

データベース検索タスクの文脈的制約を用いた 音声対話システムの実験的評価

神田 直之 駒谷 和範 尾形 哲也 奥乃 博

京都大学大学院 情報学研究科 知能情報学専攻

naoyuki@kuis.kyoto-u.ac.jp

データベース検索タスクにおける音声対話システムにおいて、音声認識誤りの棄却や意味曖昧性の解消のために文脈的制約を取り入れる手法について述べる。まず、データベース検索タスクの対話は「検索条件の指定」から「情報の提示要求」へと移行するとモデル化する。さらに、検索条件をその入力の順序に従って木構造状に管理する。言語理解部ではこれらのモデルから得られる特徴を決定木学習により文脈的制約として取り入れる。提案手法をレストランデータベース検索システムとして実装し、20名の被験者による評価実験を行った。実験の結果、提案手法に基づく文脈的特徴を加えることで、13.4%の意味理解誤り削減が認められた。

Experimental Evaluation of Spoken Dialogue System using Contextual Constraint in Database Retrieval Task

NAOYUKI KANDA, KAZUNORI KOMATANI, TETSUYA OGATA and HIROSHI G. OKUNO

Dept. of Intelligence Science and Technology, Graduate School of Infomatics, Kyoto University

We describe how to introduce contextual information in spoken dialogue systems for database retrieval task. In this paper, we model dialogues in the database retrieval task as consisting of two modes: “specifying retrieval conditions” and “requesting detailed information about specific entries”. Furthermore, we manage retrieval conditions as a tree structure. Based on those models, we introduce decision tree learning using features reflecting the situations in the task as well as those derived from current utterances. By using the output of the decision tree, the system can appropriately select words from a speech recognition result even when it contains some errors. The experimental result showed that our method could identify users’ intentions 13.4% better than that without the contextual information.

1. はじめに

近年、音声対話システムの研究が盛んに行われ、実用的なシステム構築のためにさまざまな課題が取り扱われている。なかでも、音声認識誤りや言語理解誤りによる誤動作への対処は音声対話システムの構築には不可避であり、発話中の音声認識誤りを棄却したり、意味曖昧性の解消を行う研究が行われている。

音声認識誤りの棄却を目指した研究としては、音声認識結果の事後確率から内容語ごとの信頼度を算出するもの¹⁾や、音響尤度と言語尤度を組み合わせたもの²⁾、GMMによって非音声の入力による不必要な音声認識結果を棄却するもの³⁾などがある。これらの研究では、

受理/棄却の判定をその発話から得られる情報のみから行っており、対話レベルの情報を用いていない。これに対し本研究では、対話文脈レベルの情報も併せて内容語の取捨選択を行うことを考える。

対話文脈レベルの情報をを用いた研究として、意味理解曖昧性の解消のために対話行為の接続確率をコーパスから学習して利用するもの⁴⁾⁵⁾がある。これらの研究では対話行為を「SET-START-TIME(会議室の使用開始時間の設定)」や「駐車場の検索の依頼」のように、ドメイン固有のレベルで設計している。対話行為を詳細に設計するほど、詳細な制約を得ることが可能となるが、対話行為の種類が多くなるため必要な学習データ数は多くなり、また汎用性も低くなる。我々はデー

店名(キー属性) : クスノキ食堂
 タイプ : 和食
 説明 : 学生の町, 京都らしく安くて...
 住所 : 京都市左京区吉田...
 電話 : 555-5555
 営業時間 : 18:30-23:30
 休業日 : 木
 交通 : 京阪出町柳駅より徒歩 10 分
 カード : JCB, VISA, アメックス
 駐車場 : 2 台
 下限予算 : 400 円
 上限予算 : 1000 円

図 1 レストランデータベースの例

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。
 S2: 三条という条件ですと 259 件あります。
 U2: 和食の店で何かないですか?
 S3: 三条, 和食という条件ですと 51 件あります。
 U3: 3000 円以下
 S4: 三条, 和食, 3000 円という条件ですと 15 件あります。
 U4: 1000 円以下だったら?
 S5: 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。
 U5: クスノキ食堂の住所を教えてください。
 S6: クスノキ食堂の住所は京都市左京区吉田... です。

図 2 想定する対話の例

データベース検索タスクの概略をモデル化することで, このタスク一般で成り立つ文脈的制約を実現する。6) では人手で記述したルールを用いて, 文脈を反映させた言語理解を行っている。本研究ではこのルールを包括する, より多くの特徴を定義する。さらに決定木学習を行うことで, 多数のルール間の関係を自動的に学習する。7) や 8) では, 音響的な特徴や対話における特徴を用いて機械学習を行い, 発話の受理/棄却を判定している。我々は, より多様な対話的な制約を得ることを目指して「対話の進行モデル」「履歴の構造モデル」を導入する。対話の進行モデルは, 対話が「検索条件の指定」と「情報の提示要求」から成るとモデル化したものである。履歴の構造モデルでは, 検索条件をその入力順序により木構造状に管理する。これらから得られる制約を決定木学習により統合し, 音声認識誤りの棄却や言語理解精度の向上を目指す。

2. データベース検索タスクのモデル

本稿では, 図 1 のような関係データベース に対する音声での検索を扱い, 以下ではレストランデータベースを対象ドメインと設定して議論を進める。データベース検索タスクではユーザごとに目的や必要となる項目が異なり, また検索結果に応じてユーザの目的が変化し, いわゆる「相談型」の対話⁹⁾が行われる。この場合, システムが対話の流れを一意に規定するシステム主導の対話ではユーザの意図に適切に対応できない。したがって, 場面ごとに言語モデルを切り替えて音声認識誤りを減少させる手法は用いず, 言語理解部で湧き出し誤りを除去することを考える。

2.1 対話の進行モデル

本研究ではデータベース検索タスクにおける典型的な対話の進行モデルとして, まず希望する検索条件を指定して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ことで, データベ

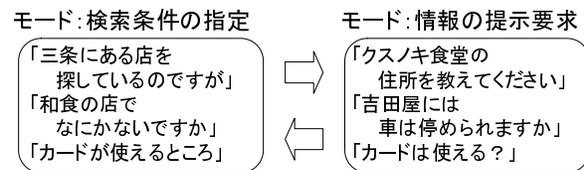


図 3 データベース検索タスクでの 2 つのモード

スからユーザが必要とする情報を取得するものを想定する。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。なお, 店の名前, 店についての情報はそれぞれ関係データベースにおけるキー属性, キーでない属性に相当する。図 2 では U1~U4 が「検索条件の指定」モードにおける発話であり, U5 が「情報の提示要求」モードにおける発話である。このように, 本研究ではデータベース検索タスクにおける各発話は 2 つのモードのいずれかに属するとし, 対話はその間を遷移することで行われるとする(図 3)。

2.2 履歴の構造モデル

対話の履歴として検索条件(属性・値ペア)を木構造として管理する(履歴の構造モデル)。検索条件は最近に入力されたものほど木の低位にくるように保持する。また同一属性の検索条件が追加された場合, 対象ノードの弟の位置に新たなノードを追加する。これにより木構造の上位にある条件の変更は対話の大きな転換を表すことになり, その大きさを文脈的制約として言語理解に取り入れることができる。さらに, ユーザがすでに入力済みの条件を上書きする場合には, ユーザの意図に変化が生じたときのみ, その変化の影響を受けないポイントに戻って対話を再開することもできる¹⁰⁾。

図 4 の (a) は【地名: 三条】【タイプ: 和食】【上限予算: 3000 円】【上限予算: 1000 円】の順に入力が行われたときの履歴構造を表しており, 現在の検索条件は木の最も右側の子を順に辿ることで得られる。図 4 の (a) の状態で【タイプ: 洋食】という入力がある

各エンタリは, 属性名とその値の組で表され, 各データは 1 つのキー属性 (その値が全データで一意) を持つ。

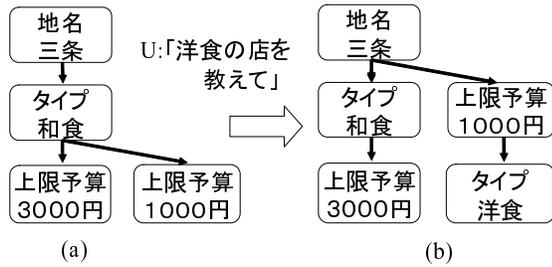


図4 対話履歴の木構造上の管理

と【タイプ：和食】の弟ノードの位置にノード【タイプ：洋食】が生成される．この際，最近に入力されたものほど木の低位にするという制約により，【上限予算：1000円】は【タイプ：洋食】よりも上位に移動される．この結果履歴の構造は図4の(b)のようになる．

3. 文脈的制約を用いた言語理解

提案するシステムの処理の全体像を図5に示す．システムは大きく分けて音声認識部，言語理解部，対話管理部からなる．まずユーザ発話は音声認識部によって単語列に変換される．言語理解部では音声認識部の出力と文脈的な制約とから，ユーザの意図(対話行為と内容語)を推定する．具体的には，まず音声認識結果と想定質問文とのマッチングで対話行為推定を行う．さらに各内容語に対して決定木を用いて対話行為の信頼度と棄却すべき信頼度を求め，最後にそれらを統合して対話行為の決定や内容語の取捨選択を行う．対話管理部では，言語理解部の出力を用いて対話の状態を変更し，それに基づいてデータベースにアクセスし応答を行う．また，履歴の構造を参照し，ユーザの意図が大きく変化したと思われるポイントでは，履歴の構造モデルを用いて検索条件の取捨選択を適切に行うための確認を行う¹⁰⁾．

言語理解部では，7種類の対話行為を取り扱う．本研究ではデータベース検索での対話の状態を「検索条件の指定」と「情報の提示要求」の2モードに大別しており，これら2つをまず対話行為として設定する．これに，実際に対話を行ううえで必要となる「条件の削除」「肯定回答」「否定回答」「条件一括削除」「やり直し」の5つを加えた．また，内容語は関係データベース中の属性名と値とする．

以下では言語理解部での処理を順に述べる．

3.1 マッチングによる対話行為の推定

各対話行為ごとに複数用意した想定質問文と音声認識結果のマッチング¹¹⁾を行い，その類似度をマッチングの尤度とする．これは，音声認識結果が文としてどれだけその対話行為らしいかを表す．想定質問文には「TYPEがおいしい店を教えてください」といったも

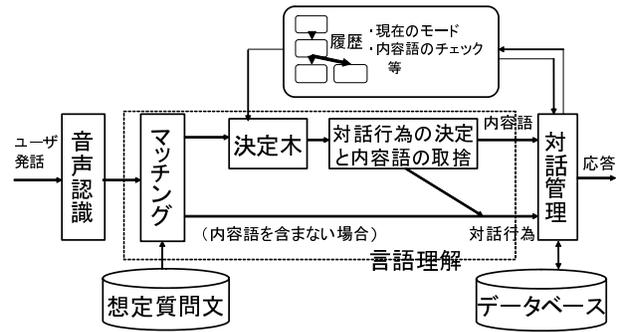


図5 データベース検索システムの処理の概略

のを590文用意した．なお想定質問文中のTYPEはタイプ属性の値に対応している．類似度の点数計算の際には内容語 w_i に対する重みとして，(構文上の重み $weight_i$) \times (音声認識器が出力する信頼度 cm_i ¹²⁾) を用いる．

内容語を含まない「肯定回答」「否定回答」「条件一括削除」「やり直し」である尤度が最も高かった場合，それを発話の対話行為として決定する．

内容語を抽出する必要がある「検索条件の指定」「情報の提示要求」「条件の削除」の尤度が最も高かった場合，さらに以下の処理を行って当該発話の対話行為を決定する．

3.2 対話行為の決定と内容語の取捨

内容語を含む対話行為を持つと判定された発話に対しては，決定木により対話行為の決定と内容語の取捨選択を行う．このために，各内容語ごとに，各対話行為の信頼度に加えて，その内容語を棄却すべき信頼度を同時に出力する決定木を作成する．この決定木の構成法については次節で述べる．言語理解処理では，まず発話中の各内容語に対する決定木の出力を統合して発話としての対話行為を推定する．その後，推定した対話行為と当該内容語を棄却すべき信頼度に基づき各内容語の取捨を決定する．具体的には以下の手順に従う．

- (1) 各内容語 w_j に対して決定木によって対話行為の信頼度を得る．それを対話行為ごとに全ての内容語に関して足し合わせ，その総和が最も大きくなる対話行為を発話 R_i の対話行為 S_i とする．具体的には次式で与える．

$$S_i = \arg \max_s \sum_j CF(s|F_i, w_j)$$

ただし $CF(s|F_i, w_j)$ は，ある対話状態 F_i における内容語 w_j に対する対話行為 s の信頼度で

CF は，学習サンプル中での正解率を平滑化した値であり， w_i が決定木のある葉に分類されたときに，その葉のもつ要素数を N ，そのうち s であるものの要素数を M として $\frac{M+1}{N+P}$ で与えられる (P は対話行為の数とする)．また， CF は学習データ中でその葉に出現した対話行為全てに与えられる¹³⁾．

S1: マッチングの尤度が最も高い対話行為
 S2: マッチングの尤度が最も高い対話行為の尤度
 S3: マッチングの尤度が 2 番目である対話行為の尤度
 S4: (2 の値) / (3 の値)
 S5: 内容語の種類 (属性, 値, キー属性, キー属性値)
 S6: 単語信頼度
 S7: 対になる内容語の存在の有無 (「上限予算」と「1000 円」など)

図 6 一発話のみから得られる特徴 (7 個)

あり, ここでは $s \in \{$ “ 検索条件の追加 ”, “ 情報の提示要求 ”, “ 条件の削除 ” $\}$ である.

- (2) 各内容語 w_j の棄却すべき信頼度 $R_j = CF($ “ 棄却 ” $| F_i, w_j)$ とし, 以下により内容語の取捨選択を行う.

- $CF(S_i | F_i, w_j) \geq R_j$ ならば w_j を受理
- それ以外の場合, 棄却

ただし, 当該内容語の棄却すべき信頼度が低く ($R_j < \alpha$), かつ $CF(S_i | F_i, w_j) \neq 0$ の場合は確認する. 実験時 $\alpha = 0.9$ とした.

3.3 文脈的特徴を用いた決定木の構成

各内容語に対して対話行為の信頼度と棄却すべき信頼度を同時に出力する決定木を構成する. このために, 学習時には正解として対話データ中の各内容語に対話行為を付与しておき, また音声認識誤りである内容語に対しては「棄却」というラベルを与えておく. これらから決定木学習を行うことによって, 出力として対話行為と棄却の信頼度を得ることができる.

判別に用いる特徴には一発話のみから得られる特徴 (図 6) に加えて, データベース検索タスクのモデルに基づく文脈的特徴 (図 7) を新たに定義し, 使用する. 1 発話のみから得られる特徴としては, マッチングの結果 (S1 ~ S4) や単語信頼度¹⁾(S6) などを用いる. なお, マッチングの尤度が最も高い対話行為が複数存在した場合は「曖昧」というラベルを付与して学習を行う. また, 対話の進行モデルに基づく特徴として, 現在の発話が「検索条件の指定」「情報の提示要求」のどちらのモードで行われているか (C1) や, 現在の検索条件で絞りこまれたキー属性の数 (C2) などを設定する. ここで, C10 は情報の提示要求モードに入ってからチェックされたキー属性の割合, C11 は情報の提示要求モードで既にチェックされたキー属性の割合を表す. さらに, 履歴の構造モデルから変更対象ノードの深さ (C12), 子ノードの数 (C14) などを設定する. なお, C22, C23, C24, C25 は変更対象ノードの深さを現在の木の深さとの差および比で正規化したものである.

C1: 現在のモード (初期状態: 検索条件の指定)
 C2: 現在の検索条件に合致するキー属性の数
 C3: 「情報の提示要求」モードに入ってから発話された店のうち, 現在の検索条件に合致するキー属性の数
 C4: これまでに発話されたキー属性の数
 C5: 現在の検索条件に合致するキー属性で, 発話されたことのある数
 C6: 現在の木の深さ
 C7: ユーザの前発話での対話行為
 C8: 直前のシステム発話が質問かどうか
 C9: 参照するキー属性が一意に絞れるか
 C10: (C3 の値) / (C2 の値)
 C11: (C5 の値) / (C2 の値)
 C12: 履歴の構造モデルで, 当該内容語と同一属性をもつノードの現在の検索条件における深さ (同一属性のノードが存在しない場合, 現在の木の深さ+1)
 C13: 履歴の構造モデルで, 当該内容語と同一属性をもつノードの深さの平均 (同一属性のノードが存在しない場合, 木の深さ平均+1)
 C14: 履歴の構造モデルで, 当該内容語と同一属性をもつノードの子の数
 C15: 当該内容語が属性・値の場合, 現在の検索条件に同一のものがあるか
 C16: 当該内容語が店名 (キー属性) の場合, 既に言及されているか
 C17: 当該内容語が店名 (キー属性) の場合, 現在の検索条件と合致するか
 C18: 当該内容語が値の場合, 確認されたことがあるか
 C19: 当該内容語が値の場合, 否定されたことがあるか
 C20: 当該内容語が値の場合, 削除されたことがあるか
 C21: (C12 の値) / (現在の木の深さ+1)
 C22: (現在の木の深さ+1) - (C12 の値)
 C23: (C13 の値) / (現在の木の深さ+1)
 C24: (現在の木の深さ+1) - (C13 の値)
 C25: 現在の検索条件に合致するキー属性の数が 0 かどうか
 C26: 現在の検索条件に合致するキー属性の数が 1 かどうか

図 7 データベース検索タスクのモデルに基づく文脈的な特徴 (26 個)

4. 評価実験

4.1 評価用データの収集

提案手法を評価するために, レストランデータベースを検索する音声対話システムを実装した. データベースは図 1 のような構造で, その件数は 1217 である. キー属性は「店名」, キーでない属性は「タイプ」「住所」など図 1 に示される 11 種類である.

音声認識エンジンには Julius を用いた. 言語モデルとして, 想定質問文から作成した言語モデル (語彙サイズ 2185) と, 大量のテキストコーパスから学習した一般的な言語モデル (グルメリシピドメイン; 語彙サイズ 19447¹⁴⁾) を言語モデル融合ツール¹⁴⁾ を用いて混合した. 混合比は 9:1 とした. 得られた言語モデルの語彙サイズは 21565 である.

決定木の学習データには, 本研究室の学生 6 名による予備実験のデータ (内容語数: 748) を使用した. 決定木学習には C5.0¹³⁾ を利用した.

システムからの応答はコンソール上の文字出力と音声出力の両方で行った. 検索した結果が 8 件より多い場合は検索された件数だけを示し, 8 件以下の場合は

<http://julius.sourceforge.jp/>

表 1 対話行為ごとの発話数と内容語数

発話・内容語の内訳	検索条件の指定	情報の提示要求	条件の削除	その他	計
発話数	1220	1013	279	503	3015
実際に発話された内容語数	1388	1253	307	0	2948
正しく認識された内容語数	1133	1037	244	0	2414
認識された内容語数	1279	1177	287	60	2803

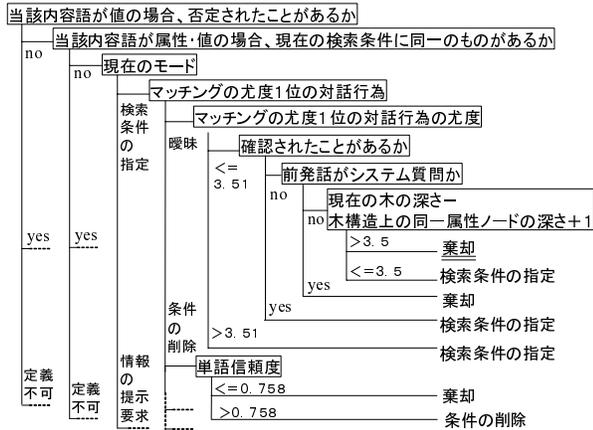


図 8 生成された決定木の一部

検索された件数と検索された店名を出力した。

上記のシステムを用いて、音声対話システムを利用したことの無い20名から対話データを収集した。ユーザにはまず、システムの説明とシステムが理解できる発話例を示し、音声入力のタイミングに慣れてもらうように5分ほど練習をしてもらった。その後「和食が食べたくなくなりましたが、あいにくVISAカードしか持っていません。」といった状況を提示し検索を行ってもらった。ユーザには自分の判断で満足できる店が見つかれば取得した情報を書き取ってもらったうえで、対話を終了してもらった。ユーザは同様の設定で3対話を行った後、自由な設定でさらに1対話を行った。

4.2 評価用データの詳細

実験により得られた発話は総計3015発話(151発話/人, 38発話/対話), 認識された内容語は2803内容語であった。また、全ての単語の単語認識率は83.4%であった。得られたデータの対話行為ごとの発話数と内容語数を表1に載せる。このうち「検索条件の指定」と「条件の削除」が同時に発話されたものが8発話存在し、表中ではいずれも1発話として二重に計数している。またこれ以外に「肯定回答」「否定回答」「条件を全て削除」「やり直し」の発話が計342発話、タスク外発話が161発話存在した。

全ての評価用データにより、全ての特徴を用いて学習した決定木の一部を図8に示す。ここでは、対話の

進行モデルに基づく特徴である「現在のモード」や、検索条件の状態から得られる「否定された語であるか」などの文脈的な特徴が上位に現れている。一例として、ユーザの「居酒屋無門」の住所をお願いします」という発話の音声認識結果が「居酒屋を、もんの住所をお願いします」であった場合を取り上げる。ここで「居酒屋無門」は固有名詞で店名である。このときこの音声認識結果に対する想定質問文とのマッチングでは、一意に対話行為が定まらず「曖昧」という結果が得られた。しかし、内容語「居酒屋」は履歴の構造モデルの上位部分にあったことから、図8の決定木をたどることにより、音声認識の信頼度が比較的高かったにも関わらず棄却された。この結果ある店の住所を質問しているという正しい解釈が得られ、「どの店の住所ですか?」という応答ができた。

また、ユーザ意図の変更へのシステムの確認発話¹⁰⁾の成否を調べた。システムはユーザの意図が大きく変化した時点で、必要でないと判定された検索条件の削除を行うかどうかをユーザに確認する。この確認は94回行われ、そのうちの37応答において検索条件の削除の必要性が認められる応答を得た。これは、ユーザの意図が大きく変更された可能性のある状況で効率よく確認が行われており、システムの確認によってユーザが検索条件の棄却を行う労力を軽減できることを示している。

4.3 言語理解精度の評価

言語理解部での内容語ごとの意味理解の精度と誤りの総数を表2と表3に示す。なお、内容語が確認と判定されたものはここでは棄却として計数している。提案手法の評価のために、以下の3種類の手法を試行した。

手法1: マッチングの尤度が最も高かった対話行為をその発話および発話に含まれる内容語の対話行為とし、各内容語ごとにその単語信頼度が閾値未満ならば棄却を行う。(ベースライン)

手法2: 提案手法に従い対話行為推定と内容語の受理/棄却を行う。ただし、文脈的な特徴を用いないで学習した決定木を用いる。

手法3: 提案手法に従い、全ての特徴を用いて、対話行為推定と内容語の受理/棄却を行う。

手法1では、棄却の閾値を20通り試行し、意味理解誤り数が最小となった0.05を選択した。また、マッチングでは一意に対話行為が定まらなかった内容語の78.2%が「検索条件の指定」だったため、その場合は全て「検索条件の指定」と判定し計数している。手法2、

表中のF値=(2*再現率*適合率)/(再現率+適合率)である。

表 2 言語理解部による対話行為ごとの意味理解精度 (F 値)

手法	(1) マッチング +信頼度	(2) 決定木 (文脈的特徴なし)	(3) 決定木 (全ての特徴)
検索条件の指定	0.862	0.876	0.876
情報の提示要求	0.900	0.921	0.940
条件の削除	0.737	0.758	0.837
棄却	0.172	0.499	0.583
計	0.869	0.900	0.923

表 3 言語理解部による意味理解誤り数 (内容語ごとの計数)

手法	(1) マッチング +信頼度	(2) 決定木 (文脈的特徴なし)	(3) 決定木 (全ての特徴)
意味理解誤り数	536	456	395

3では、収集した内容語のうち18人分を学習用データ、残りの2人分を評価用データとして学習・評価を行い、これを10通り繰り返すことによって全ての内容語の評価結果を得た。また、決定木の枝刈りのパラメータを20通り試行し、意味理解誤り数が最小となる値を選択している。

手法2で生成された決定木の上位には「マッチングの尤度が最も高い対話行為」「単語信頼度」が現れており、手法1とほぼ同じ情報から判別を行っていることがわかる。手法2ではこれらに加えて、マッチングの尤度が2番目のものとの比較を行う特徴が導入されているため、棄却の精度が向上し、全体としてF値が3.1%改善している。

手法3では、手法2と比べて意味理解誤りが61語減少しており、誤り数は13.4%(=61/456)削減された。F値は2.3ポイント上昇した。ここでは文脈的特徴を導入することにより棄却に関して最も大きな改善が見られ、手法2と比較してF値で8.4ポイントの改善を得た。手法3で得られた決定木では、検索条件の値にあたる内容語に対しては「当該内容語が否定されたことがあるか」「現在のモード」や木構造状に管理した履歴に基づく特徴(木構造上での同一属性ノードの深さなど)が上位に特徴として現れていた。それ以外の属性名やキー属性の値(店名)に対しては「現在の検索条件の数」や「キー属性ならば現在の検索条件に合致するか」「キー属性ならば既に言及されているか」などが特徴として有効であった。

ベースラインと比較して提案手法では627語の意味理解誤り削減を行えたが、未だに意味理解誤りを起こすものが395語存在している。これらのうち最も多かった誤りは音声認識誤り(「棄却」が正解)を「検索条件の指定」と判定してしまうものであり、133語存在した。このうち「四条」が「七条」と認識されるといったような同属性間の誤りが31語存在した。これらは提案手法の枠組みでは全く同一の属性を持つため、取

り除くことができない。また、対話の初期では対話や文脈の情報が十分得られない。これに起因すると見られる誤りは18語見られた。

5. ま と め

データベース検索タスクにおいてドメイン非依存に成り立つ特徴を用いて文脈的な制約を学習し、言語理解に取り入れる方法を提案し、評価を行った。評価実験の結果、文脈的制約を加えることで意味理解精度の向上が認められた。

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. 3078-3986, 2002.
- 2) 中川聖一, 堀部千寿. 音響尤度と言語尤度を用いた音声認識結果の信頼度の算出. 情報処理学会研究報告, 2001-SLP-36-13, 2001.
- 3) 李晃伸, 山田真士, 西村竜一, 鹿野清宏. 公共音声情報案内システム「たけまるくん」の運用および収集発話の分析. 情報処理学会研究報告, 2004-SLP53-9, 2004.
- 4) 東中竜一郎, 中野幹生, 相川清明. 複数文脈を用いる音声対話システムにおける統計モデルに基づく談話理解法. 情報処理学会研究報告, 2003-SLP-45-17, 2003.
- 5) 入江友紀, 松原茂樹, 河口信夫, 山口由紀子, 稲垣康善. 意図タグつきコーパスを用いた発話意図推定手法. 人工知能学会研究会資料, SIG-SLUD-A301-03, 2003.
- 6) 由浅裕規, 水野智士, 伊藤敏彦, 甲斐彦彦, 小西達裕, 伊東幸宏. 状況と文脈を利用した音声対話型車載インタフェースの構築と評価. 情報処理学会研究報告, 2003-SLP-49-34, 2003.
- 7) M. Gabsdil and O. Lemon. Combining acoustic and pragmatic features to predict recognition performance in spoken dialogue systems. In *Proc of ACL'04*, pp. 343-350, 2004.
- 8) M. Walker, I. Langkilde, J. Wright, A. Gorin, and D. Litman. Learning to predict problematic situations in a spoken dialogue system: Experiments with how may i help you? In *Proc. of NAACL00*, pp. 210-217, 2000.
- 9) 藤江真也, 坂本直宏, 清水健二, 宗近純一, 福島健太, 瀬戸祐介, 小林哲則. 音声対話ロボット ROBISUKE による相談型対話の実現. 情報処理学会研究報告, 2004-SLP53-10, 2004.
- 10) 神田直之, 駒谷和範, 尾形哲也, 奥乃博. データベース音声対話システムにおける履歴を考慮した検索条件の管理. 情報技術レターズ(情報科学技術フォーラム(FIT)講演論文集), LG-001, pp.131-132, 2004.
- 11) 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検索システム. 電子情報通信学会技術研究報告, SP2001-113, 2001.
- 12) 李晃伸, 河原達也, 鹿野清宏. 2パス探索アルゴリズムにおける高速な単語事後確率に基づく信頼度算出法. 情報処理学会研究報告, 2003-SLP-49-48, 2003.
- 13) C5.0. <http://rulequest.com/index.html>.
- 14) 河原達也, 住吉貴志, 李晃伸, 坂野秀樹, 武田一哉, 三村正人, 山田武志, 西浦敬信, 伊藤克巨, 伊藤彰則, 鹿野清宏. 連続音声認識コンソーシアム2001年度版ソフトウェアの概要. 情報処理学会研究報告, 2002-SLP-43-3, 2002.