社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2007－ＨＩ－122 (13)
2007－SLP－ 65 (13)
2007／2／10

文脈情報と応答内容を用いた単語生起確率の動的生成手法に関する検討

岩崎 祥範† 小暮 悟† 伊藤 敏彦‡ 甲斐 充彦†† 小西 達裕† 伊東 幸宏†

†静岡大学情報学部 〒432-8011 静岡県浜松市城北 3-5-1

††静岡大学工学部 〒432-8561 静岡県浜松市城北 3-5-1

‡北海道大学情報科学研究科 〒060-0814 北海道札幌市北区北 14 条西 9 丁目

E-mail: †cs1014@s.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 近年、音声認識技術や言語処理技術、コンピュータ性能の向上により、音声操作可能な高度情報システムの車内利用が現実のものとなっている。しかしながら、カーナビゲーションシステムに代表される車内音声対話システムの多くは、ユーザの発話形式が自然発話であることや、走行ノイズなどの様々な雑音環境下での利用により、認識誤りは避けて通れない問題となっている。そこで、本稿では対話状況に応じた文脈情報と応答内容を用いて次発話において発話される可能性の高い単語を予測し、それらの単語の単語生起確率を上昇させ、認識結果のN-best中に正解単語を出現させやすくすることで認識精度の向上を目指す。評価実験の結果、手法を用いない場合に比べて提案手法を用いた場合に単語正解率が83.5%から85.1%に上昇し、提案手法の有効性が示せた。

A method of dynamically generating word occurrence probabilities according to the contextual information and the system response

Yoshinori Iwasaki†, Satoru Kogure†, Toshihiko Itoh‡, Atsuhiko Kai††,
Tatsuhiro Konishi†, Yukihiro Itoh†

†Faculty of Informatics, Shizuoka University 3-5-1 Johoku, Hamamatsu, Shizuoka, 432-8011 Japan

††Faculty of Engineering, Shizuoka University 3-5-1 Johoku, Hamamatsu, Shizuoka, 432-8561 Japan

‡Graduate School of Information Science and Technology, Hokkaido University
Kita-14, Nishi-9, Kita-ku, Sapporo, Hokkaido, 060-0814 Japan

E-mail: †cs1014@s.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 Recently, the technology of speech recognition and natural language processing, and the performance of computer calculation ability has been highly improved, so we can utilize speech interface to handle information service in car. Cars' spoken dialogue systems like existent navigation system, however, often misrecognized user utterances. In this paper, the system predicts the frequency uttered word using the contextual information and the system response, and raise word occurrence probabilities of those words. As a result, we make the correct answer word appear in the recognition result easily. As a result of the evaluation experiment, the word recognition rate rose from 83.5% to 85.1% according to use the proposal method. We show the effectiveness of the method.

1. はじめに

現在、音声認識や言語処理技術の発展とコンピュータの計算能力の向上により、人間同士が日常的に行っているような自然な対話を人間とコンピュータの間でも実現しようとする研究が盛んに行われている。単純な対話を行う音声対話システムであれば、一般電話による検索予約システムや、携帯電話におけるナビゲーションなどがすでに実用化されている。そのなかでも自動車内で利用可能な音声操作カーナビゲーションシステム(以下、カーナビ)は実用化の顕著な例と言えるだろう。運転中は、運転という主となる行為があるため、リモコン操作など

の手動での操作よりも音声を用いた操作のほうが安全である。最近では走行中のカーナビ操作を音声操作のみでしか出来ないものも多く、実際多くの自動車メーカーがカーナビの音声操作の研究開発を続けている。しかし、現存する音声カーナビのほとんどは決められた単純な発話の入力を想定しており、今後求められていくと予想される自然な音声インタフェースを実現するには、自然発話(話し言葉)の理解を的確に行う必要がある。

しかし、カーナビで自然発話を理解するには、間頭詞や言い直しのような自然発話特有の現象が生じることや、走行ノイズなどの様々な外的要因から、認識誤りは避けて通れない問題となっている。また、既存のカーナビにおける現在の音声対話では、誤認識から即座に回復できるような頑健さはなく、ユーザ発話の意図を正しく理解できた場合よりも対話が円滑に進まなくなり、ユーザに不快感を与えることになる。このようなことから、ユーザの発話意図を正しく理解し、円滑な対話をするためには認識率の向上が重要と考えられる。

音声認識性能の向上に関する研究としては、ユーザ発話の言語モデル、およびユーザの内部状態に適応した言語モデルの採用による認識率改善の研究[1]や、ある時点でのシステムの理解状態を用いて、ユーザ要求の確率分布を推定する研究[2]や、音素数が少ない単語や出現頻度が低い単語は認識誤りとなりやすいという傾向から、単語正解確率のモデル化を行い、このモデルを用いて認識単位を最適化する手法の研究[3]などがある。また、次発話の予測を文単位で行い、認識対象語彙サイズを抑制した認識手法による研究[4]などもされている。

一方、我々はユーザ発話について、単語が発話された可能性を示すスコアを、それまでに話された内容(文脈情報)を使って増減する手法を提案している[5][6][7]。音響的確率と言語的制約から求まる尤度(確からしさ)で順位付けされた、複数候補文の集まり(以下、N-best)から単語信頼度[1]を計算し、さらに対話の文脈を使ってユーザ発話を理解することで、音声対話全体の精度を向上させている。しかし、この手法では、N-best中に出現しない単語はシステムの理解結果に含まれないため、いくら高精度の言語理解処理が実装されていたとしても、そのような単語を理解結果として復活させることは難しい。

そこで、本研究では対話状況を保存している文脈情報と前回発話に対する応答内容を用いて次発話において発話される可能性の高い単語を予測し、それらの単語の単語生起確率を相対的に上昇させ、認識結果のN-best中に正解単語を出現させやすくすることで認識精度の向上を目指す。提案手法を用いない場合と用いた場合で、音声認識率、単語信頼度などを比較し、提案手法の有効性を調べた。

2. 文脈情報と応答内容

2.1. システムのタスク

本研究のタスクは、目的地検索・設定を行うカーナビである。検索対象となる目的地は、静岡県、および静岡県に隣接する県のインター・駅・飲食店・コンビニ・遊園地などの43,437施設である。これらの施設を目的地に設定する際には、「静岡県」、「浜松市」のような地理情報や、「コンビニ」、「ローソン」のようなジャンル情報を用いて検索すること、および「浜松市役所」、「ガスト浜松住吉店」のように施設名を直接発話して目的地に設定することも可能である。また、「静岡県浜松市のファミレス」のように、地理情報、ジャンル情報を一度に入力する発話や、「静岡県」、「浜松市のファミレスを検索して」のように複数回に分割した発話も理解可能である。なお、システムのYES/NO回答要求に対して、肯定語(「はい」など)や否定語(「いいえ」など)を発話することが可能である。

発話可能な単語のうち、地理情報、ジャンル情報、施設名はそれぞれクラスに属し、いくつかのクラスをまとめてカテゴリを定義している。本研究で用いたカテゴリの概念階層図を図1に示す。また各カテゴリの単語例を図2に示す。

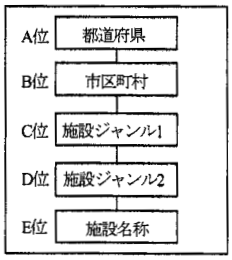

図1：概念階層図

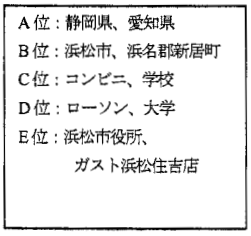

図2：カテゴリ別単語例

2.2. システム概要

現行システムは、図3のように音声認識部、信頼度生成部、言語理解部、応答生成部、GUI表示部から構成されている。

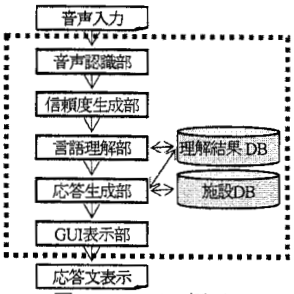

図3：システム概要

2.3. 音声認識部

音声認識部では、あらかじめ用意されている文法と語彙辞書を使って音声認識する。音声認識の出力は、音響的確率と言語的確率から求まる尤度で順位付けされた、複数候補文の集まりであるN-bestとして得られる。本研究では音声認識器に豊橋技術科学大学で開発されたSPOJUSを用いて、20-bestまでの結果を利用している。

[1] ある単語が発話された可能性を示す値

2.4. 信頼度生成部

信頼度生成部では、認識結果として受け取ったN-best中に出現する全ての単語について、その単語が出現した文の音響的な尤度の総和から事後確率を推定した値(以下、単語信頼度)として生成する。つまりN-best中の出現頻度が多い単語ほど単語信頼度が高くなる。ここで生成された信頼度は言語理解部で利用される。

2.5. 言語理解部

言語理解部では、単語信頼度付きのN-bestと文脈や単語の依存関係を利用して、単語ごとに、その単語がそれまでの対話で発話された可能性を示す単語スコアを生成する。この単語スコアは、応答生成において応答戦略を決定するための重要なパラメータである。このスコアは、(1)対話履歴中の単語と、(2)最新の認識結果中の単語のそれぞれで異なる戦略を用いてスコアの生成、更新が行われる。(2)の単語は、最新認識結果のN-bestに含まれる全単語が対象となる。単語スコア生成は、言語理解部が最新の認識結果を獲得するたびに、(1)(2)の順に単語スコア生成が行われ、(2)の単語スコアは対話履歴中の単語スコアへと更新される。この時(1)(2)の両方でスコアが生成された単語については、高いスコアの方を対話履歴中の単語スコアに更新する[5]。

さらに、単語スコアと発話タイプ[2]によってカテゴリ理解を行う。カテゴリ理解とは、それまでにどのカテゴリの組み合わせ(以下、カテゴリパタン)が発話されたかを求める処理である。例えばユーザが「浜松市のローソン」と発話した場合の正解カテゴリパタンは「B+D位」となる。

カテゴリ理解では、まず、(A)対話履歴中の単語スコアを用いて「B位」や「B+D位」などのカテゴリパタンに、そのカテゴリパタンがこれまでの文脈の中で発話されている可能性を示す値(以下、カテゴリスコア)をそれぞれ求める。例えば本研究では、カテゴリ数は 5 なので、1 つ以上のカテゴリが発話されている場合の全組み合わせは、${}_5C_1+{}_5C_2+{}_5C_3+{}_5C_4+{}_5C_5=5+10+10+5+1=31$ 通りとなり、カテゴリパタンごとにカテゴリスコアを求める[6]。カテゴリパタンCPのカテゴリスコアS(CP)は式(1)によって求める。

$$S(CP)=\frac{\sum_{i=1}^{5}\delta_i^{CP}\max_j S(w_j^{c_i})}{1+\alpha(\sum_{i=1}^{5}\delta_i^{CP}-1)} \quad \cdots\cdots\cdots(1)$$

$$\delta_i^{CP}=\begin{cases}1 & \cdots c_i\in CP\\ 0 & \cdots c_i\notin CP\end{cases}$$

$\max_j S(w_j^{c_i})$は認識結果中に存在する、カテゴリ c_i に所属する単語 w_j の中で一番単語スコアが高い単語の単語スコアである。ただし、各カテゴリの w_j 同士は言語制約を満たすものとする(例えばA位が「静岡県」でB位が「豊橋市」は言語制約を満たさない)。

次に、(B)対話履歴の単語スコアや前回の応答内容を参考に、最新のユーザ発話が前回の応答で使用した単語(式(1)の w_j)それぞれに対してどのような発話タイプ(詳細化や回答など)であるのかを判定する。ただしこの発話タイプの判定で、判定用のパラメータが基準に満たない場合は判別不能という状態とする。(C)発話タイプが詳細化や回答などに判別されたカテゴリを含む組み合わせには、(A)で計算されたカテゴリスコアの判別結果に沿った増減を行う[6]。このカテゴリスコアの増減によってカテゴリ理解精度を上げている。最後に、(D)全カテゴリパタンの中から最も高いカテゴリスコアを持つカテゴリパタンをカテゴリ理解結果とする。この単語スコアとカテゴリ理解結果を次の応答生成部で利用する。

2.6. 応答生成部

応答生成部では、言語理解部で生成された単語スコアとカテゴリ理解結果、および可能ならば検索から得られる検索結果を用いて応答戦略を決定することで最終的にユーザに提示するための応答文が生成される。本研究における応答生成部では、以下の戦略に基づいて応答を生成している。以下の単語が信頼できるかどうかには、単語スコアの値を用いて判定しているが、単語スコアによる単語正解分布から、80%が正解となる単語スコア 1.2 を信頼度が高いと判定する閾値として採用している。

1. ある単語が十分信頼できる場合、確認無しで対話を進める。
2. ある単語がある程度信頼できる場合、暗黙的確認[3]をすることで対話を進める。
3. ある単語が信頼できない場合、応答には使用しない。

また、タスクに検索が含まれるため検索結果件数によって応答を変える必要がある。各検索件数における戦略は以下のようになっている。

- **None（0 件）**：検索結果が得られない場合、実際に検索した結果を提示して条件変更を促す。
- **Only（1 件）**：ユーザの要求に対し唯一の結果が得られた場合、その結果を提示してユーザの意向を問う。
- **Some（2〜3 件）**：検索結果がある程度絞り込まれているため、結果を提示してユーザに選択を促す。
- **Many（4〜10 件）**：検索結果が多くユーザに選択を促すのは現実的ではないが、結果を提示して更なる絞込みを行うように促す。
- **Too Many（11 件以上）**：検索結果が多過ぎるためユーザに提示するのは現実的ではない上、結果自体に重要な意味を見出せないため、結果は提示せず更なる絞込みを行うよう促す。

さらに、絞込みを行うように促す応答では、ユーザが何を言ったら良いか迷わないように、システムが必要としている絞込み条件をユーザに提示する。提示する条件を決定するためのアルゴリズムは以下のようになる。

1. 言語理解結果から未知のカテゴリを判定する。
2. 未知のカテゴリが絞込み条件として使えるか判断する。

---

[2] ユーザの発話を、システム応答との関連から分類したもの。(詳細化、訂正、回答、再入力)

[3] 例えば「浜松市」のように、発話されたと予想される単語を復唱する。

3. 絞込み条件として使えるカテゴリが 1 つの場合はその条件を聞く「○○を追加してください」。条件が無い場合は、基点となる情報を聞く「どこの近くですか？」。
4. 理解内容が A 位のみの場合はジャンル情報を聞く「施設のジャンルを追加してください」。C 位のみの場合は地理情報を聞く「市区町村を追加してください」。A+C 位の場合は両方を聞く「市区町村か詳細なジャンルを追加してください」。

3. 単語生起確率の動的生成

3.1. 現行システムの問題点

現行システムでは、単語スコアの増減はN-best中に含まれる単語に対してのみ行われるので、発話した単語がN-bestに出現しなければその単語を理解することは不可能となる。先行研究[5][6][7]において、頑健な音声対話システムの構築を目的に様々な研究がされてきたが、すべての研究で利用されているのが"単語信頼度"である。このシステムの主要素とも言える"単語信頼度"を得るためには、入力発話中の単語をいかに正確に認識するか、ということが重要であると言える。そこで、本研究では音声認識精度向上のための認識手法を提案する。

3.2. 従来の認識手法と提案手法

通常、確率文脈自由文法(以下、SCFG)を文法制約とする音声認識において、ある単語 W が発話された可能性を示す対数尤度は、音響対数尤度 logP(A|W) と言語対数尤度 logP(W) の重み付き和で求められる[8]。

$$L(W^{\sim}) = \log P(A \mid W) + \alpha \log P(W) \quad (2)$$

式(2)の α は言語重みである。音響モデルの最小単位が音素または音節、言語モデルの最小単位が単語であることから、言語尤度に重みを付けるのが一般的である。本実験では言語重み α は20を用いている。

現行システムでは、SCFGではなく、CFGを文法制約に用いていた。この認識手法では、ある単語が"受理可能"であれば式(2)の P(W) (以下、生起確率)は1となる。これは対話の文脈を考慮していないため、どのような対話状況においても、受理可能な単語については等確率で認識されることを意味する。しかし文脈情報と応答内容によって、次に発話される単語をある程度予測できるのではないかと考えた。ユーザ発話がシステムの応答に対してある程度協調的であれば、次の発話でユーザが発話しそうな単語を予測することができる。ただし、予測できる状況だからといって、他の単語が全く認識されないということは避けなければならない。ユーザは必ずしもシステムの応答に対して協調的な発話をしてくれるとは限らないからである。さらに、応答に使った単語群が信頼できるほど、次に発話されやすい単語の予測が信頼できると考えられる。

そこで、音声認識において文脈情報と応答内容を利用して、発話される可能性が高い単語を予測して認識性能を向上させる手法を提案する。具体的には、予測単語の生起確率を1とし、予測単語以外の単語の生起確率は、単語スコアやカテゴリスコアなどの文脈情報を用いて計算される 0<P(W)≦1 の値を用いる。これによって、予測単語以外の単語は認識されにくくなり、予測単語の認識精度が向上すると考えられる。

本研究では、特に、以下の4つの応答タイプ時に、次に発話されやすい単語を予測することにする。

応答タイプ 1：
・目的地設定の最終確認をする応答
例：「浜松駅を目的地に設定してよろしいでしょうか？」
・予測単語：YES/NO 単語(はい、いいえ)

応答タイプ 2：
・検索結果 some の選択要求の応答
例：「静岡インターと清水インターがあります。どの施設を目的地に設定しますか？」
・予測単語：選択肢の単語(静岡インター、清水インター)

応答タイプ 3：
・検索条件追加を促す応答(カテゴリ指定)
例：「浜松市ですね。施設のジャンルを追加してください」
・予測単語：指定カテゴリに所属する単語(コンビニ、ファミレス)

応答タイプ 4：
・検索条件追加を促す応答(基点情報)
例：「どこの近くですか？」
・予測単語：ランドマーク(浜松駅、浜松インター)

また、本手法では生起確率を決定するための指標として、文脈情報を総合的に反映した値を用いる。ここでは、カテゴリスコアと検索単語の平均単語スコア(以下、平均スコア)を用いた手法をそれぞれ検討した。まず、それぞれのスコアを確率相当(0～1 の範囲)の値に変換して、予測単語以外の単語の生起確率とする(予測単語の生起確率は 1 のままにする)。また、本手法を実行する場合の指標となるカテゴリスコアや平均スコアが小さい場合は、信頼性が低いと判断できるので、本手法を用いると悪影響が大きいと考えられる。よって、単語スコアの信頼できる値である 1.2 を閾値に採用して、単語スコアが 1.2 以上の場合に式(3)を用いて、予測されなかった単語の生起確率を推定した。

$$P(S(CP)) = \frac{\alpha}{1+e^{\beta(S(CP)-\gamma)}} + \delta \quad (\alpha,\beta,\gamma,\sigma：定数) \quad (3)$$

式(3)はカテゴリスコアや平均スコアに関して単調減少となる関数であるが、これはそれぞれのスコアが高い場合に、予測単語以外の単語が認識されにくくなるような生起確率を推定するためである。

このように本手法では、前回発話の応答に応じて予測単語を判定し、全辞書単語の生起確率を動的に生成していく。なお、ここで用意した単語生起確率は、認識時に単語選択時に対象となる単語集合の総和が 1 になるように正規化している。

4. 評価実験

本手法と通常の認識手法の結果を比較するために、評価実験を行った。今回の実験では、音声認識精度を評価するために、単語正解率、単語信頼度、N-best における正解単語の出現数と順位の変動を対象として評価を行った。

4.1. 実験方法

評価データとして、これまでにわれわれが行ってきた研究で収集した音声データから模擬対話を作成した。対話の形式は、第 1 発話-システム応答 1-第 2 発話(前回発話：第 1 発話、評価対象発話：第 2 発話)の形式と、第 1 発話-システム応答 1-第 2 発話-システム応答 2-第 3 発話(前回発話：第 2 発話、評価対象発話：第 3 発話)の形式の 2 種類がある。本実験では、前回発話の応答生成後に予測単語の決定と生起確率の付与を行い、評価対象発話の認識結果を調査する。

また、本評価の対象となる発話は、「前回発話の検索結果が正解かつ次発話が応答に対して協調的な発話」と、「前回発話の検索結果が正解かつ次発話が応答に対して協調的ではない(以下、非協調的)発話」となる。前述した応答タイプをもとに分類すると、8 つの発話状況に分類される。

表 1：発話状況

応答タイプ	協調的	非協調的
1	A	A'
2	B	B'
3	C	C'
4	D	D'

それぞれについて 30 個の模擬対話を用意し、全 240 発話を評価の対象とする。

4.2. 実験結果

本実験で得られた結果を以下に示す。

4.2.1. 単語正解率と単語信頼度

まず、単語正解率と正解単語の単語信頼度について調査した。ここで言う単語正解率とは、正解単語が信頼度付 N-best の 1 位の文中に含まれている割合を示している。(表中、CS：カテゴリスコア、AS：平均スコア)

表 2 を見ると、本手法を用いた認識の方が、全体的に単語正解率が上がっていることがわかる。協調的発話(A,B,C,D)では予測単語が発話されるため認識精度が向上し、非協調的発話(A',B',C',D')では予測単語以外が発話されるため認識精度が低下すると予想していたが、実際には、非協調的発話(A',B',C',D')において悪影響が小さく、協調的発話(A,B,C,D)において有効的な結果が得られた。悪影響が小さかったことの要因としては、本手法において非協調的である予測単語以外の単語というのは、予測単語に比べて認識しにくくなるが、全く認識できないわけではなく、N-best 中の単語の信頼度が全体的に下がる傾向にあるが、正解単語を含む候補文の順位が下がるほどの影響力が無かったからだと考えられる。

A と A'で正解率に大きな差があるのは、A の正解単語である「はい」や「いいえ」といった単語は、単語長が短く誤認識しやすい傾向があり、一方の A'の正解単語は、比較的認識しやすい施設名であったことが原因となっている。

表 2：単語正解率[%]

発話状況	手法なし	CS	AS
A	53.3	53.3	**56.7**
B	86.7	**93.3**	**93.3**
C	73.3	73.3	**76.7**
D	70.0	70.0	70.0
協調的	70.8	72.5	**74.2**
A'	100.0	100.0	100.0
B'	97.1	97.1	97.1
C'	**93.3**	90.0	90.0
D'	94.4	**97.2**	**97.2**
非協調的	**96.2**	96.1	96.1
全体	83.5	84.3	**85.1**

表 3：正解単語の平均単語信頼度

発話状況	手法なし	CS	AS
A	0.7468	0.7601	**0.8008**
B	0.7892	0.8040	**0.8070**
C	0.7453	0.7454	**0.7524**
D	**0.3624**	0.3579	0.3531
協調的	0.6609	0.6669	**0.6783**
A'	**0.7907**	0.7818	0.7872
B'	0.8256	**0.8312**	0.8257
C'	**0.7426**	0.7413	0.7322
D'	0.7230	0.7302	**0.7306**
非協調的	0.7705	**0.7711**	0.7689
全体	0.7157	0.7190	**0.7236**

表 3 を見ると、平均信頼度に関しても全体的に向上している。特に A において効果が顕著であり、上述したように誤認識しやすいと考えられる「はい」や「いいえ」などの短い単語の認識において効果的であることがわかった。

認識精度が悪くなると予想していた D'で、単語正解率が上がった要因としては、本手法が単語単体を予測しているため、2 単語以上含まれる N-best 候補文は順位が下がりやすいという傾向があり、2 位にあった正解単語単体の候補文が 1 位に上昇したデータがあったことが挙げられる。また、単語信頼度が上がった要因としては、本研究で用いている単語信頼度が、単語が出現した文の音響的な尤度の総和から事後確率を推定した値であるため、上述した傾向から正解単語を含んだ候補文が増えているデータがあったことが挙げられる。

また、協調的発話で認識精度が悪くなってしまった D については、手法を用いない場合に単語信頼度が 1 で、手法を用いたもの(カテゴリスコアと平均スコア共に)で

約0.65に落ちているデータがあった。これも上述したような傾向の影響により、手法を用いない場合に出現していた正解単語を含む単語数5個の候補文が、本手法では出現しにくくなったことが原因となっている。しかし、極端に単語信頼度が落ちてしまったデータを除いた場合、本手法を用いた方が、単語信頼度が上がっていることを確認している。

4.2.2. N-best中の正解単語の出現状況

次に、正解単語がN-best中に新規出現した文数と文順位の変動を調べた。文順位の変動は、正解単語を含む最上位の候補文の順位を基準に、順位が1つ上がるごとに「+1」、1つ下がるごとに「-1」をした値で評価した(ただし、同発話状況で、N-best中の正解単語の新規出現と脱落、あるいはN-best中の正解単語の文順位が上がった文と下がった文は同時に出現していない)。

表4：正解単語の出現状況

	手法なし	CS		AS	
発話状況(総文数)	出現文数	出現文数(新規)	順位変動	出現文数(新規)	順位変動
A (30)	18	**20(2)**	**1**	**20(2)**	**1**
B (30)	29	29(0)	**3**	29(0)	**3**
C (30)	24	24(0)	0	**25(1)**	0
D (30)	28	28(0)	0	28(0)	0
協調的 (120)	99	**101(2)**	**4**	**102(3)**	**4**
A' (30)	30	30(0)	0	30(0)	0
B' (30)	29	29(0)	0	29(0)	0
C' (30)	29	29(0)	**-1**	29(0)	**-1**
D' (30)	30	30(0)	**1**	30(0)	**1**
非協調的 (120)	118	118(0)	0	118(0)	0
全体 (240)	217	**219(2)**	**4**	**220(3)**	**4**

表4を見ると、本手法を用いた場合、N-best中の正解単語の新規出現文数と順位の変動に関して、C'において、順位が1つ下がるデータがあったが、全体として効果的であることがわかる。特に、新規出現した正解文はカテゴリスコアを用いた手法で、出現順位2位が1個、4位が1個あり、平均スコアを用いた手法で、出現順位2位が2個、1位が1個といずれも上位に出現していることがわかった。

4.2.3. カテゴリスコアと平均スコア

本研究では、文脈情報を総合的に反映した値としてカテゴリスコアと平均スコアの2つを用いて、それぞれ評価結果を得たが、全体の結果を比較すると平均スコアの方が有効的な結果となった。カテゴリスコアは対話が進むごとに値が大きくなり易いということと、理解カテゴリ数が多い場合に値が大きくなり易いという傾向があり、結果として分散が大きいと考えられる。カテゴリスコアと検索正解率(ある発話において正解単語が検索に用いられた割合)の相関は0.47であった。一方、平均スコアは、先行研究によってある程度信頼できると示されている単語スコアを平均したものであり、カテゴリ数や発話数による影響が小さいと考えられる。検索正解率との相関は0.56であった。

5. まとめ

本稿では、対話状況に応じた文脈情報と応答内容を用いて次発話において発話される可能性の高い単語を予測し、それらの単語の単語生起確率を上昇させる認識手法を提案し、提案手法を用いた音声認識と、手法を用いない通常の音声認識の比較を行うための評価実験を行った。実験の結果、提案手法を用いた場合に単語正解率が83.5%から85.1%に上昇し、N-best中の正解単語の出現状況に関しても新規出現文数が増え、順位も上昇したことが確認できた。

今後の課題として、本研究の評価対象発話が、ユーザ発話の2発話目と3発話目のみであったため、さらに対話が進んだ状況を考慮した生起確率の算出方法を検討する必要がある。また、本研究では単語の予測のみ行っているが、文法の予測を組み合わせることによって、さらに認識精度が良くなるのではないかと考えられるため、文法の予測についても検討する必要がある。

参考文献

[1] 藤崎博也, 阿部賢司, 黒川一滋, 武田和也, 成澤修一, 大野澄雄:"話者の心の状態遷移モデルに基づく対話音声認識", 情報処理学会研究報告, SLP-35, pp.79-84, 2001

[2] 安田宜仁, 堂坂浩二, 相川清明:"タスク適応型高効率対話制御法", 情報処理学会報告, SLP-35, pp73-78, 2001

[3] 篠崎隆宏, 古井貞熙:"話し言葉音声認識における認識率の変動要因の分析と認識単位の設計", 話し言葉の科学と工学ワークショップ 講演予稿集, pp.59-64, 2002

[4] 河上まきほ, 西田昌史, 堀内靖雄, 市川熹:"予測文と部分単語認識の併用による音声対話システムの検討", 情報処理学会報告, SLP-64, pp.43-48, 2006

[5] 藤原敬記, 伊藤敏彦, 荒木健治, 甲斐充彦, 小西達裕, 伊東幸宏:"認識信頼度と対話履歴を用いた音声言語理解手法", 電子情報通信学会論文誌D, Vol.J89-D No.7, pp.1493-1503, 2006

[6] 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, SLP-55, pp.77-82, 2005

[7] 高木浩吉, 小暮悟, 伊藤敏彦, 甲斐充彦, 小西達裕, 伊東幸宏:"単語信頼度と検索結果を利用した協調的応答戦略", 人工知能学会研究会, SIG-SLUD-A503-07, pp.33-38, 2006

[8] 中川聖一:"確率モデルによる音声認識", 電子情報通信学会, 1988