

複数のエキスパートに基づく ロボット対話行動制御モデル構築ツール

A Tool for Building Dialogue and Behavior Control Model Based on Multiple Experts

中野 幹生*¹ 船越 孝太郎*¹ 長谷川 雄二*¹ 辻野 広司*¹
Mikio Nakano Kotaro Funakoshi Yuji Hasegawa Hiroshi Tsujino

*¹ (株) ホンダ・リサーチ・インスティテュート・ジャパン
Honda Research Institute Japan Co., Ltd.

This paper describes RIME-TK, which is a toolkit for building the dialogue and behavior control subsystem in conversational robots. It enables building the intelligence modules of robots that can perform tasks by performing physical actions and engaging in dialogues in multiple domains. It employs modules called experts each of which is specialized to perform certain kinds of tasks by engaging in dialogues and performing non-verbal actions. Experts in charge are responsible for understanding user utterances and selecting agent utterances and actions to execute their tasks. RIME-TK facilitates building experts by providing templates for experts. Robotic systems that can engage in various kinds of tasks have been implemented using RIME-TK.

1. はじめに

音声対話機能を持つロボットは、人間と対話しながら人間の要求を理解し、物理的な行動を行ったり対話によって情報を授受したりしながらタスクを遂行する必要がある。また、人の発話に応じてタスクを動的に変更したり、実行中の動作への中止の要求に対処したりすることが必要になる。

このようなさまざまなタスクを行うロボットの対話行動制御部には複雑なモデルが必要である [Asoh 01]。したがって、対話行動制御部をできる限り容易に構築・改変・拡張できるようにするためのツールが必要である。

本稿では、我々が作成している対話行動制御部構築ツールキット RIME-TK について述べる。これは、特定のタスクに特化した知識と内部状態を持つエキスパートと呼ぶモジュールを複数用いるモデル RIME (Robot Intelligence based on Multiple Experts)*¹ [Nakano 05] に基づいている。本稿ではこのツールキットの構成について述べ、さらに、このツールキットを用いて構築した、いくつかのアプリケーションを紹介する。

2. 対話行動制御のマルチエキスパートモデル

2.1 モジュール構成

RIME は、ロボットの知能システムの中で、記号レベルでの状況理解と行動選択を司るモジュール(対話行動制御部と呼ぶ)のモデルである。対話行動制御部は、音声・画像認識や他のセンサ解釈モジュールによってシンボル化された入力をうけとり、それを基に人の意図や状況を推測し、適切な適切な行動を選択し、ロボットハードウェア制御や音声合成などの行動実行モジュールに送る。RIME のモジュール構成を図 1 に示す。

現在我々が開発しているシステムでは、入力として音声認識の結果のみを用いているため、以下では音声認識についてのみ言及するが、画像認識結果と組み合わせたマルチモーダル認識結果を扱うように拡張することも可能である。

2.2 エキスパート

RIME で用いるエキスパートは、特定の種類のサブタスクに特化した知識と内部状態を持つ。たとえば、天気予報に関する質問に答えられるロボットであれば、「天気予報に関する質問を理解する」というサブタスクのためのエキスパートや「天気予報を人に伝える」というサブタスクのためのエキスパートを持つ。また、「特定の場所に移動する」という物理行動を行うサブタスクのためのエキスパートや、「人の発話に単純に反応する」ためのエキスパートなども用いることができる。これらのエキスパートを順次利用することにより、複雑なタスクを遂行することができる。たとえば、ある物を説明するタスクは、その物のところに人を案内して、言葉で説明するという二つのサブタスクを順次遂行することによって行うことができる。

各エキスパートは、発話を理解するための知識や行動選択をするための知識を持ち、サブタスクの遂行の状態を内部的に保持している。たとえば、「天気予報に関する質問を理解する」エキスパートであれば、理解の途中結果などが内部状態として保持されている。

エキスパートには、システム主導タスク遂行エキスパートとユーザ主導インタラクションエキスパートの 2 種類がある。前者は、サブタスクを遂行するために、対話や行動を行う。これに対し後者は、遂行すべきサブタスクをあらかじめ持たず、ユーザのタスク要求を理解したり、ユーザの発話に反応したりするためのエキスパートである。

すべてのエキスパートが共通に持つべきインタフェースが定義されており、それを実装しさえすれば、内部で知識や状態をどのような形で保持しているか、また、どのようなアルゴリズムで理解や行動選択を行うかに関わらず、どのようなエキスパートでも導入することができる。

エキスパートのインタフェースにはタスクタイプ依存の部分と非依存の部分がある。タスクタイプ非依存の部分はエキスパートのアクティベートを行う関数や現在実行している行動を返す関数などからなる。タスクタイプ依存の部分は、システム的设计者が実装しなくてはならないもので、以下のような関数が含まれる。

understand 音声認識結果を受け取り、現在の内部状態に基づいて理解するとともに、その発話がどのくらいその工

連絡先: 〒 351-0188 埼玉県和光市本町 8-1 (株) ホンダ・リサーチ・インスティテュート・ジャパン, 中野 幹生, E-mail: nakano@jp.honda-ri.com

*¹ [Nakano 05] では MEBDP と呼んでいた。

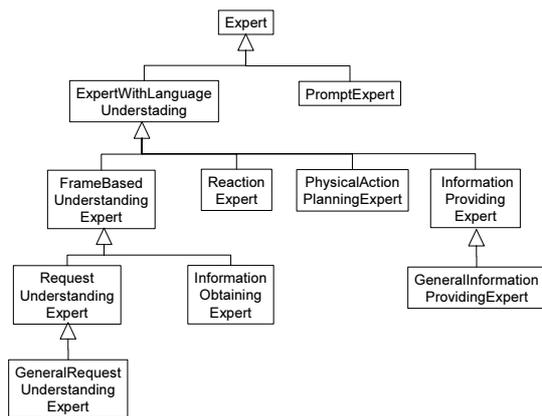


図 2: エキスパートの抽象クラス階層 (一部分)

```
<keyphrase-class name="person">
  <keyphrase>
    <orth>すずき さん</orth>
    <sem>鈴木さん</sem>
  </keyphrase>
  <keyphrase>
    <orth>たなか さん</orth>
    <sem>田中さん</sem>
  </keyphrase>
</keyphrase-class>

<action type="specify-attribute">
  <sentence> *person の でんわ ばんごー
    (わ |お おしえて [ください]) </sentence>
</action>
```

図 3: 音声理解用知識の例

また行動実行部は行動が成功したかどうかの報告を行動選択プロセスに返すようにする必要がある。

3.2 RIME-TK が提供する機能

上記の作業を容易にするため、RIME-TK は以下のような機能を備えている。

エキスパートテンプレート

エキスパートの構築を容易にするために、頻繁に用いられる機能を実装したテンプレートを用意している。これらはオブジェクト指向の抽象クラスの階層として用意されている。その階層の一部を図 2 に示し、各クラスの説明を表 1 に示す。

入力インタフェースのサンプル

音声認識システム Julius/Julian[Kawahara 04] の入力を受け付けるためのインタフェースが用意されている。また、デバッグやシミュレーションのために、ユーザ発話をキーボード入力やファイルから入力できるようにするためのインタフェースも用意されている。Julius/Julian を使う場合、ひとつの音声入力ストリームに対して複数の音声認識器を用いることが可能である。各エキスパートがどの音声認識器の出力を用いるかを指定する必要がある。

音声認識用ネットワーク文法・音声理解部の構築

ExpertWithLanguageUnderstanding を具体化しているエキスパートは、図 3 の例のような XML 形式で書かれた言語理解知識記述を読み込み、これを言語理解のための有限状態トランスジューサ [Potamianos 00] に変換する。言語理解知識記述はキーフレーズの定義と文型の定義からなる [Glass 01]。図 3 の知識記述を用いた場合「たなか さんの でんわ ばんごー わ」という音声認識結果を理解した結果は、[type=specify-attribute, person=田中さん] のようになる。

また、各エキスパートの言語理解用知識から、音声認識器 Julian 用のネットワーク文法を生成するツールが用意されている。

各エキスパートは、複数の言語理解知識を用いることができ、文脈に応じてその一部のみを用いて理解することによって理解の精度を高められるようになっている。

出力インタフェースのサンプル

出力インタフェースを実装する行動実行部として、ロボットの発話・行動を制御するシステムが用意されている。このシステムは、NTT-IT 製の FineVoice を用いた音声合成サーバ、Honda ヒューマノイドロボット Asimo や、全方向移動台車ロボット、ロボットのシミュレータなどへの行動命令を授受する

サーバ [鳥井 06]、ロボットがスクリーンを用いてプレゼンテーションができるようにしたスクリーンコントロールサーバとの通信機能を持つ。音声合成の代わりにテキスト表示も可能である。

テンプレートベースの行動表現生成

自然言語生成に用いられるテンプレートベースの文生成技術を用いて、属性-値のペアで表現された意味表現から MADL を生成することが可能である。

簡易な知識記述による対話機能の実現

GeneralRequestUnderstandingExpert と GeneralInformationProvidingExpert は、設計者がプログラミングをしなくても、テーブル形式のデータベースの内容を、ユーザの要求に応じて提供することができることを可能にするエキスパートである。設計者は、テーブルの属性値を指定するユーザの発話パターン、ユーザへの問い返しや確認要求をするロボットの発話パターン、情報を提供する場合のロボット発話のパターンなどを記述することによって、容易に情報提供タスクが実現できる。これらのエキスパートが実現している対話管理は非常に単純であるが、これらを拡張することによって、対話管理のバリエーションを増やすことができる。

4. アプリケーション

RIME-TK を用いてさまざまなタスクを行うロボットシステムを実現している。以下にタスクの例を示す。

- 天気予報・内線番号などの情報提供: 情報提供の要求理解は GeneralRequestUnderstandingExpert を具体化したエキスパートで、情報提供は GeneralInformationProvidingExpert を具体化したエキスパートで行っている。
- 挨拶などのリアクション: ReactionExpert を具体化したエキスパートを用いている。
- 呼んだ人のところへの接近: 呼ぶ要求の理解は ReactionExpert を具体化したエキスパートが、移動は、PhysicalActionPlanningExpert を具体化したエキスパートが行う。
- 場所の案内: 案内の要求は GeneralRequestUnderstandingExpert を具体化したエキスパートを、案内する場所への移動は PhysicalActionPlanningExpert を具体化したエキスパートを、移動後に行う説明は InformationProvidingExpert を具体化したエキスパートを用いる。

表 1: エキスパートの抽象クラスの機能

抽象クラス	機能
ExpertWithLanguageUnderstandin	有限状態トランスジューサを用いた言語理解を行う
PromptExpert	発話をひとつだけ生成する
FrameBasedUnderstandingExpert	フレームを用いた対話管理 [Goddeau 96] を行う
ReactionExpert	相手の発話内容に単純に反応して発話・行動を行う
PhysicalActionPlanningExpert	物理行動の計画を行い、必要な行動を選択する。人の発話による割り込みに対応する
InformationProvidingExpert	発話により人に情報提供をする。問い返しに対応できる
RequestUnderstandingExpert	人が実行を要求しているタスクを対話により理解する
GeneralInformationProvidingExpert	情報提供対話の機能を簡易な知識記述から実現する
GeneralRequestUnderstandingExpert	要求理解対話の機能を簡易な知識記述から実現する
InformationObtainingExpert	タスクの実行に必要な情報を人との対話によって取得する

- 非タスク指向型対話 (雑談): 大語彙連続音声認識の結果中のキーワードや文のパターンを用いて、データベースから応答パターンを抽出して出力する。データベースはXML形式での記述から構築できる。トップレベルの抽象クラス Expert を具体化したエキスパートを用いる [星野 06] .
- スクリーンを用いたプレゼンテーション: ユーザからの割り込みに応じて説明の内容を変更することができる。プレゼンテーションの内容は MPML-HR(Multimodal Presentation Markup Language for Humanoid Robots) ver.3 を用いて記述することができる [Nishimura 07] . プレゼンテーションの開始の要求を理解するエキスパートは GeneralRequestUnderstandingExpert を具体化したもので、プレゼンテーションを行うエキスパートは、ExpertWithLanguageUnderstanding を具体化したものである。

これらのタスクを組み合わせることにより、様々なロボット対話システムが構築できている。また、ロボットと接続せず、音声のみでユーザと対話を行うマルチドメイン音声対話システムの構築にも用いている [神田 07] .

5. おわりに

本稿では、ロボットの対話行動制御部を構築するツール RIME-TK について述べた。RIME-TK は、複数エキスパートモデルに基づいており、さまざまなタイプのタスクを行うことができるように拡張することが可能である。今後は、エキスパートの構築をさらに容易にするため、ツールを改良していく予定である。

謝辞

RIME-TK の作成にご協力いただいた方々に感謝します。特に、(株) ホンダ・リサーチ・インスティテュート・ジャパン 竹内誉羽氏、鳥井豊隆氏、元京都大学奥乃研究室 神田直之氏、元東京大学石塚研究室 西村義隆氏、東京大学石塚研究室 簗津真一郎氏に感謝します。

参考文献

[Asoh 01] Asoh, H., Motomura, Y., Asano, F., Hara, I., Hayamizu, S., Itou, K., Kurita, T., Matsui, T., Vlassis, N., Bunschoten, R., and Kroese, B.: Jijo-2: An Office Robot that Communicates and Learns, *IEEE Intelligent Systems*, Vol. 16, No. 5, pp. 46–55 (2001)

[Glass 01] Glass, J. and Weinstein, E.: SPEECHBUILDER: Facilitating Spoken Dialogue System Development, in *Proc. 7th Eurospeech*, pp. 1335–1338 (2001)

[Goddeau 96] Goddeau, D., Meng, H., Polifroni, J., Seneff, S., and Busayapongchai, S.: A form-based dialogue manager for spoken language applications, in *Proc. 4th ICSLP* (1996)

[Kawahara 04] Kawahara, T., Lee, A., Takeda, K., Itou, K., and Shikano, K.: Recent progress of open-source LVCSR engine Julius and Japanese model repository, in *Proc. 8th ICSLP*, pp. 3069–3072 (2004)

[Nakano 05] Nakano, M., Hasegawa, Y., Nakadai, K., Nakamura, T., Takeuchi, J., Torii, T., Tsujino, H., Kanda, N., and Okuno, H. G.: A Two-Layer Model for Behavior and Dialogue Planning in Conversational Service Robots, in *Proc. IROS-2005*, pp. 1542–1547 (2005)

[Nishimura 07] Nishimura, Y., Minotsu, S., Dohi, H., Ishizuka, M., Nakano, M., Funakoshi, K., Takeuchi, J., Hasegawa, Y., and Tsujino, H.: A Markup Language for Describing Interactive Humanoid Robot Presentations, in *Proc. IUI'07* (2007)

[Potamianos 00] Potamianos, A. and Kuo, H.-K. J.: Statistical recursive finite state machine parsing for speech understanding, in *Proc. 6th ICSLP*, pp. 510–513 (2000)

[神田 07] 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: マルチ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, 情報処理学会論文誌, Vol. 48, No. 5 (2007)

[星野 06] 星野 厚, 中野 幹生, 竹内 誉羽, 長谷川 雄二, 加藤和彦, 辻野 広司: 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, 情報科学技術レターズ, Vol. 5, pp. 127–130 (2006)

[鳥井 06] 鳥井 豊隆, 長谷川 雄二, 中野 幹生, 中臺 一博, 辻野 広司: 人・ロボットインタラクションシステムの為のモデルウェアの開発, 計測自動制御学会第 7 回システムインテグレーション部門 講演会 (SI2006), pp. 2D2–1 (2006)