

頑健な言語理解のための文法とワードスポッティングを 併用した音声認識手法の検討

鈴木 貞之[†] 小暮 悟[†] 伊藤 敏彦[‡] 甲斐 充彦[†] 小西 達裕[†] 伊東 幸宏[†]

[†] 静岡大学 〒432-8011 静岡県浜松市城北 3-5-1

[‡] 北海道大学 〒060-0814 北海道札幌市北区北 14 条西 9 丁目

E-mail: [†] cs0047@s.cs.inf.shizuoka.ac.jp, [†] {kogure,konishi,itoh}@inf.shizuoka.ac.jp,

[†] kai@sys.eng.shizuoka.ac.jp, [‡] t-itoh@media.eng.hokudai.ac.jp

あらまし 本稿では、カーナビゲーションの目的地設定をタスクとした自由発話において、文レベルでの文法制約を用いた音声認識で得られる N-best 候補と、ワードスポッティングによる音声認識で得られる単語ラティス情報とを組み合わせ、先行研究での N-best 候補からの音声理解の枠組みにおける性能改善のため、N-best 候補の精度(正解候補の順位や含有率など)を高める手法を提案する。まず、ワードスポッティングから各発話における単語の確からしさを計算し、その確からしさの情報を利用して、文レベルでの文法制約による N-best 候補の単語尤度の上昇や単語の交換等を行う。評価実験において、特に、間投詞や未知語が存在する文において、本手法を用いることによって正解単語の N-best 候補の出現順位が上昇し、本手法の有効性を示せた。

キーワード 音声認識, ワードスポッティング, 自由発話, 未知語処理

Speech Recognition using CFG in combination with Word Spotting for Robust Language Understanding

Sadayuki SUZUKI[†] Satoru KOGURE[†] Toshihiko ITOH[‡] Atsuhiko KAI[†]

Tatsuhiko KONISHI[†] and Yukihiro ITOH[†]

[†] Shizuoka University 3-5-1 Johoku, Hamamatsu, Shizuoka, 432-8011 Japan

[‡] Hokkaido University Kita-14, Nishi-9, Kita-ku, Sapporo, Hokkaido, 060-0814 Japan

E-mail: [†] cs0047@s.cs.inf.shizuoka.ac.jp, [†] {kogure,konishi,itoh}@inf.shizuoka.ac.jp,

[†] kai@sys.eng.shizuoka.ac.jp, [‡] t-itoh@media.eng.hokudai.ac.jp

Abstract In this paper, we propose the technique for improving the N-best candidates accuracy in a spontaneous utterance by destination setting task with car navigation, combining the N-best candidates using a grammatical constraint at sentence and the word lattice using word spotting, to improve performance in the framework of speech understanding from the N-best candidates in early research. The system calculates the reliability of the word of each utterance by using the word lattice. We use the reliability to raise the word likelihood and to exchange the word of the N-best candidate by grammatical constraint. As a result of the evaluation experiment, we raise the appearance order in the N-best candidate of the correct word by using this technique in the sentence where the interjection and OOV words existed. We show the effectiveness of our proposed method.

Keyword speech recognition, word spotting, spontaneous utterance, unknown word processing

1. はじめに

近年、音声認識処理技術の向上により、さまざまな音声対話システムが実用化されている[1]。音声対話カーナビゲーションなどはそのいい例である。カーナビゲーションシステムを利用する状況を考えた場合、その入力形式はリモコンよりも、音声入力・音声出力の方が、操作性・安全性の面から望ましい。しかしなが

ら現在の音声認識処理技術では、誤認識を完全に回避することは困難である。誤認識が生じると、システムがユーザの発話を正しく理解できないために、ユーザの意図に反した動作・応答を行うことになる。その結果、システムがユーザ発話を正しく理解した場合よりも対話がスムーズに進まなくなり、ユーザに不快感を与えることになる。

先行研究のカーナビゲーションにおける目的地設定をタスクとした音声対話システム[5]の研究において、文レベルの N-best 候補から計算される各単語の信頼度やユーザの発話意図の推定結果、対話履歴などを利用することで言語理解精度の向上を行ってきた。

この研究では、発話可能なキーワードが属するカテゴリを設定して、システムがユーザ発話の発話タイプを推定し、カテゴリレベルでユーザ発話を正しく理解することでシステム全体の精度向上を目指した。しかしながら、この研究における問題点として、文レベルの N-best 候補に出現しない正解単語はシステムの理解結果に含まれないことが挙げられる。

この問題への対応として、本研究では文レベルの N-best 候補とワードスポッティングによる単語ラティス情報を統合して、最終的な N-best 候補を生成することを目標としている。

そこで本研究では、文レベルでの文法制約を用いた音声認識器では認識できない、非文であるが発話したキーワードが既知語であるような発話に対して、ワードスポッティングによって補完することを考える。

ワードスポッティングとは、検出対象の単語について任意の区間での音響尤度を求め、当該単語の存在可能性が高い区間の開始・終了フレームと、その音響尤度の三つ組からなる複数候補(単語ラティス)の情報を出力する音声認識手法である[2]。音声認識の分野では、ある程度大きなタスクで、かつそのタスク(あるいはそのタスクに関連したタスク)における大量の対話想定文が用意できる場合は、統計情報を用いた bi-gram や tri-gram を用いる手法が主流である。これら N-gram を用いた音声認識は統計情報のみを用いた認識であり、非文も認識結果に現れる可能性があるため、本研究では有効な手法の一つである[3,4]。しかし、現在はカーナビゲーションにおける目的地設定タスクのみを想定しており、また語彙サイズ 1 万単語程度を対象としているので、それに対応した大量の対話文を収集するにはコストがかかりすぎることから、本研究では N-gram を用いる音声認識については対象としない。

キーワードスポッティングを用いた単語認識は、語彙数やタスクが非常に小さい場合に有効であるが、一般的に、ある程度大きなタスクにおいては、CFG 等の文法制約を用いた音声認識の方が、全体の単語正解率は高くなる。文レベルでの文法制約を用いた音声認識では、想定された文型の発話に対して雑音が付加されたデータであっても、ある程度の認識精度を得ることが可能である。しかし、想定されていない文型(非文)の発話に対しては、発話したキーワードが語彙内に入っていたとしても認識されない可能性が高い。結果として、発話されたキーワードが認識結果に出現しない

可能性がある。

本研究では、この文レベルでの文法制約を用いた音声認識における、非文あるいは想定文型であっても一部が未知語であるために本来既知語である部分の認識にも失敗してしまう欠点を補うため、文レベルでの文法制約により得られる N-best 候補に含まれる単語を、ワードスポッティングの認識結果を使って更新する。

2. システム概要

本研究で扱うシステムは、音声認識部、信頼度生成部、言語理解部、応答文生成部、GUI 表示部、音声合成部の 6 つの構成要素から成る。図 1 にシステムの構成図を示す。

音声認識部では、入力音声を音声認識して、音響的な尤度で順位付けされた複数候補 N-best からなる認識結果を信頼度生成部に渡す。

信頼度生成部では、音響的な尤度と N-best 候補での出現頻度を利用して単語の信頼度を生成する。生成された信頼度付き N-best 候補は言語理解部に渡される。

言語理解部では、まず生成された信頼度付き N-best と対話履歴を利用して、各単語の単語スコアを生成する。次に、最新発話を含めた対話において、発話されてきたカテゴリを求め、そのカテゴリ理解の結果と対話履歴から発話された単語列(以下言語理解結果)を生成する。生成された言語理解結果は応答文生成部に渡される。

応答文生成部では、言語理解結果の単語スコアに基づいて、応答パターンを切り替えて応答文を生成する。生成された応答文は、GUI 表示部・音声合成部へと送られる。GUI 表示部では応答文を表示し、音声合成部では生成された応答文の音声合成を行い、ユーザに音声出力を行う。

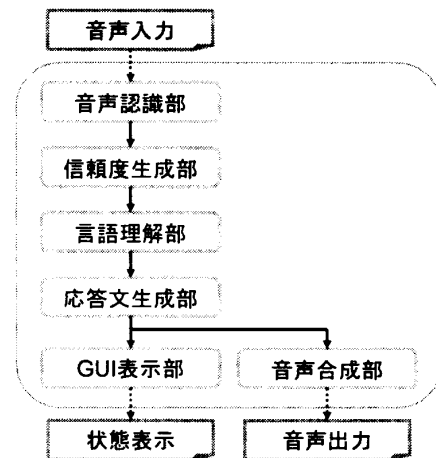


図 1. システム構成

先行研究では、言語理解部でのカテゴリ理解精度の向上を実現した[5]。しかし、これ以上のカテゴリ理解精度の向上を見込むことは難しいことが確認され、シ

システム理解精度をより向上させるためには、単語正解精度や単語スコアの改善に着手する必要がある。

そこで本研究では、音声認識部において正解単語をできるかぎり N-best 候補に出現させること、および正解単語の N-best 出現順位を上位に更新することを目指す。

3. 提案手法

本研究では、ワードスポッティングを用いた音声認識結果に出現するキーワードに対して、あるルールを適用し、利用するキーワードの絞込みを行う。これにより取り出されたキーワード群と、文レベルでの文法制約を用いた音声認識結果中のキーワードとを照合して N-best の更新を行うことで、頑健な言語理解を目指す。

3.1. 音声認識器

本研究では、音声認識器として、豊橋技術科学大学の中川研究室で開発された日本語連続音声認識システム SPOJUS[6]を用いる。本研究で用いた 2 つの音声認識器の特徴とそれらの統合手法について以下に示す。

3.1.1. 文レベルでの文法制約を用いた音声認識

文レベルでの文法制約を用いた音声認識器(以下、CFG 版 SPOJUS)では、音響モデルとして、出力確率を確率密度関数であるガウス分布で近似した 5 状態 4 出力の音節 HMM を用いる。単語の認識においては、それらの音節 HMM を連結したものを単語 HMM として利用する。

最終的な音声認識結果は、CFG による文法制約に沿って予測される次の単語についての音響尤度を次々に求め、最終フレームまでの認識において、累積の音響スコアが高い順に N 個の単語列が N-best 候補として得られる。なお、本研究では 20-best までの認識結果を利用している。

3.1.2. スポッティングによる音声認識

ワードスポッティングを用いた音声認識器(以下、WS 版 SPOJUS)では、前述の CFG 版 SPOJUS を改良し、フレームごとに単語の確からしさを計算している。

文レベルの N-best 候補が持つ音響尤度に対して、ある区間(仮に、 t_1 フレームから t_2 フレームとする)に注目したとき、その区間に単語 w が認識された確からしさ(どれだけ、その単語らしいか: R_w)を求める。入力音声の開始フレームから $(t_1 - 1)$ フレームまでの全体の音響尤度を P_1 、入力音声の開始フレームから t_2 フレームまでの全体の音響尤度を P_2 とする。このとき、 $\log(P_2) - \log(P_1)$ がその単語の音響尤度となる。

さらにその対象区間において、任意の音節接続を許して認識を行い、結果として得られた音節列による音響尤度 P_{freesyl} とでフレーム正規化された尤度比 R_w は次

の式で求められる。

$$R_w = \frac{\log(P_2) - \log(P_1) - \log(P_{\text{freesyl}})}{t_2 - t_1 + 1}$$

つまり、 R_w が 0 に近ければ近いほど、その単語が確からしいと判定する。なお P_{freesyl} の定義上、 R_w は 0 以下となる。

3.2. 統合

文レベルの N-best 候補の情報と単語ラティスの情報を統合して、N-best 候補の更新を行うアルゴリズムを図 2 に示す。

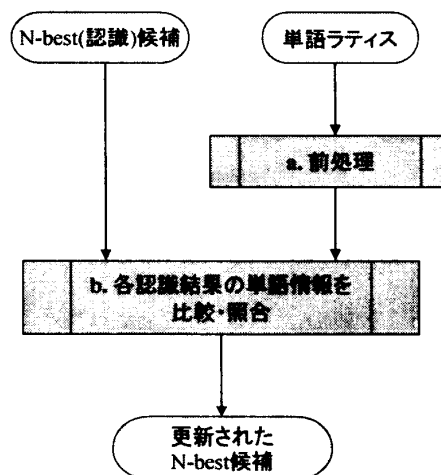


図 2. 統合アルゴリズム

まず、図 2 の「a. 前処理」では、単語ラティスの情報に存在する単語情報をルールに従って選出する。例を図 3 に示す。

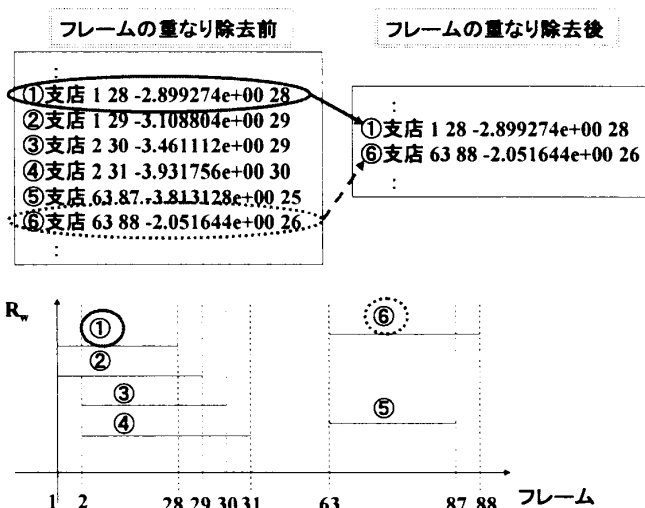


図 3. 前処理

図 3 中の①～⑥の単語は全て同単語名であるが、それぞれ認識されたフレーム区間の開始フレームや終了フレーム、音響尤度比(R_w)などが異なる。ここで、①～④および⑤・⑥は、それぞれ単語の存在するフレーム区間に重なりが見られる。ここで、①～④について

は④の持つ音響尤度比が最大となっているため、④が選出される。同様に⑤・⑥についても、⑥の持つ音響尤度比が最大となっているため、⑥が選出される。

このようにして、フレームの重なりを除去することで、単語境界の曖昧さに捕らわれずにキーワードを選出することができる。

続いて、図2の「b.各認識結果の単語情報を比較・照合」の詳細を、図4に示す。

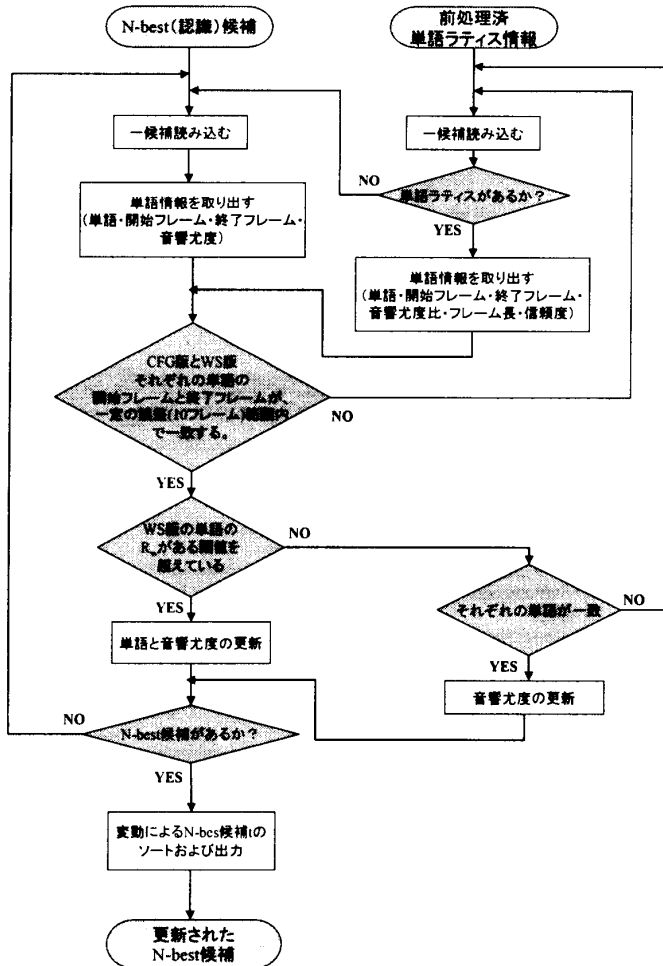


図4. N-best 候補中の単語情報の更新

ここでは、文レベルでの N-best 候補における単語情報と、単語ラティスにおける単語情報とを比較・照合する。

まず、文レベルの N-best 候補に含まれる単語が持つ開始・終了フレームと、単語ラティスに含まれる単語が持つ開始・終了フレームが、10 フレーム(約 87msec)の誤差範囲内で一致するかどうかを判定する。

その後、単語ラティスに含まれる単語が持つ音響尤度比が閾値を超えていれば、文レベルの N-best 候補の単語名・音響尤度と単語ラティスの単語名・音響尤度を置換する。また、単語ラティスに含まれる単語が持つ音響尤度比が閾値を超えていなくても、それぞれの単語名が一致すれば、単語ラティスに含まれる単語の

信頼度を文レベルでの N-best 候補の音響尤度に加算する。

ここで、文レベルの N-best 候補と単語ラティスの情報とを統合する際、文レベルの N-best 候補の単語スコアが単語レベルでの音響尤度であるのに対し、単語ラティスの単語スコアがフレーム正規化された音響尤度比であるので、単語ラティスの音響尤度比をシグモイド関数を使って、0~1 の確率空間に変換して利用することを考える。単語ラティスの尤度比 R_w について、以下の式を用いてシグモイド変換を施すことで、尤度比 R_w の確率値である $\text{sig}(R_w)$ を求めることができる。ここで、 a , b , c は定数とする。

$$\text{sig}(R_w) = \frac{1-c}{1+be^{-(R_w-a)}} - c$$

また、全ての単語について $\text{sig}(R_w)$ を求めるのではなく、 R_w が閾値 $R_{\text{threshold}}$ を超える単語について信頼度 $\log\{\text{sig}(R_w)\}$ を求め、閾値 $R_{\text{threshold}}$ を超えない単語については $\log\{\text{sig}(R_{\text{threshold}})\}$ を求める。

以上より、 $\text{sig}(R_w)$ の \log 値を単語ラティスの信頼度として、文レベルの N-best 候補の音響尤度に加算(確率空間では掛け算)することが可能となる。

ただし、このままではいずれの場合にも該当しなかった文レベルの N-best 候補の音響尤度には、何の処理もなされない。すると、単語ラティスの信頼度が加算された文レベルの N-best 候補の音響尤度は、何の処理もなされない音響尤度よりも低くなる。

そこでその対応として、何の処理もなされない単語については、 $\log\{\text{sig}(R_{\text{threshold}})\}$ を加算する。こうして、文レベルの N-best 候補に含まれる単語全てに対して音響尤度を更新している。

4. 評価実験

CFG 版 SPOJUS と WS 版 SPOJUS との併用による有効性を確認するために、カーナビゲーションにおける目的地設定をタスクとした実験を行った。

今回の実験では、文レベルの N-best 候補に対して単語ラティスの情報を統合した場合の最大誤り改善率、単語ラティスにおける一定条件下での正解単語、文レベルの N-best 候補における正解単語の出現順位の変動を対象として調査を行った。

4.1. 実験条件

実験で使用した音声データについての説明を以下に示す。

まず発話例については、カーナビゲーションにおける目的地設定タスクを前提として、情報系大学学部生・大学院生 10 人を対象にして収集された対話例 84 文を採用した。ただし、この 84 文には想定された文型

から外れた文(非文)や、認識辞書に登録されていない単語(未知語)も含まれる。また、このような非文・未知語を含む文を修正して、文受理率 100%の発話例 68 文を用意した。さらに 84 文と 68 文のそれぞれについて、「え～」や「あの～」といった間投詞を含むタイプを用意し、計 304 発話を用意した。ここで、304 発話について 9 つの発話分類を行ったので、表 1 にその概要を示す。

表 1. 発話分類

間投詞	含まない		含む	
	有り	無し	有り	無し
非文・未知語	○	—	—	—
A	○	—	—	—
B	—	○	—	—
C	—	—	○	—
D	—	—	—	○
E(=A+B)	○	○	—	—
F(=C+D)	—	—	○	○
G(=A+C)	○	—	○	—
H(=B+D)	—	○	—	○
I(=A+B+C+D)	○	○	○	○

発話文に含まれる単語数は 1,728 単語で、話者は 3 名である。なお、音声データのサンプリング周波数は 11.025[kHz]である。

また音声認識器については SPOJUS を用いる。CFG 版 SPOJUS で使用した単語辞書は、語彙サイズ 9,555 単語で、施設名称ドメインが静岡県を対象としている。WS 版 SPOJUS で使用した単語辞書は、語彙サイズ 9,799 単語で、施設名称ドメインが静岡県を対象としている。

それぞれの辞書の語彙サイズが異なるのは、WS 版 SPOJUS の単語辞書には、CFG 版 SPOJUS に登録されている単語に加え、登録されている各単語の部分単語も登録してあるためである。

4.2. 実験結果

今回の実験で得られた結果を以下に示す。

4.2.1. 最大誤り改善率

単語ラティスの情報を用いることで、文レベルの N-best 候補の単語正解率が、最大でどの程度改善することが見込めるかについて調査を行った。

表 2. 最大誤り改善率

		CFG 版		
		正解	不正解	合計
WS 版	正解	1887/5184 [36.4]	1189/5184 [23.0]	3076/5184 [59.4]
	不正解	236/5184 [4.6]	1873/5184 [36.0]	2109/5184 [40.6]
	合計	2123/5184 [41.0]	3062/5184 [59.0]	—

話者 3 名の認識結果について、CFG 版 SPOJUS においては認識した単語列と正解単語列(実際に発話された文)とを照合し、正解・不正解を判定している。WS

版 SPOJUS においては正解単語(実際に発話された単語)が認識結果に含まれているかないかで判定を行った。その結果を表 2 に示す。

表 2 より、文レベルの N-best 候補で不正解となった単語が、3,062 単語ある。そのうち、単語ラティスでは正解となった単語が 1,189 単語あるので、予想される WS 版 SPOJUS での最大誤り改善率は 38.8%となる。

4.2.2. 単語ラティスの分析

WS 版 SPOJUS を利用することで、CFG 版 SPOJUS の単語正解率を改善できる可能性があることを確認できた。次に、単語ラティスにおける、全単語の音響尤度比ごとの正解単語の割合および、WS 版 SPOJUS で認識された単語フレーム長を 50 フレームで区切った場合における正解率を図 5 に示す。なお、ここで言う正解率とは、フレームの重なりを除去した総単語数における、正解単語の割合を示す。

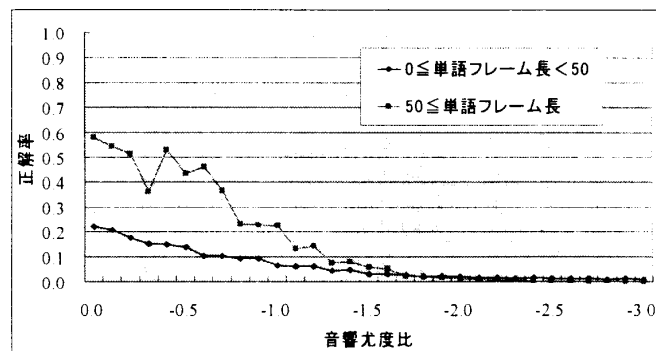


図 5. 正解率

図 5 より、全単語ラティスの中で、正解単語の含まれる割合が音響尤度比-1.0 以上の場合において正解単語の割合が高いと推測される。音響尤度比-1.0 以上の単語のみを考慮すると、過剰検出率は 193.2%となっている。つまり、本実験で用いたスポッティングの単語候補の持つ尤度比の閾値の設定では、のべ発話単語数の約 1.9 倍の候補を含む程度の条件となっている。さらに、図 5 から分かるように単語フレーム長が 50 フレーム以上の場合において正解率が高くなっているため、本研究では音響尤度比の閾値を-1.0、使用する単語の単語フレーム長を 50 以上とした。

4.2.3. 文レベルの N-best 候補出現順位の更新

単語ラティスの分析をもとに閾値を設定し、文レベルの N-best 候補を更新した。更新前と更新後における正解単語の平均出現順位(N-best 候補に含まれる正解単語の、最も高い順位の総和に対しての、N-best 候補に含まれる正解単語数)とそれらの差を表 3 に示す。表 3 中のランクとは、例えば更新前の出現順位が第 10 位であったとき、更新後に第 5 位に上昇すれば「-5」となる。単語数は、出現順位が変動した正解単語数である。

また文レベルの N-best 候補に含まれる単語と単語ラ

ティスに含まれる単語との置換によって、N-best 候補に新規に出現した正解単語数とその割合(N-best 候補に含まれない正解単語の総数に対しての、新規出現した正解単語数)、N-best 候補から欠落した正解単語数とその割合(N-best 候補に含まれる正解単語の総数に対しての、欠落した正解単語数)を表 4 に示す。

表 3. 出現順位の変動

発話分類	更新前	更新後	変動(ランク/単語数)
A	1.30	1.31	0.01(4/1)
B	1.03	1.03	0.00(0/0)
C	1.49	1.49	0.00(0/0)
D	1.81	1.80	-0.01(-5/2)
E	1.13	1.13	0.00(4/1)
F	1.69	1.68	-0.01(-5/2)
G	1.38	1.38	0.00(4/1)
H	1.34	1.33	-0.01(-5/2)
I	1.35	1.35	0.00(-1/3)

表 4. 単語置換による正解単語の出現と欠落

発話分類	新規出現 単語数 (割合)	欠落 単語数 (割合)
A	5 (0.4%)	1 (0.2%)
B	0 (0.0%)	2 (0.2%)
C	4 (0.3%)	1 (0.3%)
D	2 (0.6%)	1 (0.2%)
E	5 (0.4%)	3 (0.2%)
F	6 (0.4%)	2 (0.2%)
G	8 (0.4%)	2 (0.2%)
H	2 (0.6%)	3 (0.2%)
I	11 (0.4%)	5 (0.2%)

表 3 より、発話パターン D(間投詞含む、非文・未知語なし)において、正解単語 2 単語の出現順位の変動の和が「-5」となり、出現順位を向上することができた。また発話パターン A(間投詞含まない、非文・未知語あり)において、正解単語 1 単語の出現順位の変動の和が「4」となり、正解単語の出現順位が低下した。全体(表中の発話パターン I)では、正解単語の出現順位が上昇した。ただし、更新前後の N-best 候補の中で正解単語に該当する 2,366 単語のなかで、出現順位が変動したのは 3 単語のみである。

表 4 より、更新前の N-best 候補に出現していなかった正解単語に該当する 2,813 単語のうち、更新後の N-best 候補では 11 単語が新たに正解単語として出現した。これらは発話パターン B(間投詞含まない、非文・未知語なし)を除いた全ての発話パターンで新たに出現しており、いずれも N-best 第 1 位の単語と置換されている。なお、発話パターン B は、発話パターン A,B,C,D のなかで単語正解率が一番高かった。一方、更新前の N-best 候補に出現していた正解単語に該当する 2,371 単語のうち、5 単語が N-best 候補から欠落した。いずれも N-best 第 1 位の単語と置換されている。

本結果より、間投詞を含む発話や非文・未知語のある発話に対して、本提案手法が有効であることを確認できた。今後は、単語ラティスにおける正解単語の割

合を高めることが必要となる。また、単語ラティスにおける信頼度の生成や、フレーム重なりでの除去におけるパラメータの調整などを行うことで、正解単語の N-best 候補出現順位の向上や、文レベルの N-best 候補に正解単語を新規に出現させることのできる、有効な単語ラティスを選出する必要がある。

5. まとめ

頑健な言語理解のための文法とワードスポッティングを併用した音声認識手法について検討した。提案手法では、単語ラティスの情報を文レベルの N-best 候補と統合することで、N-best 候補の正解単語出現順位の上昇を行った。20-best での正解単語出現順位の変動を調べ、正解単語のランクが向上したことが確認できた。今後はより多くの N-best 候補を対象としていく。これにより多くの単語出現順位の変動を確認することができると予想されるため、それに対して単語ラティスの有効性を確認する。

また、CFG 版 SPOJUS で用意されている CFG から、サンプル対話を逆生成し、生成された擬似対話文を用いて bi-gram を学習させ、それを用いて統計情報を用いた音声認識を行い、文レベルの N-best 候補と単語ラティスともに、統計情報による認識結果を用いた手法についても検討する。

最後に、WS 版 SPOJUS の単語辞書に、登録されている各単語の部分単語(準キーワード)を登録し、その部分単語が認識された場合に、その部分単語を含む単語の尤度を上昇させることなどを検討する必要がある。

文 献

- [1] 安田宜仁, デネック・マティアス, “未知語を許容する対話システムにおける対話状態予測”, 情報処理学会研究報告, 2005-SLP-057, pp.105-109, 2005.
- [2] Frank K SOONG, Wai-Kit LO, 中村哲, “連続音声認識候補受理/リジェクションのためのワードスポッティング仮説検証手法”, 情報処理学会研究報告, 2003-SLP-049, pp.287-292, 2003.
- [3] 駒谷和範, 鹿島博晶, 田中克明, 河原達也, “複合的言語制約に基づくキーフレーズ検出を用いた汎用的なデータベース検索音声対話プラットフォーム”, 情報処理学会研究報告, IPSJ-JNL4405018, pp.1333-1342, 2003.
- [4] 李晃伸, 河原達也, 堂下修司, “単語 N-gram と段階的探索を用いた大語彙連続音声認識”, 情報処理学会研究報告, 1997-SLP-016, pp.19-24, 1997.
- [5] 水野智士, 高木浩吉, 小暮悟, 甲斐充彦, 伊藤敏彦, 小西達裕, 伊東幸宏, “頑健な意味理解のための音声認識信頼度と対話履歴を利用した発話意図推定手法”, 情報処理学会研究報告, 2005-SLP-055, pp.77-82, 2005.
- [6] 中川聖一, 甲斐充彦, “文脈自由文法制御による One Pass 型 HMM 連続音声認識法”, 電子通信学会論文誌, Vol.J76-D-II, No.7, pp.1337-1345, 1993.