

## 公共音声対話システムにおける N-gram と Grammar の融合によるタスク拡張

北村 任宏<sup>†</sup> 戸田 智基<sup>†</sup> 川波 弘道<sup>†</sup> 李 晃伸<sup>††</sup> 猿渡 洋<sup>†</sup>  
鹿野 清宏<sup>†</sup>

<sup>†</sup> 奈良先端科学技術大学院大学 情報科学研究科 〒630-0192 奈良県生駒市高山町 8916-5

<sup>††</sup> 名古屋工業大学 大学院情報工学研究科 〒466-8555 愛知県名古屋市昭和区御器所町

E-mail: †{taka-kit,tomoki,kawanami,sawatari,shikano}@is.naist.jp, ††ri@nitech.ac.jp

あらまし 音声認識, 音声対話システムを実環境で利用する場合, 設置する環境に適したタスクドメインを設定することが重要である. システムをより普及させるためには簡易に所望のタスクへと適用, もしくは拡張する技術が必須である. 既存の音声対話システムに新たなタスクを追加する際, 発話に対するドメインの判別や対話戦略の構築が必要となる. その際に, 統計的言語モデルの作成時は新たに大量のコーパスを必要とし, 製作のために多大な時間を要する. 本稿では, 統計的言語モデルに基づく既存の対話システムにおいて, ネットワーク文法を併用することで簡易にタスクを拡張し, 発話に対して高精度にドメインの判別, 音声認識を行う手法を提案する. 実験的評価より, ドメインの判別率は90%以上で, それに基づいた単語認識率もほぼ90%以上の値を示し, 高精度に判別, 認識が行えた. 以上より, 提案手法の有効性が示せた.

キーワード N-gram, Grammar, 融合, 音声対話システム, タスク拡張

## Task extension with fusion of N-gram and grammar in public speech dialogue system

Takahiro KITAMURA<sup>†</sup>, Tomoki TODA<sup>†</sup>, Hiromichi KAWANAMI<sup>†</sup>, Akinobu LEE<sup>††</sup>, Hiroshi SARUWATARI<sup>†</sup>, and Kiyohiro SHIKANO<sup>†</sup>

<sup>†</sup> Graduate School of Information Science, Nara Institute of Science and Technology  
8916-5, Takayama-cho, Ikoma, Nara, 630-0192 JAPAN

<sup>††</sup> Department of Computer Science and Engineering, Nagoya Institute of Technology  
Gokiso-cho, Showa-ku, Nagoya 466-8555 JAPAN

E-mail: †{taka-kit,tomoki,kawanami,sawatari,shikano}@is.naist.jp, ††ri@nitech.ac.jp

**Abstract** A suitable task domain for speech recognition and speech dialogue systems used in a real environment is important. It is indispensable to popularize such systems so that its technology can easily be applied or enhanced to a required task. When a new task is added to an existing speech dialogue system, discrimination of the domain and construction of new dialogue strategy are needed. In using a new statistical language model, a large amount of corpus is needed and it takes much time to produce it. We propose a technique that can add a new task easily by introducing a network grammar to an existing statistical language model-based speech dialogue system. The proposed method can realize domain discrimination and speech recognition with high accuracy. In the experimental evaluation, more than 90% of domain discrimination and word recognition rate are realized. The effectiveness of the proposed technique is confirmed.

**Key words** N-gram, Grammar, fusion, spoken dialogue system, Task extension

## 1. まえがき

近年、音声認識技術の発展や計算機の高性能化により、様々な音声対話インタフェースが実用化され、注目を集めている。音声が目される一因として音声は人と人がコミュニケーションする際、最も意思を伝えやすく、基本的な動作であることから、キーボードやマウス、タッチパネルなどのインタフェースよりも使いやすく、人に優しいインタフェースであることが挙げられる。音声認識技術実用の例として、IBMのVia Voice [1]などのすでに商用レベルに達しているディクテーションソフトや、また音声入力による簡易なコマンド操作を行うカーナビゲーションシステムや、家電などが挙げられる。さらに今後、音声認識の応用として先の万博でも登場していた人型ロボットなどに用いられることが考えられる。ロボットが人間の実社会で活躍、共存するためには人とロボットがコミュニケーションを容易にとることが必要不可欠であり、音声対話への期待がより高まってきている。

音声認識、音声対話システム [2] [3] [4] [5] [6] を実環境で利用する場合、設置する環境に適したタスクドメインを設定することが重要である。しかし、実環境においてユーザーの対話システムへのタスクの要求は多岐に渡り、狭く限られたタスクドメインではユーザーを満足させることは難しい。そこで、幅広いタスクドメインを扱うために、統計的言語モデル（以下 N-gram）を用いた大語彙認識を行う方法が考えられる。N-gram による認識は自然発話を持つ文章表現の多様性に対応できることから、実環境に設置する場合、非常に有効である。我々は N-gram に基づく音声情報案内システム「たけまるくん」[11]を開発し、2002年11月より奈良県生駒市にある「北コミュニティセンター ISTA はばたき」で稼働させている。対話戦略は一問一答で主に館内施設、生駒市観光案内に関する応答や、また天気予報の応答など幅広いタスクドメインを設定している。

たけまるくんのような N-gram に基づく既存の音声対話システムは、新たにタスクドメインを追加する際、ユーザーの発話のドメイン判別や対話戦略の構築が必要となる。また、大語彙による音声認識が可能であるが、モデルの作成時に大量のコーパスを必要とすることから、新たにタスクを追加する際は追加するタスクドメインのコーパスを収集し、再学習しなければならない。よって製作のために多大な時間を要し開発のためのコストが増大する。システムをより普及させるためには簡易に所望のタスクへと適用、もしくは拡張できる技術が必須である。

そこで本稿では、N-gram に基づく既存の対話システムにおいて、ネットワーク文法（以下 Grammar）を併用することで簡易にタスク拡張を行う方法を提案する。Grammar は人手で作成することから所望のタスクドメインのモデルを簡易に作成でき、想定内発話に対して高精度な認識を行うことができる利点がある。これらの特徴を活かし、対話システムに追加したい新たなタスクドメインを Grammar で記述し、既存の N-gram と並列に認識を行うことで、簡易に音声対話システムを拡張できることが考えられる。

以下、2. 節では各言語モデルの特徴を述べる。3. 節では提案

手法について述べる。4. 節では提案手法の実験的評価によりその有効性を示す。5. 節で本稿のまとめを述べる。

## 2. 各言語モデルの特徴

音声認識は、入力された音声  $x$  に最もよく一致する言語表現  $\omega$  を推定する過程で、次式によって表される。[7]

$$\hat{\omega} = \underset{\omega}{\operatorname{argmax}} P(\omega|x) \quad (1)$$

$$= \underset{\omega}{\operatorname{argmax}} p(x|\omega)P(\omega) \quad (2)$$

ここで、 $P(\omega)$  が言語モデルにより表される言語確率であり、言語表現に関する事前知識（事前確率）である。

一般的に、言語モデルは N-gram と Grammar の 2 種類に大別される。各言語モデルの主な特徴を表 1 に示す。

### 2.1 N-gram の特徴

N-gram は大量のテキストデータから自動的に学習されることが大きな特徴である。主な利点として、大量のコーパスから単語の並びを統計的に学習するので大語彙化が容易であることが挙げられる。また、言語モデルの作成時に想定していない発話に対してもある程度柔軟に対応ができる利点がある。広く一般的なドメインの発話に対しては、非常に頑健で有効なモデルであるといえる。

学習に用いたコーパスに含まれる語彙や語尾様式、発話表現などの言語的特徴が学習後の言語モデルに反映されるものの、必ずしも認識結果が正確な文章表現で出力されるとは限らない。そのため、後段により高度な自然言語処理が必要となる。また、言語モデルに新たに認識できるタスクドメインを追加する際、対象タスクに応じたコーパスを大量に収集し再学習する必要がある。そのために想定されるあらゆるタスクドメインごとの大量のコーパスの収集や、収集したテキストの整形など N-gram におけるタスクドメインの追加は非常に高いコストを伴う作業が必要となる。また、N-gram を用いて複数のタスクを取り扱うことは可能であるが、タスクドメインの判別は難しい。複数の N-gram のモデルを用いてドメインの判別を行う手法 [8] が提案されているが、システムにかかる負荷が増大する。

### 2.2 Grammar の特徴

Grammar は、人手で作成する有限状態文法であることが大

表 1 各言語モデルの特徴

Table 1 Features of each language model

	コーパスから統計的に学習
N-gram	○大語彙化が容易
	○想定外発話に柔軟に対応
	×出力単語列の制御が不可
	×学習に大量のコーパスが必要
	人手で記述
Grammar	○出力単語列の制御が可能
	○想定内発話を高精度に認識
	×想定外発話は認識困難
	×発話表現の網羅が困難

きな特徴である。利点として、想定した発話に対しては高精度な認識が可能で、簡易に作成できることが挙げられる。よって特定の狭いタスクドメインの発話において有効である。また、人手で記述するため不要な単語列の出力を抑えることができる有限状態文法であることから、認識結果がどのような遷移を経て得られたか、作成する際に記述した単語カテゴリーからどのようなタスクドメインの単語かといった情報が詳細かつ正確に得ることができ、対話システムに用いる際の制御がしやすい。またシステムにかかる負荷も少ない。

しかし、記述文法は想定内の発話のみを受理するので、想定外発話の認識は困難である。実環境における多様な発話表現を網羅するのも現実的ではない。つまり限定したタスクに大しては向いているが、広いタスクドメインに対しては不向きであるといえる。

### 3. N-gram と Grammar の併用によるタスク拡張

#### 3.1 提案システム

##### 3.1.1 N-gram と Grammar の併用

N-gram を用いた既存の音声対話システムに、新たにタスクドメインを追加するための提案手法として Grammar を併用することで、簡易にタスクドメインの追加を行う。専門的で狭いタスクドメインは Grammar で認識し、広く全般的なタスクドメインを N-gram で認識することを行う。互いの言語モデルの利点を活かすことで上述した問題が緩和できることが考えられる。具体的には、各タスクドメインを設定した言語モデルを用いて並列に認識を行い、認識された結果より、まずユーザーの発話のドメインを判別する。そして判別されたタスクドメインの言語モデルの認識結果を音声対話システムとしての最終的な認識結果として利用する。提案システムを図 1 に示す。ここで、音声認識システムに N-gram, Grammar のそれぞれに、Julius, Julian [9] を用いる。

##### 3.1.2 ドメイン判別

対話システムを運用する上で複数のタスクドメインを扱う際、適切な応答を返すために発話に対して正確にドメインを判別することが重要である。ドメインを判別するために以下に示す各

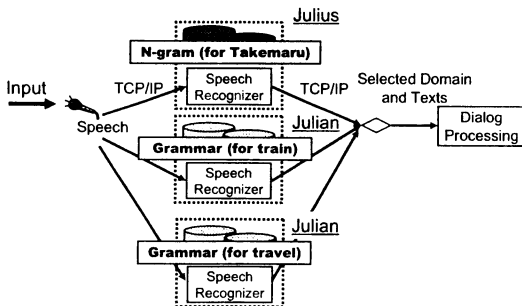


図 1 各言語モデルを併用するシステム構成図

Fig. 1 System configuration chart where each language model is used together

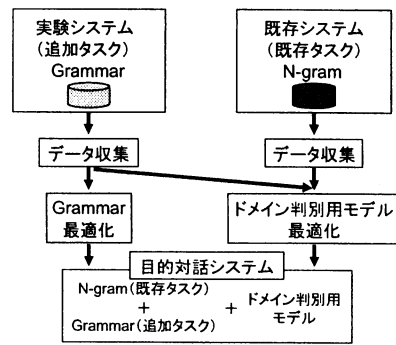


図 2 タスクドメインの追加手順

Fig. 2 Procedures of adding task domain

パラメータを用いる。各パラメータを用いてドメイン判別用モデルを作成し、図 2 に示す手順で判別用モデルの最適化を行う。本研究では、Julius, Julian が出力する以下のスコア [10] を用いてドメインの判別を行うことを検証する。

- 音響スコア (AM)
- 言語スコア (LM)
- 単語信頼度 (CM)
- 単語探索数の逆数 (1/nodes)
- 音響スコア+言語スコア (AM+LM)

ユーザーの発話が各言語モデルの想定タスク内である場合、上記の各パラメータのスコアが高い値を算出し、想定外タスクである場合は低い値を算出することが考えられる。よって発話に対して対話システムが用意したタスクドメインのいずれかに判別が可能となる。ここで、Julian は言語スコアを算出しないが、言語スコアを 0 として扱う。

#### 3.2 タスク追加手順

新たなタスクドメインを追加するための手順を図 2 に示す。製作者が想定したタスクドメインに対して何も参考にせず、簡易に Grammar を作成する。これを初期モデルとする。次に、実験対話システムとして、既存のタスクドメインの対話システムに加え、Grammar の初期モデルを用いた追加タスクドメインの対話システムを作成し、実際にユーザーに使用してもらう。タスク別ごとに発話データを収集し、収集した発話データを参考にして順次モデルを更新し、タスクに対する Grammar モデルの最適化を行う。また、ドメイン判別用モデルの最適化も行う。そして最適化された各モデルと既存の N-gram を目的対話システムに用いる。

### 4. Grammar 最適化の実験的評価

#### 4.1 実験条件

実験条件を表 2 に示す。N-gram に用いるタスクドメインは「たけまるくんタスク」、Grammar に用いるタスクは「路線探索タスク」に設定する。ここで、「路線探索」は路線案内、時刻案内、料金案内などの応答をタスクドメインに設定した。

N-gram と Grammar を併用する上で、Grammar が実環境における発話に対してどの程度の認識精度を示すのかを把握す

表2 実験条件

Table 2 Experimental conditions

音声認識システム	Julius-3.4.2 (N-gram)	Julian-3.4.2 (Grammar)
想定タスク	たけまるくん	路線探索
語彙数	42,605 語	415 語 (駅名: 394 語)
音響モデル	収集発話: 4,197 文 JNAS 読み上げ音声: 40,086 文	
更新用データ	・たけまるタスク: 500 発話 ・路線探索タスク: 500 発話	
テストセット	・たけまるタスク: 100 発話 ・路線探索タスク: 100 発話	

る必要がある。そして、Grammar を記述する際の製作のためのコストを知ることが重要である。そのために、初期 Grammar を作成し、最適化するまでを図3に示す手順で行う。

図2より、収集した発話データから評価用のテストセットデータをあらかじめ抜き出しておき、残りのデータを一定量ずつ参考にしなが Grammar を順次更新し、評価を行う。これにより、タスクドメインに対するユーザーの発話の傾向を調査することができ、Grammar に反映させることができる。また参考にするデータ数が少量で、高精度な認識率を算出する Grammar が作成できれば、製作コストが抑えられ、開発者の負担を減らすことが期待できる。よって更新、評価に用いた各データはそれぞれ「たけまるくん」と、「たけまるくん」を基に路線探索用に構築した対話システムから収集した実発話を用いる。テストセットデータは更新用データには含まれていない。更新用データセットの数は1回の更新につきそれぞれ100文で、計5回更新を行った。

また、対話システムとしての頑健性の向上のために、Julius に含まれていない、Julian のタスクに関する主要な単語を Julius 側に未知語として単語辞書に追加した。Julian の想定タスク内発話に対して、ドメイン判別に失敗した際においても、Julius での認識結果を有効に利用し適切な応答文を返すことが期待できる。

4.2 Grammar 作成のためのコストの検討

4.2.1 更新用データによる評価

表3, 4, 5, 6 はモデルの更新用データセット1から5を用

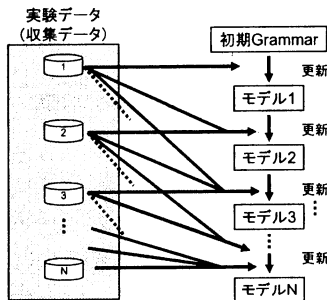


図3 Grammar の更新手順

Fig. 3 Grammar's upgrade procedures

いて初期モデルからモデル5までそれぞれの評価方法で評価した値を示している。なお、各評価はモデルの更新が全て終わってから評価した。

更新前の初期モデルは、各更新用データに対して単語認識率や文法に対するデータの受率率など精度が低いことが分かる。この初期モデルから、順次データを参考し、モデルの更新を行うと、まず1回目の更新で単語認識率が約36%上がり、その他の評価スコアの精度も上昇している。そして、参考データ数を増やし、モデルを更新するにつれて、単語認識率、受率率が向上し、未知語率が減少しており、モデルの更新とともに実環境に使用できうるモデルに更新されていることが分かる。しかし、モデルの更新により語彙数が増加するため、パープレキシティは更新の都度増大している。

4.2.2 テストセットデータによる評価

表7と表8に路線探索ドメインのテストセットを用いて、更新した Grammar の各モデルを評価した値を示す。

表7より、Grammar の初期モデルにおいて路線探索ドメインは想定内タスクドメインであるが単語認識率が約48%と精度が低い。逆に N-gram は想定外タスクドメインであるにも関わらず約71%の認識率を算出している。これは N-gram の単語辞書に未知語として駅名を追加したことから想定外タスクドメイン発話に対しても柔軟にある程度対応できていることが考えられる。

しかし、1回目の更新で初期モデルから約38%の認識率の向上が見られた。また4回目までの更新ではスコアに大きな差はなく、5回目の更新で4回目のモデルより3.3%向上した。全体的に2回目以降の更新に伴う評価スコアは徐々に一定の値をとっていく傾向が見られた。つまり、1回目の100文の実データを参考にモデルを更新するだけで十分な認識率の精度が得

表3 各更新用データにおけるモデル別の単語認識率 (%)

Table 3 Word recognition rates (%) for individual data sets when using each model

Model	Initial	1	2	3	4	5	
Data set	1	44.5	90.3	86.7	87.6	86.4	85.9
	2	52.6	85.4	90.1	90.7	90.7	87.3
	3	50.8	81.8	86.1	91.4	91.2	91.7
	4	48.8	80.5	84.7	84.2	90.0	90.0
	5	44.0	80.4	83.7	82.7	86.0	89.8
Average	48.1	83.7	86.3	87.3	88.9	88.9	

表4 各更新用データにおけるモデル別の受率率 (%)

Table 4 Accept rates (%) for individual data sets when using each model

Model	Initial	1	2	3	4	5	
Data set	1	27.9	89.5	89.5	89.5	89.5	89.5
	2	33.3	80.6	94.6	95.7	95.7	95.7
	3	28.9	71.1	75.6	87.8	87.8	87.8
	4	28.6	75	86.9	88.1	97.6	97.6
	5	29.3	78.3	82.6	82.6	83.7	91.3
Average	29.6	78.9	85.8	88.7	90.9	92.4	

表 5 各更新用データにおけるモデル別の未知語率 (%)

Table 5 Out of vocabulary rates (%) for individual data sets when using each model

Model	Initial	1	2	3	4	5	
Data set	1	4.7	0.4	0.4	0.4	0.4	0.4
	2	4.1	0.7	0.0	0.0	0.0	0.0
	3	4.1	0.7	0.5	0.5	0.5	0.5
	4	4.9	1.3	0.9	0.9	0.2	0.2
	5	4.7	1.1	0.9	0.9	0.9	0.7
Average	4.5	0.8	0.5	0.5	0.4	0.4	

表 6 各更新用データにおけるモデル別のパープレキシティ

Table 6 Perplexity for individual data sets when using each model

Model	Initial	1	2	3	4	5	
Data set	1	34.9	51.3	71.0	73.8	75.1	78.4
	2	30.0	52.9	81.1	87.6	88.3	91.1
	3	38.7	63.0	96.0	93.2	94.4	98.5
	4	27.5	54.2	85.0	89.4	84.9	87.5
	5	31.4	62.3	88.8	92.8	94.1	96.5
Average	32.5	56.7	84.4	87.4	87.4	90.4	

られており、より精度の高いモデルを要する場合はより多くのデータを参考にすればよいことが分かった。また表 8 より、モデルの更新に伴う文法に対するテストセットの受率率は順次向上し、未知語率は減少する傾向がある。しかし、パープレキシティに関してはモデルの更新に伴い増大する傾向が見られた。これは更新用データの参考に伴い、語彙数が増加することに要因があると考えられる。

よって、4.2.1 節で示した傾向がテストセットにおける評価に対しても反映されていることも分かった。

## 5. 提案システムの実験的評価

### 5.1 スコアの大小比較による評価

Julius と Julian の認識結果から提案するパラメータの差分値を用いてドメインの判別を行う。まず、単純にスコア比較の場合の各ドメインのスコアの分布図を図 4 に示す。

図 4 より、各ドメインの分布が x 軸を中心に上下に分布すれば単純なスコア比較のみでドメインの判別が可能であるが、ド

表 7 Grammar の更新に伴う単語認識率 (%)

Table 7 Word recognition rates (%) according to upgrade of grammar

テストセット	路線探索		たけまるくん		
言語モデル	Grammar	N-gram	Grammar	N-gram	
Model	Initial	47.8	71.2	4.4	93.8
	1	85.3		26.6	
	2	84.4		25.7	
	3	84.7		26.0	
	4	84.7		32.5	
	5	88.0		31.6	

表 8 Grammar の更新に伴うその他の評価

Table 8 The other evaluation according to upgrade of grammar

テストセット	路線探索			
言語モデル	Grammar			
評価方法	受率率 (%)	未知語率 (%)	パープレキシティ	
Model	Initial	23	3.0	27.5
	1	75	0.6	43.5
	2	79	0.6	64.8
	3	82	0.6	67.9
	4	84	0.6	68.1
	5	85	0.4	69.7

メインの分布が混在していることから正確にドメインを判別できない。そのため適切な認識結果を選択できていないことから単語認識率の精度も低いことが分かる。

よってドメインの境界面を適切に引く必要がある。そこで SVM [12] を用いてドメインの識別面を求め、判別を行う。

### 5.2 閾値最適化による評価

SVM を用いてドメインの識別面を求めた。線形カーネルを用いて学習した。結果の一部を表 9, 10 に示す。まず表 9 より、高精度な結果が得られていることが分かる。スコアの大小比較では、ある一つのドメインの判別の精度は高いがもう一つのドメインに関しては精度が低い結果が得られている。しかし、表 9 より、閾値を最適化した場合の結果ではどちらのドメインに関してもほぼ 90% 以上の判別率を示しており、バランスよくかつ高精度に判別できている。

次にドメインの判別に伴う単語認識率においても、表 10 よりどちらのドメインにおいてもほぼ 90% 以上の認識率を示している。しかも路線探索ドメインにおいて表 7 に示されている想定内ドメイン時の発話における上限の値よりもモデル 5 を比較すると最大約 6% の単語認識率の向上が見られた。これは、Julius に行った未知語処理により、ドメインの判別に失敗しても N-gram 側である程度認識ができたことによると考えられる。

よって、提案するパラメータを用い、また判別用モデルを最適化することで、発話に対するドメインの判別と認識が正確に行え、提案手法の有効性が確認された。

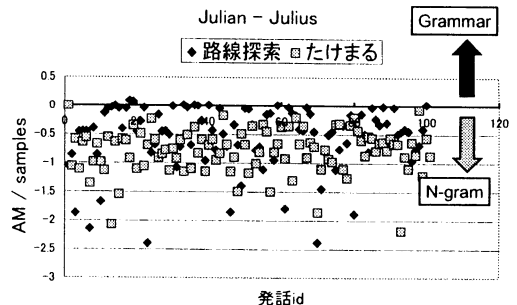


図 4 AM スコアによるドメインの分布

Fig. 4 distribution of domain in AM score

表 9 閾値最適化による各パラメータにおけるドメイン判別率 (%)

Table 9 Domain recognition rates (%) according to every parameter when being most suitable threshold

テストセット Model	路線探索			たけまるくん		
	Initial	1	5	Initial	1	5
AM	65	87	90	66	75	75
LM	95	95	95	88	88	88
CM	75	77	83	66	77	72
1 / nodes	78	96	95	35	36	39
AM+LM	79	92	93	82	90	88
AM, AM+LM	86	98	94	89	90	90
AM, AM+LM, CM	90	99	94	90	91	91
ALL	89	95	95	94	93	91

表 10 閾値最適化による用いた各パラメータにおける単語認識率 (%)

Table 10 Word recognition rates (%) according to every parameter when being most suitable threshold

テストセット Model	路線探索			たけまるくん		
	Initial	1	5	Initial	1	5
AM	70.7	85.5	89.9	78.5	86.7	85.0
LM	49.9	90.7	92.0	90.9	90.9	90.9
CM	53.7	86.5	90.4	75.5	86.7	85.0
1 / nodes	60.4	85.3	90.2	55.2	62.8	67.6
AM+LM	65.5	84.7	90.2	89.1	91.5	91.5
AM, AM+LM	53.9	84.7	90.2	88.5	91.5	91.2
AM, AM+LM, CM	57.8	84.9	90.2	90.3	91.5	91.7
ALL	49.9	91.7	93.8	92.6	92.6	91.5

## 6. まとめ

N-gram に基づく既存の音声対話システムに新たなタスクドメインを追加する方法として、Grammar で追加したいタスクドメインを作成し、簡易に対話システムを拡張する手法を提案し、実験的評価を行った。

Grammar 作成のコストを調べるために、実発話を 100 文ずつ用いて初期モデルから順次更新、評価した。一回目の更新でテストセットデータに対して単語認識率は約 85% で 5 回目の更新では 88% の値を示した。よって Grammar モデルは、100 文から 500 文の実データを参考にすることでタスクに対して最適化されることが分かり、作成の際のコストが抑えられることが期待できる。

また、N-gram と Grammar を併用した際のドメインを判別するために、判別用モデルを最適化し評価した。どちらのタスクドメインに対しても 90% 以上の判別率を示し、ドメイン判別に伴う単語認識率においても 90% 以上の精度が得られ、正確なドメイン判別、音声認識が行えた。そして、N-gram に Grammar の主要単語を未知語として追加することで Grammar の想定内ドメインの単語認識率が上限の値より、N-gram と Grammar を併用することで値が上昇する結果も同時に見られ、未知語処理とともに提案手法の有効性が示せた。よって、対話システムを簡易に拡張できることが期待できる。

今後の課題としては、より多くのタスクドメインを扱う場合の検証を行いたいと考えている。また、どちらのタスクドメインともとれるドメインの判別が困難な発話への対応が挙げられるが、ドメインの履歴情報などを用いて対応したいと考えている。以上の提案手法を対話システムに実装し、検証を行いたい。謝辞

この研究の一部は、文部科学省のリーディングプロジェクト「e-Society 基盤ソフトウェアの総合開発」によって行われたものである。

## 文 献

- [1] <http://www-06.ibm.com/jp/voiceland/products/v10/>
- [2] V.Zue, S.Seneff, J. Glass, J. Polifroni, C.Pao, T.J.Hazen, L.Hetherington: "JUPITER: A Telephone-Based Conversational Interface for Weather Information," IEEE Trans. Speech and Audio Processing, vol.8, no.1, pp.100-112, 2000
- [3] 駒谷 和範, 翠 輝久, 河原 達也, 奥乃 博, 大規模知識ベースの検索を行う音声対話システムの確認戦略の評価, 言語処理学会 第 10 回年次大会 発表論文集, pp. 87-90, March, 2004.
- [4] 小林哲則, 藤江真也, 松坂佐佐, 白井克彦, 人間会話ロボット—パラ言語の生成・理解機能を持つマルチモーダルインタフェース—, 日本音響学会誌, vol.61, no.2, pp.85-90, Feb., 2005
- [5] 松井俊浩, 麻生英樹, John Fry, 浅野太, 本村陽一, 原功, 栗田多喜夫, 速水悟, 山崎信行, 日本ロボット学会誌, Vol.18 No2, pp.300~307, 2000
- [6] <http://hil.t.u-tokyo.ac.jp/galatea/index-jp.html>
- [7] 鹿野清宏, 伊藤克亘, 河原達也, 武田一哉, 山本幹夫 (編): "音声認識システム", オーム社, pp.53-54, 2001.
- [8] Ian R. Lane, T. Kawahara, and T. Matsui. Language model switching based on topic detection for dialog speech recognition. In Proc. IEEE-ICASSP, Vol.1, pp.616-619, 2003.
- [9] A. Lee, T.Kawahara, K. Shikano: "Julius - An Open Source Real-Time Large Vocabulary Recognition Engine," In Proc. EUROSPEECH2001, pp.1691-1694, 2001.
- [10] Akinobu Lee, Kiyohiro Shikano, and Tatsuya Kawahara, "Real-time word confidence scoring using local posterior probabilities on tree trellis search,"
- [11] 西村竜一, 他, 信学論, Vol.J87-D-II, No.3, PP.789-798, 2004.
- [12] <http://www.csie.ntu.edu.tw/>