意図タグつきコーパスを用いた発話意図推定手法

Speech Intention Understanding based on Spoken Dialogue Corpus

入江 友紀[1] 松原 茂樹[2,4] 河口 信夫[2,4] 山口 由紀子[2] 稲垣 康善[3]
Yuki Irie Shigeki Matsubara Nobuo Kawaguchi Yukiko Yamaguchi Yasuyoshi Inagaki

1)名古屋大学大学院情報科学研究科
Graduate School of Information Science, Nagoya University
2)名古屋大学情報連携基盤センター
Information Technology Center, Nagoya University
3)愛知県立大学情報科学部
Faculty of Information Science and Technology, Aichi Prefectural University
4)名古屋大学統合音響情報研究拠点
Center for Integrated Acoustic Information Research, Nagoya University

Abstract: This paper proposes a method of speech intention understanding based on spoken dialog corpus. In the method, a spoken dialogue corpus with intention tags is used. The intention tag expresses the task-dependent intention of the speaker, and therefore, the proper understanding enables the system to take appropriate actions in the dialogue. We tagged about 35000 sentences in the CIAIR in-car speech database. In our method of speech intention understanding, the intention of each input utterance is regarded as that of utterance to which it is most similar in the corpus. The degree of similarity is calculated based on the degree of the correspondence in morphemes between utterances and the dialogue context information. An experiment using this corpus has shown 70.8% accuracy.

1．はじめに

近年，音声認識技術の進歩などを背景に，音声対話システムの研究が盛んに行われている．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，それに基づいて処理を行うことは，ユーザとの間で自然なインタラクションを遂行し，タスクの目標を解決するうえで必要不可欠である．発話意図を推定する方法として，ルールを用いる手法がこれまでに報告されている[1].

しかし, ユーザがある意図に基づいて発話するとき，その影響は，音韻，形態素，キーワード，文構造，文脈など発話に関連する事象に様々な形となって現れる．これらの要素を踏まえた推定ルールは複雑になるうえ，発話の多様性に対応できるようにするには，多くの推定ルールが必要であり，その作成は困難である．人間の自然な発話に対応するための方法として，事例を用いたアプローチは有効である[2][3].

事例を用いた意図推定を行うためには，対話データが必須である．しかも形態素や係り受けなどの統語レベルのタグだけではなく，意味レベルや談話レベルなど種々のレベルのタグが付与されている必要がある．このようなタグが付与されたデータは，意図推定だけでなく，対話の特徴分析やそのモデル化にも利用できる[4]．特に近年，コーパスに基づいて対話を統計的に処理することが試みられるようになり，大規模なタグつきデータの重要性はますます増大している．このような背景のもと，発語内行為レベルの情報を表すタグとして，発話単位タグ標準化案が提案されている[5]. しかしながら，音声対話システムの動作を具体的に決定するためには，「真偽情報要求」や「未知情報要求」などといった発語内行為レベルよりも詳細な意図を推

入江　友紀
名古屋大学大学院情報科学研究科情報システム学専攻
〒464－8603　名古屋市千種区不老町
Tel: 052-789-5145
E-mail: yuki-i@inagaki.nuie.nagoya-u.ac.jp

定する必要がある．

本論文では，対話事例に基づくユーザ発話の意図を推定する手法を提案する．また，対話事例として，発語内行為レベルよりも，さらに詳細な発話意図を表すタグ（以下，意図タグと呼ぶ）を設計し，コーパスを構築したので報告する．意図タグは，発話中の諸要素（文末，文体，キーワード）と意図の関連性を考慮し，意図の抽象度に応じて階層化した．タスクに依存したレベルまで意図を詳細化することにより，システムの動作に直結した意図記述が可能となる．名古屋大学 CIAIR 車内音声対話データベースに収録されている対話から，レストラン検索をタスクとする 3641 対話を抽出し，それに含まれる約 35000 文に意図タグを人手で付与した．意図推定では，入力発話とコーパス中の発話との類似度を計算する．その結果，類似度が最大となる発話に付与された意図を入力発話の意図とする．発話間の類似度は，各発話の形態素の一致度と入力発話に至るまでの意図系列に基づいて決定する．

本論文の構成は以下の通りである．２章では，事例に基づく発話意図推定手法の概要を述べ，３章で設計した意図タグと，意図タグを付与して構築した意図タグつき音声対話コーパスについて述べる．４章では意図推定手法を述べ，５章では提案した手法の評価とその考察を述べる．最後に６章で，まとめと今後の課題を述べる．

２．事例に基づく発話意図推定

ユーザがある意図に基づいて発話するとき，その影響は発話に関連する事象に様々な形となって現れる．人間の複雑かつ多様な発話に対応できるシステムを実現するために，事例を用いるアプローチは有効である．

本手法では，発話意図を表すタグが各発話に付与された意図タグつき音声対話コーパスを用いて意図を推定する．類似している発話は意図も類似する可能性が高いという仮定に基づき，コーパス中の各発話との類似度を計算する．意図推定の流れを図 2.1 に示す．意図推定は以下の手順で行われる．

1. 入力発話を形態素解析し，解析結果を得る
2. 発話が入力された時点までの意図系列を考慮し，事例の絞込みを行う
3. 得られた形態素の情報を用いて，絞り込んだ事例との類似度計算を行う
4. 類似度が最大である発話の意図を入力発話の意図と定める

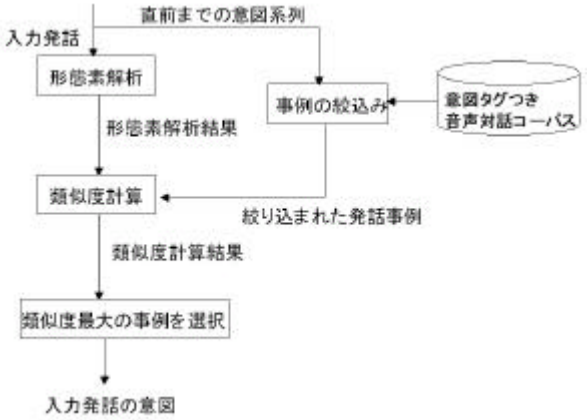

図 2.1　事例に基づく意図推定の流れ

３．　意図タグつき音声対話コーパス

事例を用いた意図推定を行うために，対話データが必須である．そこで発話意図を表すタグが各発話に付与された意図タグつき音声対話コーパスを構築した．

3.1.　意図タグの設計

意図タグつき音声対話コーパスを作成するために，名古屋大学 CIAIR 車内音声対話データベースの書き起こしコーパスを使用した[6]．すなわち，データベースに収録されている対話から，レストラン検索をタスクとする対話に意図タグを人手で付与した．作業にあたり，発話に付与すべき意図タグを設計する必要がある．

発語内行為レベルの情報を表すタグとして発話単位タグが提案されている[5]．音声対話システムの動作を具体的に決定するためには，「真偽情報要求」や「未知情報要求」などといった発語内行為のレベルよりも，さらに詳細な意図を推定する必要がある．そのため，これらのタグから意図を推定した後，どのようなシステム動作を行うかを決定する推論機構が必要となる．そこで，本手法では発語内行為のレベルよりも，さらに詳細な発話意図を表すタグを設計し，タグを推定することにより，システムの動作が決定できるようにした．意図タグは，発話の諸要素（文体，文末，キーワードなど）と意図との関連性を考慮し，意図の抽象度に応じて階層化した．それぞれの層におけるタグの例を表 3.1 に示す．

談話行為レイヤーは話者の発語内行為を表しており，動作レイヤーはドライバーまたはシステムの行為を表している．対象レイヤーは動作レイヤーで定まる動作の対象を表し，詳細レイヤーは対象に関する詳細情報を表す．

表 3.2　意図タグとその発話例

意図タグ				発話例
談話行為	動作	対象	詳細情報	
依頼	検索	店		この近くに中華の店ありますか
依頼	案内	店		マックに案内して
依頼	提示	店情報	メニュー	その店にラーメンはあるかな
依頼	提示	店情報	空席状況	今の時間は席あいてるのかな
陳述	提示	検索結果	店名	近くにガストがあります

表 3.1　階層化された意図タグの例（一部）

談話行為レイヤー	**動作レイヤー**	**対象レイヤー**	**詳細情報レイヤー**
依頼	確認	店情報	店名
提案	提示	駐車場情報	ジャンル
表明	検索	予約情報	値段
示唆	案内	検索結果	場所
陳述	選択	意思内容	空席状況

また，上の階層のタグに結び付き可能な下の階層のタグには図 3.1 に示すような制約がある．表 3.2 に意図タグとその発話例を示す．

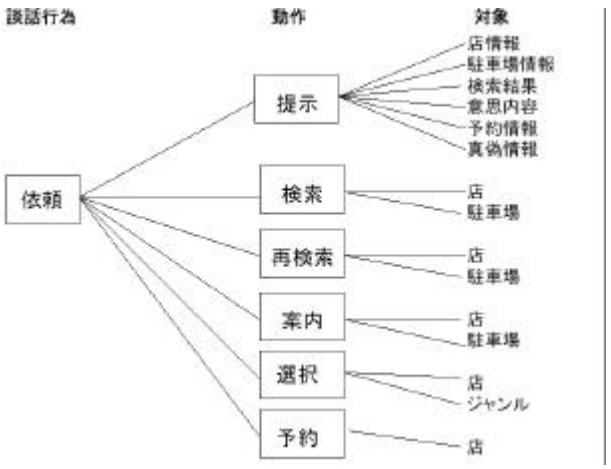

図 3.1　意図タグの制約の例

3.2. 意図タグつき音声対話コーパスの構築

名古屋大学 CIAIR 車内音声対話データベースに収録されている対話から，レストラン検索をタスクとする 3641 対話に含まれる 35421 発話に意図タグを人手で付与した．ここで付与した対話には，「オペレータ役（人間）との模擬対話」と「WOZ システムとの対話」の２種類がある．この２種類の対話に意図タグを付与することにより，対話相手の違いによる対話の遂行への影響の分析など，様々な観点からの分析が可能となる[7]．また，構築したコーパスには，話者情報や時間情報，言語現象に関するタグも含んでおり，定量的な談話分析が可能となっている．

意図タグ付与にあたって，意図タグ付与マニュアルを作成した．タグの揺れを減らすために，次のような判断基準を設けた．

- ドライバー発話の意図タグ
  ドライバー発話に対するシステム動作を一意に決定できるように，オペレータに対するドライバーの要求内容を，その発話を受けてオペレータがどのような返答をしたのかにより判断する

上記の判断基準に従うと，ドライバー発話「今あいているかな」に対するオペレータの返答が「営業時間は 9 時から 20 時までです」であれば，ドライバーは営業時間を尋ねているものとみなす．一方，オペレータの返答が「ただいま満席となっています」であれば，ドライバーは空席状況を尋ねているものとみなす．

書き起こしコーパスの作成では，200msec 以上のポーズで発話を分割している．そこで１発話単位に１つの意図タグを付与した．ただし，次の例外を認めた．

- 複数の発話単位にまたがる意図
  複数の発話単位を統合し，１つの発話意図を付与する．
- 複数の発話意図を持つ発話単位
  節を目安に発話を分割し，分割したそれぞれの発話に意図を付与する．

作成した意図タグつき音声対話コーパスの規模を表 3.3 に示す．表 3.3 において，HUM，WOZ はそれぞれ「オペレータ役（人間）との模擬対話」と「WOZ システムとの対話」であることを表す．

表 3.3　意図タグつきコーパスの規模

	HUM	WOZ
被験者数	664	592
対話数	1844	1797
ドライバー発話数	8751	7473
オペレータ発話数	9909	9278

表 3.4　出現頻度上位 5 個の意図タグ

HUM						WOZ					
ドライバー			オペレータ			ドライバー			オペレータ		
依頼+検索+店	1404	21.8%	陳述+提示+検索結果	1751	15.1%	依頼+検索+店	2537	33.9%	陳述+提示+検索結果	3250	35.0%
陳述+選択+店	1244	19.3%	表明+案内+店	1539	13.4%	陳述+提示+意思内容	1673	22.4%	表明+案内+店	1891	20.3%
陳述+提示+意思内容	977	15.0%	陳述+提示+意思内容	1314	11.4%	表明+案内+店	1651	22.1%	陳述+提示+意思内容	1859	20.0%
陳述+選択+ジャンル	693	10.8%	陳述+提示+店情報	882	7.7%	陳述+選択+ジャンル	806	10.8%	陳述+提示+店情報	1573	16.9%
依頼+提示+店情報	603	9.4%	依頼+選択+ジャンル	555	4.8%	依頼+提示+店情報	694	9.3%	依頼+選択+ジャンル	688	7.4%

構築したコーパスに現れた意図タグは，談話行為，動作，対象の 3 レイヤーの組み合わせで計 40 種類であった．表 3.4 に出現頻度上位５個の意図タグとその出現割合を示す．人との対話と WOZ システムとの対話において，頻繁に出現する意図の種類には同様の傾向が見られた．

3.3. 意図タグの評価

本論文において提案する意図タグの付与結果が，複数の作業者間で一致しないようであれば，そのタグが付与されたデータから導かれる結論は信頼できるものであるとは言えない．そこで，意図タグ体系の信頼性を検証するために評価実験を行った．作業結果がどの程度一致するかを定量的に評価するため，本研究では kappa 値を用いる[5]．Cohen の kappa 値は，観測された一致率を $P(O)$，期待される一致率を $P(E)$ とすると，

$$K = \frac{P(O) - P(E)}{1 - P(E)} \qquad (1)$$

と表される．K =1 は完全な一致，K =0 は偶然の一致程度の一致しかないことを示す．

レストラン検索をタスクとする対話を用いて意図タグ付与実験を行った．作業者数は 4 名（うち，3 名は音声・言語処理に関する知識をほとんど持っていない）で，51 対話，合計 528 発話に対して，各対話２名で意図タグを付与した．

一致率の評価の結果を表 3.5，表 3.6 に示す．表 3.5 の「意図タグ」はすべての階層のタグが一致したときにタグが一致したとみなしたときの値であり，表 3.6 の「談話行為」「動作」「対象」「詳細情報」はぞれぞれのレイヤーのタグのみを評価対象として一致率を計算したときの値である．

表 3.6 から特に対象レイヤーのタグの一致度が高いことが分かった．対象レイヤーのタグは，ある特定の単語（キーワード）に着目することで，どのタグかを判断できることが多いため，一致率が高くなったと考えられる．

表 3.5　意図タグの評価結果

	意図タグ
$P(O)$	0.705
$P(E)$	0.052
K	0.689

表 3.6　意図タグの評価結果（レイヤーごと）

	談話行為	動作	対象	詳細情報
$P(O)$	0．821	0．795	0．833	0．821
$P(E)$	0．356	0．230	0．168	0．302
K	0．722	0．733	0．799	0．744

今回の実験においては作業者のトレーニングを事前に行わなかったため，意図タグの判断がタグの名前に引きずられたことによる間違いが生じた．このような間違いは，マニュアルの整備やタグ付与前のトレーニングである程度排除できるものであると考えられる．

また，一般の作業者に対してトレーニングを事前に行わなくても，約 70%の一致率を得られたことから，今回提案した意図タグは，特別な知識を有してなくても付与することができ，ある程度の信頼性を有するデータが得られるといえる．

4．発話意図推定手法

4.1. 類似度計算

本手法では，入力発話と発話事例間の類似度から最尤の推定結果を求める．入力発話を S_i，事例発話を S_j，入力発話 S_i と発話事例 S_j 間の類似度を $SIM(S_i,S_j)$ とするとき，入力発話の意図 $\hat{I}(S_i)$ を式（2）によって求める．

$$\hat{I}(S_i) = \arg\max_j SIM(S_i, S_j) \qquad (2)$$

入力発話と発話事例間の類似度 SIM は，形態素に関する情報と意図の履歴から計算する．

4.1.1. 形態素に基づく類似度

2つの発話 S_i と S_j の間の形態素を用いた類似度 R_{ij} を式(3)で定義する．

$$R_{ij} = \frac{2M_{ij}}{M_i + M_j} \qquad (3)$$

式（3）において，M_i, M_j はそれぞれ S_i, S_j の形態素数，M_{ij} は一致する形態素数を表す．

本研究では，対話タスクに特徴的な名詞や固有名詞には単語クラスを付与し，形態素が同一クラスに属していればそれらの形態素も一致しているとみなす．また，文全体としてより類似したコーパス中の発話意図を抽出するために，自立語やキーワードだけではなく，名詞や助詞など，すべての形態素を一致の対象とした．

4.1.2. 意図系列

ユーザとオペレータの対話において，現在のユーザ発話の意図は多くの場合，それまでの発話意図に依存して生起している．そこで，本手法では，ユーザの入力発話に至るまでの意図系列を用いる．ある時点 n での発話意図 I_n の生起は直前の N-1 発話の発話意図に依存すると考える．ここで，発話意図の系列 I_{n-N+1} ・・・ I_n を I^n_{n-N+1} と書く．

4.2. 類似度

本手法では，入力発話の意図系列と同じ意図系列で生起した発話事例との間で入力発話との類似度を計算する．なぜなら，入力発話と同じ意図系列で生起した発話事例では，入力発話と同一の意図を持つ発話が生起する可能性が高いと考えられるためである．すなわち，形態素に関する情報と意図の履歴を用いた類似度 SIM を式（4）で定義する．

$$SIM(S_i, S_j) = \begin{cases} R_{ij} & (I(S_i)^{n-1}_{n-N+1} = I(S_j)^{n-1}_{n-N+1}) \\ 0 & (I(S_i)^{n-1}_{n-N+1} \neq I(S_j)^{n-1}_{n-N+1}) \end{cases} \qquad (4)$$

ただし，S_i, S_j はそれぞれ入力発話，発話事例を，$I(S_i)^{n-1}_{n-N+1}$，$I(S_j)^{n-1}_{n-N+1}$ はそれぞれ S_i, S_j の意図系列を表す．

5．発話意図推定実験

5.1. 実験に使用したデータ

3章で述べた意図タグつき音声対話コーパスのうち，1999 年度収録分 72 人 425 対話と 2000 年度収録分「人と人との対話」セッション 297 人 793 対話のドライバー発話 6137 分を発話意図推定実験に使用した．一方，一致度を計算するうえで使用する単語クラスとして，本コーパスをもとにして作成した単語クラスデータベース[8]を使用した．また，形態素解析には日本語形態素解析システム「茶筅」[9]を使用した．ただし，話し言葉に特有なフィラー，対話タスクに特徴的な名詞や固有名詞は茶筅辞書に追加登録してある．

5.2. 発話意図推定実験

5.2.1. 実験方法

1218 対話のドライバー発話に対して，1211 対話のドライバー発話 5366 発話を事例データ，残りの 7 対話のドライバー発話 171 発話を評価データとした．また，本実験では意図系列として直前の発話の意図を用いた．

本実験では，正解の意図タグをあらかじめ人手で付与した意図タグと定め，正解率（発話総数に対する正解の意図タグを出力できた発話総数の割合）を求めた．

意図系列を考慮した類似度計算の有効性を調べるため，次の2つの類似度計算方法に基づき意図推定実験を行い，その結果を比較した．

1. 形態素に基づく類似度計算
   （$\hat{I}(S_i) = \arg\max_j R_{ij}$）
2. 意図系列を考慮した類似度計算
   （$\hat{I}(S_i) = \arg\max_j SIM(S_i, S_j)$）

5.2.2. 実験結果

意図推定実験結果を表 5.1 に示す．実験結果から，形態素の情報のみを用いるよりも，形態素の情報に加え意図系列を考慮した手法の方が，高い正解率を示すことが分かった．これは，意図系列を用いて入力文と同じ意図を持つと思われる事例を絞り込んだ効果が現れていると考えられる．

また，対象レイヤーのタグは，ある特定の単語（キーワード）に着目することにより，どのタグであるかを判断することができるため，表層情報からでもある程度推定できる．そのため，対象レイヤーは意図系列

を考慮しても正解率に大きな変化は見られなかった．表 3.6 においても対象レイヤーの一致度は高く，このレイヤーのタグは推定しやすいと考えられる．一方，動作レイヤーや詳細情報レイヤーでは，表層情報のみではタグを決めることが難しい．入力文の意図系列から事例を絞り込むことにより，意図系列としては不自然となる意図が候補から除かれ，正解率が向上したものと考えられる．

表 5.1 の結果を表 3.5，表 3.6 と比較すると，人と本手法で出力する意図タグが一致する割合は，2 者間の意図タグが一致する割合と同程度であることが分かった．このことから，本手法は人間と同精度の性能を有していると言える．

表 5.1　実験結果

	正解数（正解率：%）	
意図タグのレイヤー	形態素	形態素+意図系列
談話行為	152（88.9）	151（88.3）
動作	133（77.8）	139（81.3）
対象	145（84.8）	147（86.0）
詳細情報	136（79.5）	141（82.5）
談話行為+動作	131（76.6）	137（81.1）
談話行為+動作+対象	125（73.1）	133（77.8）
談話行為+動作+対象+詳細情報	104（60.8）	121（70.8）

6．おわりに

本論文では，設計した発話意図を表すタグ（意図タグ）と，構築した意図タグつき音声対話コーパスについて述べた．意図タグは，発話中の諸要素（文末，文体，キーワード）と意図との関連性を考慮し，階層的に定義した．名古屋大学 CIAIR 車内音声対話データベースに収録されている対話から，レストラン検索をタスクとする 3641 対話に含まれる約 35000 文に意図タグを人手で付与することによって，意図タグつき音声対話コーパスを構築した．意図タグ付与実験を行った結果，特別な知識を有していなくても，ある程度の信頼性の得られるデータを構築できることを確認した．

また，構築した意図タグつき音声対話コーパスをもとに，発話意図推定実験を行った．実験の結果，70.8%の正解率が得られ，形態素，意図系列を用いた発話間類似度計算を用いた事例に基づく意図推定手法の有効性を確認した．

今後の課題としては，意図タグ付与マニュアルの記述法の工夫，タグ付与候補を表示するようなタグ付与支援ツールの構築などによる意図タグつきコーパスの信頼性の向上，形態素レベルの表層情報に加え，係り受け関係など文の構造を捉えた類似度計算の実現，事例ベース，ルールベース，統計ベースなどの発話意図推定を融合したハイブリットな意図推定手法の考案などが考えられる．

参考文献

[1] 木村，徳久，目良，甲斐，岡田：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，信学技法，TL98-15, pp.25-32 (1988)

[2] S. Matsubara, S. Kimura, N Kawaguchi, Y. Yamaguchi and Y. Inagaki：Example-based Speech Intention Understanding and Its Application to In-Car Spoken Dialogue System, Proc. of the 17th International Conference on Computational Linguistics (COLING-2002), Vol.1, pp. 633-639 (2002)

[3] 松原，河口，外山，武田：音声対話コーパスの収集と利用－より豊かな車内音声対話システムを目指して－，人工知能学会誌，Vol.17, No.3, pp.279-284 (2002)

[4] 森本，高梨：対話における行為遂行/情報授受層の区別と対応関係の分析，人工知能学会研究会資料，SIG-SLUD-A103-12, pp.69-75 (2002)

[5] 荒木，伊藤，熊谷，石崎：発話単位タグ標準化案の作成，人工知能学会，Vol.14, No.2, pp.251-260 (1999)

[6] N. Kawaguchi, S. Matsubara, K. Takeda and F. Itakura：Multimedia Data Collection of In-Car Speech Communication, Proc. of the 7th European Conference on Speech Communication and Technology (EUROSPEECH2001), pp. 2027—2030 (2001)

[7] 岸田，入江，山口，松原，河口，稲垣：大規模音声言語コーパスの高度化と走行車内音声対話の特徴分析，言語処理学会 第 9 回年次大会 発表論文集，pp.133-136 (2002)

[8] H.. Murao, N. Kawaguchi, S. Matsubara and Y. Inagaki: Example-Based Query Generation for Spontaneous Speech, Proc. of the 7th IEEE Workshop on Automatic Speech Recognition and Understanding(ASRU01) (2001)

[9] 松本，北内，山下，平野，松田，高岡，浅原：日本語形態素解析システム「茶筌」version2.0 使用説明書 第 2 版,Information Science Technical Report, NAISTIS-TR9908, 奈良先端技術大学大学院 (1999)