

トピック推定を用いたタスクドメインを選択するための発話生成

谷津 元樹 ジェプカ ラファウ 荒木 健治

北海道大学 情報科学研究科

{my, kabura, araki}@media.eng.hokudai.ac.jp

1 はじめに

音声対話インタフェースは、スマートフォン等に搭載され、広範囲に普及の兆しを見せている。しかし、いずれの製品においても、主に有効な入力として受け付けるのは、ユーザの関心が明示化され、かつ機器のもつ機能に言及している発話に限られている。質問応答やアプリケーションの動作など機器の持つ機能に直接関連のない発話や、感情を吐露した発話などは、機器の対応可能ドメイン（以下、タスクドメインとする）外と認識され、多様性のある応答を期待できないのが現状である。情報検索やタスク指向型発話以外の、従来ではドメイン外とされてきた雑談のような発話を音声対話システムが扱うことの意義が存在する。そのため、我々はドメイン外発話への対応を前提とした手法を開発を行っている。

雑談の中でタスク指向型のユーザの要求に対応するためには、明示的な要求と非明示的な関心の両方に対応する必要がある。このうち明示的な要求を 2.1 で述べる発話検出機構により処理することが可能である。一方、ユーザが常に表出するとは限らない非明示的な関心は、ユーザとの対話の継続の中でその対話履歴から推定することができると考えられる。システムが話題ドメインの推定、ドメインの粒度の収束の検出後に検出ドメインに即した提案発話を生成することで、ユーザの非明示的な関心を推測することができる。

提案手法では、2 に示すドメインの検出方法によりユーザの関心の表出を検出する。本稿では、対話とドメイン内の単語集合の類似度の算出のために用いることのできる手法のうち、WordNet を用いた手法について述べる。

1.1 関連研究

Lane ら [1] は、SVM を用いたドメイン外発話判定手法（言語要素が必ずしも明示的と限らない）を提案している。トピックのクラスタリングを行い、意味的

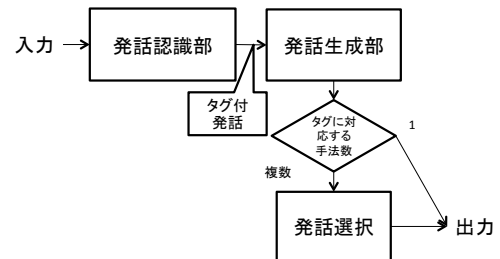


図 1: 処理の流れ。

距離が近く混同の可能性が高いトピックに対する頑健性を高めている。

Lane らの手法は、ドメイン外発話の判定によりタスク指向型システムの頑健性を高めることを目的とする。これに対して本手法では、オープンドメインな対話処理を主とすることにより、多様な応答が期待できる。

非タスク指向型対話システムにおいて対話履歴から対話戦略を獲得する手法として江頭らの研究 [2] が挙げられる。この研究では、強化学習の枠組みを用いて対話履歴より将来のユーザによる評価を最大化する発話生成モジュールの選択方針の獲得を試みている。

2 提案手法

本研究では、自由な雑談対話からタスクを推薦し遂行することのできるシステムの達成に向けて、一般的な情報検索要求発話やドメイン外発話に対し雑談発話で対応し、雑談対話を含む対話履歴から行える動作の一つを推薦するシステムを提案する。本システムは、雑談対話を行うための発話生成モジュールを有し、通常時の対話ではユーザ入力に対する雑談発話の生成を行う。システムはまた、応答を求める質問文や、あいさつなどの平叙文以外の表現を検出し専用のモジュールで応答を生成する機構を備える (図 1)。システムはこの機構を用いて、発話生成を行うモジュールを選択し、またシステムの内部状態を変更することができる。

2.1 雑談対話における応答発話生成

雑談対話における応答発話生成手法には、ルールベースによる手法と、入力発話の特徴に基づく応答生成モジュール選択による手法がある。

発話集合からの応答選択ルールによる発話生成では、検索パターンと応答発話テンプレートからなる有限個の応答生成ルールセットを用い、応答発話候補集合から直接応答発話を選択する。このような手法は高速であるが、応答生成ルールセットの大きさが発話生成にとり最終的な制約となるという課題がある。

この問題を回避するため、提案手法においては発話生成手法をモジュール化し、入力発話の特徴に基づき複数の応答発話生成手法から一つを選択する手法を用いる。明示的な特徴に対するルールを追加することにより、明示的なドメインに特化した発話生成を行う。提案手法では、質問応答の要求と考えられる入力発話の検出に対し割り当てられあつた質問応答システムが質問応答発話の生成を行う。

提案手法では、発話の特徴を認識するモジュール(フィルタ)が、入力発話に対し最適なタグと事前に人手で決定された一定のスコアを入力発話に付与する。最大スコアの付与されたタグに対応する応答生成手法が選択される。下記に述べる応答発話生成手法を用い、システムは単一の入力発話に対し一つの応答を生成する。

質問応答発話生成(タグ QA)には、Web 検索を知識源とする質問応答システム [3] を用いる。本タグに対応する応答選択モジュールでは、疑問詞のパターンマッチおよび固有表現検出による質問文認識を行っている。

非タスク指向型応答発話生成(タグ NT)には、発話選択モジュールの機能を評価するため、複数の発話生成手法 [4][5] を用い、今井ら [6] の用いた最終的な応答発話選択を行う。

また、2.2 で述べるドメイン選択に伴う提案発話の生成では、そのドメインキーワードを提示する応答発話を生成するモジュールがドメインの収束検出時に呼出され、ドメインに対応するタスクを推薦し、ユーザにドメインの選択を促す発話を生成する。

2.2 選択対象タスクドメインとキーワード集合

提案手法の選択対象ドメインは、2.1 で提案した汎用的対話機能において実装されたタスク指向型発話生成手法に人手で予め関連付けられたドメイン D である。各ドメイン d_i は、キーワード集合 $K_i = \{k|k \in$

$Natural\ Language\ Words\}$ を持つ。要素であるキーワード k は、本システムに組み込まれたタスク指向型発話モジュールと設定ファイルにおける設定により関連付けられた日本語の一般名詞であり、選択対象のドメインに属する。WordNet を利用する場合、キーワードは、語彙概念辞書の日本語 WordNet [7] 中に存在する概念 (synset) のもつ見出し語 (lemma) である必要がある。

2.3 ドメインの粒度

システムが遂行することのできるタスクの属するドメイン(タスク指向型対話を行うための個々のモジュールに結びつけられている)にキーワード集合を付与し、キーワード集合と対話履歴との類似度より、ユーザの関心の収束を検出する。キーワード集合は人間によりなされる対話全体のモデルから得られるトピックの上位-下位関係に即しているとする。キーワード集合の属するドメインは、最も近似したトピックに対応しているが、「釣り」に対する「レジャー」や、「サッカー」に対する「スポーツ」などのような、対象トピックと is-a 関係での上位-下位の関係にある概念でどれほど抽象的な概念を選択するかにより、トピック推定における焦点が異なる。この評価対象ドメインの粒度により、システムによるユーザの関心の収束の検出タイミングが異なると予想される。

2.4 目的ベクトル \mathbf{a}

目的ベクトル $\mathbf{a} = (a_1, a_2, a_3, \dots)$ はユーザの関心の方向性を表す。ここで、ベクトルの要素 a_i は選択対象ドメインのキーワード集合 K_i と発話文中の内容語との関連性を示す。 \mathbf{a} のユークリッドノルム $\|\mathbf{a}\| = \sqrt{\sum_k a_i^2}$ を求めることにより、ユーザが関心を持っているドメインの収束度を得る。

2.4.1 WordNet における意味類似度

\mathbf{a} は式 (1) で表される Leacock and Chodorow [8] による意味類似性の尺度を用いた意味類似度ベクトル $\Delta \mathbf{a}$ の累加により求める。

$$\Delta a_i = \frac{1}{N_U N_{K_i}} \sum_{u \in U} \sum_{k \in K_i} sim(u, k) \quad (1)$$

$$where\ sim(c_1, c_2) = \max \left(-\log \frac{N_p}{2D} \right) \quad (2)$$

表 1: 対話例. t はターン, User 以外の話者名は選択された発話生成手法名を示す.

t	話者, 発話内容
1	User: 日本の首相は誰ですか
1	QA: 野田佳彦だった
2	User: アメリカの大統領は誰ですか
2	QA: オバマらしいですからねえ
3	User: 今日は食堂で夕食を食べる予定です。
3	Recom: もしかして、健康、病気に興味ありますか？

ここで, N_p は 2 単語 c_1, c_2 間の語彙概念辞書における最短のグラフ距離を, D は語彙概念辞書における両者の分類の深さを表す.

2.4.2 ユーザの関心の明確度および選択発話生成

2.4.1 で述べた意味類似度の計算において $\|a\| > T$ となると, フィルタが固有のタグと正のスコアを出力し, ドメイン選択発話の生成モジュールが選択される. a における最大の値を持つ要素に関連付けられたキーワード集合 K_i を用い, ユーザにドメインの選択を促す発話が生成される (表 2 の発話例を参照).

3 印象評価実験

2.1 で提案した汎用的対話生成手法が被験者に与える印象を調査するため印象評価実験を行った. Web を用いた実験と, アンケート調査による評価について述べる.

3.1 実験方法

被験者とシステムとの対話は Web サイトを用いて, 20 ターン以上の自由な対話を行った. 20 ターン経過後に提示されるアンケート画面にて被験者は 3.2 で挙げる調査項目について回答した. 記録した被験者の属性と分布を, 表 2 に示す. 2.4.2 に定めるしきい値は, $T = 3.0$ とした

3.2 印象評価項目

被験者は下記に示す項目を用いて印象評価を行った. ベースラインとしてルールベース型対話システムにて行われた同様の実験の評価値を参考としてあわせ, 結果を表 3 に示す.

表 2: 被験者の属性および分布.

職業	数	年齢	数	
社会人	研究者 (分野外)	1	60 歳以上	1
	研究者 (分野内)	0	20 - 29 歳	8
	他	4	19 歳以下	1
大学生・院生	分野外	1	計	10
	分野内	3		
高校生		1		
計		10		

表 3: 印象評価実験の結果 (5 段階評価).

項目	評価値平均
A-E 平均	1.85
F	1.75
ベースライン平均	2.54

- (A) 対話を続けたいかどうか
- (B) 対話が文法的に自然であるか
- (C) 対話が意味的に自然であるか
- (D) システムの語彙が豊富かどうか
- (E) システムが人間らしいか
- (F) システムによる提案発話は, 興味関心に対し関連が強かったか

3.3 ドメイン選択発話生成の適切性

印象評価における評価項目 (F) はドメイン選択発話生成の適切性を問う項目である. (F) では, ドメイン選択発話がシステムにより行われなかった被験者による平均評価値 (1.83) が低くなる (2.0) 傾向がみられた.

4 発話選択精度の評価

提案手法における汎用的対話能力の有効性を計量的に示すため, 発話生成手法の選択精度の評価を行った. 被験者はシステムとユーザの対話について, 入力発話も対する応答発話の意味の妥当性を評価した.

4.1 実験方法

4 名の 24, 25 歳の男性理系大学院生を被験者とし, 3 に述べた印象評価実験にて収集された 150 ターンの対話に対してアンケートフォームを用いて評価を行ってもらった. 評価値は, 応答発話内容の主観による内容理解可能性, 文脈的容認度を 0~3 の数値として採用する. 0 はエラーを示し, 3(正応答) は文法的にも意味的にも容認でき, 2(準応答) は文法的に容認でき

るが意味的に不自然な応答, 1 は文法的にも意味的にも容認度が低い誤応答とした。

結果を表 4 に示す。全評価の平均値は 2.274, 標準偏差は 0.847 であった。

表 4: 応答精度評価値 (3 段階評価) の分布。

正応答	準応答	誤応答	エラー
302	176	104	17
50.4%	29.4%	17.4%	2.84%

4.2 考察

実験では, 汎用的対話機能およびドメイン選択発話生成モジュールにより生成された発話の評価を行った。表 5 から入力フィルタ機構による分類が正しく行われたものに正応答評価がなされる傾向があったことが読み取れる。質問応答発話の精度においては標準的なものが得られ, また非タスク指向型対話システムの中で標準的と捉えられているルールベース型対話システムとほぼ同精度で質問応答機能を対話システムに付加することが可能である点において, 本提案手法における汎用的対話機能が有効であると考えられる。

表 5: 発話タグ毎の平均精度値.*は 1 名の評価者による。

タグ	評価平均値
NT (雑談対話)	2.32
QA (質問応答対話)	2.27
提案手法による全発話	2.28
質問応答システム単体*	1.85
ルールベース型システム単体*	2.39

5 おわりに

本稿では, 自由な対話のドメインを推定しドメイン選択発話の生成を行う手法の提案を行った。発話選択精度評価実験においては, 質問応答システム単体との比較を行った場合の, 本システムにおける汎用的な対話機能による発話選択の有効性が示された。

今後の課題として, 応答速度および明示的な発話生成精度の向上, また暗黙的なドメイン推定における検出可能ドメインの粒度および検出精度の向上が挙げられる。ドメイン収束の推定において問題となる粒度, 提案手法の T に該当する収束条件, 知識源の選択において, 他の手法との比較を検討中である。

参考文献

- [1] Lane, I. R. and Kawahara, T. (2005). "Incorporating dialogue context and topic clustering in out-of-domain detection." In *Proceedings of IEEE-ICASSP*, 1, pp.1045-1048.
- [2] 江頭 勇祐, 柴田 知秀, 黒橋 禎夫 (2012). 雑談対話システムにおける強化学習を用いた応答生成モジュールの選択言語処理学会 第 17 回年次大会 発表論文集. pp. 654-657.
- [3] 亀山 恵祐, 荒木 健治, 木村 泰知 (2007). 質問応答システムにおける知識源選択規則の自動獲得の有効性について. 情報処理学会研究報告. 自然言語処理研究会報告 35, pp. 85-90.
- [4] 高橋 瑞希, ジェプカ ラファウ, 荒木 健治 (2010). Web 検索と単語 n-gram モデルを用いた文生成手法の性能評価. 言語処理学会第 16 回年次大会発表論文集, pp. 91-394.
- [5] Higuchi, S., Rzepka, R., and Araki, K. (2008). "A casual conversation system using modality and word associations retrieved from the web." In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pp. 382-390.
- [6] 今井 健太, ジェプカ ラファウ, 荒木 健治 (2010). 複数の対話システムからの応答候補文を用いた最適応答文選択手法の性能評価. 情報処理学会研究報告. 自然言語処理研究会報告 195, pp. 1-7.
- [7] Bond, F., Isahara, H., Fujita, S., Uchimoto, K., Kuribayashi, T., and Kanzaki, K. (2009). Enhancing the Japanese WordNet. In *Proceedings of the 7th Workshop on Asian Language Resources*, pp. 1-8.
- [8] Leacock, C. and Chodorow, M. (1998). "Combining Local Context and WordNet Similarity for Word Sense Identification." In *WordNet: An Electronic Lexical Database*, pp. 265-283.