

## 条件付き確率場を用いた発話内容の分類による音声認識結果の検証

大田健紘<sup>†</sup> 江原暉将<sup>†</sup>

<sup>†</sup> 諏訪東京理科大学システム工学部電子システム工学科  
〒391-0292 長野県茅野市豊平5000-1  
E-mail: †{otakenko,eharate}@rs.suwa.tus.ac.jp

**あらまし** 本研究では、条件付き確率場を用いて音声認識結果の重要語に対するタグ付および、発話の分類を行い、その結果から、音声認識誤りなどに起因するユーザ発話の曖昧さを解消するための対話戦略について検討を行っている。音声対話システムにおいて、音声認識誤りによりシステムが発話を受理できない場合、毎回確認対話を生成しては利用者にとって非常にわずらわしい。そこで、想定発話を用意しておき、それらと同じような内容を表す発話群に分類し、音声認識結果の各発話群への分類結果をもとに、発話の曖昧な箇所の同定を行っている。その結果から、発話の一部分のみの再認識により音声認識誤りの修正を行う、もしくは確認対話の生成を行い利用者に音声認識誤りの修正行わせるかを決定している。システムが得ることのできる情報をもとに、利用者の発話の推定を行い、自動的に修正をすることによって、利用者側の負担を軽減することが可能になっている。評価の結果、音声認識誤りに対して頑健に発話内容の分類を行うことが可能であり、その結果から音声認識誤り箇所およびタグ付けの誤り箇所を同定し修正することも可能であることを確認している。

**キーワード** 音声対話システム、自然言語処理、条件付確率場、発話意図、音声認識検証

## Verification of Speech Recognition Results Based on the Utterance Classification Using Conditional Random Fields

Kenko OTA<sup>†</sup> and Terumasa EHARA<sup>†</sup>

<sup>†</sup> Tokyo University Science, Suwa  
Department of System Engineering, Faculty of Electronic System Engineering  
5000-1, Toyohira, Chino, Nagano, 391-0292, Japan  
E-mail: †{otakenko,eharate}@rs.suwa.tus.ac.jp

**Abstract** A dialogue strategy for disambiguating user utterances is discussed in this study. In case where the systems cannot understand user utterances because of speech recognition errors, users are annoyed with the iterative confirmation. So, a method for automatic verification and modification of the speech recognition results is proposed. Many supposed utterances clustered in accordance with content similarity are prepared for estimating user intention. Then, a speech recognition result is classified into a cluster using conditional random fields. The system decides whether rerecognizing a part of the utterances supposed to be mistaken or generating a confirmation dialogue in order to be respoken based on an utterance classification result. Estimating and modifying speech recognition errors by the system itself based on a user utterance, users may get relief from iterative confirmation. From the evaluation results, it is possible that the proposed method can classify speech recognition results into clusters robustly and the proposed method can detect and modify the speech recognition errors automatically based on the utterance classification results.

**Key words** Spoken dialogue system, Natural language processing, Conditional random field, Utterance intention, Speech recognition verification

## 1. はじめに

音声認識率の向上により、音声対話による検索システムも実用化されつつある [1-5]。しかし、キーワードによる関係データベースからの検索が主流であり、自然言語による大規模なテキストからの検索システムはまだ少ない。キーワードによる検索の場合、決定的な文法を基にした音声認識が主に利用され、システムが理解できる発話が大きく制限される。自然言語による検索の場合は、多様な表現を認識できる必要があり、統計的言語モデルに基づく大語彙連続音声認識を用いるほうがよい。しかし、大語彙連続音声認識を用いる場合、決定的な文法に基づく音声認識と比較して、音声認識率が低下する問題がある。さらには自然言語による発話には、省略、倒置、あいまいさが存在する。そのため、検索システムはそれらに対して頑健に動作する必要がある。

これまでに音声対話システムとして、たけまるくん [1]、レストラン検索システム [2] や京都版ダイアログナビ [3] などがある。たけまるくんは、一問一答型のシステムであり、想定質問とそれに対する応答が用意されている。音声認識結果と想定質問とのマッチングには形態素の一致数を用いている。

レストラン検索システムも想定質問を用いているが、レストラン名や場所などがクラスとして表現されており、想定質問とのマッチングには構文情報も考慮して行われている。

京都版ダイアログナビは、音声認識結果の N-best を用いて検索を行い、検索されたテキストの重なりを計測し、重なりが小さい部分は音声認識誤りの可能性が高いと考え、確認対話を生成し、あいまい性の解消を図っている。

また海外の音声対話システムとしては MATCH (Multimodal Access To City Help) [6] がある。MATCH システムは、PDA 上で動作しており、ニューヨークのレストランや地下鉄の情報のガイドシステムであり、音声とペン同時入力が可能である。つまり、PDA の画面上に表示された地図にペンで範囲を指定しながら「このあたりにあるレストランを表示して」と発話することで、周辺のレストランの情報を取得することができる。MATCH では MultiModal Finite State Transducer (MMFST) を用いて、音声に入力とペン入力の結果をラティスにより統合し、N-best リストを Multimodal Dialog Manager (MDM) に渡し、MDM で最終的な結果を出力する。

本稿でも、想定発話をあらかじめ用意し、音声認識結果とマッチングをとる手法を採用する。ただし、想定発話はユーザーの発話意図を推定する手がかりとして用いるだけであり、想定発話をそのまま検索には用いない。想定発話と音声認識結果のマッチングは、想定発話から学習した条件付確率場 (CRF : Conditional Random Field) [7] を用いて音声認識結果の各形態素に発話意図を表わすラベルを付ける問題として考える。発話中の各形態素に付与されたラベルから、ユーザーの発話意図の候補を生成し、音声認識誤りの検出やあいまい性の解消のために確認対話を生成する。これにより、ユーザー発話のあいまい性、表現の揺れなどに頑健に動作する音声対話システムが期待できる。

また、音声対話システムでは、利用者は正確に発話したつもりでも、音声認識誤りなどによって発話が受理されない場合がある。そのような場合に、毎回確認対話を生成し、利用者に発話を求めるのはシステムの利用者にとっては非常にわずらわしい。そこで、本稿では、先に述べた発話意図の推定結果を用いて、システムが自動的に利用者の発話しなかった内容を推定し、音声認識誤りによって理解できなかった部分を特定して、可能な限り確認対話を生成することなく、自己解決を図る方法についても述べる。

## 2. 条件付確率場について

条件付確率場 (CRFs : Conditional Random Fields) は、生成モデルに基づく隠れマルコフモデル (HMMs : Hidden Markov Models) とは異なり、識別モデルに基づくモデルである。生成モデルに基づく HMM は、観測系列  $\mathbf{x}$  とラベル系列  $\mathbf{y}$  の同時確率  $p(\mathbf{x}, \mathbf{y})$  を、ある状態から観測値が出力される確率と、状態の遷移確率に分解してモデル化している。

$$\begin{aligned} p(\mathbf{x}, \mathbf{y}) &= p(\mathbf{x}|\mathbf{y})p(\mathbf{y}) \\ &= \prod_{t=1}^T p(x_t|y_t)p(y_t|y_{t-1}) \end{aligned}$$

ただし、HMM は、変数間の独立性を仮定しており、独立でない素性をうまく扱うことができない。

一方で、識別モデルに基づく CRF は、観測系列  $\mathbf{x}$  により条件付けされたラベル系列  $\mathbf{y}$  の条件付確率  $p(\mathbf{y}|\mathbf{x})$  を、ロジスティック回帰を基にローカルな変数間の関係をラベル間の遷移 (遷移素性) によりモデル化している。

$$\begin{aligned} p(\mathbf{y}|\mathbf{x}) &= \frac{\exp(\langle \Theta, \Phi(\mathbf{x}, \mathbf{y}) \rangle)}{\sum_{\tilde{\mathbf{y}}} \exp(\langle \Theta, \Phi(\mathbf{x}, \tilde{\mathbf{y}}) \rangle)} \\ &= \frac{\exp(\sum_{t=1}^T \langle \Theta, \Phi(\mathbf{x}, y_t, y_{t+1}) \rangle)}{\sum_{\tilde{\mathbf{y}}} \exp(\sum_{t=1}^T \langle \Theta, \Phi(\mathbf{x}, \tilde{y}_t, \tilde{y}_{t+1}) \rangle)} \end{aligned}$$

ただし、 $\Phi(\cdot)$  は素性関数、 $\Theta$  は素性に対する重み、そして  $\langle \cdot, \cdot \rangle$  は内積を表す。

CRF は HMM のように素性間の独立性を仮定しておらず、柔軟な素性設計が可能であるため、素性間の独立性を仮定できない問題 (自然言語処理の多くの問題) に対して HMM よりも有効である。

CRF と同じく識別モデルに基づくモデルとして最大エントロピーマルコフモデル (MEMM : Maximum Entropy Markov Model) [8] がある。MEMM は入力の先頭から順に問題を解いていくため、CRF とは異なり入力全体の情報を利用することができない。そのため、MEMM はラベルバイアス問題 [7] が発生する。一方で、CRF はラベルバイアス問題を解消できるため、MEMM よりも有効である。

## 3. 提案法概要

### 3.1 学習素性

素性は Mecab/Cabocha [9] により解析された結果を用いて付与した。具体的には、出現形、品詞、文節境界および主辞、機能語である。さらには、固有名 (例えば、諏訪東京理科大学のように場所を表わす単語には  $\langle \text{location} \rangle$  など) や「どこ  $\langle \text{where} \rangle$ 」,

出現形	品詞	構文情報	重要語タグ	ラベル
諏訪湖	(名詞-固有名詞-一般)//	2	<location>	1
へ行く	(助動詞-格助詞-一般)//	1	0	0
行く	(動詞-自立)/五段・カ行促音便/基本形	2	0	1
には	(助動詞-格助詞-一般)//	0	0	0
は	(助動詞-係助詞)//	1	0	1
どう	(副詞-助詞類接続)//	1	<how>	1
どう	(副詞-自立)/カ変・スル/連用形	2	<how>	1
したい	(動詞-自立)/特殊・タ/仮定形	0	<how>	1
いい	(形容詞-自立)/形容詞・イイ/基本形	0	<how>	1
いい	(形容詞-自立)/形容詞・イイ/基本形	2	0	1
ですか	(助動詞-特殊・テス/基本形	0	0	1
か	(助動詞-助助詞/並立助詞/終助詞)//	0	0	1
.	(記号-句点)//	0	0	1

1段目のCRFで付与      2段目のCRFで付与

図1 想定発話に付与した素性の例

縄文の湯に行くにはどうしたらいいですか。 車山のスキー場へはどうやって行ったらいいですか。	<location>に行くには<how>いいですか。
霧ヶ峰はどこにありますか。 ピラタスロープウェイはどこにありますか。	<location>は<where>ありますか。
お土産はどこで買えますか。 新鮮な魚を食べるにはどこへ行けばいいですか。	<mono>は<where><doushi>ますか。
スキー場はいつオープンしますか。 美術館がオープンする時間を教えてください。	<location>は<when>オープンしますか。

図2 想定発話の一部

「何<what>」などの疑問詞も素性として用いている。ただし、これらの素性を「Bag of Words モデル」として扱うのではなく、前後に出現するものも素性として含める。次節で述べるが、提案法では2段階のCRFを用いている。それぞれの段階で用いた素性数は、57980 および 79970 である。

### 3.2 CRF による発話意図の推定方法

システムではあらかじめ想定発話を用意しておき、想定発話をCRFによりモデル化したものに基き、検索要求からユーザーの発話意図を大雑把に推定する。さらには、マッチした想定発話と検索要求から、音声認識誤り箇所の推定および、検索のために不足している情報を推定することが期待される。これを、確認対話を生成するための手がかりとして用いる。

想定発話はまず、Cabochaにより解析され、前節で述べた素性を付与する。例を図1に示す。構文情報の「1」は文節境界を、「2」は主辞を表わしている。ただし、主辞と文節境界の位置が一致した場合は「2」を付与している。ラベルは似た意味を表わす発話には同じラベルが付与されている。例えば、「<location>への行き方」を尋ねる発話にはすべて同じラベルが付与されている。現在、ラベルは全部で11個である。想定発話の一部抜粋を図2に示す。ただし、想定発話はクラスを展開している。

次に検索要求に対しても同様にCabochaを用いて解析を行い、素性を付与する。ただし、固有名や「どこ<where>」などの疑問詞に付与する素性はCabochaだけでは付与することができないため、2段階にCRFを適用する。CRFにはフリーのツールCRF++[10]を用いた。まず、1段目のCRFでは、出現形、品詞、文節境界および主辞、固有名や「どこ<where>」などの疑問詞に対してラベルを付与する。そして2段目のCRFで、入力文がどういう意図で発話されたのかを想定発話から学習されたCRFモデルにより推定する。学習データに対しては手作業でラベルを付与する必要があるが、テストデータに対し

ては完全に自動的にラベルを付与することが可能となる。

通常、CRFは形態素解析などのラベル付与問題に適用され、分類問題にはSVM[11]やブースティング[12]が一般的に用いられる。しかし、本研究では発話意図の推定を各形態素へのラベル付与の問題と考え、CRFを用いた。これにより、発話途中での意図の転換や曖昧な表現へ対応が可能になると考えられる。

### 3.3 発話意図を用いた音声認識結果の検証法

本節では推定されたタグおよび、ラベルを用いて音声認識結果の検証を行う方法について述べる。推定されたラベルを元に、そのラベルが付与されている想定発話群と音声認識結果をDPマッチングにより比較し、音声認識結果に不足しているタグを見つける。音声認識を行った際に、単語のアライメントをとっておき、各単語の存在する区間を特定しておく。そして単語アライメントの結果とDPマッチングの結果を利用し、音声認識結果に不足しているタグの区間を特定する。図3に音声認識結果の検証手順を示す。

この検証が正しく動作するには、高精度にラベルが推定できている必要がある。そこで、推定されたラベル内の想定発話群と音声認識結果のDPマッチングの結果から得られる、最も距離の近い想定発話との一致率を基準としてラベル推定の正しさを検証する。これはつまり、ラベルが正しく推定されていれば、そのラベル内の想定発話群との一致率は高くなり、推定に失敗していれば一致率は低くなることに基づいている。ここでDPマッチングの一致率  $P$  は、

$$P = \frac{M}{M + R + D + I}$$

として求めている。ただし、 $M$  は一致、 $R$  は置換、 $D$  は脱落、そして  $I$  は誤挿入した形態素の個数である。ラベル推定に失敗する場合は、タグ付もしくは音声認識の誤りに起因するものと考えられる。つまり、一致率が低い場合は、タグ付もしくは音

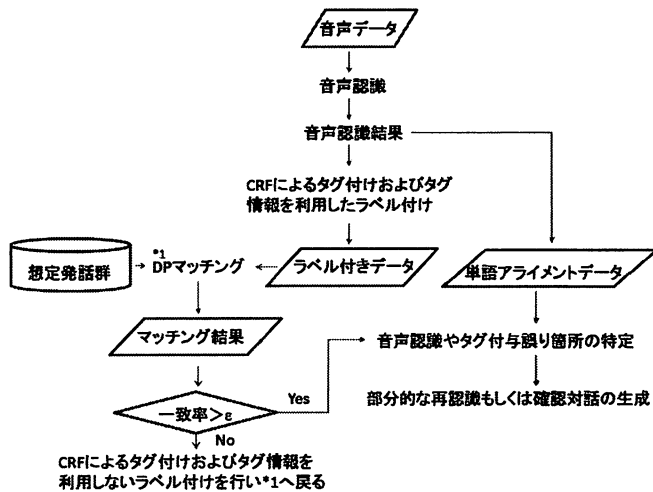


図3 音声認識結果の検証手順

声認識結果が間違っている可能性があると考えられる。そこで、そのような場合には2段目のCRFでのラベル推定に用いるモデルを学習する際にタグの情報を使わないで学習する。これにより、不確実な情報を使うことなくラベルを推定することができる。そして、タグを利用した場合の一致率とタグを利用しなかった場合の一致率の差を計算し、0以上の値の場合にはどちらのラベル推定結果も不確実であると考えられ、一方0未満の場合はタグを利用しないモデルのほうが正しいと考えられる。

一致率の差が0未満の場合には、タグを利用しないモデルによるタグ付の結果を用いて誤り箇所の特定を行い、部分的な再認識もしくは確認対話の生成を行う。

## 4. 評価

### 4.1 音声認識率について

評価に用いた言語モデルは、国立国語研究所により公開されている「日本語話し言葉コーパス」(CSJ: Corpus of Spontaneous Japanese) [13] をベースに、Web上から収集した諏訪圏の観光情報のデータを追加することにより作成した。ただし、CSJデータからフィルターや言い直しを含む文および、低頻度の語彙は除外した。音響モデルはCSJのものをそのまま用いた。音声認識エンジンにはJulius [14] を用いた。発話単位での音声認識率は51発話をテストデータとして用いて、発話正解率で約71%になった。ただし、テストに用いたデータには未登録語も含まれている。

### 4.2 重要語タグとラベルの推定結果について

#### 4.2.1 提案法を用いた場合

段階的に用いたCRFのそれぞれの性能を調べた。CRFのモデルを学習するために用意した想定発話は138発話、タグは10種類そしてラベルは11個ある。1段目で正しく推定できたのは発話単位では約69%であり、2段目で正しく推定できたのは約94%であった。推定結果の一例を図4に示す。図4は、本来「諏訪中央病院までどのくらいかかりますか。」と認識されるべ

きであるが、「諏訪中央病院今までの位掛かりますか。」と誤った例である。音声認識誤りを含んでいるが、ラベル“8”(想定質問として、「立石展望公園に行くには何分かかりますか。」や「ここから車山までどのくらいかかりますか。」などを含むラベル)と正しく推定されている。

推定誤りについて分析を行ったところ、1段目については重要語タグの付与の仕方の問題があったと考えられる。例えば、スキー場のように場所を表わす単語には<location>というタグが付与され、紅葉のような物を表わす単語には<mono>というタグを付与することを想定している。しかし、これらの単語に対して形態素解析器により付与される品詞は共に「名詞一般」である。そのため、これらのタグの付与が困難になったものと考えられる。

2段目については、音声認識誤りと重要語タグの付与の誤りが同時に起こる場合に最終的なラベル推定を誤る傾向がみられた。

#### 4.2.2 Support Vector Machine を用いた場合

ここでは、提案法と既存の手法との性能の比較について述べる。既存の手法としてSupport Vector Machine(SVM)を採用する。ただし、今回の問題は多クラス分類であるためmulti-class SVMを用いた。評価に用いたツールはSVMlight [15] である。SVMにより、2段目のラベル推定のみを行う。1段目の重要語タグの付与はCRFにより行った。SVMを用いてラベル推定を行った結果、精度は約65%となった。ただし、分類に用いている特徴数がCRFのものより少ないので、低い精度にとどまっている。

### 4.3 音声認識結果の検証について

前節での結果より、ラベル推定の精度は約94%であり、ほぼ正しく推定できていた。失敗した発話(3発話)についてDPマッチングの一致率をみると0.1~0.23と低い値になっていることが確認できた。そこで、それらの発話についてタグを利用することなくラベル推定を行った結果、1発話については一致

出現形	品詞	構文情報	重要語タグ	ラベル
諏訪中央病院今	(名詞-固有名詞-人名-姓)//	2	<location>	8/0.420326
で	(助詞-格助詞-一般)//	1	0	8/0.457868
どの	(連体詞)//	2	<much>	8/0.488432
位	(名詞-副詞可能)//	2	<much>	8/0.501729
掛かり	(動詞-自立)/五段・ラ行/連用形	2	0	8/0.504022
ます	(助動詞)/特殊・マス/基本形	0	0	8/0.492313
か	(助詞-副助詞・並立助詞/終助詞)//	0	0	8/0.474030
.	(記号-句点)//	0	0	8/0.445263

図 4 ラベル推定の結果

率が向上し、正しくラベルを推定できた。残り 2 発話の一致率については、変化なしもしくは低下するという結果になった。その結果、ラベルの正しい推定率は約 96%となった。このことから残りの 2 発話については、音声認識が正しくできていないので、再発話を促す必要がある。

音声認識誤りを正しく修正できた例としては、「ここから茅野駅まで何分。」という発話があるが、これは「ここから茅野駅まで 3 分。」と認識されている。図 5 にラベルの推定結果を用いた音声認識結果の修正手順の一例を示す。「<location>から<location>まで 3 分。」というタグが付与され、「諏訪東京理科大学へはどうやって行きますか。」などが含まれるラベルに分類されていた。この場合、DP マッチングの一致率は約 0.11 となり非常に低いため、ラベル推定が正しくないと考えられる。そこで、タグ情報を使わないラベル付けを行い、再び DP マッチングを行う。その結果、タグ情報を使った場合のラベルとは異なるラベルが付与され一致率も 0.8 に向上した。そして、音声認識結果の中で誤っている部分が特定され、単語アライメントの結果を用いてその部分の音声波形だけを切り出し、再認識を行ったところ、「何分」と認識され、正しい音声認識結果に修正できた。

また、タグ付には失敗するがラベル推定精度が高いため、ラベルの推定結果から正しいタグを推定することもできた。例えば、「新作花火大会はいつですか。」という発話に対して新作花火大会に<location>タグが付与されたが、「桜の見ごろはいつですか。(mono)の見ごろはwhen)ですか。」という想定発話にマッチしたため、<location>の部分がか<mono>に対応していることがわかる。そして、部分的に再認識しても新作花火大会と認識されるため、新作花火大会に対して正しいタグを付与しなおすことができている。このようにタグの再推定に関して、タグの推定を誤っている発話に対して、適合率 0.81、再現率 0.72 および F 値 0.76 で正しく誤りを検出している。また再認識および再発話を必要としないものについては、適合率 0.94、再現率 1、F 値 0.97 で正しく検出できている。

## 5. 考 察

発話意図推定に対する音声認識誤りや未登録語の影響について述べる。図 6 に未登録語を含む発話に対してラベルを付与した例を示す。音声認識誤りにより、出現形が異なる単語に変化したとしても、品詞や係り受けの情報を利用できるため、正しくラベル付与ができた。このことは異なる問題ではあるが[16]においても、CRF の有効性として述べられている。

次に、音声認識誤りが情報検索に与える影響について述べる。この発話は「横谷溪谷には何がありますか。」というものであるが、音声認識誤りにより読みは同じであるが表記は異なるものになった。このような場合に、表記の情報だけでなく読みの情報も用いて検索を行うことにより、未登録語を含んでいたとしても検索が可能になるので、検索への影響は小さいものと考えられる。

## 6. まとめと今後の課題

本稿では、条件付確率場を用いた発話意図の推定とその結果を用いた音声認識結果の検証を行った。小規模なデータでの評価ではあるが、音声認識誤りや未登録語などに対して頑健に動作し、音声認識誤りおよびタグ推定の誤りを修正できることを確認した。今後は、大規模なコーパスを用意し、検証を進める。

また、素性やタグの付与の仕方により結果は変わるので、最適なものを探す必要がある。

条件付確率場については、セミマルコフ CRF [17] や隠れ CRF [18] などが提案されており、これらの導入も検討する。

さらには、用いている素性ベクトルの次元数が大きいため、確率的潜在構造解析 (Probabilistic Latent Semantic Analysis : PLSA) [19] などを用いて次元圧縮の効果も検討する。

現状では、音声対話システムとして完成していないが、今後我々が取り入れるべき機能を示す。

- 未登録語と音声認識誤りを含む検索要求に対する頑健な検索
- 曖昧な質問、例えば「お勧めのレストランを教えてください」などを実現する検索
- 人間のエージェントが知識や経験を積むことにより成長していくのと同じように、対話履歴や、過去の推薦結果の採用、不採用の履歴から学習していくシステム

1 つ目の機能については、検索を行う際に出現形のみを用いて行うのではなく、読みや品詞などの情報を用いて検索を行い、検索されたテキスト群を用いて音声認識辞書および言語モデルを適応的に再構成し、音声認識を行うことで実現できるものと考えている。2 つ目の機能については、ブログなど Web 上の情報から評判情報を抽出する技術 [20] の応用により実現できるものと考えている。3 つ目の機能については、能動学習や半教師有り学習などのアプローチを取り入れる。

発話の分類を行うためのタグセットとして DAMSL (Dialog Act Markup in Several Layers) [21] がある。DRI (Discourse Research Initiative) によって構築されたアノテーションのため

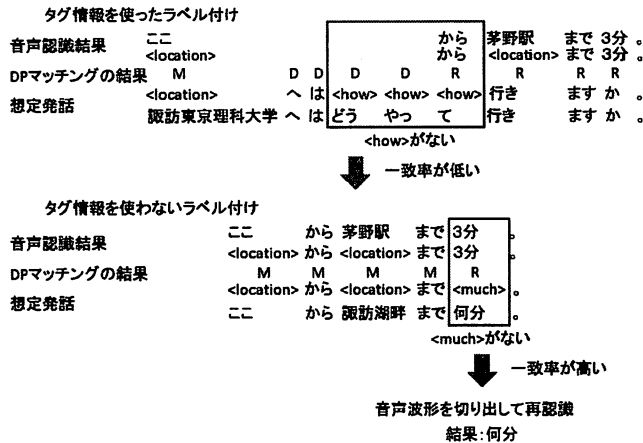


図 5 音声認識結果の検証の成功例

出現形	品詞	構文情報	重要語タグ	ラベル
横	(名詞-一般)//	2	<location>	4/0.456256
や	(助詞-並立助詞)//	1	<location>	4/0.592582
浜谷	(名詞-一般)//	2	<location>	4/0.789705
に	(助詞-格助詞-一般)//	0	0	4/0.887284
は	(助詞-係助詞)//	1	0	4/0.943094
何が	(名詞-代名詞-一般)//	2	<what>	4/0.975754
が	(助詞-格助詞-一般)//	1	<what>	4/0.984368
あり	(動詞-自立)/五段・ラ行/連用形	2	0	4/0.988463
ます	(助動詞)/特殊・マス/基本形	0	0	4/0.987820
か	(助詞-副助詞/並立助詞/終助詞)//	0	0	4/0.982929
.	(記号-句点)//	0	0	4/0.964099

図 6 未登録語に対するラベル推定の一例

の枠組みである。今後、DAMSL の導入も検討する。

謝辞 CRF++を公開してくださっている Google 株式会社 工藤拓氏、および SVMlight を公開してくださっている Prof. Thorsten Joachims に感謝いたします。

### 文 献

- [1] 西村 竜一, 西原 洋平, 鶴身 玲典, 李 見伸, 猿渡 洋, 鹿野 清宏: “実環境研究プラットフォームとしての音声情報案内システムの運用”, 電子情報通信学会論文誌, Vol.J87-D-II, No.3, pp.789-798, 2004.
- [2] 駒谷 和範, 河原 達也, 清田 陽司, 黒橋 禎夫, Pascale Fung: “柔軟な言語モデルとマッチングを用いた音声によるレストラン検索システム”, 情報処理学会研究報告, SLP-39-30, 2001.
- [3] 翠 輝久, 河原 達也, 正司 哲朗, 美濃 導彦: “質問応答・情報推薦機能を備えた音声による情報案内システム”, 情報処理学会論文誌, Vol.48, No.12, pp.3602-3611, 2007.
- [4] Heather Pon-Barry, Fuliang Weng and Sebastian Varges: “Evaluation of Content Presentation, Strategies for an In-car Spoken Dialogue”, ICSLP2006, pp. 1930-1933, 2006.
- [5] Antonio Roque, Anton Leuski, Vivek Rangarajan, Susan Robinson, Ashish Vaswani, Shri Narayanan, David Traum: “Radiobot-CFF: A Spoken Dialogue System for Military Training”, ICSLP2006, 2006.
- [6] M. Johnston, S. Bangalore, G. Vasireddy, A. Stent, P. Ehlen, M. Walker, S. Whittaker, P. Maloor: “MATCH: An Architecture for Multimodal Dialogue Systems”, Proc. of the 40th ACL, pp. 376-383, 2001.
- [7] J. Lafferty, A. McCallum, and F. Pereira: “Conditional Random Fields: Probabilistic models for segmenting and labeling sequence data”, Proc. of ICML, pp.282-289, 2001.
- [8] Andrew Mccallum, Dayne Freitag, Fernando Pereira:

- “Maximum Entropy Markov Models for Information Extraction and Segmentation”, ICML 2000, pp591-598, 2000.
- [9] <http://chasen.org/~taku/software/cabocha/>
- [10] <http://crfpp.sourceforge.net/>
- [11] Vladimir N. Vapnik: “The Nature of Statistical Learning Theory”, Springer, 1995.
- [12] Yoav Freund: “Boosting a weak learning algorithm by majority”, Information and Computation, 121(2):256-285, 1995.
- [13] <http://www.kokken.go.jp/katsudo/seika/corpus/>
- [14] <http://julius.sourceforge.jp/index.php>
- [15] T. Joachims, “Making large-Scale SVM Learning Practical”, Advances in Kernel Methods - Support Vector Learning, MIT-Press, 1999.
- [16] 齋藤 邦子, 鈴木 潤, 今村 賢治: “CRF を用いたブログからの固有表現抽出”, 言語処理学会第 13 回年次大会発表論文集, D1-3, 2007.
- [17] Sunita Sarawagi and William W. Cohen: “Semi-Markov conditional random fields for information extraction”, In Advances in Neural Information Processing Systems 17, pp.1185-1192, 2004.
- [18] S. B. Wang, A. Quattoni, L.-P. Morency and D. Demirdjian: “Hidden conditional random fields for gesture recognition”, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.1521-1527 2006.
- [19] T. Hofmann: “Unsupervised learning by probabilistic latent semantic analysis”, In Proc. of Uncertainty in Artificial Intelligence, 1999.
- [20] 杉木 健二, 松原 茂樹: “消費者の意見に基づく商品検索”, 情報処理学会論文誌, Vol.49, No.7, pp.2598-2603, 2008.
- [21] J. Allen and M. Core: “Draft of DAMSL: Dialog Act Markup in Several Layers”, 1997.