

音声対話におけるユーザに特徴的な発話フレーズの検出

西田 潤[†] 大野 誠寛^{††} 加藤 芳秀[§] 松原 茂樹[†] 石川 佳治[†]

[†]名古屋大学 大学院情報科学研究科 ^{††}名古屋大学 情報基盤センター [§]名古屋大学 情報連携統括本部

1 はじめに

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覚えることが報告されており [1], 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。本論文では、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り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。

本研究では、図 1 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 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たか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

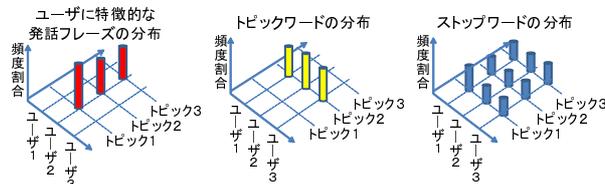


図 1: 各発話フレーズの分布

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りに発せられる

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

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) [3] が強い

3.1 検出手法

本手法の流れを図 2 に示す。本手法では、まず、話者 1 人分の発話から 2 回以上出現した形態素 n-gram を抽出する。その後、抽出した全ての形態素 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。

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、積をとって $Score_\beta(w_\alpha)$ を計算する。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: $Score_\beta^1(w_\alpha) = tf_\beta(w_\alpha) \times \frac{tf_\beta(w_\alpha)}{\sum_{\beta=1}^{N_u} tf_\beta(w_\alpha)}$

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、 u_β により発話された頻度そのものも多しとして、上記のスコア関数を

Detection of User-Specific Phrases from Conversational Speech
 Jun Nishida[†], Tomohiro Ohno^{††}, Yoshihide Kato[§], Shigeki Matsubara[†], Yoshiharu Ishikawa[†]
[†]Graduate School of Information Science, Nagoya University
^{††}Information Technology Center, Nagoya University
[§]Information and Communications Headquarters, Nagoya University

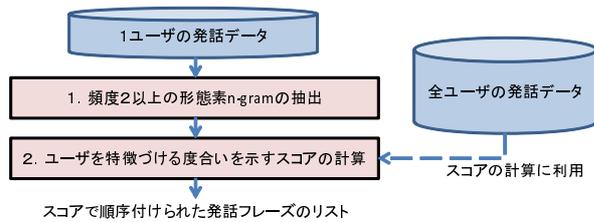


図 2: ユーザに特徴的な発話フレーズの検出手順

定めた. $tf_{\beta}(w_{\alpha})$ は, ユーザ u_{β} が形態素 n-gram w_{α} を発話した頻度を示す.

特徴 2: $Score_{\beta}^2(w_{\alpha}) = \frac{1}{H(U|w_{\alpha})+1}$
 w_{α} を発話するユーザが偏ると, 「 w_{α} が発話されたとき, そのユーザが誰であるか」のエントロピーの逆数は大きくなると考え, 上記のスコア関数を定めた.

特徴 3: $Score_{\beta}^3(w_{\alpha}) = H(T|w_{\alpha})$
 w_{α} がトピック T に関係なく発話されるのであれば, 「 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と考え, 上記のスコア関数を定めた.

特徴 4: $Score_{\beta}^4(w_{\alpha}) = \sqrt{H(L|w_{\alpha}) \times H(R|w_{\alpha})}$
 単位性を調べるためのスコア関数として, 従来研究では左右に接続する形態素の異なり語数や C-value など様々な関数利用されているが [3, 4], 本研究ではエントロピーを利用する. L は w_{α} の左側に接続する形態素, R は右側に接続する形態素を意味する.

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

本実験では, 実験データとして, CIAIR 車内音声対話コーパスを用いた [2]. このコーパスには, 497 人のユーザがオペレータとの間で遂行した車内音声対話が収録されている. 各対話には, トピックとして店検索や道案内など (全 63 種類) 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. こ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, テストデータの 10 ユーザに対する正解データに含まれるフレーズ数は合計 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, 全ユーザの合

表 1: 3 ユーザ分の検出結果 (TOP5)

順位	ユーザ ID	f2129	f2134	m2014
1		か	と	ほい
2		に	に	えー
3		じゃあ	ん	で
4		かなあ	の	ほいじゃあ
5		はい	あ	の

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, 3 ユーザ (f2129, f2134, m2014) 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.

m2014 の「ほい₁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, 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 1 の太字フレーズの他に, f2129 は「なあ」「お願い₁します」, f2134 は「かなあ」, m2014 は「んー」「あ」が各ユーザを特徴づける発話フレーズとして挙げられていた. これらのフレーズは, 該当ユーザによって頻繁に発話されているものの, 他のユーザによっても同程度に頻繁に発話されているため, 本手法では上位 5 位以内に検出でき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 No.24650066, 及び, No.22300034 により実施した.

参考文献

- [1] 西原, 松村, 谷内田. 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 回シンポジウム予稿, 発表 8, 2009. http://yans.anlp.jp/modules/menu/main.php?page_id=14&op=change_page.
- [2] N. Kawaguchi, S. Matsubara, K. Takeda, and F. Itakura. CIAIR in-car speech corpus -influence of driving status-. *IEICE Trans. Inf. Sys.*, Vol. E88-D, No. 3, pp. 578-582, 2005.
- [3] 池野, 濱口, 山本, 井佐原. Web 文章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.
- [4] 中川, 湯本, 森. 出現頻度と連接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.