2007－SLP－67 (19)
2007／7／21

予測文と音素認識の併用による発話の予測内外判定に関する検討

真柄 皓介†，西田 昌史†，堀内 靖雄†，市川 熹†
千葉大学大学院融合科学研究科 〒263-8522 千葉県千葉市稲毛区弥生町 1-33
E-mail: †makara@graduate.chiba-u.jp, {nishida, hory, ichikawa}@faculty.chiba-u.jp

あらまし 我々は，これまで車載情報機器を音声により操作するシステムの開発を目指して，発話予測に基づく音声対話に関する研究を行ってきた．こういったシステムでは，事前に予測された発話を必ずユーザがするとは限らず，予測外発話への対応が必要となる．そこで，本研究では文単位の認識と音素認識を用いて，これらの認識結果の一致度に着目した発話の予測内外判定手法を提案する．さらに，これまで我々が提案した音素列に対する DP マッチングを用いた予測文の絞り込み認識手法も適用した．カーナビにおける目的地設定の場面をタスクとした発話の予測内外判定による実験を行った結果，話者 11 名の 3399 発話に対して音声認識率 83.4%，予測内外判定率 91.9%と高い精度が得られ，提案手法の有効性が明らかとなった．

キーワード 音声対話，予測文，音素認識，予測内外判定，決定木

A Study on Utterance Prediction based on Recognition of Sentence and Phoneme

Kousuke MAKARA †, Masafumi NISHIDA †, Yasuo HORIUCHI †, Akira ICHIKAWA †
† Graduate School of Advanced Integration Science, Chiba University
1-33 Yayoi-cho,Inage-ku,Chiba-shi,Chiba,263-8522 Japan
E-mail: †makara@graduate.chiba-u.jp, {nishida, hory, ichikawa}@faculty.chiba-u.jp

Abstract We have studied on spoken dialogue using utterance prediction aiming at development of in-vehicle information system based on dialogue. It is necessary to judge whether the utterance is out of prediction because users may not do the utterance that the system predicted beforehand. In this study, we propose an utterance prediction method based on matching rate of the recognition results obtained by performing recognitions of the prediction sentence and phoneme. Moreover, we applied a recognition method by narrowing of candidate sentence using DP matching of the phoneme recognition result that we proposed. We conducted experiments using 3399 utterances by 11 speakers in setting a destination of a car navigation system. As a result, a speech recognition accuracy was 83.4% and utterance prediction accuracy was 91.9%. Therefore, we demonstrated that the proposed method was effective in a spoken dialogue system.

Keyword Spoken dialogue, Prediction sentence, Phoneme recognition,Utterance prediction, Decision Tree

1. はじめに

近年，音声対話システムは公共情報[1]，カーナビゲーションなど実環境下で利用されはじめている．音声対話においてこれらに共通する課題として誤認識の抑制および，想定外発話への対応が必要となってくる．想定外発話に関する研究としてプロジェクタの操作におけるコマンド発話を対象としてキーワードスポッティングとフィラーモデルを使用したもの[2]や，音声翻訳システムにおけるトピッククラスタリングによるドメイン外発話の検出をしたもの[3]や、キーワードスポッティングと韻律的特徴モデルを併用した発話検証法[4] が提案されている．また単語認識と音節認識を併用し尤度差を閾値処理するもの[5]、未知語，冗長語の処理については単語 N-gram とサブワード単位の音響モデルを用いた手法[6]，認識結果に正解確率の意味付けのある信頼度を設ける手法[7]が提案されている．

誤認識の抑制に対して我々は発話状態に応じてユーザの発話を予測することで認識対象語彙を絞り込む予測文認識を提案してきた[8][9]．そこで本研究ではこの予測文認識と音素認識を併用した予測内外判定手法を提案する．本手法は従来の尤度差に加えてこれらの認識結果の音素と文長一致度に着目し,決定木による予測内外判定手法を行う．さらに我々が提案した音素列に対する DP マッチングによる予測文候補の絞り込み認識[10]を適用する．

提案手法の有効性を示すために，カーナビにおける目的地設定の対話場面を想定した評価実験を行う．

2. 発話予測による音声対話システム

音声認識では大語彙を認識対象とすると，認識率が低下してしまう問題がある．これは認識候補数の増大に伴って誤認識が増えてしまう事に起因する．この問題に対してわれわれはユーザの発話を「未知情報要求」「確認」「肯定・否定」などの発話単位タグ［11］を基に対話状態を定義，状態遷移モデルを構築し，ユーザの次発話を予測することによって認識候補を絞り込み，認識精度の向上を示してきた．ここでは道案内の目的地設定を目的とした状態遷移モデルの概略図を図 1 に示す．

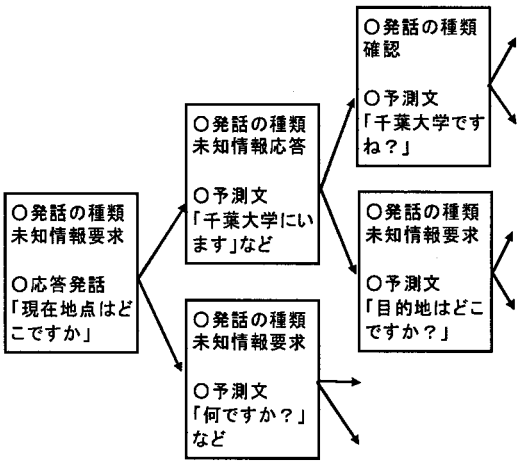

図 1．対話状態の遷移図

しかしながらユーザの次発話をすべて予測することは不可能であり，予測外の発話がきた場合，誤認識や対話が円滑に進まない可能性がある．この問題に対してユーザの次発話の予測を用いた認識を行う際に，ユーザの発話が予測内の発話か予測外の発話か，また予測内の発話であった場合認識が成功しているか否かを推定することが重要になってくる．そこでユーザの次発話を予測し認識精度の向上を図るとともに，予測内外判定と認識成否判定を行う事が重要になってくる．

予測文認識は，ユーザの次発話を予測し大語彙認識器の辞書と比較して辞書を絞り込んだ認識器である．予測文は実際の地名などを含む

文でフィラー，地名，言い回しごとに区切ってそれらの組み合わせで予測文を作成する．認識単位としては文単位での認識を行う．

3. 予測文と音素認識の尤度差に基づく予測内外判定

これまで未知発話のリジェクションとして単語認識と音節認識を併用し尤度差を閾値処理する手法が提案された[5]．本研究ではこれを予測文認識と音素認識として適用し，予測文と音素認識の尤度差により予測内外判定を行う．予測文認識と音素認識を並行に認識をおこなうシステムでの予測内外判定では，予測文認識器から得られた予測尤度 $\log P(x|\lambda s)$ と音素認識器から得られた音素尤度 $\log P(x|\lambda p)$ を用いた以下の式（1）より尤度差を求める．

$$(\log P(x|\lambda s) - \log P(x|\lambda p)) / \mathrm{L} \quad (1)$$

ここで式（1）のLは入力発話のフレーム数である．予測内外の判定としては式（1）で閾値より大きい値になったとき予測尤度が音素尤度と比べて高い尤度を持ったと考えられユーザの発話は予測内発話であるとする．また下回ったときは予測尤度が音素尤度と比べて低い値をとったと考えられるのでユーザが予測文にない発話をしたと予想されるので予測外と判定する．ここで用いられる閾値は事前に学習して設定しておく．

4. 予測文と音素認識の認識結果を併用した予測内外判定

4. 1. 音素認識に基づく予測文候補の絞り込み

これまで我々は認識精度の向上を目的として，音素列に対するDPを用いた予測文の絞込み認識手法を提案してきた[10]．まず音素認識によってユーザの発話の音素列を得る．次に予測文認識器の全予測文を音素列に変換したものと音素認識で得られた音素列を音素単位でDPマッチングを行い，認識候補の絞込みを行う．このとき比較する2文は音素列が長くなるほど絶対的な距離が増加してしまうので，比較した2つの音素列の伸縮を考慮して，伸縮後の音素列で除算をすることで正規化を行い，正規化DP距離とした．正規化DP距離が近い上位候補を新たに認識候補として辞書を更新した絞り込み認識を行い認識精度の改善を行った．図2にシステムの処理の流れを示す．

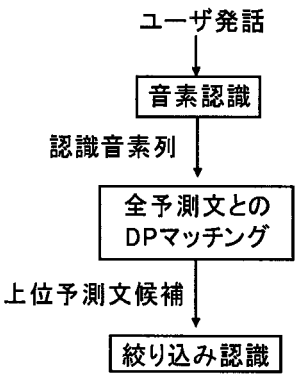

図2. 絞り込み認識の処理の流れ

4. 2. 音素と文長の一致度に着目した予測内外判定

本研究では，予測文認識と音素認識の尤度差に加えて，それぞれの認識結果における音素と文長の一致度を用いた予測内外判定手法を提案する．

音素認識器と予測文認識器から得られる認識結果の音素列を比較したとき，音素認識器では発話に即した音素列が得られ，予測文認識器では予測内発話であった場合は音素認識と同様の長さの音素列が得られると考えられる．また予測外発話であった場合は音素列の長さが異なってくると考えられるので文長一致度を用いる．文長一致度は予測文と音素認識の音素

数の比により求めた.

さらに二つの認識器から得られる音素列を比較したとき,音素認識器では発話に即した音素列が得られると考えられる.予測認識器は予測内発話であった場合は発話内容に沿った音素列が得られると考えられ,予測外であった場合は発話内容から外れた音素列が得られると考えられる.2つの認識器より得られた音素列の音素の一致数を比較することにより,ユーザの発話が予測内か予測外かを表すのに有効なのではないかと考えられ,音素一致度を用いる.音素一致度は予測文と音素認識の音素を比較し,一致した音素数を音素数の多いほうで除算することで求める.

以上の3つのパラメータにより決定木を用いて予測内外判定を行った.

5. 評価実験

予測内外判定に関して以下の 3 つの手法で評価を行う.

従来手法:従来の予測文認識と音素認識での尤度差を閾値で判定

提案手法1:予測文と音素認識の尤度差,文長一致度,音素一致度による判定

提案手法2:絞り込み認識を適用した場合の“提案手法1”での判定

5.1. 実験条件

カーナビゲーションシステムを想定した地名認識を目的として音声認識,予測内外判別を行った.被験者には予測内発話としては「千葉大学にいます」,「現在地点は千葉公園です」など発話中に予測文に登録された地名を含む発話を,予測外発話としては予測文に登録されていない地名を含むものや「大きな道路沿いです」,「ここだよ,ここ」など地名を含まない発話を行ってもらった.なお登録されている地名は約50個で予測文の数は約350文である.

被験者は男性11名であり,各被験者に309発話(予測内発話:144,予測外発話:165),全3399発話(予測内発話:1584,予測外発話:1815)取得した.被験者から得られた発話はすべて研究室内で録音し,音声は16kHz,16ビットで量子化されており,オフラインにて実験を行った.評価には交差検証法を用い,10人のデータを学習データとして1人に適用し,組み合わせを変えて11回行った.

音素認識の音響モデルには,状態数3000,混合分布数64,性別非依存の PTMtriphone を用いており,前後の調音結合を考慮している.尚,音声認識デコーダには julius-3.1[12]を用いた.

5.2. 実験結果

表1に全体の予測内外判定率を示した.ここで従来手法は予測文と音素認識の尤度差を閾値で予測内外判定をした結果を表している.提案手法1は予測文と音素認識の尤度差に文長一致度,音素一致度を加えて決定木で予測内外判定をした結果を表している.提案手法2は音素列に対して DP を用いた予測文の絞込み認識での尤度差,文長一致度,音素一致度を用いて決定木で予測内外判定をした結果を表している.

またそれぞれ従来手法での適合率と再現率を表2に,提案手法1の適合率と再現率を表3に,提案手法2の適合率と再現率を表4に表した.

表1. 全体の予測内外判定率

従来手法	90.1% (3063/3399)
提案手法1	91.6% (3112/3399)
提案手法2	91.9% (3124/3399)

表 2. 従来手法での予測内外判定精度

	再現率	適合率
予測内	84.1% (1332/1584)	94.1% (1332/1416)
予測外	95.4% (1731/1815)	87.3% (1731/1983)

表 3. 提案手法１での予測内外判定精度

	再現率	適合率
予測内	88.4% (1400/1584)	93.1% (1400/1503)
予測外	94.3% (1712/1815)	90.3% (1712/1896)

表 4. 提案手法２での予測内外判定精度

	再現率	適合率
予測内	86.1% (1364/1584)	96.1% (1364/1419)
予測外	97.0% (1760/1815)	88.9% (1760/1980)

また従来の予測文認識と，音素列に対して DP を用いた予測文の絞込み認識の音声認識精度を表 5 に示した.

表 5. 音声認識精度

予測文認識	絞り込み認識
62.0%(982/1584)	83.4%(1321/1584)

5. 3. 考察

従来手法である予測文と音素認識の尤度差による閾値処理に比べて，提案手法により予測内外判定率が改善されている事がわかった．従って音素と文長の一致度によるパラメータが予測内外判定に有効であることがわかった.

また通常の予測文認識では62.0%の認識精度が得られ，音素列に対する DP を用いた予測文の絞込み認識により 83.4%の認識精度が得られた．絞り込み認識を導入した予測内外判定は通常の予測文認識による予測内外判定に比べて判定率はほぼ同じで，高い音声認識精度が得られることから有効であると考えられる.

図３，図４に決定木の例を示す．ここで右側の濃い方（カテゴリ：１）が予測内，左側の薄い方（カテゴリ：－１）が予測外発話を表している．図３より予測内外判定には尤度差が最も有効であるのがわかる．図４より文長一致度がほぼ同じくらいの場合，音素の一致度が高ければ予測内，低ければ予測外に判別することができている.

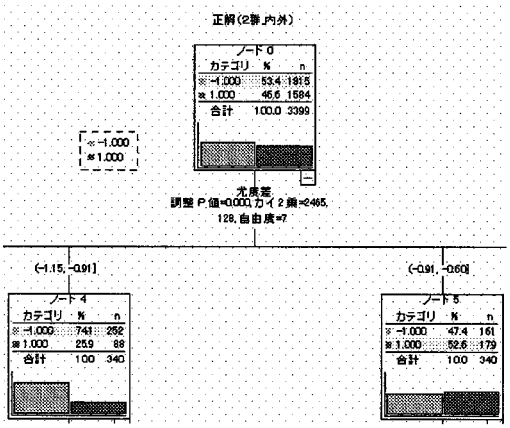

図 3. 決定木における最上位ノード

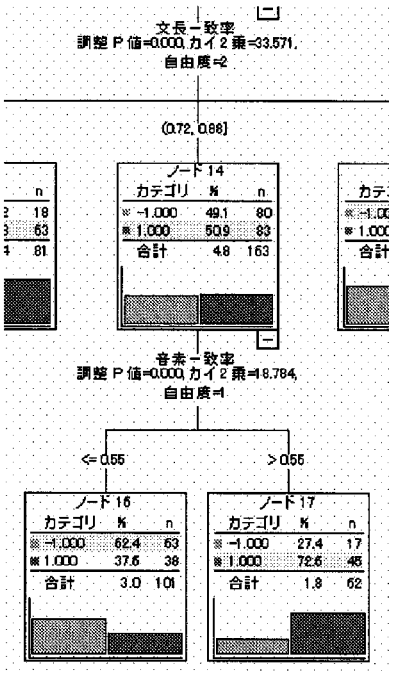

図 4. 決定木における文長一致度と音素一致度

7. おわりに

本研究では音声対話のタスクとしてカーナビゲーションの地名認識を目的とした対話場面において，予測文と音素認識を併用し尤度差に加えて音素と文長の一致度を用いた予測内外判定手法を提案した．さらに音素列に対するDPを用いた予測文の絞り込み認識を適用した結果，音声認識精度 83.4%，予測内外判定率91.9%が得られ，提案手法が有効であることがわかった．

今後の課題としては，さらに予測内外判定に有効なパラメータや認識結果の信頼度について検討を行う．また実際のカーナビゲーションを想定して辞書の語彙数を増やして，実際に提案手法をシステムに実装し評価を行っていく予定である．

謝辞

本研究は，富士重工業株式会社との共同研究により実施した．

参考文献

[1] 西村竜一，西原洋平，鶴身玲典，李晃伸，猿渡洋，鹿野清宏，“生駒市コミュニティセンター音声情報案内システムの開発と運用，”情処学研報，2003-SLP-45-6, pp. 35-40,2003.

[2] 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タ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.

[3] レーンイアン，河原達也,中村哲,” 対話コンテキストとトピッククラスタリングを用いたドメイン外発話の検出,” 信学技報.SP,音声 Vol.104, No.543, pp. 49-54,2004.

[4] 甲斐充彦,板倉雅和,” キーワード主体の頑健な音声インタフェースのための韻律的特徴を用いた発話検証,” 情処学研報，2006-SLP-61,2006.

[5] 渡辺隆夫,塚田聡,” 音節認識を用いたゆう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. D-II, Vol.75, No.12 (19921225) pp. 2002-2009,1992.

[6] 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(19990415) pp. 1383-1394,1999.

[7]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，D-II Vol. J83-D-II No.11，pp.2160-2170，2000.

[8] 玉井孝幸，堀内靖雄，市川熹，“音声対話システムにおける発話予測を利用した音声認識，”情処学研報“2002-SLP-43,pp.1-6.2002.

[9] 西田昌史，寺師弘将，堀内靖雄，市川熹，“ユーザ発話の予測に基づく音声対話システム，”信学技報，SP2004-132,pp.61-66, 2004.

[10] 寺師弘将,西田昌史,堀内靖雄,市川熹,” 音声対話システムにおける音素認識に基づく予測文候補の絞込み，” 情処学研報，2006-SLP-64, pp. 101-106,2006.

[11] 荒木雅弘，伊藤敏彦，熊谷智子，石崎雅人，“発話単位タグ標準化案の作成，”人工知能学会誌，Vol.12，No.1，pp.1-10，1997.

[12]大語彙連続音声認識システム Julius http://julius.sourceforge.jp/