4R-7

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

1. はじめに

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欠であり，音声認識誤りの棄却や意味曖昧性の解消はシステムの利便性を左右する重要な機能である．これまでに，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研究 [1] や，音響尤度と言語尤度を組み合わせた研究 [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研究 [3] などが行われている．これらは，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なっており，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利用して内容語の取捨選択の高精度化や言語理解精度の向上を狙う．

我々はこれまでに，データベース検索タスクにおける対話のモデルを定義し，それに基づく制約を用いて言語理解を行う手法を構築している [4]．ここでは，ユーザ発話を「検索条件の指定」「情報の提示要求」の 2 モードに大別し，対話をこれらのモード間の遷移で表現する (**対話の進行モデル**)．さらに，検索条件の履歴を木構造状に保持する (**履歴の構造モデル**)．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の曖昧性解消を行う．

本稿では，これらの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ので報告する．

2. システムの概要

図 1 に示したように，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．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，履歴の構造モデルを用いて，ユーザの意図が大きく変化したと思われるポイントで，検索条件の取捨選択を適切に行うための確認を行うこともできる [5].

3. 文脈的制約を用いた言語理解

言語理解部ではユーザ発話の音声認識結果から対話行為と内容語の推定を行う．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 [6] を行い，その類似度をマッチング

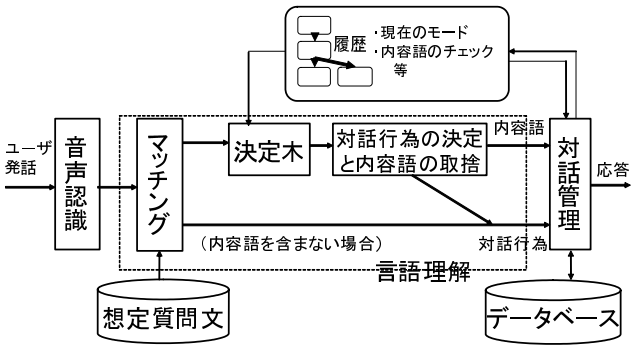

図 1: データベース検索システムの処理の概略

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．

内容語を含まない対話行為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

マッチングによって内容語を含む対話行為を持つと推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は次節で述べる．言語理解処理では，以下の手順に従って対話行為の決定と内容語の取捨を行う．

1. 各内容語 w_j ごとに決定木を用いて，現在の対話状態 F_i における各対話行為 s の信頼度 $CF(s|F_i, w_j)$ ‡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当該発話の対話行為 S_i とする．

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

2. 各内容語 w_j の棄却すべき信頼度 $R_j = CF(\text{“棄却”}|F_i, w_j)$ 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

‡ 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 [7].

Empirical Evaluation of Spoken Dialogue System using Contextual Constraint in Database Retrieval Task: *Naoyuki Kanda, Kazunori Komatani, Tetsuya Ogata, and Hiroshi G. Okuno (Kyoto Univ.)*

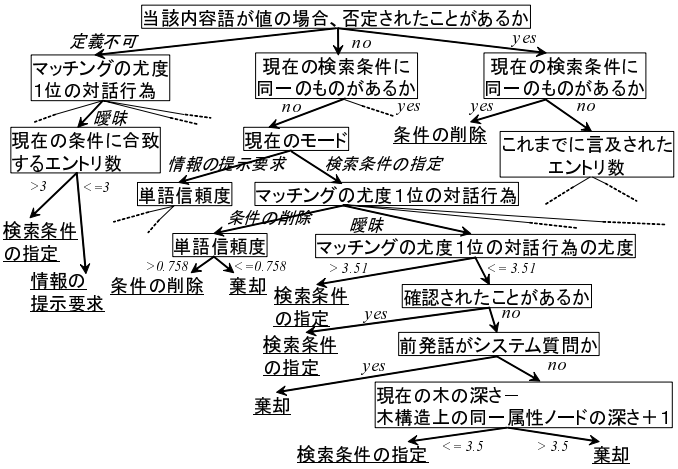

図 2: 生成された決定木の一部

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，出力として各内容語に対する対話行為と棄却すべき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**に加えて，**文脈的特徴**を使用する．一発話のみから得られる特徴としては，マッチングの結果や単語信頼度 [1] など 7 種類を用いる．文脈的特徴はデータベース検索タスクのモデルに基づいて 26 種類を定義する [4]．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や，現在の検索条件に合致するエントリ数などを設定する．さらに，履歴の構造モデルから変更対象ノードの深さ，子ノードの数などを設定する．

4. 評価実験

4.1 評価用データの収集

上記の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対話を終了してもらった．ユーザは同様の設定で 3 対話行った後，自由な設定でさらに 1 対話を行った．

実験で得られた発話は総計 3015 発話，2948 内容語であった．また，認識された内容語 2803 語のうち 389 語は音声認識誤りであった．全ての単語に対する単語認識率は 83.4%であった．

4.2 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 1 と表 2 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．本手法の評価のために，以下の 2 種類の手法を試行した．

手法 1：本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 2：本手法に従い，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．

[§]表中の F 値=(2*再現率*適合率)/(再現率+適合率) である．

表 1: 言語理解部での対話行為ごとの意味理解精度 (F 値)

手法	(1) 文脈的特徴なし	(2) 全ての特徴
検索条件の指定	0.876	0.876
情報の提示要求	0.921	0.940
条件の削除	0.758	0.837
棄却	0.499	0.583
計	0.900	0.923

表 2: 言語理解部での意味理解誤り数 (内容語ごとの計数)

手法	(1) 文脈的特徴なし	(2) 全ての特徴
意味理解誤り数	456	395

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 では，手法 1 より意味理解誤りが 61 語減少し，誤り削減率は 13.4% (= 61/456)，F 値は 2.3%上昇した．特に棄却に関して最も大きな改善が見られ，手法 1 より F 値で 8.4%の改善を得た．手法 2 で得られた決定木 (図 2) 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）が特徴として有効であった．それ以外の属性名やキー属性の値（店名）に対しては,「現在の検索条件に合致するエントリ数」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」などが上位に特徴として現れていた．

本手法を適用しても，395 語に対しては対話行為を正しく推定できない．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．これらは本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．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] 駒谷和範他．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．情処学論, Vol. 43, No. 10, pp. 3078–3986, 2002.

[2] 中川聖一他. 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. 情処研報, 2001-SLP-36-13, 2001.

[3] 李晃伸他. 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. 情処研報, 2004-SLP53-9, 2004.

[4] 神田直之他．データベース検索タスクにおける文脈的制約を用いた音声対話システム．人工知能学会研究会資料, SIG-SLUD-A401-4, 2004.

[5] 神田直之他. 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, LG-001, pp.131–132, 2004.

[6] 駒谷和範他. 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.

[7] *C5.0*. http://rulequest.com/index.html.