

## 対話シナリオ構築におけるユーザ回答候補の推定

## Predicting User Answer Candidates for Creating Dialogue Scenario

大塚 淳史\*<sup>1</sup>  
Atsushi OTSUKA片山 太一\*<sup>1</sup>  
Taichi KATAYAMA杉山 弘晃\*<sup>2</sup>  
Hiroaki SUGIYAMA東中 竜一郎\*<sup>1</sup>  
Ryuichiro HIGASHINAKA松尾 義博\*<sup>1</sup>  
Yoshihiro MATSUO\*<sup>1</sup> 日本電信電話株式会社 NTT メディアインテリジェンス研究所  
NTT Media Intelligence Laboratories, Nippon Telegraph and Telephone Corporation\*<sup>2</sup> 日本電信電話株式会社 NTT コミュニケーション科学基礎研究所  
NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation

When building a dialogue system, system builders predict user responses to the dialogue system question, and describes as the dialogue scenario. However, in the chat-oriented dialogue system, user's answer content is often diverse and it is difficult to predict all of them. In this paper, We propose a novel approach for predicting various user answers by a set expansion approach. Predicted answers by our method matches with the more user answers than previous methods.

## 1. はじめに

スマートフォンによる音声エージェントやコミュニケーションロボットの実用化に伴い、エージェントやロボットと自由な対話をするための対話技術の必要性はますます高まっている。特に近年では、エンターテインメントやカウンセリングを目的とした雑談対話に関する研究 [Ritter 11, Wong 12, Higashinaka 14] が増加している。

商用における雑談対話を実現する最も一般的な手法は、“シナリオ”と呼ばれる応答ルールを手手で大量に記述する手法である。シナリオベース型の対話システムでは、対話システムの入出力を全て手手で記述するため、ルールに合致する範囲であれば、非常に質の高い対話が可能である。しかしながら、ルールに合致しない発話が発話された場合には、対話システムは“わかりません”など簡単な応答となってしまう。そのため、シナリオベース型の対話システムにおいて、質の高い対話を実現させるためには膨大なルールを記述することが必要となる。

対話シナリオ作成において重要なことは、ユーザの発話内容をできるだけ網羅的に予測することである。しかしながら、雑談対話において、ユーザの発話内容を網羅的に予測することは非常に困難である。システム側が先に質問を投げかけることで、ユーザの発話内容をある程度限定することは出来るが、質問に対する回答を全てをシナリオ作成者が列挙することは難しい。そこで本論文では、シナリオベース型対話システムのシナリオ作成を支援を目的とした、質問に対する回答内容を予測する手法を提案する。シナリオ作成者は、質問文と、その質問に対する少量の回答集合（回答シード）を入力することで、回答シードを拡張した様々な回答候補を得ることができる。

質問に対する回答を予測するために、ニューラルネットワークによる単語ベクトルと、言語モデルを組み合わせた手法を提案する。回答シードと単語ベクトルを用いて、回答シードに対する関連語を抽出する。その後、質問文から作成した文パターンに抽出した関連語を当てはめた文の言語モデル上の尤度を算

出することで、関連語の中から、質問に対する回答として適切な語のみを抽出する。実験では、本手法によって得られた新たな回答候補によって、質問に対して、より多くのユーザの回答がカバーできるようになることを示す。

## 2. 関連研究

雑談対話システムの実現には、Artificial Intelligence Markup Language (AIML) と呼ばれるマークアップ言語を用いた手法 [Wallace 09] がある。AIML によって雑談対話の入力と出力のパターン（シナリオ）を大量に記述することにより、様々な発話に対して応答可能な対話システムを構築する。それと同時に、統計的処理をベースにした手法 [Higashinaka 14] やディープラーニングによる手法 [Vinyals 15] が提案されている。機械学習ベースの手法では、人手を介さずに多様な話題に関する雑談対話を実現するために非常に有用な技術であるが、実現できる対話の質が低く、多くの場合で対話自体が途中で破綻してしまうという課題 [Higashinaka 15] がある。目黒ら [目黒 14] は応答の質の向上と、応答バリエーションの両立を目指して、シナリオ型対話と統計型対話を組み合わせたハイブリッド型の雑談対話システムを提案している。

シードとなる幾つかのキーワードを与えて、シードに関連するキーワードを取得する手法は Set Expansion と呼ばれている。Wang ら [Wang 08] は回答に一覧を返すリスト型の QA について Set Expansion を利用した手法を提案している。Set Expansion に Web 検索や HTML の構造を用いることで、高精度で正解リストを回答できることを示している。He ら [He 11] は、Web のリストデータをクリックログデータを用いた Set Expansion の手法を提案している。この時、候補となるキーワードを採用するかどうかの閾値を動的に変化させていくことにより従来手法よりも高精度な Set Expansion が可能なることを報告している。Zhang [Zhang 16] らは、Web のクエリログからエンティティ間の意味の関係を表すグラフ構造を生成し、生成したグラフを探索することによる Set Expansion 手法を提案している。

本論文では、シナリオ型対話を対象とするが、シナリオ中のユーザの回答を統計的手法によって予測するという特徴がある。また、回答を予測する手法では、Set Expansion をベース

連絡先: 大塚淳史, 日本電信電話株式会社 NTT メディアインテリジェンス研究所, 〒239-0847, 神奈川県横須賀市光の丘 1-1, otsuka.atsushi@lab.ntt.co.jp

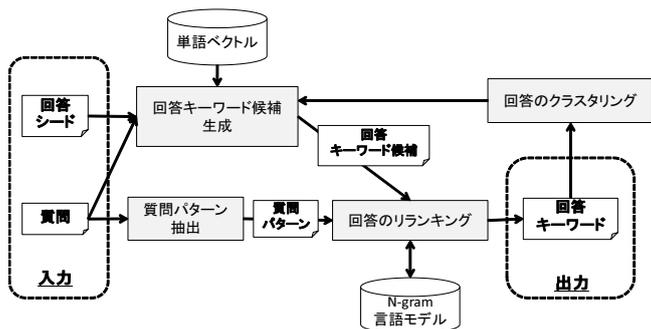


図 1: 提案手法の構成

とする。本論文の Set Expansion では、質問文と言語モデルを用いて、拡張したキーワードが質問文に対して妥当であるかを評価するという特徴がある。

### 3. 問題設定

本節では、本論文が扱う対話シナリオ、並びに回答候補推定タスクに関する定義を行う。

[定義 1] 本論文で扱う対話シナリオは、1 発話目：システムの質問文、2 発話目：ユーザの回答、3 発話目：システムの発話の形式となる 3 つ組対話の形式とする。2 発話目のユーザ回答はルール形式で記述する。なお本論文では、3 発話目については対象としない。

[定義 2] 回答候補推定は、シナリオの 1 発話目（システムの質問文）と、回答シードを入力することで、対話シナリオの 2 発話目のルールとして使用できる  $N$  個の回答キーワードを得るタスクである。

[定義 3] 質問文は、自然文で記述される。なお、本論文で扱う質問は“犬と猫どちらが好き?” の様な 2 択での選択するような質問ではなく、“好きな食べ物は?” の様に、多数の回答が想起されるタイプの質問とする。

[定義 4] 回答シードは、シナリオ作成者が入力する、質問に対する回答の例である。回答シードは質問に対して想定できる回答として、 $S = \{w_0, w_1, \dots, w_{n-1}\}$  の  $n$  個のキーワードで表現する。

### 4. 提案手法

本節では、本論文が提案する回答候補推定について述べる。図 1 に回答候補推定の構成図を示す。回答候補推定は、単語ベクトルによる回答キーワード候補生成、質問文からの文パターン抽出、文パターンによる回答キーワードのリランキング、回答キーワードのクラスタリングから構成される。以降はそれぞれについて説明する。

#### 4.1 単語ベクトルによる回答キーワード候補生成

近年、単語の意味を連続値のベクトルで表現する手法が多く提案されている。特に、Mikolov ら [Mikolov 13] のニューラルネットワークをベースとした単語ベクトルの生成法では、ベクトルの加減演算がそのまま意味上の関係性の演算となるという特徴を持つ。本論文では、単語ベクトルの加減演算を用いて、回答シードと質問文に関連するキーワードを抽出する。



図 2: 質問文からの文パターン生成

回答シードと関連するキーワードを抽出するために、回答シードを表現するベクトルを計算する。回答シードのベクトル  $\mathbf{v}_s$  は、回答シードに含まれる単語ベクトルの重心を計算する以下の式によって表わされる。

$$\mathbf{v}_s = \frac{1}{n} \sum_{w \in S} \mathbf{v}(w) \quad (1)$$

ここで、 $\mathbf{v}(w)$  は回答シード中の単語  $w$  の単語ベクトルを表している。回答シードのベクトル  $\mathbf{v}_s$  とのコサイン類似度上位の単語を取得することで、回答シードに関連するキーワードを取得することができる。

本論文では、更に質問文の情報を考慮する。質問文中の単語のベクトルを、回答シードのベクトルに加算することで、回答シード中の単語だけでなく、質問文中の単語とも共起するキーワードを得る。質問文を形態素解析し、自立語のみを取り出した単語集合を  $Q$  とする。ここで、回答キーワード候補  $w'_i$  の最終的なスコア  $score_v(w'_i)$  は以下の通り計算する。

$$score_v(w'_i) = \sum_{q \in Q} \cos(\mathbf{v}_{s+q}, \mathbf{v}(w'_i)) - \cos(\mathbf{v}_s, \mathbf{v}(w'_i)) \quad (2)$$

$$\mathbf{v}_{s+q} = \mathbf{v}_s + \mathbf{v}(q) \quad (3)$$

ここで、 $\cos(\mathbf{v}_{s+q}, \mathbf{v}(w'_i))$  は、回答シードと質問文中の単語により生成された単語ベクトル  $\mathbf{v}_{s+q}$  と単語  $w'_i$  の単語ベクトル  $\mathbf{v}(w'_i)$  のコサイン類似度を示している。

#### 4.2 質問文からの文パターン抽出

本論文では、質問文に対して妥当な回答となるキーワードを複数出力することを目的としている。回答が質問文に対して妥当であるかどうかを評価するために、質問文を平叙文に変換した形での文パターンを作成する。文パターンに回答となるキーワードを当てはめ言語モデルによるスコアを算出することで回答候補となるキーワードが質問文に対して妥当な回答であるかを判定する。

質問文から生成する文パターンを図 2 に示す。図中の  $\langle candidate \rangle$  は回答キーワードが挿入されるタグを示している。文パターンの生成ではまず、質問文を係り受け解析し、文節単位に分割する。文節中に動詞が含まれる文節を基準文節に設定する。基準文節よりも前の文節に関しては、基準文節から順番に、係り受け関係に従い、文節を接合していく。1 回文節を合成する毎に先頭に  $\langle candidate \rangle$  のタグを挿入していく。基準文節の後ろの文節に関しては、基準文節の一つ後ろの文節から係り受け関係に従い文節を接合していく。このとき、 $\langle candidate \rangle$  タグは作成した文パターンの最後尾に接続する。

文パターンの生成の際には、格助詞は全て “\*” 等のワイルドカードに変換しておく。本論文では、格助詞の違いによる意味の違いを区別しない。また、図 2 の例では、回答キーワードを挿入する <candidate> タグは文パターンの先頭または最後尾に接続詞であったが、質問文に “どこ” や、“何”、“誰” といった 5W1H の表現が含まれている場合、これらの表現を <candidate> タグに置換して文パターンを作成する。この時、文パターンの前後への <candidate> タグ挿入は行わない。

#### 4.3 文パターンによる回答キーワードのリランキング

単語ベクトルから抽出した回答キーワードと、質問文から作成した文パターンを用いて、提案法が最終的に出力する回答キーワードを決定する。4.1 節の手法によって得た、 $N_0$  個の各回答キーワード候補を、文パターンの <candidate> タグに当てはめた文を作成していく。作成した文について、N-gram 言語モデルスコアを算出する。回答キーワード候補  $w'_i$  について、質問文から作成した文パターンの集合  $T$  を与えた時、言語モデルスコア  $score_{t_i}$  は、言語モデルが生成する文の出現確率により、次式で計算される。

$$score_{t_i}(w'_i) = -\left(\sum_{t \in T} \log P(t, w'_i)\right) + \log P(w'_i) \quad (4)$$

ここで、 $P(t, w'_i)$  は、文パターン  $t$  に回答キーワード候補  $w'_i$  を当てはめたときの、N-gram 確率である。 $P(w'_i)$  は単語  $w'_i$  の uni-gram 確率である。

単語ベクトルの類似度にもとづいて算出した  $score_w$  と、言語モデルから算出した  $score_{t_i}$  に基づいて、最終的に出力する  $n$  個の回答キーワードを決定する。本論文で必要とする回答キーワードは、回答シードに関連しているかつ、質問文に対して妥当なキーワードである。最終的なスコアは、 $score_w$  と  $score_{t_i}$  について、それぞれで回答キーワードを順位付けし、その順位の平均値を用いる。回答キーワード候補  $w'_i$  についての最終スコア  $score(w'_i)$  は以下の式で計算できる。

$$score(w'_i) = \frac{1}{2} (Rank_{score_w}(w'_i) + Rank_{score_{t_i}}(w'_i)) \quad (5)$$

ここで、 $Rank_{score_w}(w'_i)$  は、回答キーワード候補を  $score_w$  の降順に並べたときの、単語  $w'_i$  の順位であり、 $Rank_{score_{t_i}}(w'_i)$  は  $score_{t_i}$  の順位である。 $N_0$  個の回答キーワード候補について、 $score$  を計算し降順に並べた上位  $N$  件を提案法が出力する回答キーワードとなる。このとき、 $N_0 > N$  である。

#### 4.4 回答キーワードのクラスタリング

質問文に対してより多様な回答キーワードを得るため、単語ベクトルによる回答キーワード候補生成と文パターンによるリランキングを複数回ループさせる。回答キーワードのリランキングによって得られた  $n$  個の回答キーワードを新たな回答シードとして利用する。しかしながら、(1) 式によって回答シードのベクトルを作成する場合、ベクトルは全単語ベクトルの重心ベクトルになってしまうため、単語それぞれの情報が十分に考慮されなくなってしまうという問題がある。

本論文では、回答キーワードを一度クラスタリングし、各クラスタ毎に 2 回目以降の回答候補推定を行う。ここで、新たな回答シードは各クラスタに属する回答キーワード集合を用いる。クラスタリングには Affinity Propagation (AP)[Frey 07] を用いる。AP は各クラスタでメッセージと呼ばれる値の交換を繰り返すことによってクラスタを決定していく。k-means のようにクラスタ数を事前に決定する必要がないという特徴を持つ。回答キーワードについて、単語ベクトルのコサイン類似度

を計算し、回答キーワード間のベクトル上での距離を求める。計算した距離に基づいて AP によるクラスタリングを行う。

AP によって作成したクラスタごとに 4.1 節と 4.3 節の計算を行う。ここで、回答キーワード候補  $w''_i$  のスコア  $score_w(w''_i)$  と  $score_{t_i}(w''_i)$  は全クラスタの合計値となる。

## 5. 評価実験

本節では評価実験について説明する。まず、実験のデータセットである質問ランキングサイトについて説明し、次に設定について説明し、最後に実験結果を示す。

### 5.1 実験データセット

データセットとして用いる質問ランキングサイトとは、日常の様々な話題についてランキング化したものである。題材は「居酒屋で必ず頼む、定番メニューは？」といった雑談に近いものであるという特徴がある。このランキングサイトの質問文を対話シナリオの質問文、ランキング結果をユーザの回答と見なすことで本論文が想定している対話シナリオの形式と合致する。

ランキングサイトのランキング結果の項目は、サイト運営者が作成したものであり、ランキング結果は、サイト運営が作成した選択肢の中からユーザが選んだ結果である。同じ質問文であったとしても、選択肢から回答する場合と自由記述によって回答する場合では回答結果が異なると考えられる。本論文が想定する対話は、自由記述による回答に近い。そのため、より対話に近いデータを得るため、クラウドソーシングを用いた応答収集作業を行った。質問ランキングサイトの質問文を 500 人のクラウドワーカーに提示し、その質問に対する回答を自由記述で回答してもらうことで、評価用の回答集合を作成した。

### 5.2 実験設定

評価実験では、質問ランキングサイトの質問文と回答をデータセットとして、ユーザの応答カバー率を評価尺度とした実験を行う。提案法が予測した回答キーワードが、ユーザの回答と一致していれば、提案法による回答予測によって、より多くのユーザの回答を想定した対話シナリオが実現できたことになる。

回答キーワードを生成するための回答シードには、質問ランキングサイトにあった選択肢を使用する。正解となる回答集合はクラウドソーシングによって収集した回答集合を用いる。クラウドソーシングによって収集した 500 人の回答から 1 回しか出現しなかった回答を除いたものを正解集合として、正解集合と、提案法が出力した回答キーワードとの一致率 (回答カバー率) を調査することで、提案法がユーザの回答をどの程度予測できたかを評価する。実験では、質問ランキングサイトから 70 の質問を抽出し、提案法により回答を 100 個生成したときの、各質問での回答カバー率を計算する。

応答カバー率の実験では、質問ランキングサイトのランキング結果の項目のみを使用したときの応答カバー率を初期値とする。比較実験では、(1) 式の回答シードベクトルの類似度のみで回答を生成したものを、提案法の 4.4 節のクラスタリング手法を用いず、1 回で回答を 100 個生成したもの、そして 4.4 節のクラスタリング手法を適用して 1 回のループで回答を 50 個生成する操作を 2 回行ったもので比較を行う。単語ベクトルモデルと N-gram 言語モデルの生成には 1 年分のブログ記事のデータを用いる。

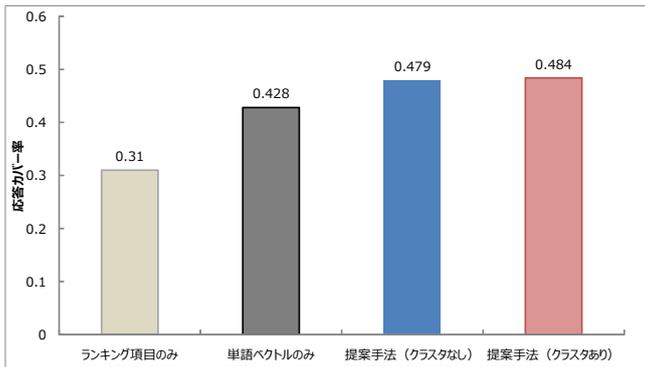


図 3: 応答カバー率実験の結果

表 1: 回答キーワードの生成例

質問文	回答シード	回答キーワード
居酒屋で必ず頼む定番メニューは?	枝豆, サラダ, フライドポテト, 唐揚げ, たこわさ	ポテトサラダ, 出汁巻き玉子, 揚げ出し豆腐, 冷やしトマト, 揚げ餃子, もろきゅう
二人で行ったら「脈アリ」なスポットといえばどこ?	夜景, スポット, 海, 遊園地, 高級レストラン	観覧車, イルミネーション, 海遊館, 山下公園, お台場, スカイツリー, 花火
ちょっとした空き時間何してる?	ゲーム, 睡眠, ネットサーフィン, メール, 読書	自習, 暇, 予約, 調べ物, 映画鑑賞, DS, ぐーたら, ヒトカラ, ナンプレ

### 5.3 実験結果

実験結果を図 3 に示す。また、実際に生成した回答キーワードの例を表 1 に示す。グラフの縦軸は回答カバー率を示している。質問ランキングサイトのランキング項目のみを使用した場合の回答カバー率は 31% であり、3 人に 1 人程度の回答しか対応出来ていないことがわかる。回答シードベクトルとの類似度のみから収集した場合の回答カバー率は 43% であり、10 ポイント程度の向上が見られる。質問文と言語モデルを考慮した提案手法 (クラスタなし) での回答カバー率は 48% であり、ベクトル類似度のみを用いた手法に比べて回答カバー率が大幅に向上していることがわかる。回答キーワード生成を 2 回に分割し、1 回目の回答キーワード集合をクラスタリングし、新たな回答シードとする手法では、クラスタリングを行わない手法を比較して僅かに回答カバー率が向上している。

## 6. 考察

評価実験の結果、提案法は単語ベクトルのみから生成した回答キーワードよりも、多くのユーザの回答に対応出来ることが明らかとなった。単語ベクトルの類似度を用いる場合、回答シードの各々の単語の意味の影響を大きく受けるため回答シード中の各単語の類義語が多く出力されるという傾向があった。提案法では、質問文を使用することで、質問文中の単語と共にしやすい語がより上位に出力される様になったと考えられる。

提案法によっても応答カバー率が向上できなかった原因とし

て、回答シードと自由記述形式の回答との乖離が考えられる。例えば“職場に導入して欲しいものは?”という質問に対して、質問ランキングサイトのランキング項目は“フレックスタイム制”、“育休制度”などであった。しかしながら、クラウドソーシングによって得た回答では“エアコン”や“ウォーターサーバー”といったものが多数を占めていた。このような「製品」に関する単語は、回答シードの「制度」に関する単語をどれだけ拡張しても得ることは難しい。話題単位での回答シードの拡張は今後の課題といえる。

## 7. おわりに

本論文では、雑談対話の対話シナリオ構築のための、質問に対する回答候補の推定について取り組んだ。質問文と回答シードを与えることで、回答シードのキーワードを拡張し、質問に対する回答となるキーワードを多数生成する技術を提案した。従来の評価実験の結果、質問文の単語と言語モデルを使用することで、回答シードのみを用いた拡張よりも、より多くのユーザの回答を予測できることが明らかとなった。

今後は回答キーワード生成の精度向上とともに、対話シナリオにおいてのシステムの 3 発話目を自動生成する技術について取り組む予定である。

## 参考文献

- [Frey 07] Frey, B. J. and Dueck, D.: Clustering by Passing Messages Between Data Points, *Science*, Vol. 315, pp. 972–976 (2007)
- [He 11] He, Y. and Xin, D.: SEISA: Set Expansion by Iterative Similarity Aggregation, *Proc of the 20th International Conference on World Wide Web(WWW 2011)*, pp. 427–436 (2011)
- [Higashinaka 14] Higashinaka, R., Imamura, K., Meguro, T., Miyazaki, C., Kobayashi, N., Sugiyama, H., Hirano, T., Makino, T., and Matsuo, Y.: Towards an Open Domain Conversational System Fully Based on Natural Language Processing, *Proc of the 25th International Conference on Computational Linguistics (COLING 2014)*, pp. 928–939 (2014)
- [Higashinaka 15] Higashinaka, R., Funakoshi, K., Araki, M., Tsukahara, H., Kobayashi, Y., and Mizukami, M.: Towards Taxonomy of Errors in Chat-oriented Dialogue Systems, *Proc of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL2015)*, pp. 87–95 (2015)
- [Mikolov 13] Mikolov, T., Chen, K., Corrado, G., and Dean, J.: Efficient Estimation of Word Representations in Vector Space, *CoRR* (2013)
- [Ritter 11] Ritter, A., Cherry, C., and Dolan, W. B.: Data-driven Response Generation in Social Media, *Proc of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2011)*, pp. 583–593 (2011)
- [Vinyals 15] Vinyals, O. and Le, Q.: A Neural Conversational Model, *Proc of ICML Deep Learning Workshop 2015* (2015)
- [Wallace 09] Wallace, R. S.: The Anatomy of A.L.I.C.E., *Parsing the Turing Test*, pp. 181–210 (2009)
- [Wang 08] Wang, R. C., Schlaefler, N., Cohen, W. W., and Nyberg, E.: Automatic Set Expansion for List Question Answering, *Proc of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2008)*, pp. 947–954 (2008)
- [Wong 12] Wong, W., Cavedon, L., Thangarajah, J., and Padgham, L.: Mixed-initiative Conversational System Using Question-answer Pairs Mined from the Web, *Proc of the 21st ACM International Conference on Information and Knowledge Management (CIKM 2012)*, pp. 2707–2709 (2012)
- [Zhang 16] Zhang, Z., Sun, , Le, , Han, , and Xianpei, : A Joint Model for Entity Set Expansion and Attribute Extraction from Web Search Queries, *Proc of the Thirtieth AAAI Conference on Artificial Intelligence(AAAI 2016)*, pp. 3101–3107 (2016)
- [目黒 14] 目黒 豊美, 杉山弘晃, 東中竜一郎, 南泰浩: ルールベース発話生成と統計的発話生成の融合に基づく対話システムの構築, 第 28 回人工知能学会全国大会論文集 (JSAI 2014) (2014)