タスク適応型高効率対話制御法

安田宜仁　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

音声対話システムは音声認識結果に基いて，ユーザの要求内容を決定していく．音声認識には限界があるため，一般にユーザの最初の発話内容からだけでは，ユーザの要求内容を決定することはできない．このため，システムは伝わった内容の確認を行ったり，不足していると思われる情報を要求したりする．これら確認や要求によって発生するシステムとユーザの一連のやりとりは確認対話と呼ばれる．

従来，多くの場合，確認対話の制御のために，人手でルールを作成していた．しかし，効率良い確認対話制御を行うためには，タスク (音声対話システムと人が行う特定の仕事) を変更する度にたくさんのルールが必要となり，タスク変更の手間を増やしていた．

本稿では，タスクに依存したルールなしでも，効率良く確認対話の制御を行う方法を提案する．

2　関連研究

効率良い確認対話戦略を自動的に決定する方法の研究はこれまでにも行われてきた．

まず，強化学習を使って対話戦略を最適化する従来方法がある [1]．しかし，現状では学習のために大量のデータを必要する．さらに，音声認識率が変わった場合などでも再び学習が必要である．

また，音声認識率を考慮し，効率的な対話戦略を選択するという従来方法も提案されている [2, 3]．しかし，これらの方法では，システムが複数のユーザ要求を扱うことは考慮されていなかった．

単一のユーザ要求しか扱わないシステムにおいては，最適な戦略を事前に決定することができる．

しかし，自然な対話を受け付ける対話システムにおいて単一のユーザ要求しか取り扱えないという制約があり，現実的ではない．

本手法は，複数のユーザ要求が受け付け可能なシステムにも適用可能である．

3　タスク適応型高効率対話制御

3.1　概要

通常，音声対話システムは1つのタスクで複数のユーザ要求を受け付けることができる．例えば，スケジュール管理であれば，スケジュールの追加，変更，確認といった複数のユーザ要求は最低でも必要であると考えることができる．

システムの理解状態は属性と値およびその値の確からしさの集合で表わされているとし，この属性のことをスロットと呼ぶ．

本稿で考えるタスク適応型高効率対話制御法は，各時点における理解状態に対するユーザ要求の確率分布と，ユーザ要求推定が正しいと仮定した場合に，ユーザ要求を確定するまでの期待ターン数を利用して，対話終了までの期待ターン数ができるだけ小さくなるようにシステムの行動を決定する方法である．

ユーザ要求を仮定した場合の期待ターン数を得るために，特定のスロット群を確定するための期待ターン数を推定する．ターン数の推定のために，確認の最中での認識率を推定する．この認識率は語彙数に依存する．

確認対話中にはシステムは要求あるいは明示的な確認しか行わないとする．また，システムは確認対象によって認識語彙を絞ることができるとする．さらに，ユーザからの Yes/No 相当の意図は正確に伝える手段があると仮定する．

3.2 ユーザ要求の確率分布

ある時点でのシステムの理解状態を用いて, ユーザ要求の確率分布を推定する方法を考える. 実際にこのような確率分布を得ることは困難なため, 各ユーザ要求と理解状態との関連度を定義し, それを近似的に確率値とする.

スロット s_i の値を v_i と表し, その値の確からしさを c_i とする. この確からしさは, 音声認識器のスコアなどを使うことができる. システムが確認を終えたスロットの確からしさは 1 とする. 対象となっているユーザ要求 G_j において必要なスロットの数を N_{G_j} とする. スロットの値 v_i が値域となりうるユーザ要求の数を, M_{v_i} としたとき, その時点で理解状態 $\mathbf{S}$ とユーザ要求 G_j との関連度 $Rel(\mathbf{S}, G_j)$ を, 以下のように定める.

G_j の値域として認められている値が入っている v_i について,

$$Rel(\mathbf{S}, G_j) = \frac{1}{N_{G_j}} \sum \frac{c_i}{M_{v_i}}$$

3.3 スロット認識率

対話を通じて同一の認識語彙を用意するのではなく, システムは確認対象に語彙を絞るようにすれば, 認識誤りの可能性が減り, より少ないターン数で対話を進行させることができると予想される. 認識語彙を絞った場合に, どの程度誤認識の可能性が減るのかを考える.

確認対象のスロットを決めれば, そのスロットの確認/要求を行う場合に必要な認識語彙は, 対象スロットに入り得る語彙と, 「はい」「いいえ」といった対話の進行に必要な一般的な語彙から構成される. この場合に期待される認識率を「スロット認識率」と呼ぶとする.

スロットに入り得る語彙の間には, 所属が決まれば名前を絞り込めるとか, あるいは名前が決まれば所属名を絞り込めるといった, 意味的依存関係があることがある. そのような場合には, あるスロットの値によって, 別のスロットのスロット認識率は決まることになる. よって, スロット認識率はスロットの種類だけから事前に決まるものではない.

語彙が与えられた場合に, 認識率を近似として, ここでは, 認識誤り率は語彙数の平方根に比例するという経験則 [4] を採用することにする.

3.4 ユーザ要求を仮定した場合の期待ターン数

ユーザ要求推定が正しいと仮定した場合のユーザ要求確定までの期待ターン数を推定する方法を考える.

そのためにまず, スロット認識率が与えられた場合に, 特定のスロット (群) についての確認/要求により, そのスロットの値が確定するまでの期待ターン数を推定する方法を考える.

ユーザの Yes/No の相当意図はシステムに正確に伝わると仮定したので, スロット認識率が r のときに, スロット確定までの期待ターン数を以下のように求めることができる.

確認が終了するまでに必要な期待ターン数 t_{conf}

$$t_{conf} = \sum_{t=1}^{\infty} tr(1-r)^{t-1} = \frac{1}{r}$$

要求が終了するまでに必要な期待ターン数 t_{req}

$$t_{req} = t_{conf} + 1 = 1 + \frac{1}{r}$$

複数のスロットを確認/要求する場合も同様に考えることができる.

次に, ユーザ要求を仮定した場合の期待ターン数を推定する方法を考える.

ある時点でのシステムの理解状態において, 特定のユーザ要求の確定までに必要な行動は, スロットとそのスロットについて必要な行動 (確認, 要求) の対の集合で表すことができる. この対の集合が決まった場合, 最も期待ターンが小さくなる確認の組合せの順序を計算することができる. なぜなら, この集合の任意の分け方について, 任意の順序で並べ替えたものについて期待ターン数を計算することが可能であるからである. この最小の期待ターン数を返す場合をユーザ要求を確定した場合の, ユーザ要求確定までの期待ターン数とする.

3.5 システムの次行動の選択

各ユーザ要求の確率と, ユーザ要求を仮定した場合の期待ターン数を使って, 対話終了までの期待ターン数を小さくするような, システムの次行動を決定する方法を考える.

実際のユーザの要求が G_i である確率を p_{G_i}, ユーザ要求が G_i であると仮定した場合の期待ター

ン数を t_{G_i} と表す. 仮定したユーザ要求が実際のユーザ要求とは異なる場合に, システムがそのことを知ることができるまでのターン数がユーザ要求確定までの期待ターン数と同じであると仮定する. さらに仮定したユーザ要求が異なっていた場合には, 今の状態から確認を繰り返すものとする. この場合, 例えば可能なユーザ要求が 2 つのシステムで, G_1, G_2 の順に対話をすすめていった場合の対話終了までの期待ターン数は, p_{G_1} の確率で t_{G_1} ターンかかり, p_{G_2} の確率で t_{G_1} ターンかけて失敗し, t_{G_2} ターンかけて実際のユーザ要求を確定することになる. よって, この場合の期待ターン数は $p_{G_1}t_{G_1}+p_{G_2}(t_{G_1}+t_{G_2})$ と考えることができ, 逆に G_2, G_1 の順に対話をすすめていった場合の対話終了までの期待ターン数は, $p_{G_2}t_{G_2}+p_{G_1}(t_{G_1}+t_{G_2})$ であると考えることができる.

一般にシステムが, 複数のユーザ要求を受け付けることができる場合でも,
$p_{G_{a(1)}}t_{G_{a(1)}} + p_{G_{a(2)}}(t_{G_{a(1)}} + t_{G_{a(2)}}) + \ldots p_{G_{a(n)}}(t_{G_{a(1)}} + \ldots + t_{G_{a(n)}})$
がもっとも小さくなるようなユーザ要求の選択順 $a(1), a(2), \ldots, a(n)$ を選択する.

この方法によって, 確率の高いユーザ要求に対応する対話を先に行うよりは, 要求の推定が不確実な場合でも, 期待ターン数が小さくなるような制御を行うことが可能になると考える.

4 評価

計算機上で実装された模擬ユーザとシステムとのシミュレーション対話による評価実験を行なった. 模擬ユーザとの対話による評価は, 新たな方式を短期間で評価し, 再調整することが可能であるだけでなく, 評価の基準を統一することができるという特徴がある [5].

模擬ユーザとシステムは意図のレベルでのやりとりを行い, 構文解析等は行っていない. システムから模擬ユーザへの意図は正確に伝わるとした. 他方, 模擬ユーザからシステムへは音声認識の誤りを模するため, 3.3 で行った推定同様, 認識語彙に依存して, 意図単位で認識誤り (置換, 脱落のみ) が起こるとした.

基本となる認識率を変更しながら, 各 500 対話ずつシミュレーションを行なった.

人手によって調整した例との比較とのため, 当研究所によって開発された音声対話システム “飛遊夢 (ひゅーむ)”[6] の確認対話戦略を模したものとの比較を行なった.

飛遊夢の確認戦略は概ね以下のようになる:

- ユーザ要求を特定できるまでは予め決められた順序に情報を要求する
- 意味的依存関係があるスロット間で, 依存関係に反する情報が入った場合には, 候補が多い方のスロットを消す
- 意味的依存関係があるスロットのどこかの値が埋まっていて, 依存関係を使って一意に他のスロットも決まる場合はその値を埋める
- 確認は一括して行う

ただし, 飛遊夢では 2 種類のコストを用いる方法 [7] を用いることにより, 最終的に確認する量を減らす場合がある. シミュレーションではこの機能は実装していない. また割込みは受け付けない.

対話実験に使用した模擬ユーザプログラムは以下の能力を備える:

- 要求を決定する. システムに要求が伝わるまでは要求を変えない
- システムに自分の要求, あるいは要求の一部を伝える
- 確認に対しては, Yes/No を必ず伝える
- No の場合にはランダムに訂正発話を行う
- システムの情報要求に対しては要求された範囲内で伝える
- 確認対象が自分の意図の範囲になければその旨を伝えて再度自分の要求を伝える

4.1 タスクの仕様

意味的制約が比較的強く現れる例として, 会社の受付のような組織名と名前 (姓) を使うようなタスクがある. 架空のタスクを作成し, 2 つのスロットは人の姓と部署の対応のような構成を作った. 人は 3000 人, 姓と部署はそれぞれ 1000 種類, 300 種類あるとし, 姓と部署から人が一意に決定されるとした. 意味の依存関係を入れるために, ある部署に含まれている姓の種類の数は正規乱数で決定し, 特定の部署には同じ姓をもつ人はいないと仮定した. その他 4 つのスロットを用意し, それらには意味の依存関係は定めていない.

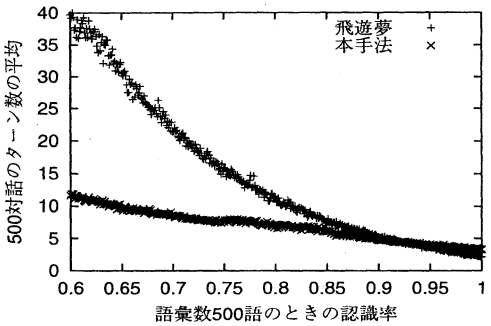

図 1: 認識率に対する平均ターン数

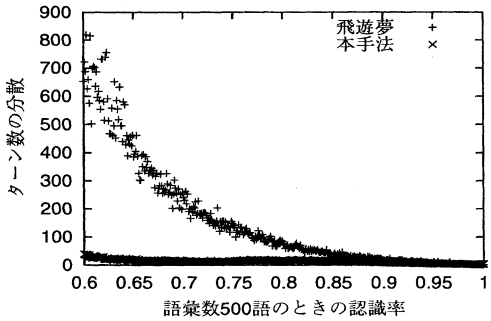

図 2: 認識率に対するターン数の分散

5 評価結果

図 1 は, 語彙数 500 語のときの認識率を 0.6 から 0.999 まで変化させた場合の, 500 対話での平均ターン数である. この結果から本手法は, タスクに依存したルールを記述することなく, 短いターン数で対話終了に達していることがわかる. また, 図 2 に同じく 500 対話での分散を示す. この結果より本手法は, 極端に長い対話になることは少ないことがわかる.

6 おわりに

タスク適応型高効率対話制御法を開発した. この方法は, 複数のユーザ要求が存在するようなタスクにも適用することができる. 語彙数から推定された, 確認時の認識率を用いて推定される, 各ユーザ要求までの確認終了までの期待ターン数と, 理解状態に対するユーザ要求の確率分布を利用して, 対話終了までのやりとりの回数を小さくするような確認手順を選択することができる. 加えて, タスク変更時にもタスクに依存したルールを記述する必要がないので, 音声対話システムのタスク移行を容易に行うことを可能にする.

模擬ユーザとの対話シミュレーションでは, さまざまな認識率を仮定した場合でも, 広い範囲で人手でルールを書いたものより短いターン数で動作することを示した.

謝辞 日頃よりご指導いただく, 当研究所メディア情報研究部 萩田紀博部長, 有益な示唆をいただくマルチモーダル対話研究グループの諸氏に感謝いたします.

参考文献

[1] Diane J. Litman, , Michael S. Kearns, and Marilyn A. Walker, : Automatic Optimization of Dialogue Management, in *COLING* (2000).

[2] Yasuhisa Niimi, and Takuya Nishimoto, : Mathematical Analysis of Dialogue control strategies, in *EUROSPEECH*, Vol. 3, pp. 1403–1406 (1999).

[3] 井本貴之, 相川清明：平均対話回数を用いた対話設計方法, 日本音響学会講演論文集, pp. 165–166 (1997).

[4] 中川聖一, 伊田政樹：連続音声認識のタスクの複雑さを表す新しい尺度, Vol. J81-DII, No. 7, pp. 1491–1500 (1998).

[5] Eckert, W., Levin, E. and Pieraccini, R.: Automatic evaluation of spoken dialogue systems, in *TWLT13: Formal semantics and pragmatics of dialogue* (1998).

[6] 音声対話システム 飛遊夢 (ひゅーむ), in *http://www.brl.ntt.co.jp/cs/dug/hyumu/*.

[7] 堂坂浩二, 安田宜仁, 宮崎昇, 中野幹生, 相川清明：システム知識制限下における効率的対話制御, 情報処理学会 SLP 研究会資料, No. 33, pp. 49–54 (2000).