

## 家電製品のマニュアルの知識と構造を利用した音声対話ヘルプシステム

伊藤 亮介 駒谷 和範 河原 達也

京都大学 情報学研究科 知能情報学専攻

〒 606-8501 京都市 左京区 吉田本町

e-mail: rito@kuis.kyoto-u.ac.jp

あらまし 家電製品のマニュアルの検索に音声対話インタフェースを適用することで、ユーザフレンドリなヘルプシステムを構築する。本システムでは、まずユーザに自由な発話による検索を許し、キーワード集合を基にマニュアルの項目とマッチングを行うが、この段階では多数の候補が得られる。そこで、家電製品のマニュアルから抽出される知識と階層的な構造を利用して対話を行うことにより、ユーザの意図を詳細化し検索結果を絞り込む。効率的に絞り込む質問を生成するために、マニュアルのディレクトリ構造を利用し、2種類のコスト関数を定義した。VTRのマニュアルに関して14人の被験者に対して実験を行った結果、本対話戦略によってシステムからの質問回数は、単純に尤度の高い候補から確認発話を行う戦略に比べて約66%に減少した。

## Spoken Dialogue Help System for Electrical Appliances using Knowledge and Structure of their Manuals

Ryosuke Ito Kazunori Komatani Tatsuya Kawahara

Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan

e-mail: rito@kuis.kyoto-u.ac.jp

**Abstract** We present a user-friendly help system for electrical appliances with speech interface that makes queries to their manuals. Users can make queries by unconstrained speech, from which keywords are extracted and matched to the items in the manual. As a result, so many items are usually obtained. Thus, we introduce an effective dialogue strategy which narrows down the items using a tree structure extracted from the manuals. We present two cost functions that minimize the number of dialogue turns. We have evaluated the system performance with 14 subjects on VTR manual task. The number of average dialogue turns is reduced to 66% using our strategy compared with a conventional method that makes confirmation in turn according to the matching likelihood.

# 1 はじめに

近年、我々を取り囲む家電製品の性能が飛躍的に高機能なものになり、我々の生活は便利なものになっている。しかしその反面各家電製品の操作法が煩雑となり、マニュアルから目的の項目をすばやく見つけ出すことは難しい。これに対して音声言語による検索は要求を直接的に伝えられ習熟が容易である。本研究では家電製品のマニュアルの検索に音声認識及び音声対話インタフェースを適用することで、よりユーザにとって使いやすいヘルプシステムの構築を目指す。従来、家電製品に音声言語インタフェースを用いる研究はあるが、基本的にはコマンドによる機器の操作であった [1][2]。本研究では、自由発声から対話によりユーザの意図を推測し、検索できるシステムを目指す。

ユーザフレンドリなシステムを実現するには自然な発話から検索結果が得られることが望ましい。しかし、ユーザが最初から検索に必要な情報を含む発話ができるとは限らない。また、誤認識や解釈誤りの問題もある。そこで認識された音声からユーザの意図を推測するため、ユーザの発した情報の不足分と認識段階における情報の欠損を、ユーザとの対話によって埋めていく枠組みを考える。またマニュアルのもつ知識とマニュアルに見られる一般的な構造を利用することで、汎用性の高い対話管理を行うシステムを構築する。

従来のヘルプシステムにおいては、各項目をカテゴリでまとめていくことにより木構造を作り、ユーザはその木構造を上から順にたどっていく方法(ディレクトリ検索型検索システム)と、キーワードを複数入力することで、そのキーワードと一致する項目を羅列する方法(キーワード検索型検索システム)が用いられている [3]。ディレクトリ検索型では、概念的に理解しやすい構造になっているが、木構造を矛盾なく、しかもユーザにわかりやすいように作り上げることが困難である。また出力が音声、もしくは家電製品の小さな表示画面を考慮すると、表示するカテゴリが多いとユーザは全てのカテゴリが提示されるのを待たなくてはならない。一方キーワード検索型では、ユーザが自由に入力できるという利点はあるが、出力の際に最尤解を一つに絞ることは難しい。検索結果が多いと音声ですべての候補を読み上げるのも効率的ではない。

そこで本研究においては、入力にはユーザに自由

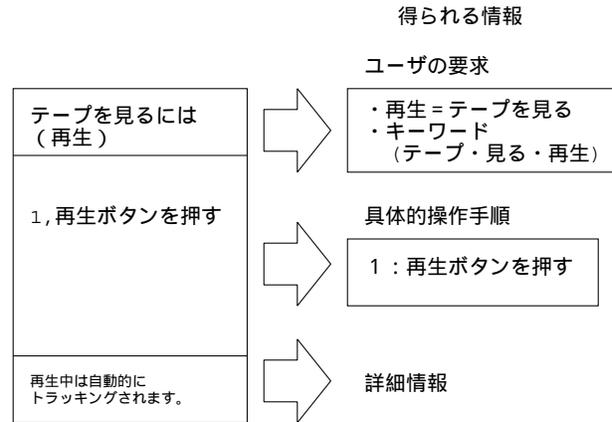


図 1: 各操作法の項目の基本的構成の例

な発話を許し、音声出力の際の問題点を解決するために、マニュアルのディレクトリ構造を用いて、検索結果が多い場合に候補を効率的に絞り込むための質問を生成する方式を提案する。また最適な対話戦略のために、質問のコスト関数についても検討する。

## 2 マニュアルからの知識と構造の抽出

ヘルプシステムは、ユーザの操作法に関する質問に対して、システムの持つ知識の中で最も適当なものを出力するものである。本研究では、対話システムの持つ知識として家電製品などの操作マニュアルを仮定する。したがって、本ヘルプシステムのタスクは、ユーザ発話に対応するマニュアル内の項目をいかに効率良く見つけるかである。

操作マニュアルには、目次に見られる知識の階層構造や、各項目における一般的なフォーマットが存在する。本研究ではこれらを利用することにより、構造を持ったシステムの知識を構築する。

### 2.1 各項目の基本的構成の利用

操作マニュアルの各項目には以下のような部分に分割できると仮定する(図 1)。

- 見出し部分
- 具体的操作手順
- 詳細情報

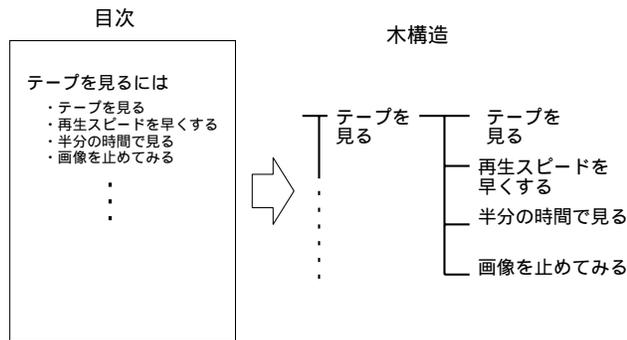


図 2: 目次からのカテゴリ化の例

見出し部分にはその項目に記述されている操作方法の要約が載せられている。各項目の内容を表すキーワードとして、この見出し部分の自立語を取りだして保持しておく。このキーワードは後段でのユーザ発話とのマッチングに利用する。また見出し部分にはそのマニュアル独自の専門用語が用いられることがあるが、通常は近くに具体的な意味が記してあるため、この同値関係も保持しておく。

具体的操作手順部分と詳細情報部分に関しては、出力すべき項目が決定された場合にユーザに対して出力する。

## 2.2 マニュアルの目次に見られる木構造の利用

マニュアルの目次は、概念の包含関係を表すように階層的に構造化されている(図 2)。この木構造をシステムの知識構造としてそのまま利用する。すなわち木構造のそれぞれの子が持つ内容の内包的概念を、その親となるノードの見出しが持っているとみなす。

入力発話とのマッチングにおいて多くの項目が得られた場合には、この構造を利用して各項目の内容の概念をシステム側から質問することによって、効率良く目的の項目を絞り込むことが可能となる(3.3 節)。

## 3 音声対話ヘルプシステムの構築

本研究では、家電製品の一例としてビデオデッキのマニュアルを対象とする。当例を用いた理由は、典

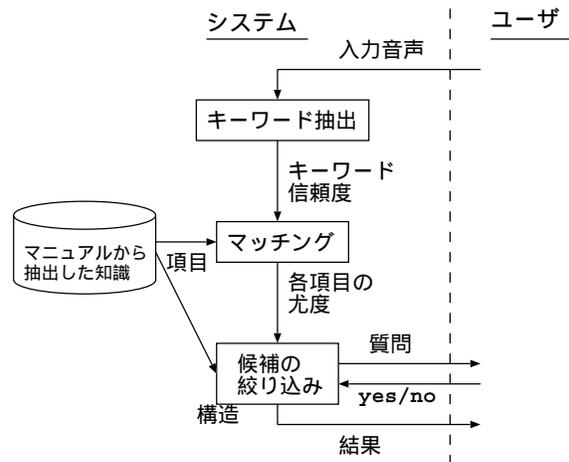


図 3: 処理の概要

型的な家電製品でありかつ多様な操作法があり、規模も適度であるためである。システムの処理の基本的な流れを図 3 に示す。以下各モジュールについて順に説明する。

### 3.1 入力音声からのキーワード抽出

ユーザの発話に対して音声認識を行い、認識結果からはキーワードを取り出す。各キーワードには音声認識結果の信頼度 [4] を計算しておく。

### 3.2 入力単語とマニュアルの知識とのマッチング

音声認識部において得られたキーワードと信頼度を用いて、ユーザの入力発話がマニュアルのどの項目に該当するかを検索する。認識されたキーワード集合に対して、マニュアルの各項目  $j$  がどの程度近いかを表す尤度  $L_j$  を定義する [5]。

各キーワードの重要さと音声認識の確実性を考慮するために、認識された各キーワード  $i$  に対して文書頻度  $df_i$  と信頼度  $CM_i$  を用いる。文書頻度  $df_i$  はキーワード  $i$  を含むマニュアルの項目数であり、この逆数をとることで、一般的に現れるキーワードの重みを小さくし、特定の項目にしか現れないキーワードの重みを大きくすることが可能である [6]。またここに信頼度を導入することで、認識誤りの可能性の高



3. ユーザの発話により以下の処理を行う。

- 肯定の場合: 2. で選択したノードの子孫以外の葉をすべて候補より削除する。
- 否定の場合: 2. で選択したノードの子孫の葉をすべて候補より削除する。

### 3.3.2 質問生成のためのコスト関数

次に質問の効率を評価するコストについて検討する。効率的な質問であるほど、この質問のコストが小さくなるような質問関数を定義する。本研究では、以下のように2つの質問関数を考える。

質問関数 1 曖昧性が大きい項目から質問するのが絞り込みに有効と考えられる。本研究では、事後確率が0.5に近いほど曖昧性が高いと考え、下式を用いる。

$$\text{質問のコスト} = |(\text{ノードの尤度}) - 0.5|$$

質問関数 2 現在の状態以後の質問の回数を考慮した対話戦略が有効であると考えられる [7]。本研究では、予測される質問の回数を残った葉の数として近似した下式を用いる。

$$\begin{aligned} \text{質問のコスト} = \\ (\text{ノードの尤度}) * (\text{yesの場合に残る葉の数}) + \\ (1 - (\text{ノードの尤度})) * (\text{noの場合に残る葉の数}) \end{aligned}$$

## 4 評価実験

### 4.1 システムの仕様及び実験条件

本システムで使用したマニュアルと音声認識システムの構成と、評価実験の条件について説明する。

ヘルプシステムの知識源であるマニュアルにはビデオデッキマニュアルを用いた。ページ数は111で、検索の結果出力する操作項目(例:「再生する」「録画する」)の総数は47である。そこから作成したマニュアルの木構造の深さは3、キーワード(操作ページのヘッダから抜き出した自立語)の総数は137である。認識エンジンは本研究室で開発した有限状態文法に基づくJulian[8]を用いた。認識語彙数は241で、文法は人手で記述した。音響モデルは性別依存PTMモデルを用いた。

実験は、まず想定場面を設定し(「ビデオを借りてきた」、「予約の時間を間違えた」等)、その状況に基

表 1: テキスト入力による評価

マッチングによる平均候補数	13.4
正解の平均順位	3.5
成功数	146/157 (93%)
提案手法による平均質問回数	2.5

づく自発的な質問を10発話収録した。次に何の設定もない自由な質問を2~3発話収録した。以上を14人(男10、女4)の被験者に対して行い、合計195発話を得た。そのうち当タスクで回答可能な質問である157発話に対して評価を行った。

システムの評価尺度として、まずこの157発話のうち、いくつの質問に対して正解項目を検索できるか(成功数)を調べた。次にシステムが正解項目を得るまでに行った(平均)質問回数を求めて、マッチングによる尤度の大きい項目から順に確認発話を行うシステムの質問回数(正解の順位)と比較した。

### 4.2 テキスト入力における評価

まず、入りに音声を用いず、書き起こしテキストから手動でキーワードを抽出してシステムを動作させ、マッチングと絞り込み部分の評価を行った。表1にテキスト入力によるシステムの評価を示す。「候補数」はマッチングの結果得られた候補の数を表す。候補中に正解項目が出現した質問が146発話で、約93%(146/157)に対して正解である項目にたどりつける。「順位」はマッチングの結果得られた候補の中での正解項目の尤度の順位を表す。正解が候補中に存在しない場合には、「候補数」を「順位」としている。この順位は検索結果を尤度の高い候補から順に確認する場合の質問回数に相当し、平均値は3.5であった。これに対して提案する対話戦略による平均質問回数は2.5であり、尤度の高い候補から順に確認する方法に比べて、質問回数が平均約1回少なくてすむことになる。質問関数1、2による質問回数の比較では、両者に全く差が見られなかった。

表 2: 音声入力における正解項目の順位と候補数

	信頼度あり	信頼度なし
マッチングによる平均候補数	13.4	
正解の平均順位	4.2	4.4
成功数	136/157 (87%)	

表 3: キーワードの信頼度と単語正解率

信頼度	1	1-0.9	0.9-0.8	0.8-0.7	0.7-	合計
正解単語数	279	15	10	18	16	338
不正解単語数	63	17	20	49	60	209
正解率	82%	47%	33%	27%	21%	62%

### 4.3 音声入力における評価

次に音声認識を適用した場合で評価した。表 2 に入力音声に信頼度を付加した場合としない場合の候補数と平均順位、成功数を示す。これより、音声入力を用いた場合でも約 87% に対して正解項目に到達できることがわかる。また平均順位の比較から、信頼度を付加した方がマッチングの精度がわずかに良い。キーワードの信頼度とその正解率（正解数 / 出現数）の分布を表 3 に示す。信頼度が 1 のとき単語正解率は 82% で、信頼度が音声認識結果の信頼性を表す尺度として機能している。

表 4 に提案する対話戦略によるシステムの質問回数を示す。「質問関数」は 3.3.2 節で述べた質問の評価関数 1、2 をそれぞれ表す。表 4 の平均質問回数と平均順位を比較すると、本システムにおける対話戦略により、尤度の高いものから順に質問する戦略と比べて質問回数が 66% までに削減できることがわかる。なお二つの評価関数による平均質問回数にほとんど差がないが、質問関数 2 の方が若干良かった。

## 5 結論

本研究では、大規模な電気機器マニュアルに音声対話インターフェースを適用することによって、よりユーザフレンドリなヘルプシステムの構築を行った。音声によるヘルプシステムを構築する際、検索において候補が複数出現した場合の出力の方法が問

表 4: 音声入力によるシステムの質問回数

	提案手法による平均質問回数	正解の平均順位
質問関数 1	2.9	4.2
質問関数 2	2.8	

題となるが、マニュアルの木構造を利用して、対話のコストが最小となるように上位概念を用いた質問の生成を行うことで、より少ない質問回数でユーザの要求している項目を見つけ出せる。また、音声認識結果の信頼度をマッチングの際に考慮することによって、誤認識による影響を抑えた。評価実験の結果から、提案した対話戦略が、尤度の高い候補から順に質問する対話戦略に比べて、質問回数を約 66% に削減できた。これにより提案した対話戦略がマッチングの結果得られた候補の絞り込みに有効であることが示された。

## 参考文献

- [1] 長田誠也, 土井伸一, 亀井真一郎. 自然言語を用いて家庭機器操作を行う対話システム. 人工知能学会研究会資料, SIG-SLUD-9802-4, 1998.
- [2] 今井裕之, 本田大介, 荒木雅弘, 堂下修司. 遠隔講義機器設定タスクにおける発話内容の抽象化. 情報処理学会研究報告, 99-SLP-25-14, 1999.
- [3] 亀田弘之, 藤崎博也. 情報検索における音声・言語処理. 情報処理学会研究報告, 97-SLP-16-11, 1997.
- [4] 駒谷和範, 河原達也. 音声認識結果の信頼度を用いた頑健な混合主導対話の実現法. 情報処理学会研究報告, 2000-SLP-30-9, 2000.
- [5] 日笠亘, 藤井綱貴, 黒橋禎夫. 入力質問と知識表現の柔軟なマッチングによる対話ヘルプシステムの構築. 情報処理学会研究報告, 99-NL-134, 1999.
- [6] 長尾真(編). 自然言語処理, pp. 416-421. 岩波書店, 1996.
- [7] 安田圭志, 堂坂浩二, 相川清明. タスク適応型高効率対話制御法. 情報処理学会研究報告, 2001-SLP-35-13, 2001.
- [8] 李晃伸, 河原達也, 堂下修司. 文法カテゴリ対制約を用いた A\*探索に基づく大語彙連続音声認識バ - ザ. 情報処理学会論文誌, 第 40 巻 第 4 号 1374-1382, pp. 1374-1382, 1999.