4S-9

一問一答型音声対話システムにおける システムからの自発的な発話生成

吉田 達平 † 駒谷 和範 † 佐藤 理史 †

† 名古屋大学 工学部 † 名古屋大学大学院 工学研究科 電子情報システム専攻

1 はじめに

一問一答型音声対話システムでは，ユーザの発話が無いとシステムは発話をしない．一方，近年普及した Siri やしゃべってコンシェルでは，むしろシステムの雑談的応答をユーザが楽しんでいる様子が，インターネット上の記事や Twitter などで見られる．そこで我々は，システムが応答した後に，さらにシステムから自発的に雑談的な発話を生成することを目指す．これにより，一問一答型音声対話システムにおける雑談対話の実現の可能性を探る．図 1 に，システムが自発的に発話を生成する例を示す．ここでは，システムからの応答後に一定時間の沈黙があった場合に，システムから話し始めることを想定している．

この自発的な発話には，その前のシステム応答に続くものとして妥当な発話を選ぶ．このような発話には，内容に関連があるものとそうでないものがある．システム発話例 A は自発的な発話として妥当だが，内容には関連が無い．システム発話例 B は前の発話の内容の一つである階段について発言している．本稿では，発話例 B のような，前の発話内容に関連した発話の生成にまず取り組む．発話例 A はほぼランダムに生成すればよく，これだけでは雑談相手として飽きられると考えるからである．

システムから生成する発話は，既存の一問一答型音声対話システムの知識ベースから選択する．本研究では，奈良先端大で開発された「たけまるくん」[1] を利用する．このシステムの知識ベースは，質問とその答えとなる応答のペアの集合である．この知識ベースを質問応答データベースと呼ぶ．Web や Twitter から発話を生成する研究もあるが [2]，本稿では以下の 2 つの理由により，質問応答データベースから選択する．まず質問応答データベース中の発話は，人手で書かれているために文法的に自然で，1 発話として適切である．次に，その音声対話システムのコンセプトや，使用しているキャラクターに合致した発話であることも保証されている．

2 発話の選択方法

発話生成は，質問応答データベースの発話にそれぞれスコアをつけて，スコアの最も高い発話を選択することで行う．このスコアは自発的な発話として妥当であるほど高い数値となるようにする．この妥当性を，二つの発話の内容の関連の強さ（以下，関連度）で近似する．

システム発話 S_1 の後に続く，システム発話 S_2 を選択する方法として，以下の 2 種類を考える．

- 手法 1：2 つの応答の関連度に注目する方法
- 手法 2：2 つの質問の関連度に注目する方法

図 2 に概念を図示する．ユーザ発話 U に対するシステム応答 S_1 は，質問応答データベース中の発話対である，質問 Q_i と応答 A_i に対応する．これに対して手法 1，または手法 2 により，応答 A_j を選択する．手法 1 では応答 S_1 に対し，最も関連度が高い応答 A_j を選び S_2 とする．手法 2 では質問 U に対し，最も関連度が高い質問 Q_j を選び，データベース中でそれに対応する応答 A_j を S_2 とする．

Spontaneous System Utterance Generation in Spoken Dialogue Systems based on Question-Answer Database: *Tappei Yoshida, Kazunori Komatani, and Satoshi Sato (Nagoya Univ.)*

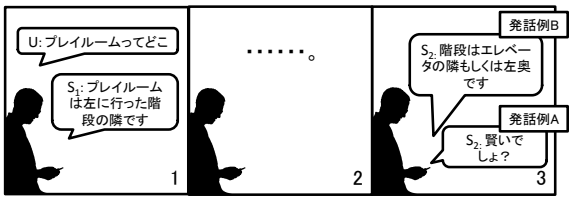

図 1: システムから生成する自発的な発話の例

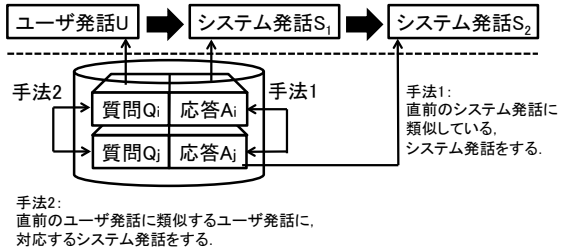

図 2: 発話の選択方法

2.1 関連度に基づく発話の選択

関連度は，以下の考えに基づいて計算する．

1. 関連度の強い発話同士は表層的に似ている．名詞が発話の内容を表す．
2. 名詞の細分類により，発話の内容を表す強さに差がある．
3. 妥当性は後続する発話（S_2）により強く左右される．

1 より，関連度の計算には名詞オーバーラップ率を基本とする．2 より，名詞オーバーラップ率に名詞の細分類に応じて重みを与える．3 より，後続する発話により重みを与える．

上記の 3 点を反映して，2 つの発話 X_1 と X_2 の関連度 $score(X_1,X_2)$ を計算する．まず式 1 により $overlap(X_1,X_2)$ を X_1 と X_2 で共通する名詞の重みの合計とする．

$$overlap(X_1,X_2) = \sum_{n \in X_1 \bigcap X_2} [w(n,X_1) + \lambda \times w(n,X_2)] \quad (1)$$

- $X_1 \bigcap X_2$: X_1 と X_2 で共通する名詞の集合
- $w(n,X)$: 発話 X における名詞 n の重み
- λ: X_2 に重みを与えるための定数 (本稿では 3 とした)

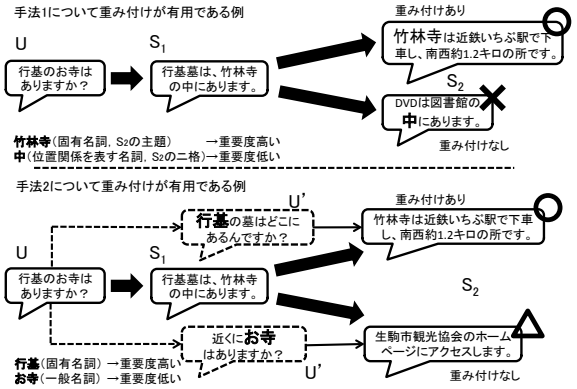

図 3: 応答（S_1）の後にシステムから行う発話（S_2）の例

次に X_1 と X_2 に出てくる名詞の重みの総和で式 1 の値を正規化し，$score(X_1, X_2)$ とする．名詞に与える重み（重要度）に関しては，次節で述べる．

2.2 名詞の重要度の定義

発話 X の中での名詞 n の重要度が大きいほど $w(n,X)$ が大きくなるよう，3 つの重みを定義する．これは (1) 名詞の品詞細分類，(2) 名詞の格要素，(3) 名詞を含む節の特徴，に基づく．この 3 つの重みを $w_1(n,X)$，$w_2(n,X)$，$w_3(n,X)$ とする．$w(n,X)$ は式 2 で計算する．

$$w(n,X) = w_1(n,X) \times w_2(n,X) \times w_3(n,X) \quad (2)$$

それぞれの重みは次のように決める．$w_1(n,X)$ は，固有名詞は 2，サ変名詞，用言語幹，非自立語は 0.1，位置関係や方向を表すもの，数詞は 0.5 とし，位置関係や方向を表すもの以外の一般名詞を 1 とする．$w_2(n,X)$ は中心化理論 [3] での，次の文（発話）で主題になりやすい名詞は，発話中の重要度が高いという仮定に基づいている．具体的には，ハ格は 1.8，ガ格は 1.6，ヲ格は 1.4，ニ格は 1.2，それ以外は 1 とする．$w_3(n,X)$ は，目的節に含まれる名詞は重要であると考え，2 とする．目的節とは，次の文の太字に示したようなものである．

- **近鉄生駒駅に行くには**、バスで富雄駅まで行き、近鉄をご利用ください。

ただし，手法 2 で用いる，質問と質問の関連度では，$w_2(n,X)$ と $w_3(n,X)$ は全て 1 とした．これは「たけまるくん」の質問データには，短い発話が多いためである．

図 3 に重み付けが有効な例を示す．手法 1 の例では，位置関係を表す「中」よりも，固有名詞であり S_2 ではハ格となる「竹林寺」に大きい重みが与えられ，適切な発話が選択されている．手法 2 の例では，「お寺」よりも，固有名詞の「行基」に大きい重みが与えられ，適切な発話が選択されている．

3 評価

システムが選択した発話を，「システムから生成する自発的な発話としての妥当性」という観点から人手で 5 段階で評価する（1 点が最も悪く，5 点が最もよい）．5 点は発話として全く問題がないもの．3 点，4 点は言い回しはおかしいが内容は妥当な発話．1 点，2 点はそもそも内容が妥当でない発話とした．

評価に用いるテストセットとして質問応答データベースから，100 対の質問と応答のペアをランダムに選んだ．

表 1: 評価結果

手法	出力無し	1 点	2 点	3 点	4 点	5 点
手法 1 の (a)	24	18	17	20	8	12
手法 1 の (b)	24	11	17	14	8	26
手法 1 の上限	36	-	-	6	11	47
手法 2 の (a)	42	25	3	4	9	17
手法 2 の (b)	41	20	4	1	11	23
手法 2 の上限	59	-	-	0	12	29

これは，図 1 の，ユーザ発話 U とシステム応答 S_1 に対応する．「たけまるくん」の大人用データベースは，質問 5881 発話，応答 368 発話からなる．

評価として以下の 4 つの手法を比較した．

1. 手法 1（応答同士の関連度に基づく選択）
   (a) $w(n,X) = 1$ とする場合
   (b) 2.2 節にしたがって重み付けした場合
2. 手法 2（質問同士の関連度に基づく選択）
   (a) $w(n,X) = 1$ とする場合
   (b) 2.2 節にしたがって重み付けした場合

表 1 中の条件を説明する．「出力無し」のものは，発話中に名詞が一つもない発話と，重要度が低い名詞のみがオーバーラップする発話である．「重要度が低い名詞」とは具体的には w_1 が 0.1 である名詞とした．「手法 1 の上限」とは，最も関連度の高い応答を人手で選んだ場合である．ただし 3 点〜5 点にあたる発話が，そもそも質問応答データベースにない場合「出力なし」とした．「手法 2 の上限」とは，同様に最も関連度の高い質問を，人手で選択したものである．

表 1 の結果より，まず，提案する重み付けは有効であったことがわかる．これは，いずれの手法でも，(a) と (b) を比較すると，後者で 5 点が増えていることから確認できる．次に，両手法での (b) の結果は，上限よりも劣る．これは，重み付けに改善の余地があることを示している．

発話選択の性能の改善には，手法 1 と手法 2 の結果の統合が考えられる．実際，両手法の上限の結果では，手法 1 のみで 5 点がつくものは 19 個，手法 2 のみで 5 点がつくものは 8 個存在した．両手法の結果から適切なものを自動で選択できれば，発話選択性能を向上させられる．

さらに表 1 より，両手法の上限で「出力なし」が多い．これは今回の手法は少なくとも 1 つの名詞が共通していないと，関連度が計算できないからである．これは，図 1 の発話例 A に相当する．名詞の共通部分に基づく手法で「出力なし」にあたる場合に適切な発話を生成する方法も，今後の課題である．

参考文献

[1] 西村竜一, 原直, 川波弘道, 李晃伸, 鹿野清宏. 10 年間の長期運用を支えた音声情報案内システム「たけまるくん」の技術. 人工知能学会誌, Vol. 28, No. 1, pp. 52–59, 2013.

[2] Alan Ritter, Colin Cherry, and William B. Dolan. Data-driven response generation in social media. *Proc. EMNLP*, pp. 583–593, 2011.

[3] Aravind K. Joshi and Scott Weinstein. Control of inference: Role of some apects of discourse structure-centering. *Proc. IJCAI*, pp. 291–331, 1981.