音声対話システムにおける話題の構造を用いた効率的な対話管理

神田 直之† 駒谷 和範‡ 尾形 哲也‡ 奥乃 博‡
†京都大学工学部情報学科 ‡京都大学大学院情報学研究科知能情報学専攻

1. はじめに

音声対話システムを実用化するには、音声認識誤りへの対処が不可欠である。音声認識誤りがあった場合、必要な部分に対してのみシステムは確認を行うべきである。従来研究では、確認を行うべき部分の同定には一発話の音声認識スコアなどから計算する信頼度を使用している[1) 2)]。しかし、一発話より大きな単位から音声認識誤りを判断していないので、文脈的に整合性のない音声認識結果に適切に応答することができない。本稿では、音声認識結果が文脈的に信頼できるかを表す尺度として、文脈的信頼度を導入する。

本稿ではデータベース検索音声対話における話題の動的な構造と静的な構造を用いて、文脈的信頼度を定義する。静的な構造による文脈的制約は、データベース検索音声対話における対話を、「検索条件の指定」「情報の提示要求」の 2 モードとした時のモード間の遷移から求める。また、動的な構造による文脈的制約は、入力された情報の履歴を表す木構造を利用して求める。この話題の動的な構造を利用することで、既に入力されたスロット・値ペアをどこまで元に戻すかというロールバックの範囲を決定することもできる。

2. データベース検索タスクにおける対話のモデル化

2.1 静的構造

本研究ではレストラン検索をタスクドメインとしたシステムを対象とする。データベース検索タスクにおける典型的な対話として、まず希望する検索条件を追加・削除することにより数件まで店を絞り込み、その後絞り込んだ店の具体的な情報に関して質問を行うという手順を想定する。ここでの前者を「検索条件の指定」モード、後者を「情報の提示要求」モードとする。本研究ではこの 2 モード間の遷移を、静的な構造による文脈的制約とする（図 1）。

各発話は 2 つのモードのいずれかに属し、表 1 に示すスロットに対して値の代入を行うものとみなせる。各モードで現れる表現や語彙は異なるため、これを利用して発話がどのモードであるかを決定する。まず、各モード s ごとにあらかじめ用意しておいた想定質問文と音声認識結果とのマッチングの尤度 $M(s)$ を計算する[3)]。次に音声認識結果からスロットに入る値 w_i を抽出し、それぞれに対して単語レベルの信頼度 $CM_w(w_i)$ を計算する[2)]。さらに、各モードごとに現発話がモード s であるとした場合の文脈的信頼度 $CM_c(s)$（次章で説明する）を求める。このとき、下式によって現在のモード S を求める。

$$S = \arg\max_s \{(2M(s) + \sum_{w_i \in s} CM_w(w_i)) \times CM_c(s)\}$$

ここで、$2M(s) + \sum_{w_i \in s} CM_w(w_i)$ は現発話から得られる尤度を、$CM_c(s)$ は文脈的な尤度を表す。

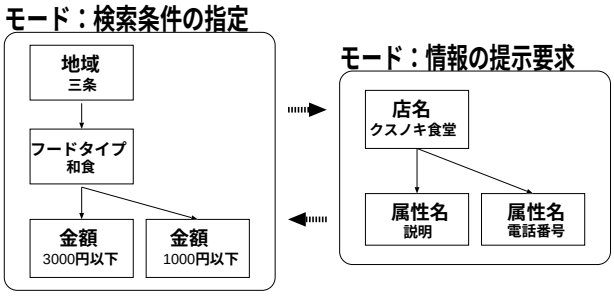

図 1 データベース検索対話における静的な構造と動的な構造

表 1 各モードにおけるスロットと値の対応

モード	スロット	値
検索条件の指定	地域、値段...	三条、3000 円...
情報の提示要求	店名	クスノキ食堂、吉田屋...
	属性名	説明、予算...

2.2 動的構造

話題の動的な構造として、対話の履歴を表す木構造を定義する。木の各ノードはスロットと値の組で構成されており、2.1 節で定義した 2 モードそれぞれに対して木を保持する（図 1）。発話によりスロットに値が追加されればノードが追加され、値が削除されるとノードも削除される。対話のある時点で話題となっているスロットと値の組の集合は、木のルートから最も右の子を辿っていくことで得られる（以下これらのノードを参照ノードと呼ぶ）。

木構造は具体的には以下の規則により作成される。

- 新たに追加するノードと同じスロットを持つノードが参照ノード中になければ、参照ノードの子の位置にノードを追加する
- 同じスロットを持つノードが参照ノード中にあれば、そのノードの弟の位置にノードを追加する
- 削除したいノードの下に部分木が存在しない場合、そのノードを削除する。
- 削除したいノードの下に部分木が存在する場合、削除するノードの弟の位置にその部分木を移動する

2.3 動的構造を用いたロールバックの範囲の決定

本研究では、ロールバックするポイントを、入力された発話と同一のスロットを持つ参照ノード中のノードとする。ロールバックする範囲が決定できることで、ロールバックするポイントの子孫ノードを利用して、ユーザに提案をおこなうなどの対話戦略が実現できる。

Efficient Dialogue Management in Spoken Dialogue Systems Using Contextual Structures
by Naoyuki Kanda, Kazunori Komatani, Tetsuya Ogata and Hiroshi G. Okuno (Kyoto Univ.)

表 2 文脈的信頼度の算出

	次状態でのモード	
現在のモード	検索条件の指定	情報の提示要求
検索条件の指定	$E_{dy}(q)$	$E_{st1}\cdot E_{st2}$
情報の提示要求	$E_{st3}\cdot E_{dy}(q)$	$E_{st1}\cdot E_{dy}(i)$

3. 文脈的信頼度の定義

文脈的信頼度 CM_cは表 2 のように定義し、現状態から次状態の尤度を算出する。表中の各評価値について以下で説明する。なお、現在の検索条件に該当する店名の集合を G、情報の提示要求モードの動的木に存在する店名の集合を M とする。$|G|$ は集合 G の要素数を表す。また現発話中に含まれる店名を a とする。ここで店名は関係データベースでのキー属性の値にあたる。

3.1 静的構造から得られる評価値

- $E_{st1} = \begin{cases} 1 & (a \in G \ or \ a \text{が存在しない}) \\ 0.1 & (a \notin G) \end{cases}$
  「情報の提示要求」モードにおいて、条件検索で絞り込んでいない候補に関する発話は文脈的整合性が低い。
- $E_{st2} = \begin{cases} 1 & (a \text{が存在する}) \\ \frac{1}{|G|} & (a \text{が存在しない}) \end{cases}$
  店名が絞りこまれていないまま「情報の提示要求」モードに遷移するのは文脈的整合性が低い。
- $E_{st3} = \frac{|M|}{|G|}$
  「情報の提示要求」モードにおいて絞り込んだ候補に言及せずに、「検索条件の指定」モードに戻るのは、文脈的整合性が低い。

3.2 動的構造から得られる評価値

- $E_{dy} = \min(E_{dy1}, E_{dy2})$
  表 2 中の $E_{dy}(q)$ は「検索条件の指定」側の動的木による評価値、$E_{dy}(i)$ は「情報の提示要求」側の動的木による評価値である。また、以下で r は追加されるノードと同一スロットを持つ、参照ノード中のノードをさす。
  - $E_{dy1} = \frac{depth(r)}{depth(n)}$
    n は参照ノード中で最も深い位置にあるノード、$depth(r)$ は r の深さである。木構造の中で上位にあるノードが入れ替わった時ほど文脈が大きく変わった可能性が高い。
  - $E_{dy2} = e^{-\alpha \cdot descend(r)}$
    $descend(r)$ は r の子孫の数である。長い間更新されていない（子ノードが多い）ノードが入れ替わった時ほど文脈が大きく変わった可能性が高い。今回 $\alpha = 0.25$ とした。

3.3 文脈的信頼度を用いた対話管理

得られた発話の文脈的信頼度が閾値以上ならば受理し、閾値未満ならば確認を行う。今回閾値は 0.4 とした。

4. 動 作 例

図 2 に実際に作成したシステムの動作例を示す。
まず、U3 ではユーザは実際には「3000 円以下で」と発話したが音声認識の結果「ランゼン以下で」と誤認識されている。ランゼンはカレーの店で【三条、和食】という条件にあてはまらないので評価値 E_{st1}が 0.1 となる。その結果文脈的信頼度が 0.1 としきい値以下になり、システムはユーザに確認を求める（S4）。また、U6 ではユーザの「えっと」という発話を「別当」という地名に誤認識した。同一スロットの検索条件である【三条】は木の上位にあるため、評価値 $E_{dy}(q)$ が小さくなる。その結果文脈的信頼度は 0.33 となり、システムは確認を行う（S7）。逆に、U8 では評価値 $E_{dy}(q)$ が高いため確認は行われない。このように、対話の流れに応じて確認の生成を適切に制御している。

S12 はロールバックが行われる例である。この対話例では、スロットの状態を元に戻したうえで確認を行うという戦略をとっている。このことにより条件【和食、1000 円以下】を削除することなく U13 のように効率よく対話が進行する。

S1:**こちらはレストラン案内システムです。どのような店をお探しですか?**
U1:三条にある店を探しているのですが [1.0]
S2:**三条という条件ですと259件あります。他に条件はありますか?**
U2:和食の店で何かないですか？ [1.0]
S3:**三条、和食という条件ですと51件あります。他に条件はありますか?**
U3:ランゼン以下で（３０００円以下で） [0.1]
S4:**ランゼンというカレーの店についてご質問ですか?**
U4:いいえ [1.0]
S5:**もう一度おっしゃってください。**
U5:３０００円以下の店は？ [1.0]
S6:**三条、和食、3000円以下という条件ですと15件あります。他に条件はありますか?**
U6:別当（えっと） [0.33]
S7:**三条を別当という条件に置き換えますか?**
U7:いいえ [1.0]
S8:**もう一度おっしゃってください。**
U8:３０００円以下だったらどうですか？ [1.0]
S9:**三条、和食、1000円以下という条件ですと2件見つかりました。**
吉田屋、クスノキ食堂です。
U9:クスノキ食堂はどんな店ですか？ [1.0]
S10:**クスノキ食堂は〜〜な店です。他にご質問はおありですか?**
U10:祇園でも店を探しているのですが。 [0.17]
S11:**三条を祇園という条件におきかえますか?**
U11:はい [1.0]
S12:**祇園という条件ですと159件あります。**
和食、1000円以下という条件はそのままでよろしいですか?
条件を変更することもできます。
U12:居酒屋を探してます [1.0]
S13:**祇園、居酒屋という条件ですと50件あります。他に条件はありますか?**

(注)ここで示しているのはユーザ発話の音声認識結果であり、 ( )内が実際のユーザの発話である。ただし、音声認識結果が正しい場合は省略している。
(注2)[ ]内の数字はユーザ発話の文脈的信頼度である。

図 2 対話例

5. おわりに

本稿では、データベース検索をタスクとする音声対話システムにおいて音声認識結果に対して文脈的な信頼度を付与し、効率的に確認を行う手法を提案した。今後、各評価値間の重み付けを含めた、文脈的信頼度の算出法についてさらに検討したうえで、評価実験を行う予定である。なお本研究の一部は、科研費、21 世紀 COE の支援を受けた。

参 考 文 献

1) 駒谷和範, 河原達也: “音声認識結果の信頼度を用いた効率的な確認・誘導を行う対話管理”, 情報処理学会論文誌, Vol.43, No.10, pp.3078-3086(2002)
2) Bouwman, G., Sturm, J. and Boves, L.: Incorporating Confidence Measures in the Dutch Train Timetable Information System Developed in the ARISE Project, Proc. ICASSP(1999)
3) 駒谷和範, 河原達也, 清田陽司, 黒橋禎夫, Pascale Fung, 柔軟な言語モデルとマッチングを用いた音声によるレストラン検索システム, 信学技報, SP2001-113(2001)