

対話履歴を利用した協調的応答生成

杉山 聡 堂坂 浩二 相川 清明

NTT コミュニケーション科学基礎研究所

1 はじめに

仕事の遂行を目的とした対話は、対話参加者による共同作業と考えられる。効率よく仕事を遂行するためには、それぞれが独自に行動するのではなく、お互いが協調して振る舞うことが大切である。ユーザとシステムが音声で情報を交換することで仕事を遂行する音声対話システムでも、システムがユーザと協調的に振る舞えることが重要である。

協調的な振る舞いの一つに、ユーザの質問に対する直接的な答えに加えて、ユーザにとって有益と思われる情報を付加して応答する協調的応答がある。従来から協調的応答を生成するための方法がいくつか提案されている。ユーザの質問から、質問の背景となっているユーザのゴールとそれを達成するためのプランを認識し、プランの遂行に有用な情報を付加して伝達する [1]。ユーザの持っている信念に間違いを発見した場合、そのことを伝えるだけでなく、それが間違いである証拠となる情報も伝える [4]。ユーザの能力を対話を通じて獲得し、ユーザの能力に応じて説明の仕方を変更する [2]。これらの手法は、ユーザのゴールやプランをプラン認識の手法によって獲得し、タスクの構造の知識や信念の構造の知識を利用して協調的応答を生成するモデルである。

本論文では、対話履歴を用いて協調的応答を生成するモデルを提案する。例えば、レストランの情報を伝達するタスクで、システムの持つデータベースの中からユーザの満足するレストランを特定する過程の対話 (図 1) を例にして説明する。システムは、ユーザの発話 (U2) の意図を U1 の条件の一部の条件の変更と解釈し、「渋谷、5000 円以下、イタリア料理」でデータベースを検索した結果を伝達している (S2)。同様に、U3 の発話の意図を U2 で条件の一部の変更と解釈し、「渋谷、8000 円以下、イタリア料理」の結果を S3 で伝達している。S3 の伝達の時に、U3 のユーザ発話の意図を U1 のときの条件に当てはめ、「新宿、8000 円以下、イタリア料理」の検索結果を付加して伝達している。ユーザが渋谷よりも新宿にあるレストランを望んでいた場合、この付加情報はユーザにとって有益である。

本論文では、form-based な対話管理の手法を用いて、ユーザにとって有益だと考えられる情報を付加した協調的応答を生成するモデルを提案する。2 節で、form-

- U1: 新宿で 5000 円以下のイタリア料理の店を知りたい。
 S1: 該当する店がありません。
 U2: それでは渋谷ではどうですか?
 S2: 渋谷で 5000 円以下のイタリア料理店は
 ありません。
 U3: 8000 円以下ではどうですか?
 S3: 渋谷ですと 4 軒あり、新宿では 5 軒あります。

図 1: レストラン情報伝達タスクでの対話例

based な対話管理の一般的手法と、これまでに提案されている協調的応答生成の方法を説明する。3 節で、対話履歴を利用して協調的応答を生成するモデルを提案する。

2 Form-Based 対話システム

プランニングの手法を使った対話管理の手法は、計算量の多い推論エンジンを必要とすることや、プランライブラリなどの大量の知識をドメインごとに記述する必要があるなどの問題がある。状態遷移で対話を管理する方法では、主導権混合型の対話システムをつくるのが容易でなく、用意していなかった遷移を要求されたときの対処が難しい。計算機への実装やドメインの移行が容易な手法として、form-based な対話管理の手法が提案され、多くの対話システムで採用されている [5, 6, 3]。

2.1 対話管理

from-based な対話管理の方法では、タスクを特徴づける属性/値のペアを保存するスロットを構成要素とするフレームを使って対話を管理する。例えば、レストラン情報伝達タスクでは、図 2 のようなフレームを用意する。システムは、ユーザの発話を理解して、フレームの各スロットの値を書き換える。例えば、図 2 は、図 1 の U1 のユーザ発話を理解した結果のフレームの値である。営業時間と設備のスロットの値は、制限なしを意味する null となっている。

最寄り駅: 新宿
 ジャンル: イタリア料理
 値段: 5000 円以下
 営業時間: null
 設備: null

図 2: レストラン情報伝達タスクのフレームの例

フレームの値に応じてシステムの行動は決まり、データベースの検索や応答を生成する。Chu-Carroll [3] は情報質問 (information-query) システムに適用するための、システムの行為を選択する一般的なアルゴリズムを提案している。このアルゴリズムでは、フレームの値に応じて Provide-Answer, Notify-Failure, Clarify, Confirm, Info-Seek の 5 つの応答の種類を選択する方法を示している。例えば、タスクの達成に必要なスロットが空の場合には Info-Seek を選択し、「何料理がお望みですか?」等の情報要求応答を生成する。必要な情報がすべてそろった時点でデータベースを調べ Provide-Answer を選択し検索結果を伝達する。

対話の過程においては、図 1 の U2 の発話のように、ユーザがすべてのスロットの情報を明示しない場合がある。この場合は、コンテキストを利用して言及されなかったスロットの値を推定する。Chu-Carroll のモデルでは、最後に埋まっていたスロットの値を継承する。Goddeau[5] のモデルでは、最後の値を継承するかしないかの判断にドメイン知識を利用する。例えば、メーカー A の車種 a を探していたユーザが車種を b に変更した場合、b は メーカー B の車であるという知識を利用して、メーカーのスロットの値である A は継承しない。AMICA[6] のモデルでは、スロット間の制約を利用して最後の値を継承するかを決める。例えば、月、日、時間のそれぞれのスロット値が指定されていて、月のスロットの値が変わった場合は、日と時間のスロット値を継承しないようにする。

2.2 協調的応答生成

データベースの検索に十分な数のスロットが埋まった場合、検索を実行して結果を伝達する。ユーザの要求に合う検索結果がなかった場合と多かった場合に、付加情報を付け加えて協調的応答を生成する方法が提案されている。

検索結果がなかった場合は、値の埋まっているスロットの値を緩和してデータベースを検索した結果を代替案として提案する。ユーザにとって有益な代替案を提案するためには、緩和するスロットを適切に選択できなければならない。AMICA のモデルでは、あらかじめ

め定義されたスロットの重要度を利用して緩和するスロットを決める。例えば、航空旅行情報タスクでは、出発地よりも時間や航空会社の方が重要度は低いと定義し、これらのスロットの条件をなくした検索結果を伝達する。Goddeau のモデルでは、あらかじめ定義してあるスロットの重要度を利用する選択方法に加え、継承したスロットの値をまず消去としている。

検索結果が多くあった場合は、スロットの値を要求する応答を生成する。Goddeau のモデルでは、値の埋まっていないスロットをまず尋ね、すべてのスロットが埋まっていた場合は、もっと特定するよう要求する。AMICA では、スロットの重要度を利用して値を要求するスロットを選択する。Chu-Carroll のモデルでは、スロットを必須なスロットと任意なスロットで区別し、必須なスロットは必ず値を要求することになっている。

従来方法では、前のスロットの値の継承/非継承を、対話システムの設計者が定義したドメイン知識に基づいて行なう。例えば、図 3 のように U2 を解釈した後の検索結果が 10 軒あったことを伝えた後に、ユーザから U3 の質問を受けた場合を考える。従来方法では、最寄り駅とジャンルの最後の値を継承して S3 の答えを生成する。しかし、ユーザが協調的に振る舞うことを前提とした場合、U3 の発話を“結果が見つかった S2 の条件を緩和する”意図だと解釈するのではなく、“U1 の条件を緩和する”意図だと解釈することもできる。この場合、「新宿、8000 円以下、イタリア料理」で検索した結果を生成することが考えられる。上記の解釈や図 1 の S3 の発話を生成する方法を次節で説明する。

- U1: 新宿で 5000 円以下のイタリア料理の店を知りたい。
 S1: 該当する店がありません。
 U2: それでは渋谷ではどうですか?
 S2: 渋谷で 5000 円以下のイタリア料理店は 10 軒あります。
 U3: 8000 円以下ではどうですか?
 S3: 渋谷で 8000 円以下のイタリア料理店は 30 軒あります。

図 3: 伝達する情報の価値が低くなる例

3 協調的応答生成

対話履歴を利用して協調的な付加情報を生成するモデルを説明する。まず、ユーザが言及しなかったスロット値の継承/非継承をフレームの値の変化に応じて判断する。このことで、よりユーザの意図に近い操作が可能

となる。次に、検索結果がゼロだった場合に、フレームの値の変化に応じて付加情報を生成する方法を説明する。最後に、1つ以上前のフレームの値を参照することで可能となる付加情報の生成方法を説明する。

3.1 スロット値の継承モデル

ユーザが言及しなかったスロットの値の従来の決め方は、もっとも新しい値を継承する (Chu-Carroll) か、もっとも新しい値を継承するかしないかをドメイン知識を使って判断する (Goddeau, AMICA) 方法である。本モデルでは、ユーザ発話の前のフレームの値での検索結果と、それから発話されたユーザ発話の内容によって、ユーザが言及しなかったスロットの値を決める。例えば、図4を例に考える。この例は、U1の条件の結果によって、その後のユーザの発話の意図の解釈が異なることを示している。従来の方法だと、U1の条件(フレームの値)での検索結果(S1またはS1'で伝達)に関わらず、タスクの定義によって、U2またはU2'を解釈結果した後、ジャンルのスロット値としてイタリア料理を継承するかしないかを決める。U1での検索結果と、その後のユーザ発話の内容からユーザの意図を推定することで、より妥当な判断が可能となる。この例では、U2の意図はU1の条件の緩和であると推定し、ジャンルを継承しない方がユーザにとって有益な情報を与えられる。U2'の場合はユーザの意図は条件の絞り込めであると推定でき、ジャンルを継承した方がよい。

- U1: 新宿で5000円以下のイタリア料理の店を知りたい。
S1: 該当する店がありません。
U2: 新宿で5000円以下の店ではどうですか?

S1': 該当する店が10軒あります。
U2': 新宿で3000円以下ではどうですか?

図4: 前の結果でスロットの継承/非継承が変わる例

本モデルでは、前のフレーム値での検索結果と、その後のフレーム値の変化の種類に応じて、言及されなかったスロット値を決める。フレーム値での検索結果は、ゼロ、1つ、複数の3種類で分類する。フレーム値の変化の分類は以下に行なう。まず、あるフレーム値での検索結果を伝えてからユーザが発話したスロット値と、もとのフレームの該当するスロット値とを比較する。スロット値の変化を、そのスロットが取得する値の集合空間上の値の変化ととらえ、“緩和”、“強化”、“変更”、“同一”の4種類に分類する。例えば、

U2の発話の場合は、最寄り駅と値段の両方のスロットとも“同一”変化であり、U2'の場合は値段のスロット値が“緩和”されていると解釈する。前のフレームから変化したすべてのスロット値の変化の組合わせで、フレーム全体での値の変化を“緩和”、“強化”、“変更”、“同一”の4種類に分類する。

結果とフレーム値の変化の組合わせで、言及されなかったスロット値を決める。基本はもっとも新しい値を継承するが、値を継承しない場合と、もっとも新しい値ではなく別のフレーム値を継承した方がよい場合がある。

値を継承しない フレーム値の変化が“同一”であった場合は、言及されなかったスロット値を継承しない。継承した場合、前回のフレーム値と同じになり、ユーザに新しい情報を伝えることができない。また、前回の検索結果がゼロだった場合、言及されなかったスロット値の条件の緩和と解釈することができる(図4のU2')。

前回の検索結果がゼロだった場合と1つだった場合に“強化”の変化があった場合も言及されなかったスロット値を継承しない。この場合、値を継承して結果を伝達しても、ユーザに新しい情報が伝達できることは見込めない。前回とは別の条件での検索を開始したと解釈した方が、有益な情報を与える可能性が高い。

別のフレームの値を継承する 図3のS2→U3のように、結果が複数あってフレーム値の変化が“緩和”だった場合、前のスロット値を継承してもしなくてもユーザに与えられる情報量は多くない。この場合、ユーザの意図は他にあると解釈し、対話履歴を調べて別の意図を推定する。まず、検索結果がゼロだったフレームを履歴を遡って探す。該当するフレームがあった場合、そのフレームに“緩和”の条件の変化が適用可能かを調べる。条件の変化の適用可能性には以下の2つが満たされることを条件とする。1) 条件の変化に含まれる“同一”のスロット値が、該当フレームでも同じ値であること。2) 条件の変化に含まれる“緩和”のスロットの変化を適用することによって、該当するフレームの条件が緩和されること。この場合、以前検索結果がゼロだったフレームの条件に戻り、そこから別のスロット値を緩和することで、別の候補の検索を開始する意図があったと解釈する。該当するフレームから、条件の変化で言及されなかったスロット値を継承することで、新たなフレームをつくる。

3.2 付加情報生成モデル

データベースの検索結果がゼロだった場合の従来の協調的応答の生成方法では、そのタスクで重要度の低

いスロット値を緩和して再検索した結果を生成する。しかし、どのスロットを重要視するかはユーザによって異なるはずである。本論文では、前の検索結果と条件の変化の情報を利用することで、より適切な緩和するスロットの選択が可能な方法を提案する。緩和するスロットを、前の検索結果とフレームの値の変化の組み合わせごとに表 1 にまとめた。前の結果がゼロだった場合は、その後の発話で条件が緩められたスロットの重要度が低いと判断して優先的に条件を緩和する。前の結果が複数あった場合は、その後の発話で条件が強化されたスロットの値を優先的に緩和する。表にない組み合わせの場合は、従来の方法と同様に、タスクごとに定義した重要度を利用して緩和するスロットを決める。

結果	変化	緩和するスロット
ゼロ	緩和	緩和されたスロット
	変更	1) 緩和されたスロット 2) 値を継承したスロット
複数	強化	強化されたスロット
	変更	1) 緩和されたスロット 2) 強化されたスロット 3) 値を継承したスロット

表 1: 優先して緩和するスロット

この方法で緩和する属性を定めることで、図 5 の S2 のような応答生成が可能となる。この例では、S1 での結果がゼロで、U2 での値の変化は“緩和”である。U2 で言及されなかったスロットの値を継承して「町田～本厚木、3000 円以下、イタリア料理」というフレームを生成して検索を行なう。その結果がゼロであったため、U2 で緩和された最寄り駅のスロットを更に緩和して検索した結果を付加する。

- U1: 町田で 3000 円以下のイタリア料理の店を知りたい。
 S1: 該当する店がありません。
 U2: 小田急の町田から本厚木間ではどうですか?
 S2: 該当する店がありません。
 新宿でしたら 5 軒あります。

図 5: 条件の変化を利用した協調的応答生成の例

3.3 対話履歴を用いた付加情報生成モデル

図 1 の対話例で示したように、対話の途中で検索結果がゼロだったフレームがあった場合 (S1 を生成したときのフレーム)、そのフレームに近い条件の解を提案

すること (S3 で提案) はユーザにとって有益であると考えられる。S3 のような付加情報を生成するために対話履歴を使う。まず、対話履歴を遡り、検索結果がゼロだったフレームを見つける。ユーザの条件の変更が“緩和”か“変更”だった場合、そのうちの“緩和”されたスロットの条件をフレームに適用する。適用したフレームの条件で検索し、結果があった場合にはその結果を伝達する。

4 おわりに

本論文では、form-based な対話管理において、対話履歴を利用して協調的応答を生成する方法を提案した。まず、前のフレームでの検索結果とフレーム値の変化との組み合わせで、ユーザが言及しなかったスロット値の継承/非継承を決める方法を示した。このことで、よりユーザの意図に近いフレームを生成することが可能となった。次に、検索結果がゼロだった場合に、フレーム値の変化に応じた付加情報を生成する方法を示した。ユーザの重要視するスロットを動的に推定することで、より適切な代替案が生成できることを示した。最後に、ユーザの“緩和”の変更要求を、過去に検索結果がゼロだったフレームに当てはめることで、従来の方法では生成できなかった代替案が生成できることを示した。

謝辞

日頃よりご指導いただく NTT コミュニケーション科学基盤研究所 メディア情報研究部 萩田紀博部長、討論して頂いた対話研究グループの皆様感謝いたします。

参考文献

- [1] James Allen. *Natural Language Understanding* (second ed.). Benjamin/Cummings, 1995.
- [2] Alison Cawsey. *Explanation and Interaction*. The MIT Press, 1992.
- [3] Jennifer Chu-Carroll. Form-based reasoning for mixed-initiative dialogue management in information-query systems. In *Proc. Eurospeech 99*, pp. 1519–1522, 1999.
- [4] Jennifer Chu-Carroll and Sandra Carberry. Collaborative response generation in planning dialogues. *Computational Linguistics*, Vol. 24, No. 3, pp. 355–400, 1998.
- [5] David Goddeau, Helen Meng, Joe Polifroni, Stephanie Seneff, and Senis Busayapongchai. A form-based dialogue manager for spoken language applications. In *Proc. ICSLP 96*, pp. 701–704, 1996.
- [6] Roberto Pieraccini, Esther Levin, and Wieland Eckert. AMICA: the AT&T mixed initiative conversational architecture. In *Proc. Eurospeech 97*, pp. 1875–1878, 1997.