

対話エージェントのための発話候補絞り込み

Utterance Candidate Narrowing for a Dialogue Agent

岩田 直之*1
Naoyuki IWATA平井 尚樹*1
Naoki HIRAI稲葉 通将*1
Michimasa INABA鳥海 不二夫*1
Fujio TORIUMI石井 健一郎*1
Kenichiro ISHII

*1名古屋大学大学院情報科学研究科

Graduate School of Information Science, Nagoya University

We design a non-task oriented dialogue agent which selects appropriate utterances from candidate sets. In this paper, we propose an utterance candidate narrowing method. The precision improvement can be expected by decreasing the number of candidates. Our method consists of the classification and the calculation using Conditional Random Fields(CRF). In order to prove validity of our method, we checked the performance. The result shows that the accuracy rate of selecting appropriate utterance has been increased about 18.9%.

1. はじめに

我々の社会を支えるロボットは数多く存在する。それは産業用ロボットにとどまらず、介護・医療・福祉を支援するサービスロボットなど我々の身近な生活の中にも増えつつある。そして、今後ロボットはさらに社会に浸透し、ロボットと人間が接する機会は多くなると考えられる。ロボットが人間社会に溶け込むためには、ロボットが人間と自然な対話を行えることが必要不可欠である。

そこで本研究では、人間と自然な対話のできるコンピュータ、すなわち対話エージェントの実現を目指す。対話としては特に「雑談」を取り上げる。我々は、事前に発話候補を用意しておき、対話ではその候補の中から発話を選択して応答する対話エージェントを目指している。人間との自然な対話を実現するためには、大量の発話候補を用意する必要があり、この大量の候補の中から適切な応答を選択しなければならない。

対話エージェントの発話選択に関する先行研究には、磯村らによる統計的発話候補選択法 [Isomura 09] が存在する。この研究では、用意した全ての発話候補を評価し、適切な発話の選択を行っている。しかし、この手法には、あきらかに不適切な発話候補も評価対象になってしまうという問題がある。

そこで本研究では、用意した全ての発話候補を発話の内容で分類し、対話の際には、応答として適した内容を推測して、その内容と一致する発話候補のみを評価対象とするという絞り込みを行う。これにより、発話選択の精度向上を目指す。

2. 発話選択

発話選択とは、事前に生成した発話集の中から発話候補を一つ選択し、対話相手の発話に対して応答を行うものである。なお本論文では、1回の発言を発話とし、会話の始まりから終わりまでの発話系列を対話と定義する。本手法では、問題を自然言語処理に特化するため、対象はテキスト対話とする。

まず、対話の1時点における状態を文脈 c とし、以下のよう表す。

$$c = \{u_1, u_2, \dots, u_l\} \quad (1)$$

各要素 $u_i (i = 1, 2, \dots, l)$ は文脈中に出現した発話を、 l は文脈中の発話数を表す。ただし u_1 は文脈 c における最後の発話を、 u_l は最初の発話を意味する。便宜上、 u_0 は文脈 c に対する応答として選択される発話を意味することとする。

次に文脈 c に対して用意する発話候補集合 A_c を以下のよう表す。

$$A_c = \{a_1^c, a_2^c, \dots, a_{|A_c|}^c\} \quad (2)$$

各要素 $a_i^c (i = 1, 2, \dots, |A_c|)$ は各発話候補を示し、文脈 c に対する応答として適切な発話 (正解発話) と、適切でない発話が混在する。また、 $|A_c|$ は文脈 c のときの発話候補数である。

ここで、正解発話集合 R_c を以下のよう表す。

$$R_c = \{r_1^c, r_2^c, \dots, r_{|R_c|}^c\} \subseteq A_c \quad (3)$$

各要素 $r_i^c (i = 1, 2, \dots, |R_c|)$ は各正解発話を示し、 A_c におけるいずれかの要素である。また、 $|R_c|$ は文脈 c のときの正解発話数である。全ての c について A_c は生成可能であり、各 c に対し A_c は1個以上の正解発話 r_i^c を含むものとする。

発話選択とは、ある文脈 c において、発話候補集合 A_c の中から正解発話集合 R_c を見出し、応答を行うものである。本論文で用いる発話選択手法は、ある文脈 c が与えられた際に、応答としての適切さをもとに発話候補に順位付けを行い、最も上位の発話候補を出力するという、磯村らによる統計的発話候補選択法 [Isomura 09] をもとにした手法である。磯村らは発話候補への順位付けを行う手法として、シグモイド関数や最大エントロピー法を用いている。これに対し、本手法では、ランキング学習手法である ListNet [Cao 07] を用いて発話候補への順位付けを行う。発話選択の際に発話の特徴ベクトルに変換するための素性は、磯村らが用いた素性と同様である。

3. 発話候補の絞り込み手法

3.1 概要

本論文では、発話集 A_c の中から発話選択を行う際に、発話候補の絞り込みを行うことを目的とする。絞り込み手法の概要を図1に示す。

まず、発話候補の分類を行う。分類基準には、発話の種類を記述するためのタグである SWBD-DAMSL (Switchboard Discourse Annotation and Markup System of Labeling) タグ [Jurafsky 97] を内容ごとにまとめた簡易 DAMSL タグを用

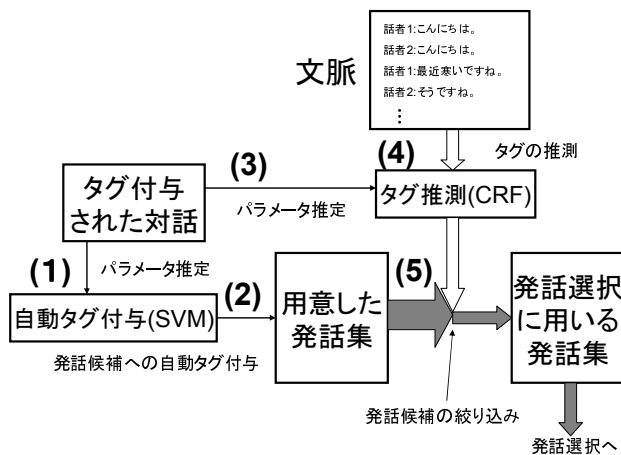


図 1: 絞り込み手法の概要

表 1: 簡易 DAMSL タグ

タグ	内容	例文
Statement	意見・説明	「この餡は甘いです」
Question	質問	「なぜですか？」
Tag.Question	付加疑問文	「一人暮らしですよね？」
Understand	相槌・了解	「なるほど」
Response	同意・拒否	「確かにその通りです」
Commit	提案・遂行	「いろいろ試してみてもは？」
Advice	指示・命令	「詳しく話して下さい」
Thanking	感謝	「ありがとう」
Apology	謝罪	「ごめんなさい」
Greeting	挨拶	「こんにちは」
Admiration	感嘆	「えっ」「へえー」
Others	その他	(独り言, 引用など)
Unknown	未知タグ	(打ち間違いなど)

いる。簡易 DAMSL タグは表 1 に示した 13 種類のタグで定義する。出現回数の少ないタグについては、その他「Others」で統一し、非言語や不明語については、未知タグ「Unknown」で表す。発話候補の分類を行う際には、磯村らによる SVM(Support Vector Machine) 自動タグ付与手法 [磯村 09] を用いる。図 1 では (1) でタグ付与済みの対話から SVM のパラメータを学習し、(2) で用意した発話候補へ自動タグ付与を行う。

次に文脈から、応答として適切な内容を推測し、推測結果に応じて使用する発話候補を決定する。推測する内容は簡易 DAMSL タグである。タグの推測には、CRF(Conditional Random Fields)[Lafferty 01] を用いる。(3) でタグ付与済みの対話から CRF のパラメータを学習し、(4) で文脈からのタグ推測を行う。推測したタグより、(5) で発話選択に使用する発話候補を決定する。これら (1)~(5) で示す手法を、本研究での絞り込み手法として提案する。

ここで、「あなたはどのように思いますか?」という発話に対する絞り込みの例を示す。この場合、相槌タグ「Understand」(「なるほど」)や感謝タグ「Thanking」(「ありがとうございます」)、謝罪タグ「Apology」(「ごめんなさい」)などは応答する内容としてあきらかに不適である。よってこれらのタグが付与された発話を、発話選択を行う前に候補から外す。

3.2 CRF を用いた推測手法

3.2.1 Conditional Random Fields

タグ推測に用いる Conditional Random Fields(CRF) は系列ラベリング問題を解く際に用いられる識別モデルである。文脈へのタグ付与では、文脈の発話系列 $u = u_1 u_2 \dots u_l$ とタグ系列 $v = v_1 v_2 \dots v_l$ の対応関係を条件付確率 $P(v|u)$ で表現する。なお、 u_i は一つの発話を表し、 v_i はその発話に付与するタグを表す。CRF は直接 $P(v|u)$ を最大化することにより、タグの自動付与を行う。素性を表現するため、CRF では $\{0,1\}$ の 2 値を返す素性関数 f を用いる。また、 f に対応する重みパラメータを λ で表し、 k 番目の素性に対する素性関数を f_k 、重みパラメータを λ_k とする。このとき、 $P(v|u)$ は式 (4) で表される。

$$P(v|u) = \frac{1}{Z_u} \exp\left(\sum_{i=1}^l \sum_k \lambda_k f_k(u, v, i)\right) \quad (4)$$

ただし Z_u は全系列の総和を 1 にするための正規化項であり、式 (5) で表される。

$$Z_u = \sum_v \exp\left(\sum_{i=1}^l \sum_k \lambda_k f_k(u, v, i)\right) \quad (5)$$

パラメータ λ は最尤推定で求めることができる。

入力発話系列 u に対して、最適なタグ系列 \hat{v} は、

$$\hat{v} = \arg \max_v P(v|u) \quad (6)$$

により求めることができる。タグ系列 \hat{v} の計算には、Viterbi アルゴリズムを用いる。

3.2.2 素性集合

ここでは、タグの性質を表現しうる特徴を複数列挙したものである素性集合について述べる。本手法では、現発話 u_i のタグを付与するための素性として、以下の 3 つを用いる。

- (1) 一つ前の発話 u_{i+1} の特徴ベクトル
- (2) 二つ前の発話 u_{i+2} の特徴ベクトル
- (3) 一つ前のタグ v_{i+1}

このうち、(1)(2) での発話の特徴ベクトルは、以下の 3 つの特徴を用いて表現する。

- 発話中の n-gram
学習データ中に、ある閾値回以上出現している n-gram を特徴として用いる。これらの特徴量は、発話中に、ある n-gram が存在するときは 1、存在しないときは 0 となる。
- 発話の語尾 3-gram
語尾の単語が発話の内容に強く関係すると考えられる。例えば「...ですか?」であれば質問、「...ですよね」であれば同意を求める等、語尾によって発話の意図が変わる。そこで、学習データ中に閾値回以上出現している発話の語尾 3-gram を特徴とする。
- 発話の形態素数
定量的に発話を表す特徴として、発話の形態素数を用いる。特徴を持つどうかは、学習データ中の発話の平均形態素数以上か否かで判別を行う。すなわち、形態素数が平均以上であれば 1、平均未満であれば 0 となる。

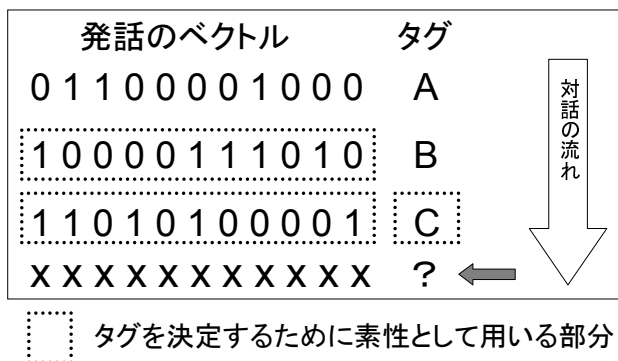


図 2: タグ推測

こうして得られる特徴ベクトルを用いて (1)(2) では CRF の素性を定義する．素性関数は発話の位置も考慮に入れ，一つ前の発話の特徴 n が 1 のときに 1 を返す素性関数，二つ前の発話の特徴 n が 1 のときに 1 を返す素性関数，と発話の位置別に素性関数が定義される．

次に，(3) の一つ前のタグでは，タグと 1 対 1 に対応する特徴をタグの種類分用意し，一つ前のタグと一致する特徴のみ 1 を返す素性関数を定義する．

これら素性集合は，タグごとに異なる素性関数で表現される．

3.2.3 タグ推測

発話系列 u の最後に特徴を持たない発話 u_0 を付け足し，新たに得られる発話系列 $u' = u_1 u_{l-1} \dots u_1 u_0$ ヘタグ付与を行う．発話 u_0 には一つ前の発話 u_1 ，二つ前の発話 u_2 ，一つ前のタグ v_1 の特徴からタグ v_0 が付与される．一つ前と二つ前の発話の内容からタグが付与されることより，タグ v_0 が文脈 c の次に来るタグと推測される．図に示すと図 2 となる．図 2 の中の点線で囲んだ部分は，CRF の素性に用いる部分である．ただし，タグ「Unknown」は推測するタグの候補から除外する．

タグの推測を行うにあたり，文脈の次の応答として適切なタグは一つとは限らない．例えば「得意なスポーツはサッカーです」という発話に対して，以下のように複数のタグが応答として考えられる．

- 意見タグ「Statement」(「サッカー選手ってカッコいいですね」)
- 質問タグ「Question」(「サッカーのどこが好きですか?」)

そこで，文脈ごとに複数のタグを推測する方法として，各タグの周辺確率の値を利用する．周辺確率の値は，CRF を用いたタグ付与で得られ，各タグがどの程度正しく付与されているかを表す．例えば，ある文脈に対し，「Statement」は 0.8，「Question」は 0.1 と各タグに与えられる．タグ推測において，この数値が高いほど，応答として次に来るタグとして正しいということを表す．以下に複数タグの推測方法の手順を示す．

1. 0 以上 1 以下の閾値 t を設定する．
2. 全タグのうち，タグ付与の結果得られる周辺確率の値が最大のタグを選択し，確率の値を閾値 t と比較する．
3. 閾値以上の場合は，そのタグを推測されるタグとする．
4. 閾値未満の場合は，次に確率の値が大きいタグを選び，2 のタグとの確率の和を閾値 t と比較する．
5. 3,4 の手順を確率の和が閾値以上となるまで繰り返す．

表 2: 重み付けによる値の変化の例

タグ	周辺確率	用いる重み (重みの値)	重み付け後の値	正規化後の値
S	0.85	$w_{(S,1 \text{ 位})}$ (0.3)	0.255	0.27
Q	0.1	$w_{(Q,2 \text{ 位})}$ (3.5)	0.35	0.38
U	0.05	$w_{(U,3 \text{ 位})}$ (6.5)	0.325	0.35

このように周辺確率の値が高い順にタグに順位をつけ，閾値 t までのタグを取り出すことで複数タグを推測する．

3.2.4 重み付け学習

簡易 DAMSL タグは性質上，対話中の出現頻度に差がある．例えばタグ「Statement」はタグの意味する内容の幅が広いいため，対話中の様々な場面で用いられる．その結果「Statement」はタグ推測の際にも周辺確率の値が大きくなりやすく，1 位に推測されることが多い．これにより，下位タグに関係なく「Statement」のみしか推測されない，という問題が生じる．

そこで本手法では，タグの推測順位をもとに周辺確率への重み付けを行う．あるタグのある順位に対する重みを $w_{(\text{タグの名前}, \text{順位})}$ と表記する．例えば，推測したタグの順位が，順に 1 位:「Statement(以下 S と表記)」，2 位:「Question(以下 Q)」，3 位:「Understand(以下 U)」である場合に，「S」には重み $w_{(S,1 \text{ 位})}$ を，「Q」には重み $w_{(Q,2 \text{ 位})}$ を，「U」には重み $w_{(U,3 \text{ 位})}$ を用いる．

表 2 に重み付けによる値の変化の例を示す．閾値 t を 0.9 と設定すると，重み付けをしない場合に推測されるタグは「S」と「Q」の二つ ($0.85 + 0.1 > 0.9$) となる．周辺確率に重みをつけ，周辺確率の和が 1 となるよう正規化を行うと，最終的な値は「S: 0.27」「Q: 0.38」「U: 0.35」となる．この場合，推測されるタグは「Q」「U」「S」の三つ ($0.38 + 0.35 + 0.27 > 0.9$) となる．このように，重み付けにより推測するタグを変え，順位が下位のタグについても推測されるようにする．

重み w の設定には，文脈 c に対する応答としてどのタグが正解を示す学習データと，文脈 c の次に来ると推測されるタグの周辺確率の値を用いる．以下に重み学習の手順を示す．

1. 周辺確率の値を重み付けし，タグ推測を行う．
2. 推測したタグが全て正解の場合は，タグ推測の順位で 1 位のタグの重みを増加させる．
3. 推測したタグが正解と不正解を含む場合は，不正解タグの重みを減少させ，正解タグの重みを増加させる．
4. 推測したタグが全て不正解の場合は，不正解のタグの重みを減少させ，全タグのうち最も上位の正解のタグの重みを増加させる．
5. 上記の手順を一定数繰り返す．

正解のタグを正しく推測するように繰り返し重みの学習を行うことにより，各タグ各順位の重みが得られる．

4. 評価実験

4.1 実験の概要と手順

提案手法の有効性を確認するための評価実験を行った．評価実験には，タグ推測 (CRF) のパラメータ推定に用いる対話データと，発話選択に用いる対話データの 2 種類を用いた．

パラメータ推定に用いる対話データは互いに面識のない被験者により行われた全 59 の対話であり、全発話に手動でタグが付与されている。

発話選択に用いる対話データは被験者 12 人による対話である。被験者は実験者と 1 対 1 のチャット形式のテキスト対話を行った。実験者は被験者のプロフィールをもとに作成した発話候補集 A_c の中から発話候補 a_i を選択し、応答を返した。全 12 の対話が得られ、この対話の中から実験者が a_i による応答をしたとき、その直前の被験者の発話までを文脈 c とした。文脈 c の総数は 185 個であり、これを対話データとして用いた。正解データ集合 R_c は、各文脈 c に対して適切な発話候補 a_i を人手で選択することにより作成した。

発話選択には、185 分割交差検証法を用いた。すなわち、185 文脈のうち 184 文脈を学習データ、残り 1 文脈をテストデータとして、発話集の中の発話候補の順位付けを行った。

この順位付けされた発話候補をもとに、絞り込み手法の性能評価を行った。性能評価は、最適な発話を選択する精度を以下の 3 つの場合において比較することにより行った。本実験では順位付けされた発話候補の上位 5 位までを評価対象とした。

- ベースライン - 絞り込みなし

順位付けされた発話候補をそのまま発話選択の結果とする。

- 手法 1 - タグ推測による絞り込み

文脈 c の次に来るタグの推測を行う。次に、順位付けされた発話候補へタグ付与を行い、推測したタグを含む発話候補のみを発話選択の結果とする。

- 手法 2 - タグ推測による絞り込み+重み学習

手法 1 のうち、タグ推測を重み w を用いて行う。

発話候補 a_i が文脈 c における正解発話 r であるか否かの判定は、人手で行った。発話候補の評価は、主観による影響を減らすため、3 人の評価者による多数決で行った。

4.2 実験結果

第 4.1 節で示した各手法における順位付けの結果を図 3 に示す。横軸は正解発話 r が初めて出現する順位を表し、縦軸はその累積である。すなわち「その順位以内に正解発話を少なくとも一つ含む」文脈 c の割合を示している。

正解発話が 1 位となる文脈 c の割合は、割合の高い順に手法 2 (絞り込み+重み) が 60.0%、ベースライン (絞り込みなし) が 41.1%、手法 1 (絞り込み) が 36.2%であった。CRF のタグ推測に重みを用いた手法が他の手法より高い値を示しており、ベースラインと比較した際に 18.9%正解の割合が向上している。5 位については、手法 2 とベースラインともに差がないが、対話エージェントに用いる際には、選択される発話は 1 位の発話のみであるため、提案手法の有効性は確認できた。

精度が向上している理由として、相槌タグ「Understand」の推測が増えたことが挙げられる。「そうですね」や「なるほど」などの相槌は、相手の発話への応答として適切である場合が多い。ただし、人間と自然な対話を行うためには相槌だけでは不自然であり、相槌の後にさらに発話を付け足して応答を行うことにより、相槌のみに比べてより自然な応答となる。

発話候補中上位 3 位以内に適切な応答が含まれている割合は 80%と高い。そこで、討論や対談の場において、対話内容や相手の情報に応じて適切な発話を提示する、という人間の応答補助への応用が可能であると考えられる。

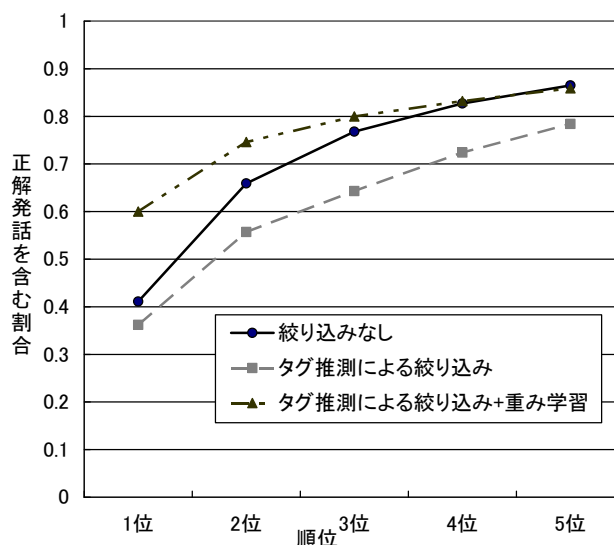


図 3: 発話選択性能の比較

5. まとめ

本論文では、対話エージェントにおける発話候補の絞り込み手法を提案した。本手法では、簡易 DAMSL タグを用いた発話内容の分類と、CRF を用いたタグの推測を組み合わせることにより、発話候補の絞り込みを行った。評価実験の結果、適切に発話候補を選択する精度は 41.1%から 60.0%へ向上しており、本手法の有効性が確認された。

今後の課題として、本手法とは別の分類手法を用いることが挙げられる。他の分類手法として、どのような話題に関する対話が行われているのかという対話の主題を把握し、その話題に関する発話を選ぶことにより発話候補を絞り込む方法が考えられる。

参考文献

- [Isomura 09] Isomura, N., Toriumi, F., and Ishii, K.: Statistical Utterance Selection Using Word Co-occurrence for a Dialogue Agent, Principles of Practice in Multi-Agent Systems, pp.68-79, (2009)
- [Cao 07] Cao, Z. Qin, T. Liu, T.Y. Tsai, M.F. and Li, H.: Learning to rank: from pairwise approach to listwise approach Proceedings of the 24th international conference on Machine learning, pp.129-136, (2007)
- [Jurafsky 97] Jurafsky, D. Shriberg, E. and Biasca, D.: Switchboard SWBD-DAMSL Shallow-Discourse-Function Annotation Coders Manual, Institute of Cognitive Science, Tech. Rep, pp.97-01, (1997)
- [磯村 09] 磯村直樹, 鳥海不二夫, 石井健一郎: 対話エージェント評価におけるタグ付与の自動化, 電子情報通信学会論文誌, Vol.92, pp.795-805, (2009)
- [Lafferty 01] J. Lafferty, A. McCallum, and F. Pereira.: Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In Proc. Eighteenth International Conference on Machine Learning Table of Contents, pp.282-289, (2001)