

対話コーパスに基づく発話意図推定

入江 友紀^{†1}

松原 茂樹^{†2†3}

河口 信夫^{†2†3}

山口 由紀子^{†2}

稲垣 康善^{†4}

^{†1}名古屋大学大学院情報科学研究科 ^{†2}名古屋大学情報連携基盤センター

^{†3}名古屋大学統合音響情報研究拠点 (CIAIR) ^{†4}愛知県立大学情報科学部

yuki-i@inagaki.nuie.nagoya-u.ac.jp

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上で必要不可欠である。

発話意図を推定する方法として、ルールを用いる方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。

一方、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が考えられる [2, 3]。著者らはこれまで事例に基づく発話意図推定手法を提案してきた [5]。この手法では、入力発話とコーパス中の発話との類似度を計算し、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。しかし、そもそも類似した事例がコーパス中に存在しない発話に対し、意図を推定することは難しいという問題がある。

そこで本論文では、対話コーパスに基づく発話意図推定手法を提案する。本手法では対話コーパスを用いて、意図推定のための決定木を作成する。決定木学習により、意図と関連性の強い特徴量を取り出すことができ、新たな事例に対してもロバストに対処できる。

著者らはこれまで、発話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、コーパスを構築してきた [4, 5]。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ている。本手法ではこのコーパスを用いて、複数の決定木を作成し、それらを組み合わせることにより発話意図を推定する。意図タグの各階層ごとに決定木を作成するため、その階層の意図と関連性の強い発話中の要素に着目した意図推定を行うことができる。

本手法の有効性を評価するために、CIAIR 車内音声

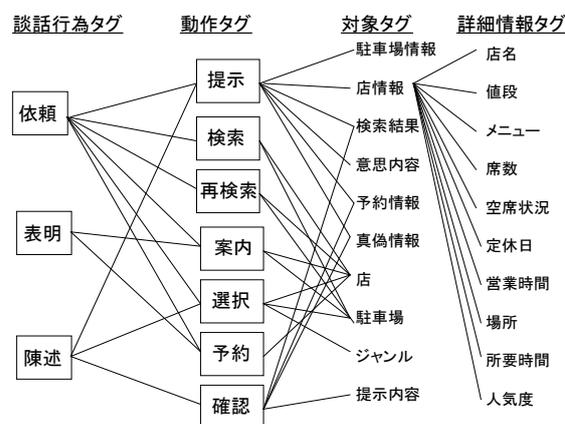


図 1: 意図タグの例

対話コーパスに収録されたレストラン検索をタスクとするドライバー発話に対して、発話意図推定実験を行った。実験では、意図タグつきコーパスの 2972 発話から得られた決定木を用いて、意図推定を行った。評価実験の結果、本手法の有効性を確認した。

2 意図タグつき音声対話コーパス

2.1 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、CIAIR 車内音声対話データベースの書き起こしコーパス [6] を利用した。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手で付与した。

付与するにあたり、発話の諸要素と意図との関連性を考慮し、意図の抽象度に応じて 4 つの階層を有する意図タグの体系を設計した。意図タグは、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って、システムの動作に直結した意図記述を行っている。そのため、談話行為、モダリティに関する情報から「店名」「値段」といった詳細な情報まで、推定する意図のレベルは多岐に渡る。したがって、意図と強く関連する発話の諸要素に着目して、タグを階層化した。談話行為レイヤは話者の発話内行為を表しており、動作レイヤはドライバまたはシステムの行為を表している。対象レイヤは動作レイヤで定まる動作の対象を表し、詳細レイヤは対象に関する詳細情報を表す。また、上の階層のタグに結びつき可能な下の階層のタグには図 1 に示すような制約がある。

表 1: 意図タグつきコーパスの例

発話例	話者	談話行為	動作	対象	詳細情報
この近くに中華の店ありますか	ドライバー	依頼	検索	店	
近くに大黒天があります	オペレータ	陳述	提示	検索結果	店名
その店にラーメンはあるのかな	ドライバー	依頼	提示	店情報	メニュー
はいございます	オペレータ	陳述	提示	店情報	メニュー
そこに案内して	ドライバー	依頼	案内	店	
大黒天までご案内します	オペレータ	表明	案内	店	

2.2 意図タグつき音声対話コーパスの構築

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 発話に意図タグを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 [4, 5].

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おり、原則として 1 発話単位に 1 つの意図タグを付与している。表 1 に意図タグつきコーパスの例を示す。

3 発話意図推定手法

3.1 発話意図推定手法の概要

本研究において推定対象となる意図タグは、意図と強く関連する発話の諸要素に着目して階層化している。また、発話によって意図と強く関連する要素も異なる。例えば、文末表現からモダリティを推測できる発話がある一方、文末が省略されており、他の要素が意図に強く関連する発話もある。また、意図と強く関連する述語が現れる発話は、述語が意図推定の重要な手がかりとなる。

このように、意図の抽象度や発話によって、意図と関連性の強い発話中の要素は異なる。本研究では、これらに柔軟に対応した意図推定を行うために、タグの階層ごとに決定木を作成する。ある階層の意図と関連性の強い発話中の要素に着目し、発話に応じ決定木間に順序付けを行うことにより、発話に強く関連する階層の意図から推定することができる。

3.2 決定木の作成

各階層に対し、その階層のタグを推定する決定木を作成する。本研究では、決定木の作成に決定木学習ソフトウェア See5¹を使用した。これは、データ(属性・属性値の集合とクラス)の集合を入力とし、決定木を出力するソフトウェアである。データの属性としては、著者がこれまでに行った意図推定実験において有効であった、現在の話者、直前の意図タグ(談話行為、動作、対象のタグ)、直前の話者、コーパスに現れた形態素を用いる。さらに、すでに推定によって得られた他階層の意図タグも属性として利用する。これは、2.1 節で述べたように、意図タグの階層間に制約が存在するため、ある階層のタグは、他の階層のタグに依存していることがあるためである。データのクラスとして、各階層のタグを与えた。表 2 に属性と属性値を、表 3 にクラスとその種類数を示

表 2: 属性と属性値

属性	属性値	
話者	ドライバー (D), オペレータ (O)	
直前の意図タグ	「陳述 + 提示 + 店情報」など 48 種類	
直前の話者	ドライバー (D), オペレータ (O)	
形態素	発話中に出現する, 出現しない	
判明したタグ	談話行為	5 種類
	動作	7 種類
	対象	12 種類
	詳細情報	36 種類

表 3: クラスとその数

	クラス	クラス数
談話行為レイヤ推定時	談話行為タグ	5
動作レイヤ推定時	動作タグ	7
対象レイヤ推定時	対象タグ	12
詳細情報レイヤ推定時	詳細情報タグ	36

す。また、図 2 に、発話「その店にラーメンはあるのかな」に対応するデータの例を示す。

対話コーパスから得られたデータ集合に対して、以下の 32 種類の決定木を作成する。

- どの階層のタグも推定されていないとき (第 1 群) 属性は話者, 直前の意図タグ, 直前の話者, 形態素を用いる。談話行為レイヤ推定用, 動作レイヤ推定用, 対象レイヤ推定用, 詳細情報レイヤ推定用の計 4 種類の決定木を作成する。
- 1 つの階層のタグが判明したとき (第 2 群) 属性は話者, 直前の意図タグ, 直前の話者, 形態素, 判明したタグを用いる。各階層に対して, 判明したタグの階層のパターンが 3 種類存在するので計 12 種類の決定木を作成する。
- 2 つの階層のタグが判明したとき (第 3 群) 属性は話者, 直前の意図タグ, 直前の話者, 形態素, 判明したタグ (2 種類) を用いる。各階層に対して, 判明したタグの階層のパターンが 3 種類存在するので計 12 種類の決定木を作成する。
- 3 つの階層のタグが判明したとき (第 4 群) 属性は話者, 直前の意図タグ, 直前の話者, 形態素, 判明したタグ (3 種類) を用いる。各階層に対して, 判明したタグの階層のパターンが 1 種類存在するので計 4 種類の決定木を作成する。

¹See5, <http://www.rulequest.com/see5-info.html>.

意図タグつきコーパス

発話	意図タグ
D: その店にラーメンはあるのかな	依頼 + 提示 + 店情報 + メニュー

データファイル

話者	意図前のタグ	直前の話者	直前のタグ	形態素	タグ	判明した	クラス
D	陳述 + 提示 + 検索結果	○	無	有 有 …	1-依頼	○	2-提示
D	陳述 + 提示 + 検索結果	○	無	有 有 …	1-依頼	○	3-店情報
D	陳述 + 提示 + 検索結果	○	無	有 有 …	1-依頼	○	4-メニュー

図 2: 発話「その店にラーメンはあるのかな」に対応するデータの例 (第 2 群の場合)

表 4: 推定結果 (訓練データ: 2972 発話, closed test)

タグ	談話行為	動作	対象	詳細情報
正解率 (%)	92.2	76.8	80.6	85.8

3.3 推定アルゴリズム

3.2 節で得られた複数の決定木を用いて、各階層のタグを推定する。どの階層から先に推定するのは、発話に応じた決定木間の順序付けにより決定する。この順序付けは次のように定めた。

[決定木順序付け規則]

1. 推定において到達した葉の再分類誤り率 (訓練データにおける誤り率) が一番低い決定木を選ぶ。
2. 誤り率が同じ場合は、その葉に到達する訓練データが多かった決定木を選ぶ。
3. 上記で決定できない場合は、「談話行為タグ」、「詳細情報タグ」、「対象タグ」、「動作タグ」の順番で決定木を選ぶ。これは、あらかじめ See5 を用いてタグ推定 (closed test) を行ったときの推定結果 (表 4) の正解率に基づいている。

この順序付けを用いて、以下のアルゴリズムに従って推定を行う。

[推定アルゴリズム]

- (1) 第 1 群の 4 種類の決定木を用いて、各階層のタグの候補を出力する。
- (2) 上記の順序付けに従って決定木を 1 つ選び、その決定木が出力したタグを推定結果とする。
- (3) (2) で得られたタグをさらに属性値として追加し、第 2 群の決定木を用いて、未決定の階層のタグを推定する。
- (4) (3) で得られたタグをさらに属性値として追加し、第 3 群の決定木を用いて、未決定の階層のタグを推定する。
- (5) (4) で得られたタグをさらに属性値として追加し、第 4 群の決定木を用いて、未決定の階層のタグを推定する。

3.4 意図推定の例

発話「いくらぐらい」の意図推定の例を図 3 に示す。まず、第 1 群の 4 種類の決定木を用いて、各階層のタグ

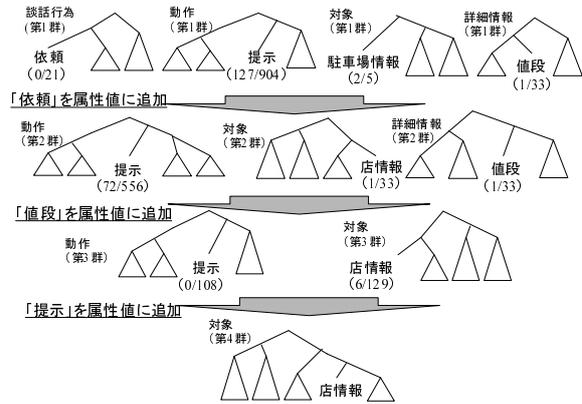


図 3: 発話「いくらぐらい」に対する意図推定の流れ

の候補を出力する。図 3 において、談話行為タグは「依頼」、動作タグは「提示」、対象タグは「店情報」、詳細情報タグは「値段」を候補として出力する。決定木順序付け規則を適用すると、規則 1 より再分類誤り率が一番低い談話行為タグの「依頼」を推定結果とする。

次に、談話行為タグを属性に追加し、他の階層のタグを推定する。第 2 群の決定木を用いて、残りの階層のタグ候補を出力すると、動作タグは「提示」、対象タグは「店情報」、詳細情報タグは「値段」を候補として出力する。規則 3 より、詳細情報タグ「値段」を推定結果とする。

同様に、談話行為タグ、詳細情報タグを属性に追加し、第 3 群の決定木から残りの階層のタグを推定する。その結果「提示」を得る。最後に、動作タグを属性に追加し、第 4 群の決定木を用いて、「店情報」を推定する。以上より、意図推定の結果は「依頼 + 提示 + 店情報 + 値段」となる。

4 発話意図推定実験

4.1 実験の概要

4.1.1 実験に使用したデータ

2 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を発話意図推定実験に使用した。形態素解析には日本語形態素解析システム「茶筌」²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筌辞書に追加登録してある。

4.1.2 実験方法

392 人分のドライバー発話から、362 人分のドライバー発話 2972 発話を訓練データ、残りの 7 人分のドライバー発話 171 発話を評価データとした。決定木の作成には、決定木学習ソフトウェア See5 を使用した (枝刈レベル: 25%, ブースティング: なし)。訓練データ中に現れた形態素の種類数は 1371 個であった。

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。本手法による決定木の組み合わせ推定の効果を確認するた

²Chasen, <http://chasen.aist-nara.ac.jp/>.

め、次の2つの決定木を作成して、意図推定実験を行い、その結果を比較した。

- 1つの決定木による一括推定
意図タグを「陳述+提示+予約情報+人数」のように1つのタグとみなす。
- 複数の決定木による組み合わせ推定（提案手法）

4.2 実験結果

推定実験の結果を表5に示す。1つの決定木による一括推定は正解率67.8%、複数の決定木による組み合わせ推定は正解率73.1%であり、提案手法の方が高い正解率を示した。

階層ごとに作成した決定木を見ると、談話行為レイヤに対する決定木では、「お願い」「って」「ほしい」「ない」「か」「の」といった文末表現が現れており、動作レイヤに対する決定木では、「案内」や「ナビゲーション」、「予約」といった動作を表すような形態素、直前の意図タグ（直前にオペレータがどのような動作を行ったか）がノードに現れている。一方、詳細情報レイヤに対する決定木では、「ありがとう」「何分」「何時」「いくら」「台」など時間や値段に関する形態素など、詳細情報タグと強い関連性をもつ形態素が現れている。このように、各階層と関連性の強い発話中の諸要素が異なっており、これらを優先的に利用することにより正解率が向上したと考えられる。また、すでに推定によって得られた他階層の意図タグも上位ノードに現れており、意図タグの階層間の制約も意図推定に活用されていると考えられる。

なお、誤った発話のうち、タグの制約違反を起こした発話が14発話、訓練データには出てこないタグのため決定木の葉に存在せず、推定ができなかったものが2発話存在しており、これらに対処することによって、正解率の向上が期待できる。

4.3 考察

4.3.1 事例に基づく手法との比較

著者らはこれまでに、意図タグつきコーパスを対話事例として利用した発話意図推定手法を提案してきた[5]。この手法では、入力発話とコーパス中の発話との類似度を計算し、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。本実験と同一のデータを用いて行った意図推定実験では、正解率は70.8%であった。この手法は、事例中の類似した発話の意図を入力発話の意図とするため、入力発話に似た事例が存在するときは、正しく意図を推定できる可能性が高いものも、発話の多様性に比べ事例の多様性が小さいときは、誤る可能性が高くなる。一方、提案手法は決定木を事例で学習してモデルを作成しているため、本実験のデータ量においては良好な結果が得られた。

4.3.2 複数作業者の一致率との比較

一方で、著者らは複数の作業者による意図タグの一致率を調べている[5]。作業者として、次の2組の実験を行っている。

表 5: 実験結果

	一括推定	組み合わせ推定
正解率 (%)	67.8 (116/171)	73.1 (125/171)

1. 一般の作業者4名が、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する。
2. 意図タグの設計者2名が28対話、合計296発話に対して意図タグを付与する。

実験の結果、一般作業者間の一致率は70.5%、設計者間の一致率は85.3%であった。表5と比較すると、本手法の推定精度は、設計者間の一致率には及ばないものの、一般の作業者間の一致率より良好な結果となっている。

5 おわりに

本論文では、対話コーパスに基づく発話意図推定手法について述べた。本手法では、複数の決定木を作成しそれらを組み合わせることにより、意図と関連性の強い発話の要素を優先的に利用して発話意図を推定する。実験の結果、73.1%の正解率が得られ、複数の決定木を組み合わせた対話コーパスに基づく発話意図推定手法の有効性を確認した。

今後の課題として、事例数を変化したときの事例ベース、ルールベース、統計ベースなどの発話意図推定の比較や、それらを融合した意図推定手法の考案、対話において重要な役割を果たす発話に対する意図推定精度の向上などが考えられる。

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, 信学技報, 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 17th Int. Conf. on Computational Linguistics*, 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] Y.Irie, N.Kawaguchi, S.Matsubara, I.Kishida, Y.Yamaguchi, K.Takeda, F.Itakura and Y.Inagaki: An Advanced Japanese Speech Corpus for In-car Spoken Dialogue Research, *Proc. of Oriental COCOSDA-2003*, pp.209-216(2003).
- [5] 入江, 松原, 河口, 山口, 稲垣: 意図タグつ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).
- [6] 河口, 松原, 山口, 武田, 板倉: CIAIR 実走行車内音声データベース, 情処研報, SIG-SLP-49, pp.139-144 (2003).