音声言語情報処理　4－2
(1994. 12. 15)

ワードスポッティングに基づく意図抽出

亀山　晋　　中里　収　　白井　克彦

早稲田大学 理工学部 電気工学科
〒169　東京都新宿区大久保3-4-1

あらまし 自由発話の音声認識手法としてワードスポッティングを用い、その認識結果であるキーワードラティスから、タスクに依存したユーザの発話意図を抽出する方法について述べる。本手法は、キーワードラティスから、単語カテゴリ間の生起順序を考慮した共起頻度を用いて可能単語列候補を推定し、その単語列候補をあらかじめ分類された発話意図のどれかに、単語カテゴリと意図との関係度により数値的に結び付けるというものである。意図解釈に関して、テキスト入力による評価実験では最高で97.2% の意図解釈成功率を示し、音声認識部のワードスポッティング結果を用いた評価実験では、最高で83.3%の意図解釈成功率を示すことができた。

キーワード　発話意図・ワードスポッティング・単語共起

Extracting User's Intention from Key-word Lattice

Susumu Kameyama　　*Shu Nakazato*　　*Katsuhiko Shirai*

Department of Electrical Engineering, Waseda University
3-4-1 Okubo, Shinjuku-ku, Tokyo, 169 Japan
e-mail: kameyama@shirai.info.waseda.ac.jp

Abstract　In this paper, we describe a method of extracting the task oriented user's intention from key-word lattice which is a recognition result of spontaneous speech by using word-spotting. The possible word sequence is presumed by co-occurrence degree between word categories, and determine most reliable one to the user utterance. Then the word sequence is classified into the user intention according to the relation degree between the word category and the intention. The intention interpretation success rate was shown 97.2% in the maximum which is obtained by the evaluation experiment using text input. And 83.3% correct rate was achieved in the evaluation experiment by using word-spotting result.

Keywords　**intention of user, word-spotting, co-occurrence of word**

1. はじめに

ユーザの発話方法に制約を与えない、自由発話音声を扱える音声対話システムの構築を進めている[1]。ユーザの発話に可能な限り制約を与えないためには、音声対話に生ずる様々な現象に対処しながら発話を認識する必要がある。このような対話システムを実現するためには、音声認識手法として、自由発話の音声認識に向いているとされているワードスポッティング法を採用することが有効である。

スポッティングの結果において、異区間でしかも長さが異なる単語同士のスコアを比較することは容易ではない[3]という問題がある。最近では、スポッティングの方法にも様々な工夫がなされるようになってきており、garbageモデルを用いる方法[4]、発話全体にわたってヒューリスティック言語モデルを用いて評価するもの[5]、認識の対象として単語ではなくより長い単位であるフレーズを用いるもの[6]などがある。

現実には、音響的なレベルの知識のみを用いてボトムアップに認識を行なうことには限界がある。一般に音声認識に誤認識はつきものであり、この誤認識の修正には言語的あるいは意味的レベルの知識を用いる必要がある。言語的な知識としては、あるタスクの中で生ずる言語の一般的制約と対話の局面で起こる談話構造を反映したものがある。また単語列からその意味する内容を調べる時には、自然言語的なアプローチにより記号処理的に処理を進めることが多い。

ところが、この言語的知識をもってしても誤認識を完全に解消することはできない。よって、ワードラティスから最尤文仮説を得るときにも、誤った単語の挿入や正しい単語の脱落が起きることは避けられず、このような誤った単語列に対してロバストにその意図を解釈できる意図解釈の方法が望まれる。さらに、そのような処理のためには文法などの知識以外に意味解釈の手段が必要となる。

本研究は、上記のような問題に対処することを最終的な目的とするが、現実的には生ずる現象や問題の全てに十分な対応をとることは不可能であるから、現段階では対話がある程度満足に成立することを目標にシステムを構成することを考える。本稿では、主にワードスポッティングによる認識結果であるキーワードラティスからユーザの発話意図を抽出する手法について述べる。

意図解釈には、言語的あるいは意味的な知識を加えたスコアリング手法を導入する。一般的言語知識としては、単語と単語同士の共起関係を用い、単語をその意味する内容に従ってカテゴリに分類することで意味を表現する。また、抽出されるべき意図の種類があらかじめ与えられるものとして、発話の意味を定量的に扱う方法を提案する。

以下、既に報告した方法[2]も含め意図解釈の方法の定式化についてと、その性能を評価した結果について述べる。

2. 意図解釈のモデル化

2.1 ユーザの意図

タスクに依存したユーザの意図は、タスクを限定して考えれば、N個に分類することが可能である。

$$I_i \ : \ i = 1, 2, \ldots, N$$

この意図は、あらかじめ人間が自然言語文などを用いて定義する。この定義には、その意味する概念が表現されていることから、何らかの方法により入力発話とその概念との類似度をとることで発話の意図を解釈することが可能だといえる。

2.2 対話過程

ここで考える対話は、上記のユーザの意図のどれかを達成しようとするものであるが、当然ここで生ずる対話の過程もある程度制約されたものとなる。

対話の過程を対話の状態の遷移として捉えることはごく自然である。対話の状態は対話例や意図記述から抽出されるものが多いが、予期しにくいものや表現しにくいものもあるといえる。

このような対話の状態により、出現可能なユーザの意図の種類は制限される。

2.3 意図の解釈

以上のようなことから、ここでは式(1)で定まる評価関数fを最大化するI_iを求めることを、「意図を抽出する」事だと定義する。

$$f(S, I_i, D_a) \tag{1}$$

$$\begin{cases} S & : \text{音声認識結果} \\ I_i & : \text{ユーザ発話の意図} \ (i = 1, 2, \ldots, N) \\ D_a & : \text{対話状態} \end{cases}$$

さらに、タスクを限定してユーザの意図を考えた場合、「意図を解釈する」ということは、システム

に用意された機能の中のどれかにユーザの発話を結び付けることに相当する。言い換えれば、ユーザの意図とはシステムに用意された機能に対する操作命令であるとみなすことができる。このユーザの発話をシステムの機能に分類する性能を最大化することが、意図解釈の能力の最大化を意味する。

よって、ユーザの発話の意図を表現する方法としては、扱う対象のタスクにのみ使用できるものを採用すれば十分であり、一般的な意味を表現する方法をタスクに適用して表現する必要はなくなる。

このような意図解釈法の問題点は、ユーザの発話がタスク外の発話であったり、用意された機能には結び付かないような要求であったりする時に、システムの動作とユーザの意図とが全く異なってしまうことである。ここでは、対話システムの対話戦略を工夫して随時意図の認識結果を修正できるような枠組を用意することでこの問題に対処できるとする。

先に述べたように、式(1)で定まる評価関数fを最大化するI_iを求めることが「意図を抽出する」ことであるが、fを一度に求めることは難しいので、式(2)に示すようにf'とf''との2段階で求めることにする。

$$f(S, I_i, D_a) = f'(S, W^n) \otimes f''(W^n, I_i, D_a) \qquad (2)$$

$$\begin{cases} S & : \text{音声認識結果} \\ I_i & : \text{ユーザ発話の意図} (i = 1, 2, \ldots, N) \\ W^n & : \text{ユーザ発話単語列} (w_1 w_2 \cdots w_n) \\ D_a & : \text{対話状態} \end{cases}$$

前者は音声認識結果からユーザの発話単語列を推定すること、後者は発話単語列からユーザの意図を解釈することにそれぞれ相当する。

3. 意図解釈の定式化

本研究では、意図解釈の際に次のものを利用する。

・単語認識スコア
・単語カテゴリ
・単語カテゴリ間の共起関係
・単語カテゴリと意図の関係度
・対話の状態

ここで意図解釈の際に目標とすることは、次の2点である。

- 単語の共起関係を言語構造とみなして単語列のスコアリングを行なう。
- 単語と意図の関係度を意味表現として定量的に意図の抽出を行なう

また、処理の流れは次の通りである。

(1) ワードラティスからユーザの発話単語列の候補を作成する
(2) 単語の共起関係を用いて文としてのスコアの高い発話単語列を選ぶ
(3) 選択された発話単語列に対し、ユーザの発話意図としてスコアの最も高いものを選ぶ
(4) 選択された発話意図に必要なパラメータを、発話単語列から抽出する

3.1 発話単語列の推定

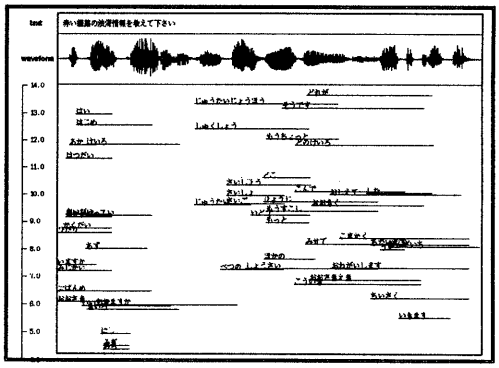


図 1: ワードスポッティングの様子

音声認識部からの認識結果は、認識単語、開始フレーム、終了フレーム、認識スコアから成るキーワードラティスで得られる。また認識結果として得られる単語の数は不定であるとする。

キーワードラティスの例

【単語】	【開始】	【終了】	【認識スコア】
あお	308	373	10.541383
あか	21	68	11.844539
おしえて	463	685	9.995821
⋮	⋮	⋮	⋮

単語のカテゴリ間の共起関係を用いてキーワードラティスから発話単語列を推定する。

1) 単語間の共起関係

一般に自然な発話を扱う場合、単語の出現順序の自由度が高くなる事や、音声対話特有の対話現象の

存在により、その文法を記述するのは困難であるとされている。また、ワードスポッティングを用いる場合、ユーザの発話中の認識対象単語のみを認識するため、その文法は特殊なものとなる。さらに、発話中に存在する認識対象単語が、誤認識により脱落することもあり得るため、単語が連続的に並ぶことを前提とするような文法を用いることが出来ない。

そこで、本方式では、文中での認識対象単語の生起順序を考慮した共起関係に着目し、これを制約として文法に利用する。

具体的には、書き起こしデータベース中での認識対象単語の単語カテゴリ同士の共起出現回数をあらかじめ求めておく。

$bigram(C_a, C_b)$: 単語カテゴリ C_a と C_b の生起順序を考慮した共起出現回数 (C_a が C_b の前方に存在)

2) 発話単語列候補の作成

意図解釈部は、音声認識の結果であるキーワードラティスが入力されると、まず発話単語列候補の作成を行なう。

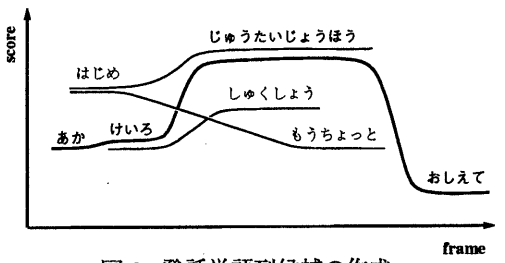


図 2: 発話単語列候補の作成

認識単語の開始・終了フレームを考慮し、ある程度の単語の重なりを許しながら、ワードラティスから構築可能な単語列を全て作成する。この際に、単語列候補 W^n の中に、$bigram(C_i, C_j)$ が 0 になる単語の組が存在する時は、その単語列候補は文として成立しないと判断して候補から削除する。

ここで、

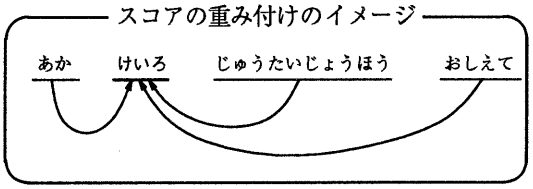
- W^n : 単語列候補 $\{w_1 \dots w_i \dots w_j \dots w_n\}$
- C_i : 単語 w_i の属する単語カテゴリ
- $bigram(C_i, C_j)$: C_i と C_j の共起出現回数 (ただし C_i が C_j の前方に存在)

3) 共起関係による発話単語列の推定

上記のように得られた発話単語列候補から、可能性の高い単語列の推定を行なう。単語のカテゴリ間の、生起順序を考慮した共起頻度を、単語カテゴリ間の親和度と定義し、この親和度を言語的な構造として用いて、発話単語を再スコアリングする。

これにより、次の 2 つの特徴を持つ単語列がより正解に近いとして、そのスコアを重み付けさせる。

- 単語列中の単語数が多い
- 親和度の高い単語の組が存在する



式 (3),(6) に従い各単語列の発話文としての可能性のスコア S を計算する。単語列候補 W^n の各単語 w_x の認識スコアを S_x とすると、S'_x は周辺の単語との共起関係に基づいた親和度により重みづけされた単語のスコアとなる。

最大のスコア $S_{sentence}$ を与える単語列候補をユーザの発話単語列と決定する。

$$S'_x = S_x \left(1 + \sum_{i=1}^{i<x} \frac{S_i}{S_x} G^T(C_x|C_i) + \sum_{j=x+1}^{k} \frac{S_j}{S_x} G(C_x|C_j) \right) \tag{3}$$

$$G^T(C_x|C_i) = \frac{bigram(C_i, C_x)}{C_i \text{の出現回数}} \tag{4}$$

$$G(C_x|C_j) = \frac{bigram(C_x, C_j)}{C_j \text{の出現回数}} \tag{5}$$

$$S_{sentence} = \frac{1}{k} \sum_{x=1}^{k} S'_x \tag{6}$$

- W^n : 発話単語列候補 $\{w_1 \cdots w_x \cdots w_n\}$
- S_x : 単語 w_x の認識スコア
- C_x : 単語 w_x の属する単語カテゴリ
- $G^T(C_x|C_i)$: $i < x$ の時の C_i と C_x の親和度
- $G(C_x|C_j)$: $x < j$ の時の C_j と C_x の親和度

3.2 発話意図の解釈

最大スコア $S_{sentence}$ を与える単語列 W^n に対し、発話意図の推定を行なう。

発話の意図は、その発話単語列と対話の状態で決まる。ここでは処理を簡単にするため対話の状態は一定であるとする。

意図の種類が決定されると、各意図ごとに必要となるパラメータを発話単語列から抽出し、ユーザの発話意図として解釈する。

ここでは発話の意図を決定する方法として、2種類の方法を考えた。

(1) 各意図毎に、単語カテゴリに対し重み係数を設定するもの

(2) 意図と単語カテゴリの間の共起関係を用いるもの

1) 重み付けによる方法

あらかじめ各意図ごとに、対話データベースの書き起こしテキストから各単語カテゴリの出現頻度を調べ、各単語カテゴリに対する重み係数 $Weight(C_x, I_i)$ を定める。

発話単語列候補に対し、式(7)によりユーザの意図の推定を行なう。最大のスコア S_{intent} を与える I_i をユーザの発話意図とする。

$$S_{intent} = \max_{i}^{intent} \sum_{x=1}^{n} Weight(C_x, I_i) S_x \qquad (7)$$

$Weight(C_x, I_i)$: I_iに対する C_xの重み係数

この重み係数としては、次の2種類を用意した。

(1) 意味重み 意味的に重要なカテゴリに対し大きな重み係数を与える方法

(2) 頻度重み 出現頻度の高いカテゴリに対し大きな重み係数を与える方法

方法(1)では、この重み係数は単語カテゴリの重要度として人間が経験的に与えた。方法(2)では、対話データベースから求めた事後確率 $P(C_x|I_i)$ を用いた。

2) 意図と単語カテゴリの共起関係による方法

情報理論的なアプローチでは、単語列 W が意図 I である確率は、式(8)のように近似される。

$$\begin{aligned} P(I|W) &= \frac{P(I)P(W|I)}{P(W)} \\ &\Rightarrow P(I)P(W|I) \\ &\doteq P(I)P(w_1|I)P(w_2|I)\cdots P(w_n|I) \end{aligned} \qquad (8)$$

ここで、$P(w_x|I)$ を $P(C_x|I)$ に置き換え、各発話意図 I_i が起きる確率 $P(I_i)$ が同じであると仮定すると、発話意図は式(9)により求まることになる。式(9)から最大スコア S_{intent} を与える I_i をユーザの発話意図とする。

$$S_{intent} = \max_{i}^{intent} \prod_{x}^{n} P(C_x|I_i) \qquad (9)$$

4. 評価実験

4.1 ナビゲーションタスクへの適用

1) 対話データベース

我々の扱うタスクは、音声入力の必要性が高いと考えられるカーナビゲーションである[1]。このカーナビゲーションをタスクとして音声対話実験を行ない、その結果をデータベース化して意図解釈部の構築に利用した。

この対話実験ではWizard of Oz方式を採用し、システムと対話する被験者にはシステムの動作を人間が操っていることを伝えないで、人間と機械との間の対話を収録した。

このデータベースは、システムとユーザの発話の書き起こしテキストと、ユーザの発話に対し、発話意図ラベルを付与したデータとからなる。また、発話意図ラベルは、システム熟練者がユーザ発話を分類して付与した。

このデータベースのユーザ発話に関する規模を表1に示す。

表1: データベース規模

被験者数	総対話数	総発話数	平均発話数
13	39	1040	26.7

2) 意図の分類

次に、ナビゲーションタスクで想定するユーザの発話意図を分類した。

システムに用意された機能としては、地図の操作(移動、縮尺変更)、経路情報検索、渋滞情報検索などがあり、これらの機能とユーザの意図を結び付ける。

ユーザの意図を表現するには次のような方法をとった。まず、システムに対するこれらの操作命令に対し、名前(ラベル)を付与し、このラベルをユーザの意図の名前(意図ラベル)であるとした。そして、この意図ラベルと操作命令に付随するパラメータとでユーザの発話意図を表現する。表2にユーザ

意図ラベルの例を示した。意図ラベル中の () 内には、意図の詳細な内容を決定するパラメータが格納される。

本実験で用意した意図ラベルは、全部で 17 個ある。この中で、ラベル名 nonsense を付与された発話は、ユーザの発話がタスクの対象外の要求や意味をなさない発話であると判断されたものである。

表 2: ユーザ意図ラベルの例

ユーザ意図ラベル	発話例
move_map_where(地名)	池袋を表示して下さい
move_map_dir(方向, 程度)	右に少しずらして下さい
move_map_zoom(縮尺)	地図を拡大して下さい
show_route(経路)	青の経路を表示して
...	...
nonsense	(タスク内では解釈不能)

3) 単語カテゴリの分類

意図解釈処理のために、認識対象の単語をその品詞あるいは意味分類によりカテゴリに分類することで、その単語と意味を結び付ける。

このナビゲーションシステムの意図解釈部が処理の対象とする単語群は、操作命令に関連する動詞、操作対象を指す名詞と操作命令や操作対象を限定する単語のグループからなる。

実際の分類に当たっては、単語をまず品詞で分類し、その後さらにその語がタスクの中で表す意味により細分類した。

その結果、単語群は次のような 23 個のカテゴリに分類された。各カテゴリにはそのカテゴリの表す意味によりカテゴリ名をつけてある。

方向、地名、経路、程度、縮尺、移動、属性、経路の色、回避、表示、説明、返事、問いかけ、行動、地図、通りの名、渋滞、選択疑問、決定、場所疑問、他の、繰り返し、何番目

図 3: 単語カテゴリーの分類

4.2 実験試料と音声認識手法

提案した意図解釈方式の性能を調べるために、まずその入力データを作り出すワードスポッティング部を 2 種類用意した。一つは単語のメルケプストラムパターンによるテンプレートマッチングに基づくもの、もう一つは音韻認識に基づくものである。

音声データ: 音声データは、対話データベースから各意図を平均的に含むように抽出した 84 文章を、話者 1 名により 1 回読み上げにより収録したもの。

読み上げ文章の例

地図を右に移動して下さい
赤の経路の渋滞情報を教えて下さい
(下線部はスポッティング対象単語である)

方法 1: 認識対象の 84 文章中に出現する単語のうち、スポッティングの対象とする 54 単語について、同一話者の 1 回発声データより辞書を作り、連続 DP マッチングでスポッティングを行なう。

方法 2: 同一話者の 216 バランス単語 1 回発声により音韻のテンプレートを作り、54 単語辞書について、音韻認識に基づきスポッティングを行なう。

なお、認識対象の 84 文章中には、累積で 206 単語のスポッティング対象単語が出現する。

4.3 ワードスポッティング性能

スポッティング結果の評価は、正しく抽出された単語数 (A) と、誤って抽出された単語数 (FA:False Alarm) の割合によって行なわれる。ここで、認識対象の文章中のキーワードの出現数 W に対する正しい単語候補の数の割合 A/W を抽出率 (%) と呼ぶ。

表 3 と表 4 にスポッティングの性能を示す。また、図 4 に出力候補数 (W の整数倍ごと) に対する抽出率の関係のグラフを示す。

表 3: スポッティング性能 (方法 1)

候補数	A/W(%)	FA/W
~W	76.3	0.24
~2 W	90.9	1.09
~3 W	95.3	2.05
~5 W	98.8	4.01
~10 W	99.1	9.01

方法 1 と方法 2 のスポッティング性能を比較すると大きな差がある。方法 2 のものはボトムアップ情報のみで認識を行なうスポッティングの性能として普通のレベルのものであり、方法 1 のものは理想的なスポッティングに近いレベルであるとみなすことができる。

このような性能を持つ音声認識の結果から発話単語列を推定し、ユーザの意図を解釈する性能を調べる。

表 4: スポッティング性能（方法 2）

候補数	A/W(%)	FA/W
～W	33.9	0.66
～2 W	44.0	1.56
～3 W	50.7	2.49
～5 W	63.8	4.36
～10 W	75.1	9.25

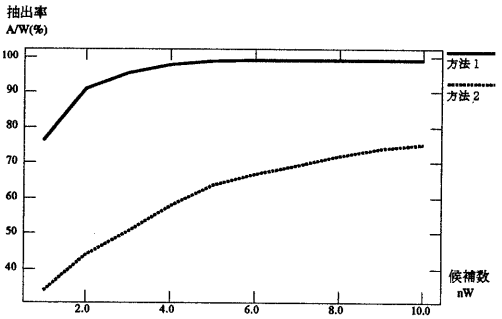


図 4: スポッティング性能

4.4 文推定能力

表 5と表 6に、図 4で示される性能を持つ認識部が出力するキーワードラティスから、単語カテゴリ間の共起関係という言語的知識を用いて発話単語列を推定する能力を示す。

ただし、スポッティングの結果は、異区間でしかも長さが異なる単語同士のスコアを比較することは容易ではない。ここでは、スコアは相互比較可能なように正規化してあるものとして扱うことにする。

比較の対象として、ラティス中の最も認識スコアの高い単語から出発して、言語的知識を用いずに、単純に各単語スコアの合計を最大化する通常の探索的な発話単語列推定方式による発話文推定の能力も示す。

ここで、ユーザの発話と判定された単語列中の正解単語数を C、付加誤り数を I、実際の発話の中に存在する認識対象単語数を W とすると、正解率 C/W、挿入率 I/W である。また、ユーザ発話と判定された単語列が実際の発話単語列と完全に一致した割合を文正解率とする。表中の w および s で示された数値は、それぞれ累積総単語数、総文章数を表す。

単語カテゴリ間の共起関係により単語のスコアリングを行なうことにより、再現率を上昇させ誤った単語の挿入も抑えて、結果として単語列が入力文と完全に一致する率も上昇していることがわかる。言語的知識が有効に働いた結果といえるだろう。

表 5: 文推定能力（方法 1）(206w,84s)

	正解率	挿入率	文正解率
共起関係	84.5(174)	18.4(38)	57.1(48)
探索的手法	80.1(165)	22.8(47)	33.3(28)

表 6: 文推定能力（方法 2）(206w,84s)

	正解率	挿入率	文正解率
共起関係	40.8(84)	63.6(131)	11.9(10)
探索的手法	31.1(64)	68.4(141)	6.0(5)

4.5 意図解釈能力

1) テキスト入力による評価

まず、スポッティング性能と文推定能力に左右されない意図解釈能力を評価するために、テキスト入力による実験を行なった。

すなわち、誤認識のない単語列を意図解釈部に入力して、この意図解釈方式による意図解釈の可能性を調べた。この評価には、対話データベース中のユーザ発話の中から、何らかの意図に分類されている 934 発話を用いた。なお、この 934 発話中に含まれるスポッティングの対象となる単語は全部で 67 単語あった。

この評価実験の結果を表 7に示す。

入力単語列の発話意図ラベルを正しく推定できたものを Correct、間違えたものを Error、どの意図にも判断できなかったものを NoAnswer に分類した。ここで解釈結果の分類の Correct の中には、操作命令を限定するパラメータを抽出できなかったものも含む。これは、その後の対話戦略によりパラメータの内容を補完することが可能だからである。このパラメータを抽出できなかったものは () 内に示した。

表 7に示された結果は、それぞれの意図解釈方式の最大の性能を示すものといえる。

表 7: 意図解釈の最大性能 (934s)

分類	Correct	Error	NoAnswer
意味重み	91.9(2.8)	8.0	0.1
頻度重み	89.8(1.8)	10.1	0.1
$P(I\|W)$	97.2(2.8)	2.8	0.1

2) 音声入力による評価

次にワードスポッティングによる音声認識結果を入力した時の、それぞれの意図解釈方式の性能を評価した。

表8と表9にキーワードラティスから発話意図を推定する能力を示す。

表 8: 意図解釈性能（方法 1）(84s)

分類	Correct	Error	NoAnswer
意味重み	82.1(1.2)	17.9	0.0
頻度重み	73.8(1.2)	26.2	0.0
$P(I\|W)$	83.3(1.2)	16.7	0.0

表 9: 意図解釈性能（方法 2）(84s)

分類	Correct	Error	NoAnswer
意味重み	32.1(0.0)	67.9	0.0
頻度重み	31.0(0.0)	69.0	0.0
$P(I\|W)$	33.3(0.0)	66.7	0.0

音声認識結果からの意図解釈能力は、音声認識性能と文推定能力と意図解釈能力によって決まるといえる。

やはりスポッティング性能自体が低い方法 2 の結果からは満足な性能を得ることができなかった。スポッティングの性能として方法 1 の結果程度の能力が必要であることがわかる。

また、意図解釈の方式としては $P(I|W)$ により評価するものがテキスト入力によるテストで最大の性能、音声入力によるテストでは意味による重み付けを用いる方法と同程度の性能を得ている。このように対話データベースから得た知識を基に定量的に意図を解釈する方式の有効性が示された。

5. おわりに

本稿では、ワードスポッティングの認識結果であるキーワードラティスから、タスクに依存したユーザの発話意図を解釈する方法について述べた。本手法は、まずキーワードラティスから、単語カテゴリ間の生起順序を考慮した共起頻度を用いて可能単語列候補を推定する。次にその単語列候補を、あらかじめ分類された発話意図のどれかに、単語カテゴリと意図との関係度により結び付ける、というものである。

発話単語列推定に関して、音声認識結果を用いた評価実験で、単純に各単語スコアの合計を最大化する通常の探索的な発話単語列推定方式に比べて、単語再現率と正解文章率を向上させ、誤った単語の挿入率を低下させることに成功した。

また、意図解釈に関して、テキスト入力による評価実験では最高で 97.2% の意図解釈成功率を示し、理想的な性能に近いスポッティング性能を示すことができる音声認識部との評価実験では、最高で 83.3%の意図解釈成功率を示すことができた。この評価実験により、意図解釈に必要なワードスポッティングの性能のある程度の指標を得ることができた。

今後は、意図解釈部構築のために予備実験を行なって、対話データベースを用意しなければならない労をなくすため、人間が意図に対する定義を自然言語により与えると、意図解釈に必要な知識や概念を自動的に抽出するような枠組を考えたい。また、対話の状態も考慮した意図解釈の方法についても考えていきたい。

参考文献

[1] 亀山晋, 中里収, 白井克彦: カーナビゲーションにおける音声対話, 春季音講論集 pp.21-22 (1994)

[2] 亀山晋, 中里収, 白井克彦: ワードスポッティングに基づく意図抽出に関する考察, 秋季音講論集 pp.33-34 (1994)

[3] 小林哲則: 対話音声の認識技術, 音響誌 Vol.50 No.7 (1994)

[4] 今村明弘, 北井幹雄: 事後確率を用いたフレーム同期ワードスポッティング, 信学技報 SP93-31 (1993)

[5] 河原達也, 宗続敏彦, 三木清一, 堂下修司: 会話音声中の単語スポッティングのための言語モデルの検討. 信学技報 SP94-28 (1994)

[6] 北岡教英, 河原達也, 堂下修司: 自由発話認識・理解のためのフレーズスポッティング, 信学技報 SP93-116,NLC93-56 (1993)

[7] 坪井宏之, 橋本秀樹, 竹林洋一: キーワードスポッティングに基づく連続音声理解, 信学技報 SP91-95,NLC91-52 (1991)

[8] 白井克彦, 竹沢寿幸: 音声対話処理, AI 誌 VOL.9 NO.1 (1994)