

2L-04 ユーザの行動目的に基づいた音声対話システム*

渡邊圭輔

石川泰

三菱電機(株) 情報技術総合研究所

1 はじめに

我々は目的指向の音声対話システムの実用化に向け開発を行っている。自由発話入力の特長を活かすためには、対話の流れを規定して受理可能な文を制限したシステム主導の対話ではなく、基本的にユーザが主導権を持った対話を実現すべきだと考える。しかし、全ての文を受理しようとする、語彙および文法的な制約力が弱くなり認識率が低下する。そこで、認識結果に対して対話文脈上の妥当性を評価してスコアリングを行う手法を提案する。

2 ユーザ主導の対話管理手法

2.1 行動目的に基づいた対話管理

ユーザ主導で対話を進める場合には、ユーザが対話目的を効率的に達成するために、対話状況に応じたシステムからの適切なインタラクションが必要である。そのためには、対話の進行に伴って変化するユーザの対話目的を推定し、システムの動作目的とユーザの対話目的を一致させることが重要となる。

そこで我々は、一発話のみからは決定できないユーザの大局的な目的を行動目的と定義し、行動目的と個々の発話との関係知識に基づいて、対話履歴から行動目的の活性度を計算して行動目的を推定することで、主導権を動的に切り替える対話管理手法 [1] を提案した。行動目的とは、ユーザがシステムとのインタラクションによって達成しようとするゴールであり、例えばホテル予約タスクでは「ホテル予約」「ホテル検索」「日時変更」などである(図1)。主導権交代は行動目的に階層関係を持たせることで実現する。

しかし、このような手法では音声認識部が常にタスクの全発話を対象とせざるを得ないため高い認識率が得られず、誤認識が生じると対話管理部が有効に動作しない。そこで本稿では、行動目的の知識源をより詳細化して、音声認識結果へのトップダウン知識として適用する手法について提案する。

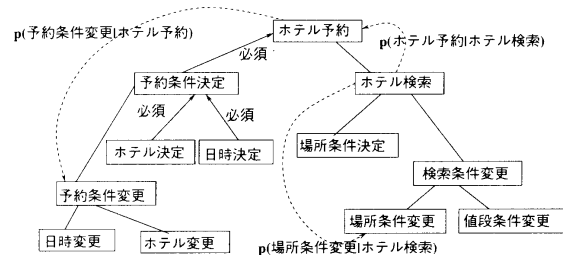


図1: ホテル予約タスクにおける行動目的知識の例

2.2 対話文脈に基づいたスコアリング

図2にホテル予約タスクでの対話例を示す。この対話文脈において、S3に続いて入力された発話の認識結果が図3に示す3つの場合、対話文脈上からは第3位の候補を選択するのが妥当と考えられる。

S1: はい、こちらは横浜市内ホテル予約センターです。
U1: 新横浜駅で12月16日に1泊できるホテル探してるんですが。
S2: その条件でしたら新横浜国際ホテル、フジビューホテル、新横浜プリンスホテルの3件ございますが。
U2: 新横浜国際ホテルはシングル10000円以下ですか。
S3: はい、新横浜国際ホテルのシングルは9300円です。

図2: 対話例

Rank	認識結果
1	じゃあ 四個お願いします
2	じゃあ えー 二個お願いします
3	じゃあ 予約お願いします

図3: N-best 認識結果例

そこで、N-best 認識結果に対して対話文脈上の妥当性を評価し、スコアリングを行うことを考える。そのために、まず行動目的 A から行動目的 A' への遷移確率 $p(A'|A)$ を考える(式(1))。

$$p(A'|A) = p(A'|A, s(A), s(v_1), \dots, s(v_M)) \quad (1)$$

*Spoken dialog system based on behavioral goals of a user. Keisuke WATANABE, Yaushi ISHIKAWA. MITSUBISHI Electric Corporation

ここで、 v はシステムの動作に必要なパラメータを、 $s(v)$ はパラメータ v の状態を表す。例えば、パラメータとしては「ホテル名」「日時」などであり、パラメータの状態は「確定、未確定、初期状態」などを規定する。 $s(A)$ は行動目的 A の状態を表し「達成、失敗、初期状態」などを規定する。遷移確率は行動目的の知識源に予め規定しておく (図 1 の点線で示した弧)。

N-best 認識結果 $I_1, \dots, I_N$ に対しては以下のアルゴリズムでスコアリングを行う。まず、入力時点の文脈における行動目的の推定を行い各行動目的 A に対する確率値 $p(A)$ を得る。次に、 I_i が入力された場合の行動目的 A_{I_i} を推定し確率値 $p(A_{I_i}|I_i)$ を得る。そして、入力 I_i に対するスコアを、音響スコア $S_a(I_i)$ と式 (2) に定める対話文脈スコア $S_c(I_i)$ の和として求める (式 (3))。ここで w は適当な重み係数である。全ての入力に対してスコア $S(I)$ を計算してスコア最大となる I_i を 1 位の認識結果として選択する。

$$S_c(I_i) = -\log\{p(A_{I_i}|I_i) \cdot p(A_{I_i}|A) \cdot p(A)\} \quad (2)$$

$$S(I_i) = S_a(I_i) + w \cdot S_c(I_i) \quad (3)$$

3 評価実験

ホテル予約をタスクとするシステムを用い、1) 対話管理部にあらかじめ正解認識系列を入力して特定の対話文脈を再現した状態での入力、2) 完全に自由な対話の 2 通りの評価実験を行った。音声認識部は不特定話者認識、認識対象語彙数は 1091 単語、構文ネットワークの平均単語分岐数は 6.84 である。式 (3) の第二項である対話文脈スコアは、簡単のため 1, 0.5, 0 の三値を行動目的の知識源に人手で付与した。また今回は、本方式の有効性確認のため対話文脈スコアのみでの性能評価を行った。いずれの場合も、対話管理部へは音声認識結果に対して意味処理を行った後の、意味的に異なる結果のみが渡される。

3.1 対話文脈を規定した評価実験

文脈上行動目的が明確な場合の評価を行うため、文脈を固定した実験を行った。被験者には対話スクリプトを提示し対話文脈を把握させた上で、提示した文脈に続く発話を自由に発声させた。成人男性 5 名による 58 発話 (認識対象語彙・構文で受理可能な発話のみ) を収集した。意味的に異なる認識結果は一発話当り平均 13 種類であった。表 1 に 58 発話に対する認識結果の内訳を示す。

	正解 1 位	正解 2 位以下	誤認識
従来手法	48 (83%)	6 (10%)	4 (7%)
提案手法	50 (86%)	4 (7%)	4 (7%)

表 1: 58 発話に対する認識結果の内訳

	正解 1 位	正解 2 位以下	誤認識
従来手法	254 (81%)	18 (6%)	43 (13%)
提案手法	257 (82%)	15 (5%)	43 (13%)

表 2: 315 発話に対する認識結果の内訳

3.2 自由対話による評価実験

自由な文脈での発話に対する評価のため、対話システムを用いた実験を行った。被験者にはホテル予約を対話目的として自由に対話を行わせた。成人男性 11 名によって収集した発話のうち認識対象語彙・構文で受理可能な 315 発話を評価対象とした。意味的に異なる認識結果は一発話当り平均 6.6 種類である。表 2 に 315 発話に対する認識結果の内訳を示す。

3.3 考察

正解が 2 位以下となった発話の多くは意味理解部の処理誤りであり、対話文脈を規定した実験で 1 発話、自由対話による実験で 8 発話であった。これらは意味理解部の翻訳処理知識の精度向上により改善が可能である。

それ以外の原因としては、意図が同一のために対話文脈スコアが同一になってしまうものが 9 発話、行動目的の推定誤りが 1 発話であった。前者は、対立候補が「一月五日から明日三泊なのですが」、正解が「一月五日から三泊なのですが」のようなものである。これに対しては、発話と行動目的の関係の詳細化、さらに一文内での言語制約の精度向上が必要である。また、行動目的の遷移確率の獲得方法、文脈スコアの重み係数 w の決定方法についても今後検討の必要がある。

4 まとめ

認識結果に対して対話文脈上の妥当性を評価してスコアリングを行う手法を提案した。評価実験により、2 位以下に正解が存在する発話で、意味理解部での処理誤りを除いた 15 発話のうち 5 発話に対して本手法が有効に働くことを確認した。

参考文献

- [1] 渡邊, 永井, 石川: “ユーザの行動目的の推定による協調的な音声対話制御”, 情報処理学会第 54 回全国大会講演論文集, 第 2 巻, pp. 2-233-2-234 (1997).