1D1-07

状況と文脈に基づいて発話を解釈するサービスロボット用対話システム

A Dialog System for Service Robot Interpreting Utterance Based on State and Context

滝澤　正夫　　槙原　靖　　島田　伸敬　　白井　良明
Masao Takizawa　Yasushi Makihra　Nobutaka Shimada　Yoshiaki Shirai

大阪大学大学院工学研究科電子制御機械工学専攻
Dept. of Computer-Controlled Mechanical Systems, Osaka University

This paper describes a dialog system for a service robot which interprets user's utterances by estimating the meanings of unknown words in case of speech recognition failure or unexpected utterance. In this paper, we define "unknown words" as the registered words which are recognized by mistake and synonyms of the registered words. The system estimates which registered word they correspond to, considering the state, the context (i.e. the preceding and following words) and the pronunciation similarity between the unknown words and the registered words.

1. はじめに

近年，高齢化社会の到来により，指定された物体を取ってくるサービスロボットが注目されている．我々は，冷蔵庫から指定した飲み物 (缶，瓶，ペットボトル) を取ってくるサービスロボットを開発している．そのようなロボットのインターフェイスには音声がふさわしいが，認識の誤りやシステムが想定していない発話をうまく処理できなければ実用的ではない．本論文では，ユーザの発話中に未知語があった場合にも，未知語の意味を推定してユーザの発話を解釈する対話システムについて述べる．未知語の推定には，単語の構成文字と文脈を利用して品詞を推定する研究 ([Nagata 99]) があるが，品詞の推定だけではロボットはタスクを達成できない．本論文では未知語を「登録語の同義語」「誤認識された登録語」の 2 種類と定義し，文中の未知語が 1 語の場合，2 語連続している場合について，状況，文脈を考慮してどの登録語と等しいかを推定する．

2. 音声処理

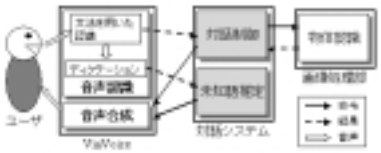


図 1: システム構成

本研究のシステム構成を図 1 に示す．音声認識には，IBM 社の ViaVoice に含まれる次の 2 つの認識エンジンを用いる．

- 文脈自由文法をサポートしたエンジン (以下，CFG)
  あらかじめ定義しておいた文法に一致した場合のみ認識可能．登録語の認識率は高い．
- ディクテーションをサポートしたエンジン (以下，DIC)
  入力された音声を文字列に変換する．新聞などの文章には強いが，短い単語や話し言葉の認識率は低い．

まず，CFG で認識を行う．文法に一致した場合はその結果を画像処理部に渡して物体認識を行う．ユーザがあらかじめ定義してある文法以外の言い方をした場合や認識がうまくいかなかった場合には，DIC で音声を文字列に変換する．システムはこの文字列と，状況，文脈ならびに CFG の文法を用いて，未知語の意味の推定を行う．物体認識と対話制御については [Makihara 2002] を参照のこと．

連絡先: 島田　伸敬，大阪大学大学院工学研究科電子制御機械工学専攻，〒 565-0871 大阪府吹田市山田丘 2-1，shimada@eng.osaka-u.ac.jp

CFG の文法は助詞以外の単語カテゴリと助詞の順列で定義される．各カテゴリには必要な単語を事前に登録しておく（表 1）．カテゴリに登録された単語を以下登録語と呼ぶ．例えば「＜商品名＞の＜相対位置＞の＜形状＞を＜他動詞＞」(＜＞内はカテゴリを表す) という文法を定義すると,「ダカラの左の缶を取って」などの文章が認識可能である．

表 1: カテゴリと登録語の例

カテゴリ	登録語
商品名	飲料の商品名 (ダカラ，コーラ，など)
形状	缶，瓶，ペットボトル
相対位置	右，左，後ろ，上，下，真ん中
他動詞	取って，とけて，入れて，見せて

3. 未知語の推定

未知語の推定手順は以下のとおりである．

1. ディクテーション文字列から未知語部分を検出する．
2. 状況，文脈，ディクテーション文字列を考慮して，各登録語の確率を計算する．
3. 最大の確率を持つ推定結果を採用する．

2. では未知語が，a-1) 登録語の同義語，a-2) 誤認識された登録語，である場合にわけ，さらにそれぞれの場合を，未知語部分が，b-1)1 語からなる場合，b-2)2 語からなる場合，にわけ，合計 4 通りの場合について確率を計算する．全ての確率が閾値以下の場合は，雑音であると推定する．未知語が 3 語以上連続することもありうるが，計算量が著しく増加する上正確な推定が困難であるので，雑音とされた場合には不明な部分をユーザに質問する．

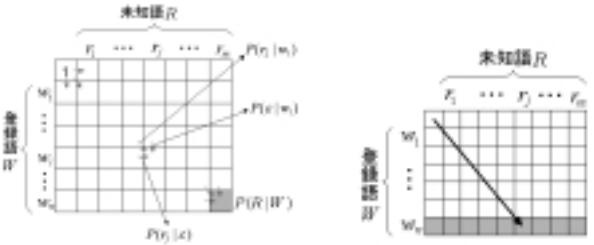
3.1 未知語の検出

表 2: 「のほほん茶の左」に対する認識結果
(下線部は登録語，括弧内は発音，ー は代替解釈がないことを表す)

第一位の認識結果	代替解釈
五 (ご)	の (の)，同 (どう)，後 (ご)
本 (ほん)	今 (こん)，黄金 (おうごん)
茶 (ちゃ)	地 (ち)，著 (ちょ)，所 (じょ)
狼 (ろう)	の (の)
左 (ひだり)	ー

ディクテーションで音声認識をすると，表 2 のように第一位の認識結果と代替解釈のリストを取得できる．これらから登録語を検出し，どちらからも登録語が検出されないものを未登録語とする．表 2 の場合,「五」「本」「茶」が未登録語である．「五」に関しては代替解釈の「の」が登録語であるが，助詞が最初に発話されることはないので，未登録語として扱う (未登録語にはさまれた登録語も未登録語とする)．未登録語をまとめたものを未知語部分とする．表 2 の例では「五 本 茶」が未知語部分である．

図 2: $P(C|S, cont)$ の計算方法



(a) 計算方法　　(b) 未知語の分割

図 3: 発音類似度

3.2 状況

ユーザの発話がどのような状況でなされたかを考慮すれば，未知語部分がどのカテゴリの単語であるかを限定することができる．本研究では，以下のような状況を用いている．

- 最初にユーザが発話する状況
- ユーザの発話に対して質問している状況
  「何を取りますか?」「何色の物体ですか?」など
- 認識結果に対して質問している状況
  「見つかりませんでした．どのへんにありますか?」「1 個見つかりました．これを取りますか?」など

本論文では状況ごとにカテゴリや単語の発生しやすさを条件付き確率として学習しておく．

3.3 確率に基づく単語推定：1 語の場合

登録語の同義語であると仮定して推定する場合は，両者の発音は似ていないと考えられるので，状況 S，文脈 $cont$ だけを考慮した登録語 W の事後確率を最大にするものを求める (式 (1))．全ての W について計算すると時間がかかるので，式式 (1) を (2) のように分解し，$P(C|S, cont)$ (C はカテゴリ) が閾値を越えるカテゴリに絞って式 (2) を計算する．

$$\hat{W} = \arg\max_{W} P(W|S, cont) \quad (1)$$

$$P(W|S, cont) = P(C|S, cont)P(W|S, cont, C) \quad (2)$$

また，誤認識された登録語と仮定して推定する場合は，ディクテーション文字列 R に発音が似た単語の可能性が高いので，以下のようにする．

$$\begin{aligned}\hat{W} &= \arg\max_{W} P(W|S, cont, R) \\ &= \arg\max_{W} \frac{P(W|S, cont)P(R|S, cont, W)}{\sum_W P(W, R|S, cont)} \end{aligned} \quad (3)$$

ここでも，式 (2) を用いてカテゴリを絞って推定する．$P(C|S, cont)$ は以下のようにして計算する．

$$P(C|S, cont) \simeq P(C|C_p, C_n, S) = \frac{P(C_p, C, C_n|S)}{\sum_C P(C_p, C, C_n|S)} \quad (4)$$

$$P(C_p, C, C_n|S) \simeq P_{c-s}(C|S)P_{c-p}(C_p|C)P_{c-n}(C_n|C) \quad (5)$$

$P(W, R|S, cont)$ についても同様の計算を行う．

$P(R|S, cont, W)$ は $P(R|S, cont, W) \simeq P(R|W)$(発音類似度) と近似し，以下のように計算する ([Ristad 98])．

$$\begin{aligned} P(R|W) &= P(r_1, r_2, \cdots, r_m|w_1, w_2, \cdots, w_n) \\ &= P(\epsilon, r_1, \cdots, r_m|w_1, w_2, \cdots, w_n) \\ &\quad + P(r_1, \epsilon, \cdots, r_m|w_1, w_2, \cdots, w_n) + \cdots \\ &\simeq P(\epsilon|w_1)P(r_1|w_2)\cdots P(r_m|w_n) \\ &\quad + P(r_1|w_1)P(\epsilon|w_2)\cdots P(r_m|w_n) + \cdots \end{aligned}$$

表 3: 成功例 (下線部は未知語)

	成功例 1	成功例 2
発話	ダカラちょうだい	青いペットボトルを取って
認識	だから ちょうだい	おもい くう とらぶる を とって
推定	ダカラ 取って	青い ペットボトル を 取って

2,3 行目では空文字列 ϵ を複数挿入して文字数を一致させることで W と R の各文字の対応付けを行い (ϵ 同士のペアは許さない)，全ての対応に対する確率の総和を計算している．これらの計算は動的計画法により効率よく計算ができる (図 3)．

3.4 確率に基づく単語推定：2 語の場合

登録語の同義語，誤認識された登録語と仮定して推定する場合は，それぞれ以下のように 1 語の場合の W を W_1, W_2 で置き換えたものを計算する．

$$(\hat{W}_1, \hat{W}_2) = \arg\max_{W_1, W_2} P(W_1, W_2|S, cont) \quad (6)$$

$$(\hat{W}_1, \hat{W}_2) = \arg\max_{W_1, W_2} P(W_1, W_2|S, cont, R) \quad (7)$$

(W_1, W_2) の組み合わせ数は W_1,W_2 の数の積に比例するので，高速化を必要とする．そこで，図 3(b) のように発音類似度の計算の一番下の行で最大かつ閾値を越えるところで未知語を分割し，残りの未知語部分を 1 語の場合と同様に推定する．計算量は W_1,W_2 の数の和のオーダとなる．

4. 実験

未知語の推定実験を行ったところ，57 発話中 53 例について推定が成功した (推定率 92.9%)．成功例を表 3 に示す．また，冷蔵庫画像 10 枚を研究室内の被験者に見せて発話してもらい，本システムの発話解釈の評価を行った．その結果，文法を用いた認識のみでは 54.5%しかユーザの発話を解釈できなかったのに対し，未知語推定の機能を備えた本システムでは 68.2%まで向上した．

5. まとめと今後の課題

本論文では，音声認識の誤認識や，システムの想定外の発話があった場合に，未知語の意味を推定してユーザの発話を解釈する対話システムについて述べた．また，この対話システムを，指定された飲料を持ってくるサービスロボットに適用し，冷蔵庫画像を用いた実験により本手法の有効性を示した．

現在は検出した登録語は正しいとしているが，間違った登録語が検出されることもあるので，検出した登録語の信頼性を評価する必要がある．今後の課題としては，未知語が推定できなかった場合や情報が不足している場合にユーザから自然に情報を引き出すための対話生成手法を検討することがあげられる．

参考文献

[Nagata 99] Nagata, M.: "A Part of Speech Estimation Method for Japanese Unknown Words using a Statistical Model of Morphology and Context", 37th Annual Meeting of the Association for Computational Linguistics, pp.277-284, 1999.

[Ristad 98] Ristad, E. S. and Yianilos, P. N.: "Learning String-Edit Distance", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 20, No. 5, pp. 522-532, 1998.

[Makihara 2002] Makihara, Y., Takizawa, M., Shirai, Y., Miura, J., Shimada, N.: "Object Recognition Supported by User Interaction for Service Robots", Proc. of 5th ACCV, Vol.2, pp. 719-724, 2002.