレーズ「そいじゃ」、トピックワード「和食」、ストップワード「が」の分布を示す。図 1~3 と同様に、あるドライバによる、あるトピックの対話で、各フレーズがどの程度発話されたか（頻度割合）を示すグラフである。「和食」は、レストラン検索に関するトピック（図 4 上のトピック ID は 1）のトピックワードである。これらの分布は、図 1~3 と同様の傾向を示している。なお、全体的に若い ID をもつトピックに分布が集中しているのは、各ドライバがすべてのトピックについて対話しているわけではないため、各トピックの ID

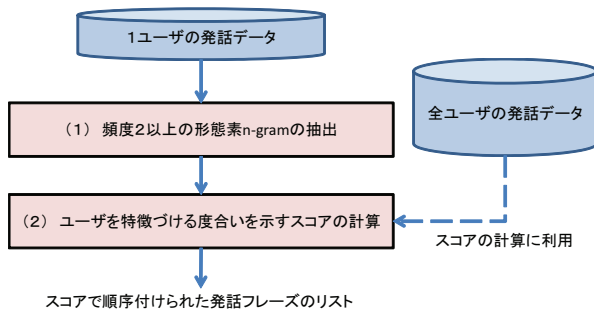


図 5 ユーザに特徴的な発話フレーズの検出手順

を、そのトピックについて対話しているドライバの数が多い順に付与したためである。

上記の特徴に加えて、ユーザに特徴的な発話フレーズは、フレーズとしてひとまとりに発せられるものとして、以下の特徴も備えているとする。

特徴 4 単位性 (ひとまとりのフレーズとして見なせる度合い) [5] が強い

3.2 検出手順

本手法では、ユーザ 1 人分の発話から形態素 n-gram を抽出し、抽出した各形態素 n-gram に対して、ユーザを特徴づける度合いを示すスコアを求め、このスコアが高いものから降順に出力する。本手法の流れを図 5 に示す。以下で、各処理について説明する。

(1) 形態素 n-gram の抽出

ユーザ 1 人分の発話から、形態素 n-gram を抽出する。フレーズは発話単位をまたがないとして、発話単位ごとに、その中の形態素列から、あらゆる形態素 n-gram を獲得する。例えば、3 個の形態素からなる発話単位からは、1-gram が 3 個、2-gram が 2 個、3-gram が 1 個、獲得される。獲得した n-gram のうち、ユーザ 1 人分の発話において、1 回しか発話されていない n-gram はユーザに特徴的な発話フレーズではないと考えて排除し、その他の n-gram を次の処理に渡す。

(2) ユーザを特徴づける度合いを示すスコアの計算

上記で抽出された各 n-gram に対して、ユーザを特徴づける度合いを示すスコアを計算し、このスコアが高いものから降順に出力する。スコアの計算式については次節で詳述する。

3.3 ユーザを特徴づける度合いを示すスコア

形態素 n-gram $w_\alpha (1 \leq \alpha \leq N_w^\beta)$ がユーザ $u_\beta (1 \leq \beta \leq N_u)$ を特徴づける度合いは、以下のスコア関数 $Score_\beta(w_\alpha)$ を用いて計算する。

$$Score_\beta(w_\alpha) = \prod_{i=1}^4 \frac{\exp(Score_\beta^i(w_\alpha))}{\sum_{\alpha=1}^{N_w^\beta} \exp(Score_\beta^i(w_\alpha))}$$

ここで、 $Score_\beta^i (1 \leq i \leq 4)$ は、前述した特徴 1 から

4 の各特徴を捉えるためのスコア関数である。これら 4 つのスコア関数をそれぞれ正規化し、積をとったものを $Score_\beta(w_\alpha)$ としている。なお、 N_w^β 、 N_u はそれぞれ、ユーザ u_β の発話データから抽出した形態素 n-gram の総数、ユーザの総数を示す。

以下では、特徴 1~4 を捉えるための各スコア関数 $Score_\beta^i(w_\alpha)$ をそれぞれ説明する。

● 特徴 1 を捉えるためのスコア関数 : $Score_\beta^1(w_\alpha)$

全ユーザにより発話された頻度に対してユーザ u_β により発話された割合が高く、かつ、ユーザ u_β により発話された頻度そのものも多いとして、以下のスコアを定めた。 $tf(w_\alpha, u_\beta)$ は、ユーザ u_β が形態素列 w_α を発話した頻度を示す。なお、割合だけで算出すると、1 人のユーザだけが発話している場合、その頻度に関係なく、一定値 (= 1) となるため、頻度との積をとっている。

$$Score_\beta^1(w_\alpha) = tf_\beta(w_\alpha) \times \frac{tf_\beta(w_\alpha)}{\sum_{\beta=1}^{N_u} tf_\beta(w_\alpha)} \quad (1)$$

● 特徴 2 を捉えるためのスコア関数 : $Score_\beta^2(w_\alpha)$

形態素列 w_α を発話するユーザが偏ると、「形態素列 w_α が発話されたとき、そのユーザが誰であるか」のエントロピーが小さくなる。そこで、このエントロピーの逆数を取り、特徴 2 を示すスコアとした。なお、このスコアに idf を用いることも考えられるが、口癖となるフレーズは、トピックワードと比べて、それを口癖としないユーザによっても低頻度であれば発話される可能性が高く、頻度を考慮しない idf ではユーザに偏りはないと算出される恐れがあるため、本研究ではエントロピーを用いた。

$$Score_\beta^2(w_\alpha) = \frac{1}{H(U|w_\alpha)+1} = \frac{1}{-\sum_{\beta=1}^{N_u} P(u_\beta|w_\alpha) \log P(u_\beta|w_\alpha)+1} \quad (2)$$

● 特徴 3 を捉えるためのスコア関数 : $Score_\beta^3(w_\alpha)$

形態素列 w_α はトピックに関係なく発話されるのであれば、「形態素列 w_α が発話されたとき、その発話されたトピックが何であるか」のエントロピーが大きくなる。そこで、このエントロピーを特徴 3 を捉えるためのスコアとした。なお、 $t_\gamma (1 \leq \gamma \leq N_t)$ はトピックを表し、 N_t はトピックの総種類数を示す。

$$Score_\beta^3(w_\alpha) = H(T|w_\alpha) = -\sum_{\gamma=1}^{N_t} P(t_\gamma|w_\alpha) \log P(t_\gamma|w_\alpha)+1 \quad (3)$$

● 特徴 4 を捉えるためのスコア関数 : $Score_\beta^4(w_\alpha)$

単位性を調べるためのスコア関数 $LScore$ や $RScore$ として、用語抽出に関する従来研究 [5], [6] では、左右に接続する形態素の異なり語数や頻度、C-value など様々な関数が利用されているが、本研究では、エント

ロピーを利用する。単語 w_α の左右に接続する単語のエントロピーが小さいほど、特定の形態素が接続しやすく単位性が低いことを意味し、逆に、大きいほど、様々な形態素が接続しやすいことになり、高い単位性を持つと考えられる。そこで、式(4)を特徴4をとるためのスコアとした。なお、 L を形態素列 w_α の左側に接続する形態素、 R を形態素列 w_α の右側に接続する形態素を示す。

$$Score_\beta^4(w_\alpha) = \sqrt{H(L|w_\alpha) \times H(R|w_\alpha)} \\ = \sqrt{-\sum_x P(l_x|w_\alpha) \log P(l_x|w_\alpha)} \\ \times \sqrt{-\sum_y P(r_y|w_\alpha) \log P(r_y|w_\alpha)} \quad (4)$$

4. 実験

本手法による検出可能性を確認するため、実際の対話データを用いて検出実験を行った。

4.1 実験概要

実験データとして、3.1節の事前調査で用いた CIAIR 車内音声対話コーパス [2] を用いた。表1のデータの中からランダムに選択したユーザ10人分の発話データをテストデータとして用いて、検出実験を実施した。なお、テストデータにおいて、ユーザ1人分の発話から抽出された頻度2以上の形態素 n-gram は平均125個であった。

評価は、ユーザごとに出力される検出結果の TOP5 (スコアが高い形態素 n-gram 上位5個) と、ユーザごとに人手で作成した正解データを比較することにより行った。正解データは、作業員2名がそれぞれ各ユーザに特徴的な発話フレーズを抽出し、それら2つの抽出結果の積集合をとることにより作成した。なお、正解データに含まれるフレーズの総数は合計37個であった。

4.2 実験結果

本手法が検出した発話フレーズの TOP5 のうち、正解データと完全に一致したものの数は、10ユーザの合計で17個であった。これは、検出すべきフレーズ(正解データに含まれるフレーズ37個)の半数近くをカバーしていることを意味しており、本手法がある程度の検出性能を有していることを確認した。

表1に、テストデータの10ユーザのうち、5ユーザ(f2117, f2129, f2134, m2014, m2047)に対する本手法の検出結果(TOP5)を示す。正解データと一致したフレーズは太字で表記している。記号「_」は形態素境界を表す。なお、ユーザIDのfとmはそれぞれ女性、男性を意味する。

m2014の「ほい_じゃあ」のように、他のユーザにはほとんど発話されないようなユーザ独特のフレーズを検出できていることがわかる。また、f2134の「と」や「ん」などのように、他のユーザによっても発話されると考えられるが、

表2 3ユーザ分の検出結果(TOP5)

順位	ユーザ ID				
	f2117	f2129	f2134	m2014	m2047
1	て	か	と	ほい	ん
2	で_いい_わ	に	に	えー	はい
3	の	じゃあ	ん	で	で
4	そこ_で_いい_わ	か_なあ	の	ほい_じゃあ	の
5	ん_だ_けど	はい	あ	の	あ

このユーザによって特に頻繁にフィルターとして発話されることによって、ユーザを特徴づけるフレーズとみなせるものも検出できた。なお、正解データには、表2の太字フレーズの他に、f2117は「いい_わ(11)」「ああ(27)」, f2129は「お_願_い_し_ま_す(10)」「なあ(19)」, f2134は「か_なあ(13)」, m2014は「んー(7)」「あ(10)」, m2047は「え(7)」「なら(8)」が各ユーザを特徴づける発話フレーズとして挙げられいた。括弧()の中に表示した数値は、各フレーズの検出結果における順位を示す。f2117の正解フレーズ「いい_わ」については、このフレーズを含むフレーズを上位2位や4位として検出できていることが分かる。その他、上位5位以内で検出できなかった正解フレーズの多くは、該当ユーザと同程度に他のユーザによっても頻繁に発話されていたため、高いスコアが与えられなかったと考えられる。

5. おわりに

本論文では、ユーザ発話から各ユーザに特徴的な発話フレーズを検出する手法を提案した。今後は、ユーザに特徴的な発話フレーズに関して詳細に分析しスコア関数を精緻化することなどにより、検出性能を向上させる予定である。

謝辞 本研究は一部、科研費挑戦的萌芽研究「音声対話における発話からの話者の個人性検出」No.24650066により実施した。

参考文献

- [1] 西原陽子, 松村真宏, 谷内田正彦: QA サイトにおける質問に適した回答の判定, NLP 若手の会第2回シンポジウム発表8予稿(2009). http://yans.anlp.jp/modules/menu/main.php?page_id=14&op=change_page.
- [2] Kawaguchi, N., Matsubara, S., Takeda, K. and Itakura, F.: CIAIR In-Car Speech Corpus -Influence of Driving Status-, *IEICE Trans. Inf. Sys.*, Vol. E88-D, No. 3, pp. 578-582 (2005).
- [3] 松本裕治, 高岡一馬, 浅原正幸: 形態素解析システム『茶釜』version2.4.3 使用説明書(2008).
- [4] 伝 康晴, 小木曾智信, 小椋秀樹, 山田 篤, 峯松信明, 内元清貴, 小磯花絵: コーパス日本語学のための言語資源: 形態素解析用電子化辞書の開発とその応用, *日本語科学*, No. 22, pp. 101-122 (2007).
- [5] 池野篤司, 濱口佳孝, 山本英子, 井佐原均: Web 文書集合からの専門用語獲得, *情報処理学会論文誌*, Vol. 47, No. 6, pp. 1717-1727 (2006).
- [6] 中川裕志, 湯本紘彰, 森 辰則: 出現頻度と接続頻度に基づく専門用語抽出, *自然言語処理*, Vol. 10, No. 1, pp. 27-45 (2003).