

データベース検索用音声対話システムにおける 意味理解部の移植性の検討

2M-05

梅田 将満 小暮 悟 中川 聖一

豊橋技術科学大学 情報工学系¹

1 はじめに

近年、音声対話システムの研究が広く行われている。我々の研究室でも富士山観光案内をタスクとする音声対話システムを開発してきた。本システムは、ユーザの発話に対してロバストな認識を行えるような開発が行われてきた [1]。

本システムはいままで Lisp で記述されていたが、性能の向上や改良を考えたとき、Lisp での開発では処理が複雑になり、可読性も悪く、Lisp での開発は効率が悪い。そこで、文字列処理に優れている Perl でシステムを再構築した。また、意味理解部の見直しに加え、新たな改良も行うことにより意味理解率の向上を目指した。

また、従来のシステムではタスクの移植性を考慮に入れての開発も行われていた [2]。そこで今回のシステムでもタスクの移植性を考慮し、文献検索タスクへの適用を行った。本稿では、これらの改良と結果について報告する。

2 従来の音声対話システム

本研究室で開発されている従来の音声対話システムは、あるタスクに対してのデータベース検索を音声入力で行い、検索結果を合成音声で応答するシステムである。このシステムは大きく 3 つの要素に分かれており、音声認識を行う音声認識部、音声認識結果文からデータベース検索を行うためのキーワードを抽出するための意味理解部、データベースにアクセスして結果を応答するための対話制御・応答生成部からなっている。

3 Perl での再構築と改良

本研究では、システムの核ともいえる発話文の意味を解析する意味理解部の再構築を重点的に行った。従来のシステムの仕様を参考に可読性や処理を改善しながら構築を行った。また、従来のシステムで行っていなかった処理などを追加して改良も図った。

3.1 意味理解部の構成と再構築

意味理解部は音声認識で得られた日本語テキスト文を形態素解析、構文解析、意味解析し、データベース検索するためのキーワードを抽出するという言語理解の重要な部分である。

形態素解析とは、入力された日本語テキスト文を形態素に分解することである。本システムでは、形態素解析器として茶筌 2.02 を用いている。これを Perl 上で呼び出して利用し、各単語に意味素を付加し、形態素、品詞情報、意味素のペアが構成される。構文解析で、この品詞情報を利用して形態素解析結果を文法 (CFG) に当てはめ、文節ごとに区切っている。今回の移植では構文解析にチャート法 [3] を用いた。そして、構文解析結果を元に、動詞の格フレーム情報から、文に対して一つの意味表現を生成し、検索のためのキーワードを抽出する。また、意味表現が生成できなかった場合、意味素のみで検索キーを作成するキーワード検索機能もある。

以上が意味理解部であるが、データベース検索から結果を簡易的に応答する部分も作成した。データベースには PostgreSQL* を用い、検索のための SQL 生成と検索処理も Perl で構築している。また、一連の対話処理を実現するために、音声入力や結果の表示、合成音声による簡単な応答生成についても構築した。

3.2 改良点

今回新しく訂正発話の処理を追加した。発話中に「いえ」「じゃなくて」といったような言い直しの発話の際、その単語と 1 つ前の単語を破棄するという処理である。図 1 に訂正発話の例を示す。

```

入力文      : 「スキーいやテニスがしたい」
形態素解析 : スキー (名詞) いや (名詞) テニス (名詞) ...
              ↓ 「いや」と前の 1 単語を除去
解析文      : 「テニスがしたい」
  
```

図 1: 訂正発話

また、今回は構文解析にチャート法を用いている。CFG による文法にあてはめ、文頭から文末までで文と定めた状態に遷移した (つまり不活性弧となったもの) ものだけを解析するのではなく、途中状態で文に遷移した不活性弧の中で最長で文になるものを採用することによって、文頭に言い淀み

Consideration on the portability of the interpreter in the spoken dialogue system for database retrieval
Masamitsu Umeda, Satoru Kogure, and Seichi Nakagawa
¹Toyohashi University of Technology

*<http://www.postgresql.org>

や誤認識が入力されたり、文法に当てはまらないような発話になってしまったときに対応できるようにした。図2に不要語の除去の例を示す。ここで、構文解析の(4)はCFG規則の4番目のルールが使われたとき、[1,6]は1個目の単語の区切りから6個目の単語区切り(この例では「ホテルはありますか」)で文になっていることを示している。

単語区切り
 入力文: 「0 ほ¹ ホテル² は³ あり⁴ ます⁵ か⁶」
 形態素解析: ほ(動詞) ホテル(名詞) は(助詞) …
 構文解析: 文法(4) 文 → 名詞句 動詞句 [1,6]
 (文(名詞句 ホテル(名詞) は(助詞))
 (動詞句(動詞句(あり(動詞)) ます(助動詞)) か(助詞)))

図2: 不要語の除去

4 システムのタスク移植性

本研究における移植性の高さとは、構築の効率の良さである。従来のシステムでも移植性について十分考慮されていた[2]。そこで、今回の再構築でも別のタスクに変更する際、変更するデータ部を明確にし、タスク依存なデータのみを差し替えられるようなシステム作りをした。表1にドメイン-タスク独立情報とドメイン-タスク依存情報の部分について示す。

表1: ドメイン-タスク依存, 独立情報

	データ
ドメイン-タスク独立	音節 HMM, 名詞と動詞以外の形態素語彙辞書, 形態素接続辞書, 構文解析文法
ドメイン-タスク依存	言語モデル bi-gram, 名詞と動詞の形態素語彙辞書, 検索対象データ中の単語の名詞意味辞書と動詞意味辞書, PostgreSQL用のデータベース, 動詞の格フレーム情報辞書, 意味表現から検索キーワードへの変換規則, 検索キー拡張データ, 検索結果表示プログラム, 検索結果表示項目データ

5 評価実験

5.1 富士山観光案内システムの評価

従来システムで試されていた4名の発話文による音声認識(CFG, bi-gram)結果文と発話書き起こし文それぞれ全404文を用いてどれだけ意味表現が正しく生成されたかを調べた。表2に結果を示す。発話の書き起こし文については、今まで対応していなかった構文にも対応でき、5.7%意味理解率が向上した。また、音声入力文でも、CFGで12%, bi-gramで9%ほどよくなった。音声入力文ではCFGを用いた方がbi-gramを用いるよりも音声認識率が悪い。CFGの方が向上率が大きいということは、頑健な理解ができるようになったと言える。

表2: 富士山観光案内システムの意味理解率

システム	評価	成功/合計(理解率)
従来:Lisp	音声 CFG	249/404(61.6%)
	音声 bi-gram	299/404(74.0%)
	書き起こし文	328/404(81.2%)
今回:Perl (改良版)	音声 CFG	299/404(74.0%)
	音声 bi-gram	335/404(82.9%)
	書き起こし文	351/404(86.9%)

5.2 文献検索タスクの評価

移植性を評価するために文献検索タスクへの変更を行い、過去のWizard of Oz法による評価実験で得られたテスト文100文のうち、比較的易しい34文を3名が発話した発話文によって評価した。音声認識の言語モデルはbi-gramのみで行った。表3に結果を示す。文献検索タスクにおいても、高い意味理解率を得ることができた。

表3: 文献検索システムの意味理解率

システム	評価	成功/合計(理解率)
従来:Lisp	音声:bi-gram	184/204(90.2%)
	書き起こし文	204/204(100%)
今回:Perl (改良版)	音声:bi-gram	196/204(96.1%)
	書き起こし文	204/204(100%)

移植の際の手間は、データの書き起こしを除けば、システムが動作するようにするには3日ほどでできた。

6 まとめ

Perlで再構築し、意味理解部を改良したことにより、意味理解率の向上が得られた。

今後は、質問に対する応答生成部や対話に関する発話に対しての対話処理部を構築していく予定である。また、今回作成した意味理解部においても、よりロバストな意味理解を目指したり、ユーザへのリジェクトを確実に実行していきたい。さらに、移植性においても検討を重ねていく予定である。

参考文献

- [1] 伊藤 敏彦, 小暮 悟, 中川 聖一: 協調的応答を備えた音声対話システムとその評価, 情報処理学会論文誌, pp.1248-1257(1998.5)
- [2] 小暮悟, 中川聖一, データベース検索用音声対話システムにおける移植性の高い意味理解部・検索部の構築と評価, 情報処理学会誌, Vol. 43, No. 3, 採録決定(2002.3)
- [3] 田中 穂積: 自然言語処理 - 基礎と応用 -, 電子情報通信学会(1999)