

音声対話システムにおける文法検証結果と発話履歴に基づくヘルプメッセージ候補のランキング

駒谷 和 範^{†1} 池田 智 志^{†1} 福林 雄 一 朗^{†1}
尾形 哲 也^{†1} 奥 乃 博^{†1}

本稿では、ヘルプメッセージを提示することにより、音声対話システムにおける想定外発話の問題に取り組む。想定外発話に対する音声認識結果では、重要な単語が誤認識されていたり、ユーザが十分な情報を発話しない場合があるため、想定外発話に対するヘルプメッセージ生成は困難な課題である。本稿ではまず、重みつき有限状態トランスデューサ (WFST) を用いて、想定外発話に対する文法検証手法を開発した。この文法検証では、音声認識結果中の重要な単語が抜けている場合でも、ユーザが意図した表現に最も近い文法を同定する。さらに、この文法検証結果やユーザの経験を表す対話履歴を特徴として用いて、RankBoost アルゴリズムによりヘルプメッセージ候補の順序付けを行った。

Ranking Help Message Candidates Based on Robust Grammar Verification Results and Utterance History

KAZUNORI KOMATANI,^{†1} SATOSHI IKEDA,^{†1}
YUICHIRO FUKUBAYASHI,^{†1} TETSUYA OGATA^{†1}
and HIROSHI G. OKUNO ^{†1}

We address an issue of out-of-grammar (OOG) utterances in spoken dialogue systems by generating help messages for novice users. Help generation for OOG utterances is a challenging problem because language understanding (LU) results based on automatic speech recognition (ASR) results for such utterances are always erroneous as important words are often misrecognized or missed from such utterances. We first develop grammar verification for OOG utterances by using a Weighted Finite-State Transducer (WFST). It robustly identifies a grammar rule that a user intends to utter, even when some important words are missed from the ASR result. We then adopt a ranking algorithm, RankBoost, whose features include the grammar verification results and the utterance history representing the user's experience.

1. はじめに

音声対話システムの研究は、研究室内で試用するシステムから、一般市民に公開されるシステムへと進んでおり^{7),10),11)}、一般市民が音声対話システムを使用する機会が増加している。この際、初心者ユーザは、従来の研究室内で実験を行う際のユーザとは異なり、一切事前教示を受けずにシステムを使用する機会が多い。このためしばしばシステムの能力を超える発話を行い、自分の発話が正しく認識されない状況を体験する。これはシステムに関するユーザの期待と実際のシステムとの間にギャップがあり、ユーザの発話がシステムの能力を超えることに起因する。本稿では、システムが受理解釈可能な範囲を超えたユーザの発話を想定外発話 (out-of-grammar (OOG) utterances) と呼ぶ。ユーザの発話が想定外であった場合には、ユーザが再発話しても音声認識誤りが繰り返される。どのような発話がシステムに受理されるのかわからない場合には、ユーザは発話表現を改めることができない。

本研究の目的は、想定外発話に対処するために、ユーザのタスク遂行に必要な言語表現をヘルプメッセージとして提示することである。例えば、ユーザが「観光施設の情報を知りたい」という意図を持っている場合、「施設情報を調べるには、施設名と項目名を指定して、『清水寺の電話番号を教えてください』などと言ってください。」のようなヘルプメッセージの提示を行う。このヘルプメッセージは、システムの持つ各文法ルールに対応させて用意する。このため、ユーザの意図に対応する文法ルールを推定できれば、適切なヘルプメッセージが提示できるとして議論を進める。ヘルプメッセージを提示するための課題は以下の2点を含む。

- (1) 想定外発話に対しても頑健なユーザ意図の推定
- (2) ユーザにとって既知のヘルプメッセージの繰り返しの抑制

課題の1点目は、システムにとって主要な入力である音声認識結果が、想定外発話に対しては誤りである場合が多いことに対応する。このため特に内容語（検索対象のデータベースに含まれる地名などに対応する）が正しく認識されない場合には、ユーザの意図に近い文法ルールを推定するのも難しくなる。このため、想定外発話に対する音声認識結果中の、任意の位置に音声認識誤りがあっても動作する必要がある。一方第2の課題は、想定外発話の音声認識結果には欠損があることも多く、複数のターンを通じて正しいヘルプメッセージ

^{†1} 京都大学大学院情報学研究所
Graduate School of Informatics, Kyoto University

を提示する機会が多いため、必要となる。この原因は、音声認識誤りとともに、ユーザが発話の一部を省略し短い発話を行うことにもある。複数のヘルプメッセージを複数のターンにわたって提示する場合、ユーザが既知っている内容のヘルプメッセージを繰り返すのは無駄である。

本研究ではまず、第1の課題に対応するために、Weighted Finite State Transducer (WFST) に基づく文法検証を行う。文法検証では、文法から外れた発話に対しても最も近い文法ルールを頑健に同定する。ここで用いられる WFST は、あらゆる位置に誤りがある可能性を考慮したうえで、音声認識結果から動的に生成する。さらに第2の課題への対処として、ブースティングアルゴリズムのひとつである RankBoost¹⁾ を用いて、ヘルプメッセージの候補を尤度順に並べ替え、順に提示することを考える。特にユーザの発話が短い場合などでは、学習のための正解ラベルとして、ヘルプメッセージを一意に決めることが人手でも難しいため、優先順位を与えた複数のラベルから学習を行えるこのアルゴリズムを採用する。さらに、当該発話のみから情報を得るのではなく、既知度²⁾ や対話文脈に対応する特徴を加えることで、ユーザ意図推定の高精度化を図る。

2. 関連研究

音声対話システムにおけるヘルプメッセージ生成はこれまでも研究が行われてきた。Gorrell らは、決定木学習を用いて想定外発話に対するヘルプメッセージの生成を行った³⁾。Hockey らは、2種類の音声認識結果を比較することで、想定外発話を endpointing errors, unknown vocabulary, subcategorization mistakes の3カテゴリに分類した³⁾。これらは Targeted Help と呼ばれ、ユーザの一発話に対してフィードバックを行う。Lee らは、例に基づく対話管理モデルにおいてヘルプメッセージを生成し、エラーからの回復を図った⁹⁾。これらの研究はいずれも、主に音声認識結果の文字列に基づいて、提示されるヘルプメッセージが決定される。このため、ユーザの発話が短い場合など、音声認識結果から情報が欠落している場合は、適切なヘルプメッセージを提示できない。また同じ音声認識結果が得られた場合には、同じヘルプメッセージが繰り返されることになる。

またヘルプメッセージ候補のランキングは、音声入力を備えたマルチモーダルインタフェースでも有効である。ヘルプメッセージは、音声を入力として用いる場合には必要だからである。例えば、City Browser⁴⁾ というマルチモーダルインタフェースでは、音声認識結果とシステムの内部状態に応じて、画面上にテンプレートに基づくヘルプメッセージが表示される。特に携帯機器の小さい画面を用いる場合には、画面に一度に表示できるヘルプメッセー

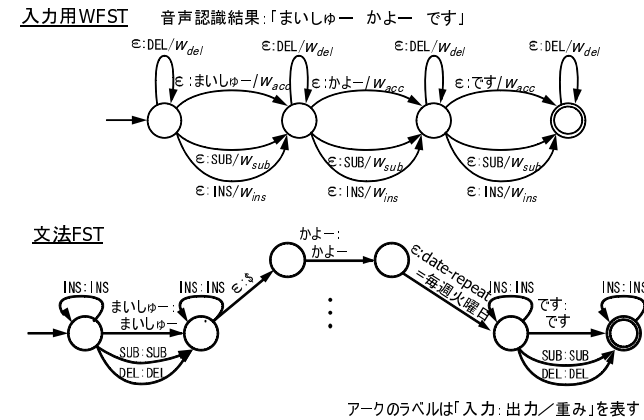


図1 入力用 WFST と文法構造を表す FST の例 (概略)

ジの数が限られるため⁵⁾、ヘルプメッセージの表示順序は重要である。大きな画面が使える場合であっても、順序なく多くのヘルプメッセージを表示するのは、ユーザの注意を散漫にしユーザビリティの低下を招く。

3. WFST に基づく文法検証

本稿では、ユーザの意図として、ユーザが発話で使おうとした表現に対応する文法ルールを、想定外発話からでも頑健に推定する。この推定を文法検証と呼ぶ。この際、タスク文法を表す finite-state transducer (FST) と、音声認識結果とその信頼度を表す weighted FST (WFST) を用いる。これら2つを以降ではそれぞれ文法 FST、入力用 WFST と呼ぶ。これらの例を図1に示す。

本稿での文法検証手法の利点は、各種の音声認識誤りを考慮している点である。つまり、入力用 WFST が、音声認識結果中のあらゆる単語が挿入または置換誤りであったり、任意の単語が削除誤りとなっている場合を表現している。WFST の重みは音声認識結果の信頼度を用いて算出する。入力用 WFST と文法 FST の合成により、入力用 WFST から出力されるあらゆる記号列が、文法 FST で受理される際の、全ての可能な出力系列とその時の累積重みが得られる。この際、最大の累積重みを持つ言語理解結果とそれに対応する文法ルールが、出力として得られる。入力用 WFST で各種誤りを考慮しているため、音声認識結果

に誤りが含まれていたり重要な単語が抜けている場合でも、音声認識結果に最も近い文法を得ることができる。

言語理解結果は、システムの応答生成に必要なスロットとその値の組の集合である。例えば、[month=2, day=22] という言語理解結果は、スロット month の値が 2、スロット day の値が 22 という組により構成される。

3.1 入力用 WFST と文法 FST の設計

図 1 の入力用 WFST と文法 FST において、各遷移のラベルは「入力単語:出力単語/重み」を表している。入力単語における ε は入力なしでの遷移、つまり ε 遷移、出力単語における ε は出力がないことを意味する。例えば「です: ε /1.0」では、「です」が入力されると出力なしで遷移し、累積重みに 1.0 が加算される。文法 FST では遷移に重みはないので重みは省略されている。

入力用 WFST は音声認識結果から自動的に、図 1 のように各単語に対応した遷移と FILLER 遷移が並列となるように生成される。INS, DEL, SUB といった FILLER 遷移が各状態に割り当てられ、挿入、削除、置換誤りの各種誤りを表す。入力用 WFST はすべて ε 遷移で構成され、これにより任意の誤りを含む全ての単語列を WFST が表現できる。例えば、図 1 の入力用 WFST は「まいしゅー かよー です」という音声認識結果に対して生成され、「まいしゅー かよー です」や「まいしゅー FILLER FILLER」や「FILLER かよー FILLER」など、まず $8(=2^3)$ 通りの入力列を表す。ここで FILLER は挿入または置換誤りを表す。さらに、誤りシンボル DEL が出力シンボル列の任意の場所に挿入され、音声認識結果の削除誤りの可能性を表現する。状態遷移ごとのそれぞれの重みは累積され、最適な結果が決定される。それぞれの重みについては 3.2 節で述べる。

文法 FST は、システム開発者が記述したタスク文法から生成される。この FST は、入力された系列がタスク文法に沿っているかどうかを判定する。また、文法 FST の各状態に対して FILLER 遷移を用意し、入力 WFST で導入された各種の音声認識誤りを取り扱う。FILLER 遷移は、INS, DEL, SUB のいずれかで、システム開発者が指定したキーフレーズ内に相当する状態を除く各状態に付加される。図 1 の例では「SUB かよー です」が入力されると「SUB \$ かよー date-repeat=毎週火曜日 です」が出力される。ここで \$ はスロット-値の組に対応する単語の範囲を特定するための記号で、「date-repeat=毎週火曜日」は言語理解結果である。

3.2 入力用 WFST に与えられる重み

入力用 WFST に対する重みづけとして以下の 2 種類を用意する。

(1) 受理単語 w_{acc} に対する報酬

(2) 各種誤り ($w_{sub}, w_{del}, w_{ins}$) に対するペナルティ

累積重み w は以下の式のように、受理単語に対する重み、文法間違いに対する重みの和で計算する。

$$w = \sum_{E_{accepted}} w_{acc} + \sum_{E_{error}} (w_{sub} + w_{del} + w_{ins})$$

ここで、 $E_{accepted}$ は文法の要素とのアライメントがとれた(受理された)単語の集合、 E_{error} はアライメントの結果文法の要素に対応せず、誤りとされた単語の集合である。なおこれらの重みは事前に与えられておらず、実行時に音声認識結果に基づいて計算され、入力用 WFST に与えられる。受理された単語 e_{asr} に対する重みは、その単語の信頼度 $CM(e_{asr})$ ⁽⁸⁾ と単語の長さを用いて定義する。単語の長さは語彙内の最長単語で正規化したモーラ長とし、これを $l(\cdot)$ で表す。

$$w_{acc} = CM(e_{asr})l(e_{asr})$$

この重み w_{acc} は、長く、かつ音声認識結果が信頼できる単語に、より大きな重みを与える。誤りに対する重みはペナルティと考え負の値を与える。具体的には以下である。

$$w_{sub} = -\{CM(e_{asr})l(e_{asr}) + l(e_{gram})\}/2$$

$$w_{del} = -\{\overline{l(e)} + l(e_{gram})\}/2$$

$$w_{ins} = -\{CM(e_{asr})l(e_{asr}) + \overline{l(e)}\}/2$$

ここで、 e_{gram} は文法中の要素、つまり単語クラスであり、 $\overline{l(e)}$ は語彙内の平均モーラ長である。置換誤りは、文法中の要素が音声認識結果中の他の単語で置き換えられた場合である。削除誤りは、文法中の要素が音声認識結果中の単語に対応しなかった場合である。挿入誤りは、音声認識結果の単語に対応する文法要素がない場合である。重みは全て、文法中の要素の長さ、音声認識結果中の対応する単語の長さに信頼度を乗じたものの平均として定義されている。挿入および削除誤りで、対応する要素が定められない場合には、 $\overline{l(e)}$ を用いる。文法中の要素 e_{gram} がクラスの場合には、当該クラスの要素の平均単語長をその長さとして用いる。

3.3 累積重みの計算例

累積重み w の計算例を図 2 で説明する。この例では、「光悦寺の連絡先(未知語)を教えてください」というユーザ発話に対して「あ 光悦寺 えと 左京区 教えて 下さい」という音声認識結果が得られている。この音声認識結果と、システムが持つ 2 つの文法「〈寺社名〉の〈項目名〉を 教えて 下さい」および「〈区名〉の〈施設タイプ〉を 教えて 下さい」と

書き起こし: 「光悦寺の 連絡先 (未知語) を教えてください。」

音声認識結果	あ	こーえつじ	えと	さきょーく	おしえて ください	累積重み	
想定文法		〈 寺社名 〉	の	〈 項目名 〉	を	おしえて ください	
WFST 出力	INS	こーえつじ	DEL	SUB	SUB	おしえて ください	
重み	-0.13	+0.18	-0.14	-0.13	-0.04	+0.18	+0.18
想定文法		〈 寺社名 〉	の	〈 項目名 〉	を	おしえて ください	
WFST 出力		SUB	SUB	SUB	SUB	おしえて ください	
重み	-0.13	-0.11		-0.13	-0.04	+0.18	+0.18
想定文法		〈 区名 〉	の	〈 施設タイプ 〉	を	おしえて ください	
WFST 出力	INS	SUB	DEL	SUB	SUB	おしえて ください	
重み	-0.13	-0.21	-0.14	-0.13	-0.04	+0.18	+0.18

図 2 重みの計算例

の間での累積重み w の計算例を示している。重みはシステムの持つ各文法ルールごとに計算されるが、ここでは例として上記の 2 つの文法に対する重みの計算を示している。前者の文法ルールについてはさらに 2 通りのアライメントの例が示されている。なお $\langle \rangle$ は文法中のクラスを表す。

重みはあらゆるアライメント結果に対して計算され、最大の累積重みを持つ文法ルールとそのアライメント結果が最終的に得られる。この例では、「〈 寺社名 〉の 〈 項目名 〉を 教えてください」に対する重みが +0.10 と最大であり、この場合の文法中の各要素と音声認識結果のアライメント結果も同時に得られる。

4. 優先度学習に基づくヘルプ候補順序付け

本章では、文法検証結果と対話の履歴などを統合し、ヘルプメッセージ候補を順序付ける手法について述べる。これにより、音声認識誤りやユーザの短い発話により欠損している情報を補い、同じヘルプメッセージが繰り返されるのを防ぐ。

4.1 順序付けに用いる特徴

本研究で用いた特徴を表 1 に示す。これらの特徴は、ヘルプメッセージの候補それぞれに対して計算される。H1 から H5 は、WFST に基づく文法検証結果がどれだけ信頼できるかを表す。H1 は文法検証のスコアである。H2 は、全ての候補の文法検証スコアの総和で H1 を正規化することで得る。すなわち、この文法検証結果が他の候補に比べて相対的にどれだけ信頼できるかを表す。H3 から H5 は、ユーザ発話が部分的にどれだけ文法に一致しているかを表す。

H6 は、対話文脈に相当する。発話が受理されない時、ユーザは同様の発話を繰り返す性

表 1 各候補ごとに使用した特徴

H1:	WFST に基づく文法検証スコア
H2:	文法検証スコアの事後確率値
H3:	音声認識結果中の受理単語の割合
H4:	文法検証結果中における最大連続受理単語数
H5:	文法検証結果の受理スロット数
H6:	当該文法ルールに沿った発話が受理されたのは何発話前か
H7:	当該文法ルールに対する現発話までの最大 WFST スコア
H8:	当該文法ルールがコマンド発話かどうか
H9:	当該文法ルールが検索条件を指定する発話かどうか
H10:	当該文法ルールが検索結果中の情報を取得する発話かどうか
H11-H17:	H8 と、H1 から H7 のそれぞれとの積
H18-H24:	H9 と、H1 から H7 のそれぞれとの積
H25-H31:	H10 と、H1 から H7 のそれぞれとの積

質があるため、定義した。H7 は当該ユーザがその文法ルールに対応する言語表現を知っているかどうかを表し、文献 2) の既知度に相当する。これはユーザが既に知っている文法ルールに対するヘルプメッセージが何度も提示されるのを抑制するために定義した。

H8 から H10 はそれぞれの文法が表す発話の種類 (コマンド、検索条件の指定、検索結果中の情報取得) を表す。一例として観光ドメインでは、H8 から H10 に属する文法ルールの数はそれぞれ 8, 4, 11 である。検索結果の情報を取得する発話や検索条件を指定する発話は連続して行われることが多い一方、コマンド発話は文脈に独立に発話されることが多いなど、発話の種類ごとに類似した性質がある。H11 から H31 は、H8 ~ H10 と、H1 から H7 までのそれぞれの積をとったもので、RankBoost では特徴の組み合わせが考慮されないため明示的に定義した。

4.2 RankBoost アルゴリズムの概略

優先度の学習には RankBoost アルゴリズム¹⁾を用いた。このアルゴリズムは、優先順位の一部を示すインスタンス x の集合から、その順序を学習するのに用いられる。本研究で対象とするヘルプメッセージは、特にユーザ発話が短い場合には人手でも唯一に正解を定めるのが難しいため、候補となる複数のヘルプメッセージに順序をつけて正解ラベルとしている。このためこのアルゴリズムは本研究での問題設定に適している。アルゴリズムの整列対象とするインスタンス x は、文法ルールに対応するヘルプメッセージの候補とする。

RankBoost は、候補を順序付けるスコア関数 $H(x)$ をブースティングを用いて学習し、スコア関数 $H(x)$ の値に従って対象を整列することで候補の順序付けを行う。ここで、 $H(x') < H(x'')$ は x'' が x' より上位にランクされることを意味する。また、 $H(x)$ は

順序付けの部分的な情報を与える弱順序付け器を線形結合することで学習される．具体的には以下の式で表される¹⁾．

$$H(x) = \sum_t^T \alpha_t h_t(x)$$

ここで T はブースティング回数， α_t は結合重み， h_t は弱順序付け器， x はランキングの候補である．本研究では，弱順序付け器 h_t は各特徴の値を閾値処理することで得た．具体的には，以下の式で表される¹⁾．

$$h_t(x) = \begin{cases} 1 & \text{if } f_i(x) > \theta \\ 0 & \text{if } f_i(x) \leq \theta \\ q_{def} & \text{if } f_i(x) = \perp \end{cases} \quad (1)$$

ここで， $f_i(x)$ は候補 x の i 番目の特徴の値， \perp は特徴の値が得られなかったことを示す．また q_{def} は $\{0, 1\}$ である．

5. 評価実験

5.1 対象データ

マルチドメイン音声対話システムを用いて，合計 30 名の被験者から 180 対話，11,733 発話を収集した．このマルチドメインシステムは，レストラン，ホテル，観光，バス，天気 of 5 ドメインを扱うものである⁶⁾．このうち 5 名の被験者から得た発話データは，音声認識用言語モデルの学習とブースティング回数の決定のみに用いた．またバスドメインと天気ドメインは文法ルール数が少ないため，評価にはレストラン，ホテル，観光の 3 ドメインに対する発話を用いた．これらのうち文法検証の結果累積重みが負であった発話を想定外発話とみなし，本手法の評価対象とした．この結果得られた 25 名による 1,349 発話を，以降評価対象データとする．発話数の内訳は，レストランドメインが 363 発話，ホテルドメインが 563 発話，観光ドメインが 423 発話であった．データ収集時には，まずシステムの使用法や使用可能な言い回しに関する事前教示は全く与えず，システムの受理可能な言語表現が限られている旨のみを伝えた．システムからの出力は音声のみで，ディスプレイは用いなかった．達成すべきタスクとして 6 つのシナリオを与えた．メモの使用は許可した．

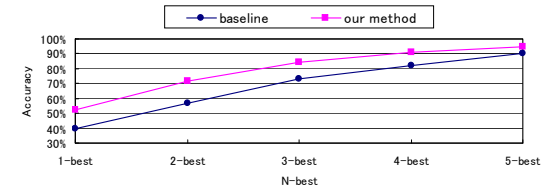


図 3 N-best 候補まで提示した場合に正解が含まれる割合 ($1 \leq N \leq 10$)

文法検証で使用する音声認識器には Julius^{*1} を用いた．言語モデルは認識文法から生成した例文 10,000 文と，評価に用いない 5 話者 600 発話の書き起こしから学習した 3-gram モデルを用いた．レストラン，ホテル，観光の各ドメインにおける語彙サイズは 3,456，2,625，3,593 であった．単語正解精度はそれぞれ 45.8%，57.1%，43.5% であった．単語正解精度が低いのは，評価対象データが全て想定外発話だからである．RankBoost の弱学習における閾値の候補 $\{\theta_j\}_{j=1}^J$ は，学習データに現れた全ての特徴量値とした．ブースティング回数は 5 名の被験者によるデータを用いて決定し，レストラン，ホテル，観光の各ドメインでそれぞれ 400，100，500 であった．

5.2 評価基準

ヘルプメッセージの正解は，ユーザの発話に関係の強いと思われる順に上位 5 つを手手で与えた．ヘルプメッセージの候補 (文法ルール) の数は，レストラン，ホテル，観光の各ドメインでそれぞれ 28，27，23 であった．

評価尺度は，ヘルプメッセージに対する順序付け結果中で，上位 N 番目の候補までに正解が含まれる発話の割合とした．これは， N 個のヘルプ候補を提示した場合に適切なヘルプメッセージが少なくとも一つ含まれる割合に相当する．この割合を 5-fold cross validation により計算した．ベースライン手法として，一発話のみから得られる文法検証スコアの値をそのままランキングスコアとして利用し，降順に整列したものを順序付け結果とする手法を用いた．

5.3 実験結果

観光ドメインにおける実験結果を図 3 に示す．全ての N の値において，本手法の精度がベースライン手法を上回った．これらの差は，マクマナー検定において統計的に有意 ($p < 0.05$)

*1 <http://julius.sourceforge.jp/>

表 2 各特徴に対する重み α の絶対値の総和

H7	H17 (H7*H8)	H19 (H2*H9)	H2	H6
9.58	6.91	6.61	6.02	6.01

であった。本手法の精度は、他の 2 つのドメインでも全ての N においてベースラインを上回り、その差の平均は $N=1$ で 11.7 ポイント、 $N=2$ で 9.7 ポイント、 $N=3$ で 6.7 ポイントであった。これらの差は、特に N が小さい場合に大きかった。以上の結果は、本手法により適切なヘルプメッセージを提示するまでに必要なヘルプ提示回数を少なくできることを示している。これらの改善は、本手法では文法検証結果に加えて、ユーザの知識推定結果に対応する特徴などを加えたことにより得られた。

順序付けの際に有効であった特徴を調査した。ここでは、RankBoost により選択された特徴に対する重み α の絶対値の総和を計算した。その値の上位 5 つの特徴を表 2 に示す。上位 5 つの特徴には、WFST に基づく文法検証結果から得られる特徴 (H2)、対話履歴を表す特徴 (H6)、既知度²⁾ に対応する特徴 (H7)、その文法の発話の性質を表す特徴 (H8, H9) が含まれている。この中で最も効果のあった特徴は、上位 2 つの特徴の両方に含まれている既知度 (H7) であった。これはユーザが既にその表現を知っている文法ルールは、以降の対話においても想定外発話となりにくいからであると考えられる。ある文法が高い文法検証スコアで受理された場合、ユーザがその表現を知っている可能性が高いとみなせる。次に有効であった特徴は、第 3 位と第 4 位に含まれる H2 であった。これは、WFST に基づく文法検証結果が有効に働いていたことを示している。

6. ま と め

本稿では、音声対話システムにおける想定外発話の問題に対して、ヘルプメッセージの生成による対処を目指した。ヘルプメッセージはシステムの持つ各文法ルールに対応して用意したうえで、ユーザの意図として、ユーザが使おうとしていた文法ルールを WFST を用いて推定した。さらに、文法検証結果に加えて様々な情報を統合して、ヘルプメッセージの候補を効率的に提示するためにランキングする手法を示した。

今後の課題として以下の各点が挙げられる。まず本稿での評価実験は、事前に収集された発話データに基づいて行われている。提示したヘルプメッセージが実際の対話でユーザにどのような影響を与えるかを、対話実験により確かめる必要がある。次に、本稿ではヘルプメッセージの表現は、各文法ルールごとに固定されており、単純にその文法ルールの例を示

すものとした。どのような表現が特に初心者ユーザにとって有用であるかも検証する必要がある。

参 考 文 献

- 1) Freund, Y., Iyer, R.D., Schapire, R.E. and Singer, Y.: An Efficient Boosting Algorithm for Combining Preferences, *Journal of Machine Learning Research*, Vol.4, pp.933–969 (2003).
- 2) Fukubayashi, Y., Komatani, K., Ogata, T. and Okuno, H.G.: Dynamic Help Generation by Estimating User’s Mental Model in Spoken Dialogue Systems, *Proc. INTERSPEECH*, pp.1946–1949 (2006).
- 3) Gorrell, G., Lewin, I. and Rayner, M.: Adding Intelligent Help to Mixed-Initiative Spoken Dialogue Systems, *Proc. ICSLP* (2002).
- 4) Gruenstein, A. and Seneff, S.: Releasing a multimodal dialogue system into the wild: User support mechanisms., *Proc. 8th SIGdial Workshop on Discourse and Dialogue*, pp.111–119 (2007).
- 5) Hartmann, M. and Schreiber, D.: Proactively Adapting Interfaces to Individual Users for Mobile Devices, *Adaptive Hypermedia and Adaptive Web-Based Systems, 5th International Conference (AH 2008)*, Lecture Notes in Computer Science, Vol.5149, Springer, pp.300–303 (2008).
- 6) Komatani, K., Ikeda, S., Ogata, T. and Okuno, H.G.: Managing Out-of-Grammar Utterances by Topic Estimation with Domain Extensibility in Multi-Domain Spoken Dialogue Systems., *Speech Communication*, Vol.50, No.10, pp.863–870 (2008).
- 7) Komatani, K., Kawahara, T. and Okuno, H.G.: Analyzing Temporal Transition of Real User’s Behaviors in a Spoken Dialogue System, *Proc. INTERSPEECH*, pp. 142–145 (2007).
- 8) Lee, A., Shikano, K. and Kawahara, T.: Real-time word confidence scoring using local posterior probabilities on tree trellis search, *Proc. IEEE-ICASSP*, Vol.1, pp. 793–796 (2004).
- 9) Lee, C., Jung, S., Lee, D. and Lee, G.-G.: Example-based error recovery strategy for spoken dialog system, *Proc. of IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, pp.538–543 (2007).
- 10) Nisimura, R., Lee, A., Yamada, M. and Shikano, K.: Operating a Public Spoken Guidance System in Real Environment, *Proc. EUROSPEECH*, pp.845–848 (2005).
- 11) Raux, A., Bohus, D., Langner, B., Black, A. and Eskenazi, M.: Doing Research on a Deployed Spoken Dialogue System: One Year of Let’s Go! Experience, *Proc. INTERSPEECH*, pp.65–68 (2006).