

音声対話システムにおけるヘルプ生成のための システム想定外発話の誤受理抑制

福林雄一郎 駒谷和範 尾形哲也 奥乃博

京都大学大学院 情報学研究科 知能情報学専攻

fukubaya@kuis.kyoto-u.ac.jp {komatani,ogata,okuno}@i.kyoto-u.ac.jp

音声対話システムにおいて、音声認識誤りによる誤動作は避けられない。特に、初心者や慣れていないユーザの発話には、システムが想定していない文法や語彙が多く含まれ、音声認識誤りを引き起こす。これへの対処として我々は、ユーザの状態推定に基づく動的なヘルプ生成を目指してきた¹⁾。しかしシステム想定外発話に対する認識誤りの誤受理は、ユーザの状態推定に多大な悪影響を及ぼす。さらに音声認識結果が信用できない発話からも、ユーザの状態を推定できる必要がある。この問題に対して、我々は発話検証技術²⁾の導入により解決を図る。この際に、検証用モデルとして、他タスクでの統計的言語モデルが使用可能かどうかを実験的に確認した。さらに検証用モデルとの尤度差が、発話の棄却判定だけでなく、語彙外発話と文法外発話との判別にも使用可能であることを示す。これにより、内容語が認識できていない場合でもユーザの状態を更新でき、ユーザへのヘルプ生成の精度向上が見込める。

Reduction of False Acceptances caused by Out-of-Domain Utterances for Generating Help Messages in Spoken Dialogue Systems

YUICHIRO FUKUBAYASHI, KAZUNORI KOMATANI, TETSUYA OGATA
and HIROSHI G. OKUNO

Dept. of Intelligence Science and Technology, Graduate School of Informatics, Kyoto University

In spoken dialogue systems, false acceptances (FA) caused by automatic speech recognition (ASR) errors are inevitable. Especially, when a novice user uses the systems, he/she often makes out-of-grammar or out-of-vocabulary utterances, which cause ASR errors. We have developed a method for generating dynamic helps by estimating the user's mental model¹⁾. However, these FAs badly affect the estimation of the model. It should be also possible that the model can be estimated even when ASR results are not reliable. To solve this problem, we incorporate a method of utterance verification. We confirmed that several statistical language models generally used are available as verification models. Furthermore, the differences of scores between two recognizers are helpful not only for rejecting ASR errors but also for distinguishing between out-of-grammar and out-of-vocabulary utterances. This result shows that user's mental model can be updated even when content words are not correctly recognized, and accordingly accuracy of the generated helps will be improved.

1. はじめに

我々はこれまで、初心者でも使いやすい音声対話システムを目指して、動的なヘルプを生成する手法を研究してきた¹⁾。初心者は必ずしもタスク遂行に適切な表現を知っているとは限らないため、システムの想定している言い回しや単語を適宜ヘルプとして教示する必要がある。しかし、ヘルプが必要となる状況、すなわちユーザがシステムの受理・解釈可能な範囲を十分

に把握していない状況では、ユーザはシステム想定外発話をしばしば行う。これを誤認識し、システムの要求として受理すると、ユーザの状態推定¹⁾やタスク遂行に多大な悪影響を及ぼす。したがって、初心者の想定外発話に対する誤受理の抑制は必要不可欠である。さらにこの場合、音声認識結果は誤りを多く含むため信頼できない。適切なヘルプを生成するためには、このような音声認識結果からでも発話を正しく取捨したうえで、ユーザの状態を推定できる必要がある。

本研究では、発話検証技術の導入によりこの問題の解決を図る。つまり、システムの音声認識器と別の言語モデルを持つ音声認識器との音響尤度差を利用して、システムの想定する文法との近さを判定する²⁾。

先行研究では、発話検証で用いる検証用モデルに対して様々な検討を行っている。2)では検証用モデルに音節接続モデルを用いている。3)では、講演調の話し言葉に頻出する文字列をコーパスから学習し、フィルモデルとして検証用モデルに用いている。4)では、大語彙音声認識において部分的に音節モデルを併用することにより、発話中の未知語部分を検出している。また音声認識結果に対する信頼度を算出し、信頼度の低い部分を棄却するというアプローチも採られている。5)では、音声認識結果の N -best 候補を利用して事後確率を計算することにより、音声認識結果に対する信頼度を算出している。6)では、検証用モデルとの尤度比に加えて、単語内の音節継続時間の分散を併用することにより、信頼度を算出している。

本研究では、想定外発話が頻出し、音声認識結果が信頼できない状況での対処を目指すため、従来用いられている信頼度⁵⁾に加え、検証用モデルとの尤度差を併用する。この際、様々なタスクメインでの音声対話システムへの適用を考慮して、検証用モデルを簡便に構成することを志向する。具体的には、タスクの異なる、新聞記事コーパスから学習された統計的言語モデル⁷⁾を検証用モデルとする。この場合の性能を、音声対話システムにおいて初心者から収集した発話データを用いて実験的に比較する。さらにこの発話検証により得られる尤度差を用いて、発話の受理/棄却の判定だけでなく、語彙外発話と文法外発話の識別についても試みる。

2. 音響尤度差を利用した発話検証

音響尤度差による発話検証では2つの音声認識器を利用する²⁾。1つは、音声対話システムが利用する認識器で、システムが認識・解釈に用いる文法に沿ったものを利用する。もう1つは検証用のモデルを持つ認識器で、検証用モデルには、システムの文法もしくは言語モデルより語彙サイズの大きいものを使用する。音響モデルには共通のものを使用する。

ユーザの発話がシステムの文法に近いかどうかは、この2つの認識器で発話を認識したときの音響尤度差を利用して決定する。ここで、検証用の認識器の音響尤度を $score_{ver}$ 、システムの認識器の音響尤度を $score_{sys}$ 、発話長を T 、閾値を θ_{score} として、以下の式で判別する。

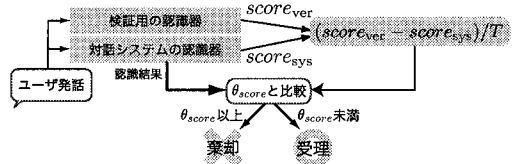
$$\begin{cases} S = (score_{ver} - score_{sys})/T < \theta_{score} & (Accept) \\ \geq \theta_{score} & (Reject) \end{cases}$$


図1 発話検証の流れ

- **京都大学から最も近いものを教えてください**
(認識結果: 長講堂の FFFFF を教えてください)
(理解結果: 受理; 寺社名=長講堂)
(正解: 棄却)
- **その中でバスで行ける ところ**
(認識結果: FFバスで行ける F)
(理解結果: 受理; 検索条件=バスで行ける)
(正解: 受理; 検索条件=バスで行ける)

F はフィルターとして認識された部分。下線部は未知語。

図2 使用した発話データの例

つまり、 S が閾値 θ_{score} 未満ならシステムの文法に近い発話として受理し、 θ_{score} 以上ならば遠い発話なので棄却する。図1に以上の発話の処理の流れを示す。

この手法は、システムの文法に近い発話に対しては、両認識器が音韻的に近い認識結果を出力し、遠い発話に対しては、システム側の認識器が無理に単語列を当てはめることで、検証用の認識器と音韻的に離れた認識結果を出力するという事実を利用している。

3. 評価実験

対話実験における書き起こしデータに対して言語理解の正解データを用意し、発話検証技術によりどれだけ誤受理を抑制できるかを調べる。

3.1 実験条件

評価データとして、京都寺社案内システムを利用した動的ヘルプ生成実験¹⁾での対話データを利用した。京都寺社案内システムでは、音声認識エンジンには Julian⁷⁾を用いた。語彙サイズは 673 である。対話データは 1518 発話 (12 人 \times 2 対話分) で構成される。なお発話データの収集時は、被験者は全員音声対話システムを利用したことがない初心者であり、システムの使用方法を知らなかった。したがって収集した発話はシステムの文法から大きく外れるものが多く、その認識結果も誤りが多い(図2)。音声認識の単語正解率は平均 42.9%であった。

検証用の認識器には、言語モデルに毎日新聞記事データから学習された 20000 語モデル⁷⁾を用いた Julius を用いた。音響モデルは、いずれの認識器でも性別非依存 3000 状態 PTM モデルを用いた。

3.2 言語理解の正解データ

言語理解の正解データをユーザ発話の書き起こしから

- 壬生寺の開館時間は
(正解:受理, spot=壬生寺 action-type=getinfo)
- 京大の近くでうどんが食える寺を検索
(正解:棄却)
(間違った正解:受理, condition=寺 action-type=search)

下線部は未知語。

図3 言語理解の例

ら自動的に作成した。具体的には、音声認識結果の代わりに発話の書き起こしを言語理解部に入力し、Finite State Transducer (FST) によるキーワードスポッティングにより正解を作成した。しかし、単純なキーワードスポッティングでは、発話の一部分にマッチして得られる言語理解結果が多数湧き出してしまう。

図3に例を挙げる。「京大の近くでうどんが食える寺を検索」という発話(の書き起こし)では、「京大の近くでうどんが食える」が文法外で、言語理解の正解は「なし(=棄却すべき発話)」である。しかし、単純なキーワードスポッティングでは「寺を検索」の部分にマッチして、「condition=寺, action-type=search」といった正解がつけられてしまう。

そこで正解作成の際には、全体の50%以上が文法外の単語列で構成される発話は正解を「なし(=棄却すべき発話)」とした。割合の計算は音素単位で行う。したがって、図3の例では「壬生寺の開館時間は」の場合は正解は「spot=壬生寺, info=開館時間, action-type=getinfo」であり、「京大の近くでうどんが食える寺を検索」の場合「なし」が正解になる。

3.3 実験結果

実験では、発話単位での誤受理率 (FA) とスロットエラー ($SErr$) を調べる。誤受理率は認識誤りによって、棄却すべき発話を誤って受理してしまう割合であり、スロットエラーは受理した発話の集合に正解が含まれない割合である。発話データのうち言語理解の正解が「なし」ではなく何らかの正解が付与されている、つまり受理すべき発話数を C とする。また、実際に音声認識結果に対して何らかの言語理解が得られる発話数を A とし、そのうち発話検証の結果が正解であった数を CA とすると、 FA と $SErr$ は $FA = 1 - (CA/A)$ 、 $SErr = 1 - (CA/C)$ と計算される。この FA と $SErr$ は、それぞれ適合率 (precision) と再現率 (recall) の補数になっており、互いにトレードオフの関係にある。

また、 FA と $SErr$ に加えて、実際に言語理解結果が正しかったかどうかを調べる。その指標として、言語理解結果の正解精度 (Accuracy) を計算する。正解精度は、 $Acc = (N - D - S - I) / N$ で計算される。ただし、 N, D, S, I はそれぞれ、言語理解結果の正解、脱落誤り、置換誤り、挿入誤りの総数である。

- 正解 action-type=search, condition=寺, condition=左京区 (コンセプト数3)
- 仮説 action-type=search, condition=寺, condition=左京区, condition=江戸時代 (コンセプト数4)

一致=2, 脱落=0, 置換=1, 挿入=1

図4 言語理解部の正解精度の算出例

図4で正解精度の計算を説明する。この例では、正解のコンセプトは3つのコンセプトを持ち、そのうち2つは一致している。また「condition=左京区」が「condition=右京区」に置換され、仮説には「condition=江戸時代」が湧き出している。この場合、この発話ではそれぞれ $N = 3, D = 0, S = 1, I = 1$ とカウントされる。これを発話データすべてについて総和をとり、正解精度を計算する。

3.3.1 損失関数による手法の評価

θ_{score} を変化させたときの結果を図5, 6に示す。なお損失関数 $cost_1, cost_2$ を $cost_1 = (FA + SErr) / 2$ 、 $cost_2 = (FA + SErr + (100 - Acc)) / 3$ と定義する。 $cost_2$ は言語理解結果の正解精度を考慮に入れた損失関数である。

図5, 6の結果から、閾値 θ_{score} を適切に設定すると、発話検証を行わなかった場合 ($\theta_{score} = \infty$) よりも損失関数 $cost_1, cost_2$ の値を低く抑えられることが分かる。すなわち、音響尤度差を利用する発話検証により、音声認識誤りによる誤受理や間違った言語理解を抑制できることを示している。 $cost_1$ は $\theta_{score} = 130$ で最小となり、 $cost_2$ は $\theta_{score} = 75$ で最小となった。

3.3.2 発話検証手法の比較

以下の4つの場合を比較する。

- 発話検証なし
- 音響尤度による発話検証
- 音声認識信頼度⁵⁾による発話検証
- 音響尤度と音声認識信頼度を組合せた発話検証

音声認識信頼度による発話検証は、単語信頼度 CM が閾値 θ_{CM} より低い内容語を言語理解部に渡さずに棄却することにより行った。表1は、その結果である。ただし、閾値は $cost_2$ が最小になるよう決定した。

$cost_2$ の最小値を比べると、音響尤度と音声認識信頼度を組み合わせると最も効果が高いことが分かる。また、音響尤度、音声認識信頼度の両方で、単独でも発話検証の尺度として効果があるが、 $cost_1$ や $cost_2$ の値を比較すると音響尤度による発話検証の方が効果が高い。この結果から、音響尤度差を利用した発話検証は誤受理の抑制に有効であり、音声認識信頼度とは異なる側面から誤受理を抑制していることが分かる。

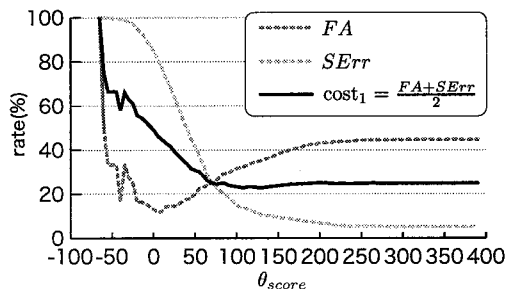


図5 θ_{score} による FA, SErr, cost₁ の変化

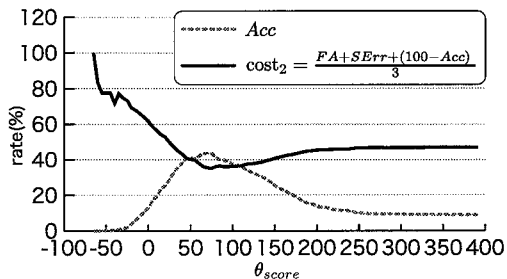


図6 θ_{score} による正解精度, cost₂ の変化

信頼度	尤度差	FA(%)	SErr(%)	cost ₁	Acc(%)	cost ₂
n/a	n/a	44.8	5.1	25.0	8.9	47.0
n/a	$\theta_{score} = 75$	26.0	23.0	24.5	43.5	35.2
$\theta_{CM} = 0.6$	n/a	40.4	8.9	24.7	22.5	42.2
$\theta_{CM} = 0.6$	$\theta_{score} = 95$	26.0	19.8	22.9	46.5	33.1

表2 言語モデルによる cost₁, cost₂ の最小値の比較

言語モデル	θ_{score}	cost ₁	cost ₂
新聞 5k	75	24.4	35.2
新聞 20k	75	24.5	35.2
新聞 60k	75	24.8	35.4
音節連接	95	24.2	34.6

3.3.3 検証用認識器の言語モデル

検証用の言語モデルを、新聞記事 5000 語、20000 語、60000 語の各モデル⁷⁾と音節連接モデル(音声タイプライタ)とした場合を比べる。音節連接モデルは任意の音節の連接を許した言語制約を Julian に与えて実装した。表 2 はその結果である。この比較でも cost₂ が最小になるよう閾値 θ_{score} を選んだ。

この結果から、cost₁ の最小値も cost₂ の最小値も、各言語モデル間でそれほど大きな差はない。これにより、ある程度自由な制約を持つ言語モデルであれば、検証用モデルとして使用可能であることを確認した。

4. 動的ヘルプ生成のための誤受理抑制

我々は、ユーザのメンタルモデルと実際のシステムとの差異を埋めるヘルプの生成手法を開発している¹⁾。メンタルモデルとは、ユーザが心的に持つシステムの動作に関する知識である。ユーザのシステムに関する知識に応じて、生成されるヘルプの内容は動的に変化する必要がある。我々の手法では、システムに関するユーザの知識を、ユーザ発話の認識結果に基づいて更新し、ヘルプを生成する。そのため、音声認識誤りの誤受理に起因する、ユーザ知識の間違った推定の抑制は必須である。さらには内容語が正しく認識されていない場合でも、ユーザ知識が正しく更新される必要が

ある。本研究で用いた発話検証結果を利用することで、ユーザの知識を正確に推定することを考える。

4.1 動的ヘルプ生成

本節では、1) に基づき、動的ヘルプを生成する手法について概説する。これを 1. システムに関する知識の表現、2. ユーザ知識の更新と管理、に分けて説明する。

4.1.1 ドメイン概念木によるシステムに関する知識の表現

システムに関する知識を表現するために、ドメインの概念の階層を表す木(以下ドメイン概念木)を考える。例として京都寺社案内システムのドメイン概念木を図 7 に示す。これは以下に述べる 4 つの層からなり、内容語層のノード名はシステム語彙にある具体的なフレーズを表す。

- **システム層** システム全体の概要を表す
- **機能層** システムが提供する個々の機能を表す
- **要素層** 機能を利用するために発話に含まれるべき要素にあたる
- **内容語層** 各要素を構成する実際の単語やフレーズにあたる

ドメイン概念木では、音声対話システムの文法制約を反映して構成される AND ノードと OR のノードの 2 種類のノードを定義する。機能層に属するノードは、その機能を利用するには子ノードの要素がすべて満たされる必要があるため、AND ノードとなる。これは、システムの音声認識器の文法が、一発話の中にその機能を利用するために必要な要素(単語)が揃うように設計されていることに対応する。一方で OR ノードは、機能層以外の層に属するノードである。これは、機能や内容語は、複数の中から一つを選んで利用すること

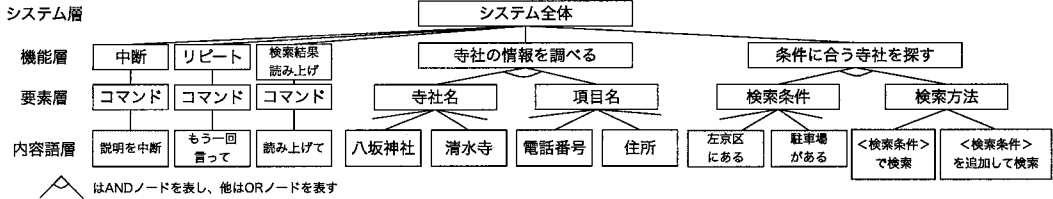


図7 京都寺社案内システムのドメイン概念木

に対応する。

4.1.2 ドメイン概念木上でのユーザのメンタルモデルの推定と管理

ユーザのメンタルモデルとシステムの世界構造との差異を、ドメイン概念木上で各ノードに付与する既知度として表現する。既知度は、ユーザがその概念をどれだけ知っているかの推定値である。既知度は0から1の値をとり、0.5より大きければその概念を知っている、小さければ知らない(=ヘルプが必要)という状態を表す。初期値は0.5とし、これは「どちらかわからない」に相当する。つまり対話開始時点では、すべての概念に関してユーザが知っているかどうかはわからないとする。

既知度は、ユーザの発話後とシステムによるヘルプの提供後に更新する。つまり、ユーザの発話に基づき、ユーザの知っていること、知らないことを推定する。また、ヘルプの提供によっても既知度は増えるとする。

具体的な既知度の更新を $renew(n, p)$ を使って定義する。 $renew$ は、 $p \geq 0$ のとき、ノード n とその先祖の既知度を増やし、 $p < 0$ のとき n とその子孫の既知度を減らす。このように、先祖や子孫の既知度とともに変化させるのは、システムの具体的な概念を知っているユーザは、その概念を内包する抽象的な概念も知っていて、逆に、抽象的な概念を知らないユーザは、具体的な概念も知らないとの仮定に基づく。更新のパラメータ p は、 $p_u^+ (> 0)$, $p_h^+ (> 0)$, $p_u^- (< 0)$ の3つを用意する ($-0.5 \leq p_{h,u}^{\pm} \leq 0.5$)。 p の添字 u はユーザ発話による更新を表し、 h はシステムの提供したヘルプによる更新を表す。

ユーザの発話により既知度が更新されるとき、概念木の更新は以下の順に実行される。ここでは内容語層に含まれる単語をキーワードとしている。

- (1) 音声認識結果に含まれるキーワードに対して既知度を更新
- (2) AND ノードが満たされているかを確かめて更新
音声認識結果にキーワードが含まれる場合、対応する内容語ノード n に対して $renew(n, p_u^+)$ が実行され、ルートノードに向かって既知度が更新される。また、認識結果にキーワードが全く含まれない場合は、ルー

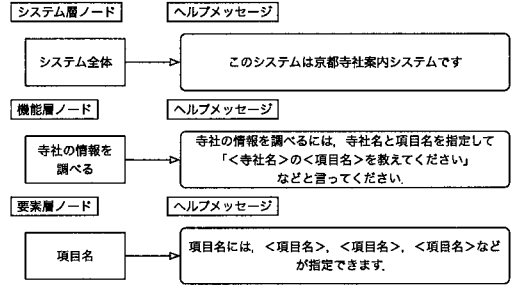


図8 ドメイン概念木の各ノードに対応するヘルプメッセージの例

トノード(システム層)を n として $renew(n, p_u^-)$ が実行され、内容語ノードに向かって既知度が更新される。また、ヘルプを提供した際には、提供した概念に対応するノード n に対して $renew(n, p_h^+)$ が実行される。

ユーザに提供するヘルプは、既知度が0.5未満のノードを概念木から探し、図8のようなテンプレートを利用し生成される。機能層では文法に関するヘルプが、要素層では各内容語に関するヘルプが出力される。

4.2 発話検証技術を利用した概念木の更新

4.1節で述べたように、これまでのドメイン概念木の更新は主に受理した発話の内容語に基づいて行われていた。しかし、内容語が正しく認識できない発話に対しても、発話内容に応じて正しくドメイン概念木を更新し、ユーザの知識状態を正確に推定できる必要がある。例えば「吉田神社のアクセスを教えてください(下線部は未知語)」といった発話を行うユーザは、システムの文法は知っているが具体的な内容語は知らない、と推定できることが望ましい。そこで発話検証結果をドメイン概念木の更新時に利用することを検討する。

3章で定義した Acc を最大にする θ_{score} を θ_1 とする。このとき、 θ_1 未満の発話への言語理解結果は正しい可能性が高く、 θ_1 以上の発話はそうでない可能性が高い。一方で、 $cost_1$ を最小にする θ_{score} を θ_2 とする。このとき、 θ_2 未満の発話はシステムの文法に沿っている可能性が高く、 θ_2 以上の発話はそうでない可能性が高い。以上のように θ_1, θ_2 を設定すると、ユーザ発話を以下の3つに分類できる。

表3 2つの閾値を設定した場合の発話の分類とその例 ($\theta_1 = 75, \theta_2 = 125$)

発話検証の正解	言語理解の出力結果	$S < \theta_1$	$\theta_1 \leq S < \theta_2$	$S \geq \theta_2$	発話例 (書き起こし, 下線部は未知語)
受理	正解 (100%)	454 [†]	50*	8	吉田神社の見学時間を教えてください
受理	誤りあり (<100%)	84	29***	34	曼珠院の連絡先を教えてください
受理	出力なし	28	13***	23	徳川時代にできた寺を教えてください
棄却	誤りあり (誤受理)	158	104	185 [‡]	吉田神社 に入るには連絡する必要がありますか
棄却	出力なし (正しく棄却)	166	86	98	京都大学から最も近いものを教えてください

- (1) $S < \theta_1$: システム文法内で、言語理解も正しいと推定される発話
- (2) $\theta_1 \leq S < \theta_2$: システム文法内だが、言語理解が正しくないと推定される発話
- (3) $S \geq \theta_2$: システム文法外で、言語理解も正しくないと推定される発話

音声認識信頼度による発話検証を併用した場合、 $\theta_{CM} = 0.6$ とすると、 $\theta_1 = 75, \theta_2 = 125$ となった。ここでは検証用モデルとして新聞記事 20000 語モデルを用いている。上記の方法で分類した発話を、さらに発話検証の正解、言語理解の出力結果により分類した結果を表3に示す。

この分類に基づいた新たな既知度の更新方法を考える。 $S < \theta_1$ の発話は従来通り認識された内容語に基づいて更新する。 $S \geq \theta_2$ の発話は従来の棄却された発話と同様に扱う。 $\theta_1 \leq S < \theta_2$ の発話はシステム文法内であるが、言語理解は正しくないと推定される発話である。つまり、内容語は得られていないものの、システムの文法には沿った発話であると推測できる。ただし、認識結果はそれほど信頼できず、どの機能もしくは要素に対する発話かは分からない。そこで、すべての機能層もしくは要素層のノードに対して一律に $\text{renew}(n, p_u^*)$ を実行し、既知度を増加させる。この更新により、内容語が正しく認識できない発話に対しても、ユーザがシステム文法は知っているという情報をドメイン概念木に反映させることができる。これにより、機能層の既知度が向上し、文法に関するヘルプが抑制される。この結果、このユーザにはより具体的なヘルプが生成されるようになる。

以上の手法の効果を $S \geq \theta_2$ の発話だけを単に棄却する場合と比べる。表3の**は、一部の言語理解誤りにより誤動作していた発話で、***は、言語理解が得られず単に棄却していた発話である。したがって、**と***の42発話において、ユーザがシステムの文法を知っているという新たな情報を得ることができる。ただし、*の50発話は、もともと正しい言語理解が得られていた発話で、その言語理解は誤棄却されることになる。したがって、現状では新たな更新方法を導入すると**と***の改善以上に誤棄却が増えてしまう。この要因のひとつとして、閾値 θ_1, θ_2 はそ

れぞれ†と‡の数が最適になるように設定されているので、*や**, ***の発話に対しては最適であるとは限らない点が挙げられる。また棄却の精度についても改善が必要である。これを含め、さらに詳細な分析を行う予定である。

5. まとめ

本研究では、音声対話システムにおいてシステム想定外発話の誤受理を抑制するために、発話検証技術を導入した。音声認識信頼度を併用した発話検証を利用することで、発話検証を行わない場合より定義した損失関数 cost_2 の値において 13.9 ポイント改善が得られた。また、いくつかの検証用モデルを実験的に評価し、ある程度自由な制約を持つ言語モデルであれば、検証用モデルとして使用可能であることを確認した。

さらに検証用モデルとの尤度差が、発話の棄却判定だけでなく、語彙外発話と文法外発話との判別にも使用できる可能性を示した。これにより、内容語が認識できていない場合でもユーザの状態を更新でき、ユーザへのヘルプ生成の精度向上が見込める。

参考文献

- 1) 福林雄一朗, 駒谷和範, 尾形哲也, 奥乃博. 音声対話システムにおける発話パターンを教示するヘルプの動的生成. 人工知能学会研究会資料, SIG-SLUD-A601-03, 2006.
- 2) 渡辺隆夫, 塚田聡. 音節認識を用いたゆ度補正による未知発話のリジェクション. 電子情報通信学会論文誌, Vol. J75-DII, No.12, pp. 2002-2009, 1992.
- 3) 河原達也, 石塚健太郎, 堂下修司. 発話検証に基づく音声操作プロジェクトとそれによる講演の自動ハイパーテキスト化. 情報処理学会論文誌, Vol.40, No.4, pp. 1491-1498, 1999.
- 4) 甲斐充彦, 廣瀬良文, 中川聖一. 単語 N-gram 言語モデルを用いた音声認識システムにおける未知語・冗長語の処理. 情報処理学会論文誌, Vol.40, No.4, pp. 1383-1395, 1999.
- 5) 駒谷和範, 河原達也. 音声認識結果の信頼度を用いた効率的な確認・誘導を行う対話管理. 情報処理学会論文誌, Vol.43, No.10, pp. 3078-3086, 2002.
- 6) 北岡教英, 赤堀一郎, 中川聖一. 認識結果の正解確率に基づく信頼度とリジェクション. 電子情報通信学会論文誌, Vol. J83-DII, No.11, pp. 2160-2170, 2000.
- 7) 河原達也, 武田一哉, 伊藤克互, 李晃伸, 鹿野清宏, 山田篤. 連続音声認識コンソーシアムの活動報告及び最終版ソフトウェアの概要. 情報処理学会研究報告, 2003-SLP-49-57, 2003.