

# 対話システムにおけるユーザの発話意図推定

岩下 薫<sup>†</sup>      嶋田 和孝<sup>‡</sup>      遠藤 勉<sup>‡</sup>

<sup>†</sup> 九州工業大学大学院情報工学研究科情報科学専攻

<sup>‡</sup> 九州工業大学情報工学部知能情報工学科

## 概要

音声対話システムにおいて、ユーザの発話意図を正しく理解し、それに基づいて自然なインタラクション処理を行うことは非常に重要である。ユーザの発話意図を推定する手法には、ルールを用いる手法や事例に基づく手法がある。本稿では、事例に基づく2つの発話意図推定手法の概要と、それらを用いて行った発話意図推定実験について報告する。発話意図推定手法の1つ目は発話文間の類似度を用いる手法、2つ目は最大エントロピー法を用いる手法である。この2つの手法を用いて発話意図推定実験を行い、その有効性を確認する。

## Case-based Estimation of Speech Intention

Kaoru IWASHITA<sup>†</sup>, Kazutaka SHIMADA<sup>‡</sup>, and Tsutomu ENDO<sup>‡</sup>

<sup>†</sup> Graduate School of Computer Science and Systems Engineering, Kyushu Institute of Technology

<sup>‡</sup> Department of Artificial Intelligence, Kyushu Institute of Technology

## Abstract

Dialogue analysis is one of the most important tasks for human-machine interaction. It is important for dialogue systems to recognize the intention of an utterance and the intentional structure of a discourse. If a system estimates the intentions of each utterance in a dialogue correctly, it can recognize the structure of the dialogue.

This paper describes a method for estimation of an utterance. We use a corpus for this method. We compare two methods for intention estimation: (1) a similarity measure between an input utterance and utterances in the corpus and (2) the maximum entropy method. We evaluate the effectiveness of our method in the experiment.

## 1 はじめに

近年、音声認識技術の進歩などから音声対話システムの研究が盛んに行われている。我々の研究グループでも、対象を小学1年算数ドリルテキスト[1, 2]に絞り込み、人間との対話とドリルテキストとを相互参照しながら問題解決と知識獲得を行う対話システムの構築を行っている。この対話システムでは、教師であるユーザが発話とジェスチャを用いて生徒であるシステムに算数の問題を解かせるというタスクを想定している。このような対話システムにおいて、ユーザの発話意図を正しく理解し、それに基づいて自然なインタラクション処理を行うことは非常に重要である。したがって、本研究では対話システムにおけるユーザの発話意図を正しく推定することを目的とし、研究を行っている。

ユーザの発話意図を推定する手法として、ルールを用いる手法[3]や対話事例に基づく手法[4, 5]がある。ユーザがある意図に基づいて発話する場合、その影響は形態素、キーワード、音韻、文脈など発話に関連する事象に様々な形で現れる。これらを基に発話意図推定のルールを作成すると、ルール自体が複雑になる上、発話の多様性に対応するためには非常に多くの推定ルールが必要となる。それに対し、対話事例に基づく手法は、自然な発話に対応するアプローチとして有効である。事例に基づく意図推定を行っている研究には入江ら[4]や乾ら[5]がある。入江らは、発話文間の類似度と入力発話に至るまでの意図系列を用いて発話意図推定を行っている。また乾らは、最大エントロピー法(ME法)を用いて自由回答アンケートの意図分類を行っている。

著者らはこれまでに、タスクである小学1年算数ドリルテキストを対象とした対話システムにおけるユーザの発話意図推定として、発話文の類似度を用いた事例に基づく発話意図推定手法を用いて推定実験を行っている[6]。ここでは複数の類似度尺度を用いて入力発話と発話事例の類似度を求め、各尺度の正解率の比較を行うとともに、類似度を用いる手法、ME法を用いる手法それぞれの有効性を確認する。

## 2 意図タグ付きコーパス

事例に基づく発話意図推定を行う場合には、ユーザの意図を表す意図タグが付与された事例データが必要となる。そこで、テキストとして書き起こした対話データに意図タグを付与し、意図タグ付き対話

コーパスを構築した。コーパス作成のために用いた対話データは、対話システムで扱う算数のドリルテキストを用いて行われた模擬対話のデータであり、3種類の算数の問題に対して各27対話ずつ、計81対話ある。また、意図タグは<情報要求:Howto>や<同意表示:Correct>、<不同意表示>など全部で18種類ある。このコーパスの規模を表1に示す。

表 1: 発話意図タグ付きコーパス

	ターン数	発話数
教師	1390	1509
生徒	1449	1651
計	2839	3160

タグ付きコーパス作成にあたっては、タグ付とマニュアルに従って意図タグ設計者とタグ付と経験者の2名でタグ付とを行った。この意図タグ付きコーパスの信頼性については、意図タグの設計およびタグ付ととマニュアル作成後にを行った発話意図タグ付とと実験により確かめられている[6]。

## 3 発話意図推定

本研究では、2つの発話意図推定手法を用いて対話におけるユーザ発話の意図推定を行う。1つ目は発話文の表層の類似度を用いたシンプルな手法(以下、手法1とする)、2つ目は、テキスト自動分類などでよく用いられている最大エントロピー法を用いた手法(以下、手法2とする)である。本節では、この2つの手法を用いた発話意図推定の概要について述べる。

### 3.1 発話文の類似度を用いる手法

手法1では、入力として与えられる発話と、タグ付きコーパス中の発話事例の表層の類似度を用いて発話意図推定を行う。この手法は、「類似した発話は意図も類似する可能性が高い」という仮定に基づいている。発話意図推定の処理の流れは以下の通りである。

1. 茶釜<sup>1</sup>を用いて発話文の形態素解析を行う。
2. 形態素情報を用いて、発話文のベクトル空間を作成する。
3. 入力発話  $U_x$  と発話事例  $U_y$  の類似度  $SIM(U_x, U_y)$  を求める。

<sup>1</sup>茶釜 <http://chasen.naist.jp/hiki/ChaSen/>

#### 尺度 1 2 値の内積

$$SIM(U_x, U_y) = \sum_{i=1}^T x_i \cdot y_i$$

ただし,  $x_i, y_i$  は, それぞれ  $U_x, U_y$  における形態素  $i$  の存在を表す 2 値変数

#### 尺度 2 2 値の余弦

$$SIM(U_x, U_y) = \frac{\sum_{i=1}^T x_i \cdot y_i}{\sqrt{\sum_{i=1}^T x_i^2 \times \sum_{i=1}^T y_i^2}}$$

ただし,  $x_i, y_i$  は, それぞれ  $U_x, U_y$  における形態素  $i$  の存在を表す 2 値変数

#### 尺度 3 出現頻度の内積

計算式は尺度 1 と同様

ただし,  $x_i, y_i$  は, それぞれ  $U_x, U_y$  における形態素  $i$  の出現頻度

#### 尺度 4 出現頻度の余弦

計算式は尺度 2 と同様

ただし,  $x_i, y_i$  は, それぞれ  $U_x, U_y$  における形態素  $i$  の出現頻度

#### 尺度 5 出現頻度を補正した内積

$$SIM(U_x, U_y) = \frac{\min(M_x, M_y)}{\max(M_x, M_y)} \times \sum_{i=1}^T x_i \cdot y_i$$

#### 尺度 6 出現頻度を補正した余弦

$$SIM(U_x, U_y) = \frac{\min(M_x, M_y)}{\max(M_x, M_y)} \times \frac{\sum_{i=1}^T x_i \cdot y_i}{\sqrt{\sum_{i=1}^T x_i^2 \times \sum_{i=1}^T y_i^2}}$$

#### 尺度 7 Dice 係数

$$SIM(U_x, U_y) = \frac{2 \sum_{i=1}^T x_i \cdot y_i}{\sum_{i=1}^T x_i^2 + \sum_{i=1}^T y_i^2}$$

#### 尺度 8 Jaccard 係数

$$SIM(U_x, U_y) = \frac{\sum_{i=1}^T x_i \cdot y_i}{\sum_{i=1}^T x_i^2 + \sum_{i=1}^T y_i^2 - \sum_{i=1}^T x_i \cdot y_i}$$

#### 尺度 9 形態素の一致度 (関連研究: 入江ら [4])

$$SIM(U_x, U_y) = \frac{2M_{xy}}{M_x + M_y}$$

#### 4. 類似度最大の事例に付与された発話意図タグの中で最も数が多いものを推定結果とする.

ここで,  $T$  はコーパス中出现する形態素の異なり数である. また, 尺度 1 と 2 を用いる場合の  $x_i, y_i$  は, それぞれ  $U_x, U_y$  における形態素  $i$  の存在を表す 2 値変数, それ以外の尺度を用いる場合の  $x_i, y_i$  は, それぞれ  $U_x, U_y$  における形態素  $i$  の出現頻度である.  $M_x, M_y$  はそれぞれ  $U_x, U_y$  の形態素数,  $M_{xy}$  は  $U_x$  と  $U_y$  で一致する形態素数を表す.

#### 3.2 各尺度の比較実験

発話意図推定の手順 3 において, 9 つの尺度を用いて発話意図推定を行う. それぞれの正解率の平均を求め, 発話意図推定の類似度計算においてどの尺度が有効であるかを確かめる. ここでの正解率とは, 実験データの発話文に予め付与された意図タグを正解タグとした場合の推定結果の意図タグとの一致率を表す.

実験に用いたデータは, 2 節で述べたタグ付きコーパスに収録された全 81 対話である. 1 対話を実験データ, その実験データと同じ算数問題を扱う別の 26 対話を事例データとして交差検定を行い, 正解率の平均を求めた. 実験に用いるのは対話中出现する全教師 (ユーザ) 発話であり, 推定の対象となるのも教師発話のみである.

手法 1 の手順 3 における類似度計算尺度の比較実験の結果を表 2 に示す. 3 種類の算数問題に対し対話データがそれぞれ 27 対話あるため, 各問題で 27 対話それぞれの正解率を求め, その平均を問題毎の正解率として表した.

実験の結果, 2 値の内積・余弦, Dice 係数, Jaccard 係数, 形態素の一致度を尺度として用いた手法が最も高い正解率を示し, 2 値の内積 (尺度 1) と 2 値の余弦 (尺度 2) の組, Dice 係数 (尺度 7) と Jaccard 係数 (尺度 8) の組ではすべてのケースにおいてまったく同じ正解率を得た.

高い正解率を示した 2 値の内積 (尺度 1) と同様に内積を用いているにも関わらず, ベクトルの値として出現頻度を用いた場合 (尺度 3) では最も低い正解率を示した. 同じ出現頻度を用いた場合でも, 余弦を尺度としている手法 (尺度 4) では他の手法と大差ない結果となっていることから, 出現頻度による補正が行われない内積では, 出現頻度の高い形態素がノイズとなって不正解を導いていると考えられる. しかし, 出現頻度を単純に補正した尺度 5 および 6 においても, それぞれ 2 値のベクトルを用いた尺度 1 および 2 には及ばない結果となった. したがって, 短い発話文が大半を占めるような対話コーパスでは, 出現頻

表 2: 比較実験結果 (正解率: %)

	尺度 1	尺度 2	尺度 3	尺度 4	尺度 5	尺度 6	尺度 7	尺度 8	尺度 9
問題 1	<b>76.47</b>	<b>76.47</b>	47.00	74.46	72.40	74.69	74.79	74.79	76.38
問題 2	85.10	85.10	74.30	85.89	82.36	82.98	85.87	85.87	<b>86.02</b>
問題 3	84.67	84.67	78.32	84.13	80.31	82.59	<b>84.78</b>	<b>84.78</b>	84.13
平均	82.08	82.08	66.54	81.49	78.35	80.09	81.81	81.81	<b>82.18</b>

度よりも各素性が生じたか否かの方が重要であり、文書分類などで用いられるような出現頻度はかえってノイズになりやすいと考えられる。

また、問題毎の正解率を見ると、どの尺度を用いた場合でも、問題 2 と 3 に比べて問題 1 の正解率が低くなっている。これは、問題 1 の難易度 (システムが問題を解く上での難易度) が他の 2 つの問題に比べて高いことが原因であると考えられる。問題 1 は蝶の数を数えて同じ数のカードと線で結ぶ問題、問題 2 は枠の中のオブジェクトの数を数えて隣の四角に答えを書き込む問題、問題 3 は大きな枠の中に 3 種類あるオブジェクトそれぞれの数を数えて、その数の分だけ下の枠に丸印を書き込む問題である。問題 2 と 3 は数を数えて答えを書き込むだけであるが、問題 1 では数を数えるだけでなく、同じ数のカードを見つけるという行為も必要となる。ゆえに教師であるユーザの発話もその説明のために長く複雑になりやすく、発話意図推定が難しくなっていると考えられる。

### 3.3 ME 法を用いる手法

手法 2 では、最大エントロピー法 (以下、ME 法 [7]) を用いて発話意図推定を行う。発話意図推定の処理の流れは以下の通りである。

1. 茶釜を用いて発話文の形態素解析を行う。
2. 事例データから素性を抽出する。
3. 抽出した素性をもとに訓練データを作成する。
4. ME 法を用いて学習し、確率分布を求める。
5. 入力発話を解析し、最大の確率値が得られた発話意図を推定結果とする。

本研究では、コーパス中の発話事例に出現する全形態素を素性として用いる。手順 3 で作成される訓練データは、事例データ中の各発話文に対して付与

されている発話意図タグと出現する素性をまとめたものである。

次に、ME 法を用いて発話文に付与すべき発話意図タグの尤もらしさを確率分布として求めるモデルについて述べる。

発話文の集合を  $B$ 、出力である発話意図の集合を  $A$  とするとき、発話文  $b (\in B)$  で出力値  $a (\in A)$  となる事象  $(a, b)$  の確率分布  $p(a, b)$  を ME 法により推定する。出力となる発話意図  $a$  は  $n$  個の発話意図  $a_i (1 \leq i \leq n)$  のいずれかであるとし (本研究では  $n = 18$ )、発話文  $b$  は  $k$  個の素性  $f_j (1 \leq j \leq k)$  の集合として表す。素性として用いるのは、事例データ中に出現する全形態素である。さらに、発話文  $b$  において素性  $f_j$  が観測されかつその発話意図が  $a_i$  となるときに 1 を返すような素性関数を以下のように定義する。

$$g_{i,j}(a, b) = \begin{cases} 1 & (\text{if } \text{exist}(b, f_j) = 1 \ \& \ a = a_i) \\ 0 & (\text{Otherwise}) \end{cases}$$

ここで  $\text{exist}(b, f_j)$  は、発話文  $b$  において素性  $f_j$  が観測される場合に 1 を返し、そうでない場合には 0 を返す関数である。

ME 法によって得られる推定すべき確率分布は唯一存在し、次のような式で表される。

$$p(a|b) = \frac{\prod_{i,j} \alpha_{i,j}^{g_{i,j}(a,b)}}{Z_{\alpha}(b)} \quad (0 < \alpha^{g_{i,j}} \leq \infty) \quad (1)$$

$$Z_{\alpha}(b) = \sum_a \prod_{i,j} \alpha_{i,j}^{g_{i,j}(a,b)} \quad (2)$$

ただし、

$$\alpha_{i,j} = e^{\lambda_{i,j}} \quad (3)$$

であり、 $\lambda_{i,j}$  は、素性関数  $g_{i,j}(a, b)$  が発話文  $b$  の発話意図が  $a$  となることを推定するのに素性  $f_j$  がどれだ

け重要であるかを表す重みである。本研究では、ME法におけるパラメータ推定プログラム Amis<sup>2</sup>を用いて、GIS アルゴリズムによるパラメータ  $\lambda_{i,j}$  の推定を行う。

### 3.4 意図系列の利用

入江らの関連研究 [4] では、発話が入力された時点までの意図系列を考慮し、発話事例の絞り込みを行う手法を用いて発話意図推定を行っている。これは、入力発話と同じ意図系列で生じた発話事例は、入力発話と同じ意図を持つ可能性が高いと考えられるためである。

入江らは、レストラン検索をタスクとするユーザとオペレータの車内音声対話を用いて発話意図推定実験を行い、その結果、形態素情報のみを用いるよりも意図系列を併せて用いた手法の方が高い正解率を示すと述べている。ただし、入江らの設計した発話意図は 4 階層あり、上位の階層から下階層までを合わせて評価した場合には正解率の上昇が見られたが、最も上位の階層のみを評価した場合には形態素情報のみを用いた場合と意図系列を併せて用いた場合で差が見られなかった。

本研究で用いる発話意図は、入江らの発話意図定義における最上位の階層に属する発話意図であると考えられる。したがって、意図系列を用いた発話意図推定を行っても正解率の上昇は見込めない可能性があるが、これについては実験を行った後、改めて考察する。

## 4 発話意図推定実験

本節では、前節で述べた 2 つの手法を用いて行った発話意図推定実験の概要と結果、およびその考察について述べる。

### 4.1 実験概要

2 節で作成した意図タグ付きコーパス、3 節で述べた手法 1 および手法 2 を用いて発話意図推定実験を行った。

手法 1 では、3.2 節の実験結果において正解率の高かった 5 つの尺度のうち、まったく同じ正解率を示した 2 値の内積と余弦、Dice 係数と Jaccard 係数をそれぞれ同じ尺度グループと見なし、類似度の尺度として 2 値の内積、Dice 係数、形態素の一致率の 3 つを採用した。したがって、本実験で用いる発話意図推定手法は以下の 4 つである。

- 手法 1：発話文の類似度を用いる手法  
(2 値の内積, Dice 係数, 形態素の一致度)
- 手法 2：ME 法を用いる手法

実験に用いたデータは、意図タグ付きコーパス中の全 81 対話である。3.2 節で行った実験では事例データとして同じ算数の問題を扱う対話のみを用いたが(テストデータ 1 対話: 事例データ 26 対話)、今回の実験では算数の問題の種類に拘らず、コーパス中のテストデータ以外の全対話を事例データとして用いる。すなわち、1 対話をテストデータ、残り 80 対話を事例データとする。

以上のようにして、全 81 対話について交差検定を行い、ユーザ発話の発話意図推定結果の正解率の平均を求める。

### 4.2 実験結果

実験の結果を表 3 に示す。この実験では、手法 2 の ME 法を用いる手法が最も高い正解率を示した。また、手法 1 では、事例データを問題毎に分けて用いた場合(26 対話)に比べ、テストデータを除くコーパス中の全対話(80 対話)を事例データとして用いた場合の方が、全体として 4~5%程度良い結果となった。

表 3: 発話意図推定実験結果 (正解率: %)

手法	手法 1			手法 2
	2 値内積	Dice 係数	一致度	
問題 1	80.23	81.53	83.21	<b>84.98</b>
問題 2	90.43	88.72	89.11	<b>92.32</b>
問題 3	89.13	88.43	89.14	<b>92.42</b>
全体	86.60	86.22	87.16	<b>89.91</b>

## 5 考察

実験の結果、手法 2 が手法 1 に比べて高い正解率を示した。この結果から、単純に発話文間の類似度を求めるよりも、ME 法を用いて各発話意図における有効な素性を学習することで、より正確に、一意に発話意図を推定することが可能であると言える。また、手法 1 においては、事例データを増やしたことで正解率が全体として 5%程度向上した。事例データを増やしたことにより、各問題に共通する「数を数える」という行為を要求する発話や、問題の種類に関係なく同意表示:Correct > や < 不同意表示 > のような

<sup>2</sup>Amis - A maximum entropy estimator for feature forests, <http://www-tsujii.is.s.u-tokyo.ac.jp/amis/>

発話の正解が増えたため、全体として正解率が向上したと考えられる。

ただし、手法1には根本的な問題も存在する。正しい推定結果が候補として得られているにも関わらず、手順4で多数決を用いることによって不採用となってしまう可能性があるのである。例えば<同意表示: Response>は、多くが<同意表示: Correct>と同じような発話文に付与されており、コーパス中の総数も<同意表示: Correct>の10分の1程度しかないため、類似度最大となる推定候補の中に<同意表示: Response>よりも<同意表示: Correct>の方が多くなってしまふ。したがって、多数決をとると<同意表示: Correct>を出力してしまうため、候補の中に<同意表示: Response>が含まれているにも関わらず不正解となる。実際に、<同意表示: Response>が付与されたテストデータのほとんどが<同意表示: Correct>を出力して不正解となっている。

### 5.1 追加実験1

以上の結果を踏まえて、手法1の手順4で多数決をとらず、類似度最大の候補が複数ある場合(手法2では確率値が最大の候補が複数ある場合)には、その中に1つでも正解が含まれていれば正解とするという評価基準を用いて再度発話意図推定実験を行った。すなわち、どの手法が最も候補正解率が高いかということを確認する実験である。ここでの候補正解率とは、推定結果の候補の中に正解が1つ以上含まれているテストデータの数を、テストデータの総数で割ったものである。

2値の内積、Dice係数、形態素の一致度を手法1における類似度の尺度とし、手法1と手法2それぞれで事例データ数を80対話とした場合について発話意図推定を行い、各対話における候補正解率の平均を求めた。この再実験の結果を表4に示す。

表4: 追加実験1の結果(候補正解率: %)

手法	手法1			手法2
	2値内積	Dice係数	一致度	
問題1	88.74	84.38	86.19	85.21
問題2	96.01	92.59	93.00	92.32
問題3	93.96	90.11	90.82	92.42
全体	92.92	89.03	90.00	89.98

実験の結果、手法2では前回の実験の正解率と比

べ問題1グループのみわずかに正解率が上昇したが、問題2と問題3のグループではまったく同じ正解率であった。手法2の場合には各発話意図に与えられた確率値の大小で推定結果を決定しており、またその確率値が完全に等しくなることは稀であるため、ほぼ一意に結果が求められていると考えられる。

手法1においては、2値の内積を尺度として用いた場合、4つの手法の中で最も高くなるという結果となった。これはすなわち、2値の内積は他のどの手法よりも正しい推定結果を候補として得られているということであり、この推定結果候補の中から正しい解を選択することができれば、ME法を用いた場合よりも正解率が高くなるということである。

### 5.2 追加実験2

現在、複数候補からの選択の際には多数決を採用しているが、対話コーパス中に出現する各発話意図の総数には偏りがあるため、総数の大きい発話意図が推定結果として選択される確率が高い。そこで、多数決によって得られる値を出現頻度によって単純に正規化する次のような式を考えた。

$$v(a) = \frac{\text{類似度最大候補中の発話意図 } a \text{ の数}}{\text{対話事例中の発話意図 } a \text{ の出現頻度}} \quad (4)$$

この式を用いて再び発話意図推定実験を行った。手法1(2値内積)を用い、事例データ数を26対話とした場合と80対話とした場合について発話意図推定を行い、多数決を用いた場合の正解率と比較する。この再実験の結果を表5に示す。

表5: 追加実験2の結果(正解率: %)

手順4	多数決		式(4)を用いた多数決	
	26対話	80対話	26対話	80対話
問題1	76.47	80.23	43.97	45.30
問題2	85.10	90.43	60.72	62.36
問題3	84.67	89.13	50.87	55.61
全体	82.08	86.60	51.85	54.42

表5を見てわかる通り、式4を用いたことにより正解率が大幅に低下した。このような大幅な正解率低下の原因として、<同意表示:Correct>の存在が挙げられる。例えば「はい。」という発話に対して手法1(2値内積)を用いて発話意図推定を行うと、<同意表示:Correct>と<同意表示:Response>が類似度最大の候補として得られる。この2つの発話意図

候補に対して多数決をとると、先の考察で述べたように <同意表示:Correct> が推定結果として選択される。しかし、式 4 を用いて多数決をとると、使用している対話コーパスでは  $v$  (<同意表示:Correct>) よりも  $v$  (<同意表示:Response>) の方が値が大きくなるため、すべての「はい。」という発話に対して <同意表示:Response> が推定結果として出力されることになる。これにより、各対話に多く出現する <同意表示:Correct> が正しく推定されず、結果として正解率が大幅に減少してしまったと考えられる。

<同意表示:Correct> と <同意表示:Response> を正しく分類する方法について考察する。 <同意表示:Correct> と <同意表示:Response> の絶対的な違いは、「発話対をなす発話意図が異なる」という部分である。つまり、 <同意表示:Correct> は <確認要求:Correct> のみと対をなすのに対し、 <同意表示:Response> は <行為要求:Other> などと対をなすこともあれば、単なる相槌として単独で出現することもある。したがって、発話文の表層で見分けることができない以上、このような対話構造を用いて候補の絞り込みを行う必要があると考えられる。他の発話意図の場合は絞り込みを行うことでデータが少なくなり正解率が下がる可能性があるが、 <同意表示:Correct> のように多数出現するものに対しては事例の絞り込みは有効に働くと考えられる。次節では、この事例の絞り込みを用いて行った実験について報告する。

### 5.3 追加実験 3

3.4 節で述べたように、入江らは意図系列による事例の絞り込みを用いて発話意図推定を行っている。そこで、この意図系列による絞り込みを加えた手法 1 および手法 2 を用いて発話意図推定実験を行った。手法 1 では 2 値の内積と形態素の一致度の 2 つを類似度の尺度として用い、手法 1 と手法 2 それぞれにおいて絞り込みを行う場合と行わない場合の正解率を比較する。手法 1 と手法 2 における発話意図推定の手順を以下に示す。

#### 手法 1

1. 茶釜を用いて発話文の形態素解析を行う。
2. 形態素情報を用いて、発話文のベクトル空間を作成する。
3. 入力発話の直前の生徒（計算機）発話を用いて事例の絞り込みを行う。入力発話の前にユーザ

発話がある場合（1 ターンに複数の発話がある場合）には、それらも考慮する。ただし、絞り込みによって候補 0 となった場合には全発話事例を用いる。

4. 絞り込まれた発話事例と入力類似度を求める。
5. 類似度最大の事例に付与された発話意図タグの中で最も数が多いものを推定結果とする。

#### 手法 2

1. 茶釜を用いて発話文の形態素解析を行う。
2. 事例データから素性を抽出する。
3. 抽出した素性をもとに訓練データを作成する。このとき、入力発話の発話意図とその直前の生徒（計算機）発話の発話意図を組として表す。
4. ME 法を用いて学習し、確率分布を求める。
5. 入力発話を解析し、最大の確率値が得られた発話意図を推定結果とする。

手法 2 では事前に事例を絞り込むことが難しいため、入力発話と直前の生徒（計算機）発話の発話意図をペアとして表し、絞り込みの代替手法としている。

以上のような手法 1 および手法 2 を用いて、1 対話をテストデータ、残り 80 対話を事例データとして交差検定を行い、絞り込みを行う場合と行わない場合の発話意図推定結果の正解率を比較する。この実験を追加実験 3a とし、実験の結果を表 6 に示す。また、追加実験 3b として、手法 1 の手順 3 において多数決をとらない手法を用い、事例の絞り込みを行う場合と行わない場合の発話意図推定結果の候補正解率を比較する。1 対話をテストデータ、80 対話を事例データとして交差検定を行い、各問題毎の正解率の平均を求めた。この結果を表 7 に示す。

実験 3a の結果、手法 1 および手法 2 の両方において、全体の平均を見ると事例の絞り込みを行わない場合よりも絞り込みを行う場合の方が正解率が低くなった。手法 1 では、類似度の尺度に関わらず、絞り込み有りの手法の方が絞り込み無しよりも全体の正解率が低くなっている。これは、事例の絞り込みを行うことにより、推定されるべき発話意図が付与された事例が少なくなったことが原因と考えられる。しかし、前節で考察したように、 <同意表示:Correct> と <同意表示:Response> については、意図系列によって事例を分類したことにより正解率が上昇することが確かめられた。

また、実験 3b においても、絞り込みを行わない場合よりも絞り込みを行う場合の方が候補正解率が低

いという結果になった。これも実験 3a と同様に，事例の絞り込みによって事例数が減ってしまい，結果として類似度最大の事例の中に正解発話意図が含まれにくくなったことが原因と考えられる。

以上の結果から，事例の絞り込みを用いる場合には，<同意表示:> などのように絞り込みを行わなければ正しく意図を推定できない場合など，絞り込みが有効である場合を適切に判断して利用する必要があると言える。しかし，今回の実験で事例の絞り込みによって正解率が低下したのは，本研究で用いたタグ付きコーパスの規模が小さく，絞り込みを行うには事例数が十分でなかったことが主な原因である。したがって，大規模コーパスを構築し，大量の事例データを用いて発話意図推定を行う場合には，事例の絞り込みは有効に働くのではないかと考えられる。

表 6: 追加実験 3a の結果 (正解率: %)

手法	手法 1				手法 2	
	2 値内積		一致度		有	無
絞り込み	有	無	有	無		
問題 1	80.12	80.23	80.52	83.21	82.51	84.98
問題 2	88.16	90.43	87.82	89.11	90.33	92.32
問題 3	89.04	89.13	89.39	89.14	92.55	92.42
全体	85.77	86.60	85.91	87.16	88.47	89.91

表 7: 追加実験 3b の結果 (候補正解率: %)

手法	手法 1			
	2 値内積		一致度	
絞り込み	有	無	有	無
問題 1	84.32	88.74	82.03	86.19
問題 2	90.92	96.01	89.04	93.00
問題 3	90.78	93.96	90.09	90.82
全体	88.67	92.92	87.05	90.00

## 6 まとめ

本論文では，まず 2 節で発話意図推定に用いる意図タグ付き対話コーパスについて述べた。3 節および 4 節では，2 つの発話意図推定手法とそれらを用いた発話意図推定実験について述べた。手法 1 において類似度をはかる 9 つの尺度について発話意図推定の比較実験を行った結果，本研究で用いる対話コーパス

においては，2 値の内積・余弦，Dice 係数，Jaccard 係数，形態素の一致度が有効であることがわかった。手法 1 と手法 2 を用いて発話意図推定実験を行った結果，一意に推定結果を求める場合には手法 2 が最も高い正解率を示した。しかし手法 1 では多数決を用いることにより推定候補から正しい発話意図を選択することができず正解率を下げているという問題がある。これを解決するためには多数決に替わる候補の選択手法が必要となるが，多数決の値を事例中の出現頻度だけで簡単に正規化しても，期待されるような良い結果は得られなかった。また，手法 2 においても同様に，付与確率のランキングから上位数個を候補とし，その中から推定結果を選択するような手法を用いることで正解率が向上する可能性がある。

追加実験 3 においては，事例の絞り込みを用いて発話意図推定を行った結果，絞り込みを行う場合よりも絞り込みを行わずに形態素情報のみを用いた場合の方が正解率が高くなった。ただし，<同意表示:Correct> などのように事例の絞り込みを行うことで正解となりやすい発話意図もあるため，単純な絞り込みではなく，対話構造なども考慮した上で事例を絞り込むなど，適切な手法を用いることで正解率が向上すると思われる。

したがって，今後の課題としては，複数候補からの選択手法の考案・実装や，対話構造を利用した事例の絞り込みなどを行う手法の検討，また，複数の推定手法の組み合わせなどを実装し，その有効性を確認することなどが挙げられる。

## 参考文献

- [1] しょうがく 1 年 さんすう塾，教育書籍，1989。
- [2] 小 1 算数 5 分間トレーニング，数学研究社，1990。
- [3] 木村 宏，徳久 雅人，目良 和也，甲斐 郷子，岡田 直之，“対話における相手意図の理解と応答のためのプランニング”，TL98-15，pp.25-32，1988。
- [4] 入江 友紀，松原 茂樹，河口 信夫，山口 由紀子，稲垣 康善，“意図タグ付きコーパスを用いた発話意図推定手法”，人工知能学会研究会資料，SIG-SLUD-A301-03，pp.7-12，2003。
- [5] 乾 裕子，村田 真樹，内元 清貴，伊佐原 均，“表層表現に着目した自由回答アンケートの意図に基づく自動分類”，自然言語処理，Vol.10 No.2，2003。
- [6] 岩下 薫，嶋田 和孝，遠藤 勉，“対話履歴と事例に基づく発話意図の推定”，FIT2006。
- [7] 北 研二：確率的言語モデル，東京大学出版会，1999。