

対話事例を利用した音声対話システム

村尾浩也 (三洋電機・ハイパー研/名大 CIAIR)
Email: mura@hr.hm.rd.sanyo.co.jp

河口信夫 松原茂樹 (名大 CIAIR) 稲垣康善 (名大工)

あらまし 本稿では、対話事例を用いる対話手法についての提案を行う。本手法は人間の自由な発話に対応可能な音声対話システムの実現を目指したものである。はじめに、情報検索対話のモデル化を行い、そのモデルに従って対話手法の定式化を行う。次に、本手法を用いて設計した、情報検索対話を対象とする車内情報検索システムについて述べる。このシステムは、オペレータ対ユーザの対話事例を蓄積したデータベースを用いて、過去の類似した事例に基づいて検索式生成を行うものである。検索式生成性能の評価を行ったところ、本手法の有効性が確かめられた。

キーワード 音声対話, 音声認識, 言語理解, 自動車

Example-based Spoken Dialogue System

Hiroya Mura (Sanyo Electric Co., Ltd./CIAIR, Nagoya Univ.)
Nobuo Kawaguchi, Shigeki Matsubara(CIAIR, Nagoya Univ.),
Yasuyoshi Inagaki(Nagoya Univ.)

Abstract This paper describes a new dialogue control method that is based on examples from human to human dialogue. Along with modeling and formalization of dialogue, a system for information retrieval in a car has been designed. The system refers the corpus of dialogue for an example that is similar to input speech, and makes a query by the example. We then give results of experiments to see the effect of this method.

key words Spoken dialogue, Speech recognition, Language understanding, Car environment

1 はじめに

我々は現在、多様な入力発話に対応可能な音声対話システムの実現を目指し研究を行っている。音声対話システムにおける対話制御方式としては、これまでに状態遷移モデルを使用するもの[1]や、フレーム駆動型のもの[2]などが報告されている。これらの対話モデルでは、対話状況の遷移形態や、状況を規定するためのフレームの構造や内容をあらかじめ定義しておく必要がある。しかし、自由対話の多様な現象をすべて想定し、対話モデルの記述を行うことは難しい。これらの現象を扱い、人間の自然な発話に対応できる対話システムを実現するための方法として、実際の人間の行動事例に基づいたアプローチが有効である。

本稿ではまず、タスク指向対話である情報検索対話を対象とし、事例に基づく音声対話処理手法の提案を行う。その後、本手法を用いて設計した車内情報検索システムについて述べる。このシステムは、対話事例として我々が収集した車内音声対話コーパス[3][4]を用い、レストラン情報検索をタスクとするシステムである。過去の類似した事例に基づいて検索式生成を行うことにより、情報検索オペレータの経験やノウハウを取り入れた処理を行うことができると期待される。最後に、このシステムの検索式生成性能の評価実験結果について述べる。

2 事例に基づく対話

2.1 情報検索対話のモデル化

人間のオペレータが情報データベースを検索し、ユーザに対して情報を提供する状況での、オペレータとユーザ間の対話は、次のようにモデル化することができる。(図1)

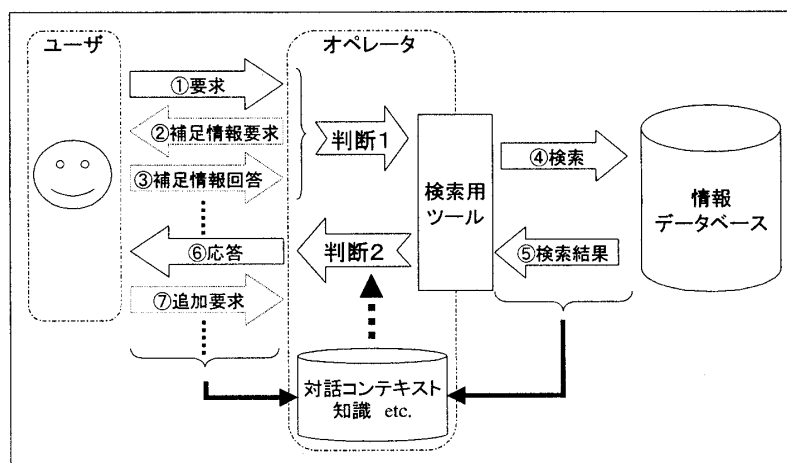


図1: 情報検索対話のモデル

- ① 要求 ユーザの要求発話を受けたオペレータは、現在の対話コンテキストを参照しながら検索式を生成する。検索式の生成は、一般にはコンピュータなどの検索用ツールを操作することで間接的にされる。
- ② 補足情報要求 検索式生成に必要な情報が不足している場合は、補足情報をユーザに対して要求する。
- ③ 補足情報回答 ユーザは補足情報要求への回答を行う。
- ④ 検索 生成された検索式により、検索が実行される。
- ⑤ 検索結果 検索結果が生成される。
- ⑥ 応答 オペレータは、検索結果と対話コンテキストに基づいて応答を行う。
- ⑦ 追加要求 応答を受けたユーザは、追加質問を行ったり、別の要求を発したりする。

このように情報の流れを整理すると、次のように考えることができる。図1で、オペレータは対話の進行のために次の2つの判断を行っている。

判断1 ユーザの発話を受け検索式を生成する

判断2 検索結果を受け、応答を生成する

熟練したオペレータは過去の経験や知識などを利用し、発話に現れる表層的な情報以外の種々の情報を援用した「判断」を行い、ユーザの目的に合った検索を行っていると考えられる。

2.2 対話モデルに基づく定式化

我々は、このような対話を進行するための「判断」を、熟練した人間のオペレータが行った行動事例を参照して行うことが有効であると考えてい

る。情報検索システムを利用した人間同士の対話と、その際に行われた検索操作を蓄積し、それを利用して入力発話に対する検索操作を決定する。

対話事例データを、次のように対話の基本要素 X_n の時系列として定義する。^{1,2}

$$U = \{X_1, X_2, \dots, X_n, \dots, X_N\} \quad (1)$$

$$X_n = (S_n, Q_n, A_n, R_n) \quad (2)$$

U : 対話事例データベース全体
 X_n : 対話の基本要素
 N : 事例データベース中の基本要素数
 S_n : 入力発話テキスト
 Q_n : 検索式
 A_n : 検索結果
 R_n : 応答発話テキスト

ユーザからの入力発話 S に対して、検索式 Q を、次のようにして決定する。

$$Q = F(Q_k, S) \quad (3)$$

$$k = \arg \max_{1 \leq n \leq N} (D_S(S, S_n)) \quad (4)$$

$D_S(S, S_n)$: 発話 S, S_n 間の類似度

関数 F は、式(4)で得られた最類似事例中の検索式 Q_k を入力発話 S に合わせて修正するものである。

生成した検索式 Q を用いて検索を行い、検索結果 A を得る。(f は検索操作をあらわす)

$$A = f(Q) \quad (5)$$

入力発話 S に対して検索結果 A が得られたとき、応答発話 R を次のようにして求める。

$$R = G(R_k, S, A) \quad (6)$$

$$k = \arg \max_{1 \leq n \leq N} (D_S(S, S_n) + \omega \times D_A(A, A_n)) \quad (7)$$

$D_A(A, A_n)$: 検索結果 A, A_n 間の類似度
 ω : 重み係数

関数 G は、得られた最類似事例中の応答発話を、入力発話と検索結果に合わせて修正するものである。

3 事例に基づく車内情報検索システム

前節の式に基づき、検索式生成部のプロトタイプを実装した。図2にシステムの構成を示す。このシステムは、図1のモデルにおける①要求④検索⑤検索結果生成を行うもの、つまり「判断1」を対話事例に基づいて行うものである。

3.1 システム構成

対話事例データベース 走行車内において収集されたオペレータ対ユーザの対話コーパス[3][4]を元に作られている。具体的にはユーザ発話テキスト、検索式、検索結果、オペレータ発話テキストのセットである。発話テキストは形態素解析³されており、対話の遂行に重要な語(店名、施設名、食品名など)はあらかじめ意味別にクラス化され、単語クラスタグが付与されている。

単語クラスデータベース 本タスクにおける重要語を抽出し、意味別にクラス化したデータベース。単語クラスは、対話コーパス中の対話を元に経験的に求めた。

情報データベース 名古屋市内のレストラン、店舗など約800件が収録された店舗情報データベース。

音声認識部 日本語ディクテーション基本ソフトウェア99年版[5]を使用。言語モデルは対話コーパスの書き起こしから作成した。

検索式生成部 入力発話に対し、その時点の対話状況に最も近い事例を対話事例データベースから抽出し、その事例で用いられた検索式を現在の状況に合うように修正して出力する。

検索実行部 検索式を用いて情報データベースにアクセスし、検索結果を得る。

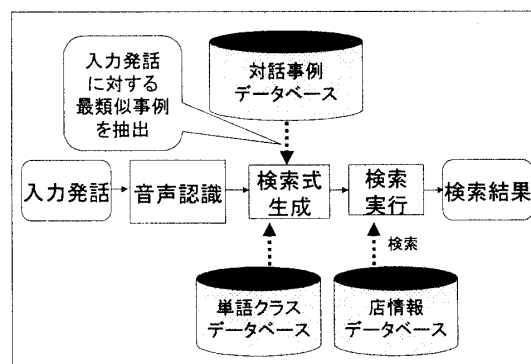


図2：システム構成

¹入りに情報が不足するなどして検索操作が行われない時、応答が生成されないときなどは、 I_n, A_n, R_n 等は値なしとなる。

²ここでは、簡単のために対話コンテキスト（文脈）を考えないとして定式化を行う。文脈を含む場合は式(2)の右辺が文脈情報を含む項組となる。

³ 形態素解析は茶筌 2.02 [8]を使用した。

3.2 動作

本システムの動作を図3の例に沿って述べる。

【ステップ1：類似事例抽出】

音声認識結果テキスト S に対し、式(4)により対話事例データベース中から最類似事例を抽出する。走行雑音や運転中発話に対する音声認識・理解のロバスト性を考慮[6][7]し、式(4)における、2発話間の類似度関数 Ds としては単語クラスを考慮したキーワードマッチングの手法を採用している。キーワード抽出は、形態素解析結果から自立語を抽出して行った。

入力発話 S から抽出したキーワード列を S' とする。

$$\begin{aligned}
 S &= \text{“えーとスパゲティのお店に行きたいな”} \\
 S' &= \{W_1, W_2, \dots, W_i, \dots, W_I\} \\
 &= \{\text{スパゲティ, 店, 行く}\} \\
 W_i &: \text{キーワード} \quad I : S' \text{のキーワード数}
 \end{aligned} \tag{8}$$

同様に、事例データベース中のユーザ発話 S_n からキーワード列 S'_n を求める。

$$\begin{aligned}
 S_n &= \text{“カレーのお店に行きたいんですけど”} \\
 S'_n &= \{w_1, w_2, \dots, w_j, \dots, w_{J_n}\} \\
 &= \{\text{カレー, 店, 行く}\} \\
 w_j &: \text{キーワード} \quad J_n : S'_n \text{のキーワード数}
 \end{aligned} \tag{9}$$

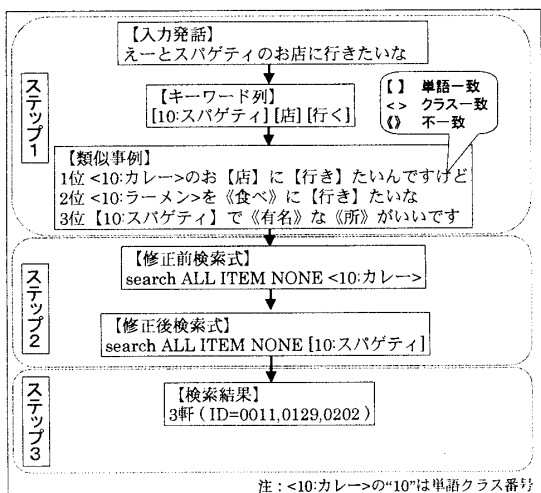


図3：動作例

表1：類似度計算時のキーワード分類基準

		単語クラスに	
		属する	属しない
w_i と w_j が	一致	分類M ₁	分類M ₂
	不一致	分類M ₃	分類M ₄

S と S'_n の類似度を次式の定義に従って求める。

$$Ds(S, S'_n) = \alpha M_1 + \beta M_2 + \gamma M_3 - \mu M_4 - \Delta \tag{10}$$

$M_1 \sim M_4$ は、入力発話中の各キーワード W_i について、事例中発話のキーワード w_j との関係を表1に従って4つに分類した際の各分類に属する語数であり、

$$M_1 + M_2 + M_3 + M_4 = I \tag{11}$$

である。

Δ は、 S' と S'_n の語数の差によるペナルティ値で次式のように定義される。

$$\Delta = \begin{cases} \lambda(J_n - I) & J_n - I > 0 \\ 0 & J_n - I \leq 0 \end{cases} \tag{12}$$

$\alpha, \beta, \gamma, \mu, \lambda$ は正の値をとる重み係数で、現在は予備実験の結果を元に、

$$\alpha = 5.0, \beta = 2.0, \gamma = 3.0, \mu = 1.0, \lambda = 1.0$$

としている。

図3の例では、

$$\begin{aligned}
 M_1 &= 0, M_2 = 2 \text{ (「店」と「行く」の2つ)}, \\
 M_3 &= 1 \text{ (「スパゲティ」と「カレー」が同じクラス)}, \\
 M_4 &= 0, \Delta = 0 \quad (I = J_n = 3)
 \end{aligned}$$

となり、

$$\begin{aligned}
 Ds(S, S'_n) &= 5.0 \times 0 + 2.0 \times 2 + 3.0 \times 1 - 1.0 \times 0 - 0 \\
 &= 7.0
 \end{aligned}$$

である。 $M_1 \sim M_4$ の各分類に対して異なる重み係数を掛けることにより、単語の重要度や一致度を反映した類似度計算を行うことができる。

事例データベース中のすべてのユーザ発話（つまり、すべての n ）について式(10)を計算し、類似度が最も高い発話を最類似事例とする。

【ステップ2：検索式修正】

抽出した最類似事例 X_n に対応する検索式 Q_n を、式(3)により入力発話 S に合うように修正する。

検索式 Q_n は、次のような形式である。⁴

$$Q_n = (\text{search ALL ITEM PAR}_n K_n) \tag{13}$$

⁴ 現在は対話コンテキストのない発話を対象としているため、最初の3項 (search ALL ITEM) は固定である。

PAR_n は、検索結果の表示順を表すパラメータである。例えば、 $PAR_n = NEAR$ のときは、検索結果を現在位置から最も近い順に出力する。

検索式の修正は、検索式中のキーワード列 K_n 中の語を入力発話 S に応じて入れ替えることにより行う。

検索式 Q_n 内のキーワード列 K_n を、

$$K_n = \{V_0, V_1, \dots, V_p, \dots, V_P\} \quad (11)$$

P : キーワード数

とあらわす。図3の例ではキーワードは1語で、

$$K_n = \{\text{カレー}\}$$

である。

入力発話 S 中のキーワード列 $S' = \{W_i\}_{i=1, I}$ について、次の①~③の処理を行う。

- ① $W_i = V_p$ なら V_p はそのまま使用
- ② ①に該当しない W_i で、 $C(W_i) = C(V_p)$ なら、 V_p に代えて W_i を使用
 $C(W)$: 単語 W の属する単語クラス
- ③ ①②で置き換えられなかった全ての V_p を除去

図3の例では「スパゲティ」が規則②に、「店」「行く」が③に該当し、キーワード「カレー」が「スパゲティ」に置き換えられる。

【ステップ3：検索】

得られた検索式で検索を実行。検索結果を得る。

4 評価

検索式生成部の評価を行った。

4.1 評価方法

実験条件について表2に示す。

事例データベースは、事例数と性能との関係を調べるため、表3に示すように話者数を変えた4種類のデータベースを作成した。4つのデータベースにおける話者の重なりは図4のようにした。

評価は以下の手順に基づいて行った。

- 1) 評価データの発話テキストを検索式生成部に入力し、検索式を生成する。
- 2) 得られた検索式を表4の4種に分類する。

表2：実験条件

評価データ	CIAIR車内対話コーパスの書き起こし18名分 (発話履歴を用いない発話のみを抽出：計264発話)
事例データ	CIAIR車内対話コーパスの書き起こし28名分に検索式を事後的に付与したもの。(発話履歴を用いない発話のみを抽出) 表2のように、話者数を変えて規模の異なるデータベースを使用
情報データベース	CIAIR車内対話コーパスの収集に使用したものと同一項目を持つ。
単語クラスデータベース	事例データ中に出現する単語を元に作成。クラス数25。

表3：事例データベースの話者数

	話者数		総発話数
	女性	男性	
DB1	2	5	109
DB2	4	10	226
DB3	6	15	349
DB4	8	20	470

注：話者の重なりは図4のようにした。

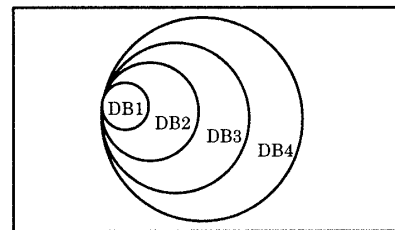


図4：事例話者の重なり

表4：評価基準

分類	内容	例	
		入力発話	検索式
分類1	正解	正しい内容を検索できる。	えっと洋食の店 <search ALL ITEM NONE 洋食>
分類2	部分正解	一部情報が欠落しているが、誤った内容を検索してはいない。	オレンジのチューリップのある花屋さんはないかなあ <search ALL ITEM NONE 花屋> →「オレンジのチューリップのある」店は検索できていないが、花屋は検索できている。
分類3	誤り	誤った内容の検索を行う。	花屋さんとケーキ屋さんあるかなこの近くに <search ALL ITEM NEAR ケーキ> →花屋が検索できていない
分類4	検索式生成失敗	キーワード抽出失敗、類似する事例がない、など。	こちら辺にスーパーはありますか (生成されない)

4.2 結果

評価結果を図5に示す。ここでは、分類結果のうち【分類1、2】を正解としている。事例話者数28名の場合で、平均約70%の発話に対して正しい検索結果を抽出することができた。また、事例の数を増やすことにより、正解率が向上することも確認できた。

4.3 誤りの分析

本手法の問題点を調査するため、DB4に対して検索式生成を誤った（【分類3、4】）評価用発話77発話について、誤りの原因を調査した。

その結果、誤りの主な原因は、次の5つに分けられた。（括弧内は誤り全体に占める割合）

【A】事例不足（31%）

（例）入力：「こちら辺にスーパーはありますか」
→「スーパー」の対話事例がない

【B】入力文が曖昧で検索式が作れない（29%）

（例）入力：「さっぱりしたものとか食べたいんだよねえ」

【C】最類似事例抽出法の問題（18%）

キーワード抽出誤りや、式(10)、(12)における定数 $(\alpha, \beta, \gamma, \mu, \lambda)$ の調整不足など

【D】コーパスの書き起こし（12%）

対話の書き起こしと茶筌の辞書における漢字やかなの表記の違いのため、形態素解析に失敗する。

（例）「チャンポン」と「ちゃんぽん」

【E】辞書登録の問題（10%）

固有名詞、あるいは辞書登録が必要な語
（例）冷やしうどん、鍋焼きうどん、など

上記分類において、分類【A】【D】【E】についてはデータの追加により誤りが改善される可能性がある。また【C】については、最適なパラメータ設定を求めることで性能改善の可能性があるといえる。

5 おわりに

人間の自然な発話に対応できる対話システムの実現を目指し、ユーザ対オペレータの対話事例を利用した対話手法に関する提案を行った。本手法を車内情報検索システムとして実装し、対話コンテキストを用いないレストラン検索発話に対して検索式性能評価を行った結果、発話の約70%に対して妥当な検索式を生成できることがわかった。さらに、誤りの内容を調査したところ、誤りの約70%が事例や語彙の不足が原因であることがわかり、蓄積する事例数をさらに増加することにより性能向上が期待できることがわかった。

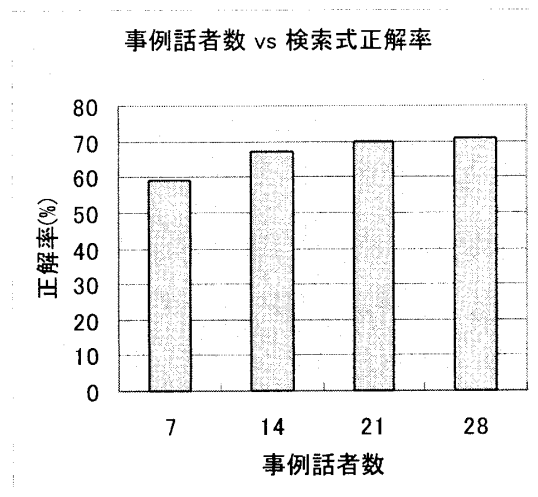


図5：検索式生成性能

今後の課題としては以下のような事項が挙げられる。

- ・ 応答部の実装、対話性能評価
- ・ 類似事例計算法の改良
- ・ 実車内での評価
- ・ 対話コンテキストの取り扱い法の検討
- ・ 検索式が生成失敗時の補足対話法の検討

【参考文献】

- (1) 内藤他：“大規模内線電話受け付けシステムの試作”，信学技報 SP94-90(1995-01)
- (2) Goddeau et al.：“A Form-Based Dialogue Manager for Spoken Language Applications”，Proc. ICSLP '96, pp.701-704(1996)
- (3) 河口ほか：“実走行車内における音声データベースの構築”，99-SLP-30-12(2000.2)
- (4) Kawaguchi et al.：“Construction of Speech Corpus in Moving Car Environment”，Proc. ICSLP2000, pp. 362-365,(2000).
- (5) 河原ほか：“日本語ディクテーション基本ソフトウェア(99年版)の性能評価”，2000-SLP-31-2(2000.6)
- (6) Kuhn et al.：“The Application of Semantic Classification Trees to Natural Language Understanding”，IEEE Trans. Pattern Analysis and Machine Intelligence, Vol.17, No.5(1995.5)
- (7) 河原、松本：“音声言語処理における頑健性”，情報処理, Vol.36, No.11, pp.1027-1032
- (8) 松本ほか：“日本語形態素解析システム「茶筌」version2.0 使用説明書 第2版”，Information Science Technical Report NAIST-IS-TR99008, 奈良先端科学技術大学院大学,1999