

タスク指向対話とチャットを統合した

対話ロボットインターフェース

An Integrated Conversation Robot Interface

for Task-Oriented and Non-Task-Oriented Dialogues

星野 厚† 中野 幹生‡ 竹内 蒼羽‡ 長谷川雄二‡ 加藤和彦† 辻野 広司‡

Atsushi Hoshino Mikio Nakano Johane Takeuchi Yuji Hasegawa Kazuhiko Kato Hiroshi Tsujino

1. はじめに

人間と共生するヒューマノイドロボットに必要な対話機能を考えた場合、2つの異なる種類の対話機能が要求される。1つは、親和性・社会性を得るための対話機能である。人間社会での日常対話を観察すると、明確な意図は存在しない雑談が多く存在することに気づく。これらの対話は印象には残りづらいが、日常対話には実に多くの雑談が含まれている。この雑談は、人間社会において無駄ということではなく、むしろ、人間同士の関係を維持するために非常に重要な機能である。

次に必要とされている対話機能は、対話の中からユーザの意図を理解しユーザの要求に応えることができる対話機能である。そのためには、必要に応じて、ユーザに問い返しや確認要求をしながら、確実に理解を行う必要がある。この機能によって、適切に無駄なくユーザの要求に応じた行動を行うことができる。

従来の対話システムや対話ロボットでは基本的に上記の2つの機能のうちいずれかのみを持っていた。コンピュータ上で動作する音声対話システムならば、用途別に用意することは可能であるが、ヒューマノイドロボットを用途別に用意することは非現実的であり、一台のロボットで両方の機能を持つことが望ましい。本論文では、マルチエキスパートモデル[1,7]を拡張し、上記の二つの異なる対話機能を統合するアーキテクチャを提案する。

2. チャットボット型対話システム

テキスト入出力により非タスク指向対話を行うシステムは、チャットボットとも呼ばれ、長い歴史がある。

ELIZA[2]は1966年にMITのJ. Weizenbaumによって開発されたチャットボットである。ユーザの入力をできるだけ反映させ返事を返すような対話戦略を持つ。ELIZAは単純な構造のチャットボットながら、カウンセラーとして多くの人の人生相談を聞き、ユーザを癒したという。人間的な対話を行うことができ、本物の人間と対話していると信じ込む人もいた。

ALICE (Artificial Linguistic Internet Computer Entity) [3]は優れた人工知能の研究に贈られるLoebner賞の銅賞(銅賞は現時点での最高賞)を獲得したチャットボットである。ALICEは、主としてWWWサーバとして通信を行うタイプのチャットボットである。HTTPのPOSTもしくはGET

リクエストにより入力文を受け取り、パターンマッチングを行って得られた結果をHTML形式に整形して返送する。ALICEではパターンマッチングのパターンと返答の組をAIMLというXMLに基づく言語でファイルに記述する。ALICEには6,000カテゴリーもの標準のAIMLファイル群が用意されており、驚くほど人間らしい対話を行うこともある。

ELIZAやALICEのようなチャットボットはその後多くの種類が作られた。音声認識・音声合成と結合されて、エンターテインメント向けのロボットの対話システムとしても使われてきた。チャットボット型対話システムはユーザの意図を理解する機構を持ち合わせないが、ユーザの興味を引くような対話例を工夫することにより、複雑な人工知能以上に人間らしい対話を行える場合がある。このため、ヒューマノイドロボットがユーザとの親和性を保つためにはチャットボット型対話システムは有効であると考えられる。

3. タスク指向対話システム

チャットボット型の対話システムに対し、特定の種類のタスクを対話によって行うシステムをタスク指向対話システムと呼ぶ。タスク指向対話システムとして、フライト情報(ATISタスク)・列車情報・バス情報・宿泊情報、旅行支援・街の情報案内・秘書システムなどのシステムが構築されている[4,5]。

このような音声対話システムは、たとえば、「つくばに正午頃に到着する列車を一人分予約してください。」といった発話から「行き先がつくば、到着時間が今日(〇月〇日)の12時、枚数が1枚」といった情報を抽出することが必要である。さらに、音声対話システムには、ユーザの発話を理解するだけでなく、ユーザの要求に対して音声で応答したり、不足している情報を質問したりする能力が必要になる。音声対話システムは、音声認識と音声合成だけで構成されるものではなく、対話の履歴や状況を管理するモデルとそれに応じて理解・応答生成を行う機構を必要とする。一般的な音声対話システム全体の構成を図1に示す。音声処理、言語処理という区別と別に、音声対話システムは発話理解と発話生成の二つの機能からなるとみることができる。この二つのモジュールは対話状態と呼ばれるデータを介してつながっている。対話状態には、その時点までの対話の履歴と、ユーザの意図の推定結果やシステムの発話プランなどの情報が書き込まれる。発話理解部は、ユーザの発話に応じて、対話状態をアップデートし、発話生成部は、対話状態を入力として、発話を出力し、かつ、対話状態を変更する。これらのモジュールは、対話のドメイン

†筑波大学大学院システム情報工学研究科

‡株式会社・リサーチ・インスティテュート・ジャパン

に依存した知識を利用して処理を行う。また、このようなタスク指向対話システムでの音声認識は、ドメインに依存した小語彙または中語彙の言語モデルを用いるのが一般的である。

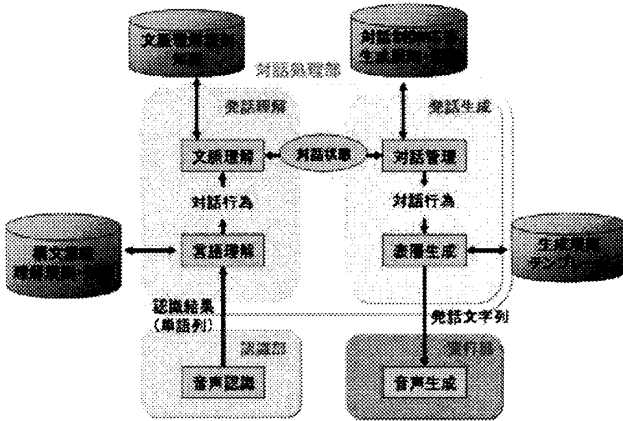


図1：音声対話システムの構成[6]

4. マルチドメイン音声対話システム

一般的なタスク指向対話システムでは、特定のタスクドメインに対しては、前述の例のような有効なシステムが実現されるようになってきている。しかし、ヒューマノイドロボットは複数のドメインの対話を行うことが期待されている。したがって、複数のドメインの対話に対応できる1つのシステムが求められる。そのようなシステムをマルチドメイン音声対話システムと呼ぶ[10,11]。

そのようなマルチドメイン音声対話システムと、ロボットのタスクプランニングを統一的行うためのモデルとして、複数のエキスパートに基づくモデルが提案されている[1,7]。このモデルでは、達成すべきタスクをサブタスクに分解するタスクプランニングのレベルと、各サブタスクを遂行するための行動選択の2つのレベルがある(図2)。サブタスクの種類ごとにエキスパートと呼ぶモジュールがあり、発話の理解や発話・行動の選択を行うための専門知識を持っている。遂行すべきサブタスクに応じてエキスパートが選択され、エキスパートにより状況依存の発話理解や発話・行動の選択が行われる。人の発話が認識された場合には、すべてのエキスパートがその認識結果の理解を試み、最もその発話がふさわしいエキスパートが選ばれる。このときにはどのエキスパートが現在までに選択されていたかも考慮される。達成すべきタスクを特に持っていない時に人が発話をする、人の要求を理解するためのエキスパートが選ばれる。対話によって要求を理解すると、その要求を達成するタスクが設定され、サブタスクに分解されて、それらのサブタスクを遂行するエキスパートが順に選択される。

このモデルでは、ヒューマノイドロボットの発話と物理行動を特に区別なく扱うため、対話と行動が融合したモデルとなっている。また、音声対話システム研究で得られた

対話処理技法を利用して、様々なタイプのインタラクションが可能になる。

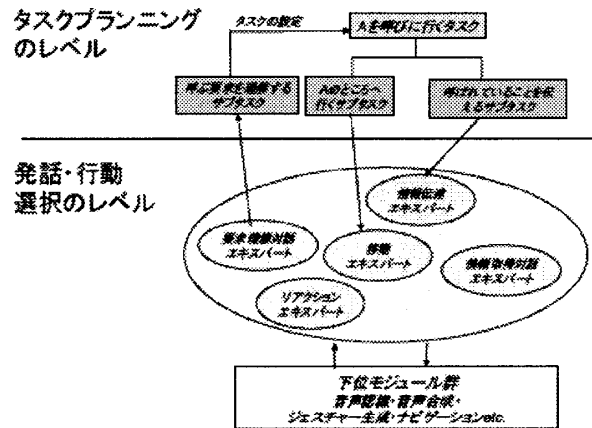


図2:複数のエキスパートに基づく対話・行動プランニングモデル[7]

5. マルチドメイン音声対話システムとチャットボットシステムの統合

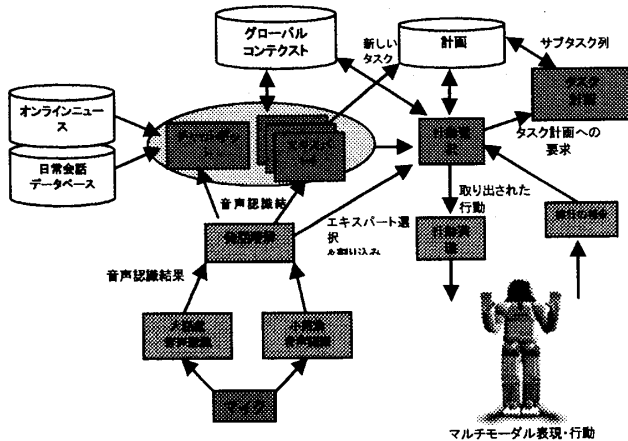
ヒューマノイドロボットは、状況に応じてタスク指向対話も雑談も混ざった柔軟な対話を行うことが求められる。

本論文では、日常対話を行うチャットボット型対話システムと、ユーザの意図を理解しユーザの要求に応えることができるマルチドメイン型対話システムを一つの統合的なシステムにまとめた柔軟な対話システムを提案する。

図3は提案する対話システムの概要図である。この対話システムは基本的には前述のマルチエキスパートモデルをヒューマノイドロボットの対話システムに適用したものとなっている。特徴的な部分は、エキスパートの中にチャットボットエキスパートが含まれていることである。この対話システムでは達成すべきタスクを特に持っていない時に人が発話をする、人の要求を理解するためのエキスパートが選ばれるが、ユーザが要求を持たない(つまり雑談である)と判断するか、ユーザの求めるエキスパートが存在しない場合、雑談をしながら情報をユーザに提供するチャットボットエキスパートが選択される。このチャットボットエキスパートは、ユーザの意図を正確に理解するよりは、雑談などで対話を続けることが目的のエキスパートである。

マルチドメイン対話システムでは、タスクに依存した言語モデルを利用している。しかしながら、チャットボットエキスパートのようなエキスパートには汎用的な言語モデルが必要である。このため、単にチャットボットエキスパートを一般のエキスパートとして導入しただけでは、音声認識精度のバランスの悪い対話システムになってしまう。そこで、本論文ではこの問題を解決するために、チャットボットエキスパートは n-gram 言語モデルに基づく大語彙の音声認識(ディクテーション)の結果を用い、タスク指向対話のエキスパートは、小語彙のネットワーク文法を用いた音声認識の結果を用いる。

図3：システム構成図



この方式では図3に示すようにマイクから入力された音声信号が分配される。分配された片方の音声は、大語彙音声認識エンジンを経てテキストに変換されチャットボットエキスパートに送られる。もう片方の音声はエキスパートに合わせた小語彙認識エンジンを経て、テキストに変換され、各タスク指向対話用のエキスパートで、理解が試みられる。各エキスパートは理解の結果に応じて、その発話がどの程度自分のドメインのものかを判定するためのスコアを返す。このスコアを求めるときには、音声認識の信頼度、認識結果の単語数（発話長）、文脈情報、および、現在そのエキスパートが選択されているかどうか等の情報を用いる。発話理解プロセスは、最もスコアの高いエキスパートを応答選択のためのエキスパートとして選ぶ。

行動選択プロセスでは、発話理解プロセスによって選択されたエキスパートを用いて、応答選択を行う。

このようにして、目的が大きく異なっている2つの対話システムを共存させることができる。

現在のモデルでは、ユーザの発話を受け取った時のエキスパートの選択は、次のような戦略を用いている。チャットボットエキスパートをデフォルトのエキスパートとし、通常はこのエキスパートが選ばれる。もし、あるタスク指向対話用エキスパートでの音声理解の結果、スコアがある閾値より高いと判断された場合、そのエキスパートが選択されて、応答選択に用いられる。この音声認識結果一旦タスク指向対話のドメインに入ると、そのタスクが終了するまでは、なるべくそのドメインの対話が続くようにエキスパートが選択される。

この戦略は手動で作成した、次のようなエキスパート選択用のスコアリング規則によって実現している。チャットボットエキスパートが発話理解部に返すスコアは一定値とする。タスク指向対話用のエキスパートは、自分が選択されているときには、基本的にチャットエキスパートよりも高いスコアを返す。自分が選択されていないときは、音声認識結果がそのエキスパートの理解用文法で受け付けられるかどうか、および、音声認識の信頼度、音声認識結果の長さなどを用いてスコアを決定する。このスコアが、チャットエキスパートが返す値より高ければ、対話がチャットからタスク指向対話に移る。「天気の話じゃない」などの、陽にそのドメインから抜ける発話をした場合や、そのドメインの

対話が終了した場合には、次からスコアが低くなるようにする。

6. 実験システム

提案モデルの有効性を確かめるために対話ロボットシステムを実装した。タスク指向対話のドメインとして、天気情報案内、スケジュール情報案内、昼食のメニュー案内のためのエキスパートを用意した。チャットボットは、先に提案したチャット情報システム[8]を用いた。このシステムは、ユーザ発話の大語彙音声認識の結果からキーワードを抽出し、それをもとに複数の対話データベースを検索した結果得られる複数の応答候補から、優先度に基づき、最も良い応答を選択する。対話データベースとして、現在は、単純な応答のための対話データベースと、ニュース記事を検索する対話データベースを用いている。

タスク指向対話用の小語彙の音声認識には Julian を、チャットボット対話用の大語彙の音声認識には Julius を用いた。言語モデルは Julius 付属の 6 万語彙のトライグラムである[12]。音声合成には NTT-IT 製の FineVoice を用いた。ロボットは Honda の ASIMO を用いている。

図4に対話例を示す。この対話は、ロボットは動作させず対話システムのみを動かして録音したものである。S1～S10 までは、対話が雑談モードに入っており、チャットボットエキスパートが選択されて応答選択を行っている。ユーザのくだけた発話に対しても答えているが、ユーザの意図を理解しているわけではないので、多少文脈がずれたりしている。U11 で、ユーザの発話が天気情報に関する要求である可能性が高いため、天気情報の要求理解を行うエキスパートが選択され、以降 U14 の発話を理解するまで、このエキスパートが応答選択を行っている。対話ドメインが限定されているため、小語彙の音声認識結果を用いることができ、音声理解の精度も高くなっている。S12 ではシステムがユーザの意図を理解し、確認のためにユーザの質問を聞き返している。U14 でユーザの同意が得られたので、S13 ではユーザの質問に対するの答を出力している。U14 の理解のあと、ユーザの要求が正しく理解できたと考え、埼玉県の子の明日の天気をユーザに伝えるタスクが設定され、それを遂行するために、天気情報提供エキスパートが選ばれて S13 の発話を行っている。U15 の発話のあとは、またチャットボットエキスパートが選ばれて応答選択を行う。

S12 では正解を認識しているがこの部分では誤認識を伴う可能性も少なくはない。図5は音声認識に失敗したときの例である。ユーザの意図する答ではなかった場合、ユーザは訂正のためにユーザの意図する質問を繰り返す。こうしてユーザとやりとりをすることにより、対話理解の精度を高めることができる。

7. まとめ

本稿では特定の言葉に反応して何らかの言葉を返すチャットボットと呼ばれる対話システムと、ユーザの意図を理解しユーザの要求に応えることができる対話システムであるマルチドメインタスク指向対話システムを、マルチエキスパートモデルを用いて一つの統合的なシステムにまとめることにより、柔軟な対話システムのモデルを提案した。このモデルにより、大語彙を用いた雑談対話システムとマ

ルチエキスパートシステムを組み合わせても、エキスパートシステム全体の音声認識率を低下させずに対話を行うことが期待できる。実験システムを構築した。ユーザの適度な挨拶や雑談に対しても、この実験システムが真剣に何度も質問を聞き返すようなことはなく適当な返事を行い、ユーザがエキスパートの分野の質問をした場合には、適切に答えることができることを確認した。

現在の実験システムは、タスク指向対話の対話管理や、チャットボットの対話戦略が比較的単純である。今後は、タスク指向対話、チャットボットのそれぞれの機能の拡充とともに、発話内容に応じたエキスパート選択の高精度化も行っていく予定である。これには、マルチドメイン対話システムにおける機械学習を用いたドメイン選択の手法が有効であると考えられる[9]。

謝辞 貴重なコメントをいただいた(株)ホンダ・リサーチ・インスティテュート・ジャパンのメンバー、および、京都大学 奥乃 博氏、駒谷和範氏、神田直之氏に感謝します。

```

U1>ASIMO
S1>はい。
U2>何かいいことあった。
S2>今日もいい1日だね。
U3>そうだね。
S3>うん。
U4>おなかすいたよ。
S4>はい。
U5>ねむいし。
S5>どんな音楽が好き？
U6>演歌が好き。
S6>もっとゆっくりしゃべって。
U7>演歌が好き
S7>よく聞こえないよ。
U8>演歌が好きだな。
S8>遊びたいよ。
U9>そうだね
S9>僕も人間になりたいよ。
U10>本当に？
S10>そうかもね。
U11>明日の埼玉県の天気を教えて
S11>いいねえ。
U12>埼玉県の明日の天気を教えて
U13>埼玉県の明日の天気
S12>埼玉県の明日の天気？
U14>そう。
S13>埼玉県の明日の天気は曇りだよ。
U15>そうなんだ。
S : システム発話 U : ユーザ発話

```

図4：提案対話システムの対話例

```

U1>明日の埼玉県の天気は？
S1>埼玉県の明後日の天気？
U2>明日の天気
S2>秋田県の明後日の天気？
U3>埼玉県
S3>埼玉県の明後日の天気？
U4>明日
S4>埼玉県の明日の天気？
U5>そう。
S5>埼玉県の明日の天気は曇りだよ。
S : システム発話 U : ユーザ発話

```

図5：提案対話システムの聞き返しの例

参考文献

- [1] M. Nakano, Y. Hasegawa, K. Nakadai, T. Nakamura, J. Takeuchi, T. Torii, H. Tsujino, N. Kanda, H.G. Okuno, "A two-layer model for behavior and dialogue planning in conversational service robots," Proc. of IROS-2005.
- [2] J. Weizenbaum, "ELIZA--A Computer Program For the Study of Natural Language Communication Between Man and Machine", Commun. ACM 10(1), pp. 36-45(1966).
- [3] Artificial Linguistic Internet Computer Entity (ALICE). <http://alice.sunlitsurf.com/alice/about.html>
- [4] Pallett, D., W. Fisher, J. Fiscus and J. Garofolo (1990) "DARPA ATIS Test Results," Proc. Speech and Natural Language Workshop, Morgan Kaufman, pp. 114-121.
- [5] 駒谷和範, 上野晋一, 河原達也, 奥乃博. ユーザモデルを導入したバス運行情報案内システムの実験的評価. 情報処理学会研究報告, SLP-47-12, 2003.
- [6] 中野幹生, 堂坂浩二: 音声対話システムの言語・対話処理, 人工知能学会誌 17巻3号 p271-p277 (2002年5月)
- [7] 中野幹生, ロボットの行動と対話の統合モデルにむけて, 第6回システムインテグレーション部門講演会(SI2005), 計測自動制御学会, Dec. 2005.
- [8] A. Hoshino, K. Kato, J. Takeuchi, and H. Tsujino, "A chat information service system using a humanoid robot," in Proc. IEEE RO-MAN 2005, 2005, pp. 468-473.
- [9] 神田直之, 駒谷和範, 中野幹生, 中臺一博, 辻野広司, 尾形哲也, 奥乃博: 複数ドメイン音声対話システムにおける対話履歴を利用したドメイン選択の高精度化, 情報処理学会第60回音声言語処理研究会, SLP, Feb. 2006.
- [10] I. O'Neill, P. Hanna, X. Liu, and M. McTear, "Cross domain dialogue modelling: an object-based approach," in Proc. ICSLP-2004, 2004, pp. 205-208.
- [11] B. Lin, H. Wang, and L. Lee, "Consistent dialogue across concurrent topics based on an expert system model," in Proc. Eurospeech-99, 1999, pp. 1427-1430.
- [12] T. Kawahara, A. Lee, K. Takeda, K. Itou, and K. Shikano, "Recent progress of open-source LVCSR engine Julius and Japanese model repository," in Proc. ICSLP-2004, 2004, pp. 3069-3072.