

実走行車内における運転者との自然な音声対話方式の開発 Spoken dialogue processing for in-car speech conversational system

研究代表者 松原 茂樹 名古屋大学情報連携基盤センター・助教授
Shigeki Matsubara, Associate Professor of Information Technology
Center, Nagoya University

アブストラクト

走行車内では、運転者は手足や視覚が制限された状況下におかれるため、対話カーナビゲーションなどの、優れた音声インタフェースを備えた情報システムの搭載が期待される。今後、対話処理技術が進み、車内音声対話システムとして実用化されれば、動く情報空間として自動車が新たな価値を提供することが予想される。本研究では、ユーザフレンドリな車内音声インタフェースの実現を目指し、人間の対話事例を活用する新たな対話処理方式の開発を目的とする。実際の対話事例を使用することにより、(1) 類似した事例の活用によるロバストな対話理解、(2) 人間の発声に近い自然な言い回しでの音声出力、(3) 事例の追加による対話システムの拡張、が可能であるという利点がある。本方式は、外国語初学者の学習方法を模倣している。不慣れな外国語で会話するとき、ある程度の量の文を蓄積し、それを応用して話す方が現実的である。

Abstract:

In a moving car, in order that a driver may set under the situation that hand/foot and vision are restricted, loading of the information system equipped with the outstanding speech interfaces, such as conversational car navigation, is expected. If spoken dialogue processing technology progresses and it will be put in practical use as an in-car speech dialogue system, that a car offers new value as moving information space will be expected. This research aims at development of the new dialogue processing system which utilizes human's dialogue examples. By using actual dialogue examples, there are the following advantages: (1) robust dialogue understanding by use of similar examples. (2) The output in the natural expression near human's speech. (3) Extension of the dialogue system by the addition of examples. This framework has copied a foreign language beginner's study method. It is more realistic to accumulate a certain amount of quantity of sentences, and to apply and speak about it, when talking in an unfamiliar foreign language.

1 研究の目的

近年、音声認識技術の進歩などを背景に車内音声対話システムの研究が盛んに行われている。車内音声対話システムにおいてユーザの発話を正しく理解し、それに基づいて処理を行うことは、運転者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不可欠である。運転者が発話するとき、その意図は、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。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。

事例を用いた発話理解を行うためには、大量の対話データが必須である。しかも構文レベルのタグだけではなく、意味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、発話理解だけでなく、対話の特徴分析やモデル化にも利用できる。

そこで本研究では、対話事例に基づいて運転者の発話を理解する手法の開発を目的とする。また、そのような理解を実現するために、発話意図を表すタグを付与した大規模コーパスを構築する。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、車内音声対話システムの動作に直結した意図記述が可能となる。

2 研究経過

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。

本研究では、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手法を開発した。本手法では、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、事例の絞り込みを行う
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

2.1 発話意図タグコーパスの構築

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。そのために、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。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。

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」などといった発話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、本手法では発話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、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。談話行為レイヤーは話者の発話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。

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 発話に意図タグを手手で付与した。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。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。

本研究において提案する意図タグの付与結果が、複数の作業員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。

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。作業員数は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。

今回の実験においては作業員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、一般の作業員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。

2.2 コーパスに基づく発話理解手法

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

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。

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。

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。

2.2 コーパスに基づく発話理解の評価

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「茶筌」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。

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
2. 意図系列を考慮した類似度計算

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。

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。

3 研究成果

本研究により、発話意図を表すタグ（意図タグ）を設計し、意図タグつき音声対話コーパスを作成した。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。

4 今後の課題と発展

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。

5 発表論文リスト

(1) Itsuki Kishida, Yuki Irie, Yukiko Yamaguchi, Shigeki Matsubara, Nobuo Kawaguchi, and Yasuyoshi Inagaki: An Advanced Japanese

- Speech Corpus for In-car Spoken Dialogue Research, Proceedings of 8th European Conference on Speech Communication and Technology (2003).
- (2) Hiroya Murao, Nobuo Kawaguchi, Shigeki Matsubara, Yukiko Yamaguchi and Yasuyoshi Inagaki: Example-based Spoken Dialogue System using WOZ System Log, Proceedings of 4th ACL SIGDIAL Workshop on Discourse and Dialogue, pp. 140-148 (2003).
- (3) Nobuo Kawaguchi, Shigeki Matsubara, Itsuki Kishida, Yuki Irie, Yukiko Yamaguchi, Kazuya Takeda and Fumitada Itakura, Construction and Analysis of the Multi-Layered In-car Spoken Dialogue Corpus, Proceedings of Workshop on DSP in Mobile and Vehicular Systems, Nagoya, Japan (2003).
- (4) Y. Irie, Nobuo Kawaguchi, Shigeki Matsubara, Itsuki Kishida, Yukiko Yamaguchi, Kazuya Takeda, Fumitada Itakura and Yasuyoshi Inagaki: An Advanced Japanese Speech Corpus for In-car Spoken Dialogue Research, Proceedings of Oriental COCOSDA-2003, pp. 209-216 (2003).
- (5) Tomohiro Ohno, Shigeki Matsubara, Nobuo Kawaguchi, and Yasuyoshi Inagaki: Spiral Construction of Syntactically Annotated Spoken Language Corpus Proceedings of IEEE International Conference on Natural Language Processing and Knowledge Engineering, pp. 477-483 (2003)
- (6) 入江友紀, 松原茂樹, 河口信夫, 山口由紀子, 稲垣康善: 対話コーパスに基づく発話意図推定, 言語処理学会第 10 回年次大会, pp.83-86 (2003)
- (7) 林啓太, 入江友紀, 山口由紀子, 松原茂樹, 河口信夫: コーパスに基づく発話理解・対話制御・発話生成, 情報処理学会全国大会 (2004).
- (8) 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), pp.7-12 (2003).
- (9) 河口信夫, 松原茂樹, 山口由紀子, 武田一哉, 板倉文忠: CIAIR 実走行車内音声データベース, 情報処理学会研究報告(SIG-SLP-49), pp.139-144 (2003).
- (10) 村尾浩也, 河口信夫, 松原茂樹, 山口由紀子, 武田一哉, 稲垣康善: WOZ によるオンライン修正が可能な事例ベース音声対話システム, 情報処理学会研究報告(SIG-SLP-49), pp.269-274 (2003).