

擬人化エージェントを用いた汎用対話システムの開発

4R-10

森 真史 土肥 浩 石塚 満

東京大学 工学部 電子情報工学科

e-mail: mori, dohi, ishizuka @miv.t.u-tokyo.ac.jp

1 はじめに

我々は新時代のヒューマンインタフェースの形態として擬人化エージェントをもちいた対話システムを研究しており、既にVSA(Visual Software Agent)とよぶシステムを開発している¹⁾²⁾(図1)。これはコンピュータ利用者からの音声入力に対してコンピュータ画面上の顔画像と合成音声を実時間で応答するというものである。

なお、本論文での「対話」とは従来の「音声対話」とは異なり、音声のみならず画像をも融合したマルチモーダルなやりとりであると定義する。



図1: VSA システム

2 汎用性

これまでに開発したシステムは特定の応用のみを使用するように設計されているために、別の応用に用いるためにはシステムの再設計が必要となり、またシステムが記述されているプログラミング言語に精通している必要があった。このことはシステムの構築、保守に要するコストが膨大であることを意味し、これらのシステムを実用化するに際して非常に大きな障害となっていた。

我々はこの問題を解決するために対話システムの構造を汎用的にすることが必要と考え、適用する内容

に依存しないシステム構造と汎用対話エンジンをもつ汎用対話システムを構築した。これにより、プログラミング言語の知識なしにシステムを作成することができる。

このことによって得られる利点は、個別の対話システムを構築する際に専門のプログラマや専用の開発環境が不要となり、対話システムの実用が極めて容易になるということである。

3 システム構成

本汎用対話システムの構成を図2に示す。

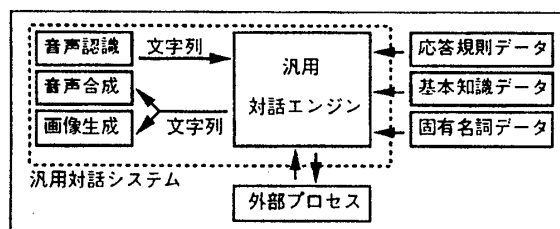


図2: システムのブロック図

音声認識、音声合成、画像生成の各ユニットは既に本研究室で動作していたものである。音声認識ユニットはキーワードスポッティングによってユーザの発話中に含まれるキーワード列を出力する。音声合成、画像生成ユニットは応答すべき文字列を受けとって画像の口を動かすと同時に発声する。

汎用対話エンジンは音声認識ユニットで認識されたキーワード列を受けとり、外部プロセスと協調しつつデータファイルの内容にしたがって応答を決定し、音声合成、画像生成ユニットに出力する。

4 外部プロセスとの協調

システムの汎用性を高めればそれだけある応用に特有な処理を作り込むことはできなくなる。そこで汎用対話システムに外部プロセスを呼び出す機構を設けておき、その外部プロセスとやりとりすることによって応用固有の処理を実現する。

例えば、案内システムであれば案内する場所の写真を画像データとして蓄えておき、ユーザが「そこはどこなところですか」という発話をしたら外部のビューアを呼び出して写真を表示する。

あるいは、運転監視システムであれば通常は対話シ

システムは外部プロセスからの情報を受信しておき、なにか緊急事態が発生して外部プロセスから信号が送出された時のみ状況に応じてユーザーに適切な指示を出すことができる。

5 データ構造

汎用対話システムの動作を規定するのは固有名詞データ、基本知識データ、応答規則データである。我々はこのデータの作成のコストを抑えるために、ごく単純な文法に従って容易に記述できるデータ構造を採用した。

基本知識データは応用で取り扱う概念と、それが所有する属性値を規定する。図3の形式で与えることで、内部に図4の階層を構築する。

この場合、「食堂」に分類される固有名詞は「メニュー、営業時間、立地」の属性を所有することを示す。

建物/
立地
店舗:建物
営業時間
食堂:店舗
メニュー
購買部:店舗
取扱商品
立地
所在地
営業時間
開店時間 閉店時間
メニュー-LIST
取扱商品-LIST

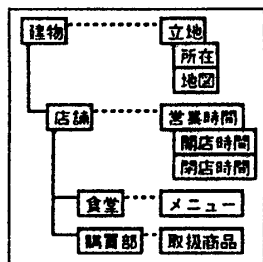


図3: 基本知識データ

図4: 基本知識データの内部表現

固有名詞データは基本知識データに対応した固有名詞を規定し、図5の形式で与えることで、内部に図6のデータを構築する。

中央食堂:食堂
立地=(所在=安田講堂下 地図="central.gif")
営業時間=(開店時間=(時間=(時=8 分=00)) 閉店時間=(時間=(時=18 分=30)))
メニュー=(「定食, ランチ」)
第二食堂:食堂
立地=(所在=ロータリー前 地図="second.gif")
営業時間=(開店時間=(時間=(時=10 分=10)) 閉店時間=(時間=(時=16 分=00)))
メニュー=(「ハンバーガー」)
第一購買部:購買部
立地=(所在=安田講堂南 地図="first.gif")
営業時間=(開店時間=(時間=(時=10 分=10)) 閉店時間=(時間=(時=16 分=00)))
取扱商品=(「文房具, 生活雑貨」)

図5: 固有名詞データ

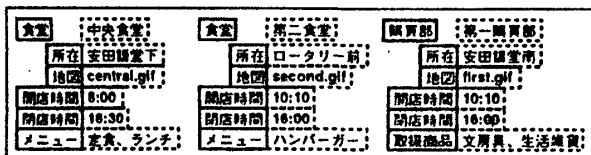


図6: 固有名詞データの内部表現

応答規則データは汎用対話エンジンの応答を規定するものであり、ユーザーの発話中に含まれるキーワード

の位置関係とシステムが返す応答の形式を記述する。

図7の形式で与え、内部に図8の検索エントリを生成する。

この場合キーワード「何時から」と「何時まで」がこの順番で現れ、かつ「店舗」に分類される固有名詞が存在したらその営業時間を答えることになる。

Exist:どこ Target:建物
Exist:行き方 Target:建物
Answer:(立地, 所在) があります | 地図をお見せしましょう (立地, 地図)

Relation:何時から-何時まで Exist:営業時間 Target:店舗
Relation:何時から-何時まで Target:店舗
Exist:営業時間 Target:店舗
Answer:(開店時間) から (閉店時間) までです

図7: 応答規則データ

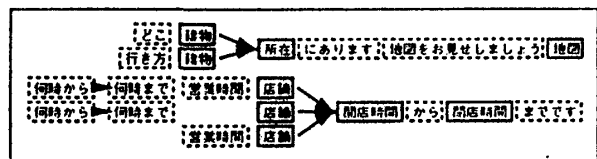


図8: 応答規則データの内部表現

例えば、「中央食堂の営業時間は何時から何時までですか」というユーザーの発話があった場合、キーワードスポッティングによって「中央食堂→営業時間→何時から→何時まで」というキーワード列が得られ、応答規則と固有名詞の中央食堂の営業時間という項目を参照することで「8時ちょうどから19時30分までです」という応答が得られる。

従ってこの汎用対話システムを別の応用に用いるにはシステムの変更は一切必要なく、動作を規定する3個のデータファイルを変更するだけでよい。

なお、実際のデータファイルには音声認識、合成のための発音データ等も記述されている。

6 おわりに

本研究では顔をもつ擬人化エージェントをもちいたマルチモーダルな対話システムにおいて、汎用対話エンジンを開発した。これによってデータの書換えのみであらゆる応用のインタフェースとして擬人化エージェントを用いることができる。今後は応答決定についてより有効な手法を考案していく予定である。

参考文献

- 1) Y. Hiramoto et al.: "A Speech Dialogue Management System for Human Interface employing Visual Anthropomorphic Agent", Proc. RO-MAN'94, Nagoya(1994)
- 2) H. Dohi et al.: "Realtime Synthesis of a Realistic Anthropomorphic Agent toward Advanced Human-Computer Interaction", Human-Computer Interaction: Software and Hardware Interfaces, Proc. HIC'93, Florida, ELSEVIER(1993)