

E-27

ユーザとの音声対話を通して問題を解決する エージェント対話システム

Agent Dialog System Which Solves Problems through Spoken Dialog with Its User

多胡順司†

Junji Tago

広瀬啓吉‡

Keikichi Hirose

峯松信明§

Nobuaki Minematsu

1. はじめに

多くの対話システムは情報検索などの静的な情報であるテキストを扱うものが多く、行動と直接結びつくような対話システムの研究は少ない。一方で、近年、後者の研究にあたるソフトウェアロボットの行動を自然言語で制御する研究が盛んになってきている[1]。しかし、これらの研究ではソフトウェアロボットとの対話は行なわれず、ユーザがソフトウェアロボットに指示を与えるのみである。そこで筆者らはエージェントを用いて仮想空間中の物を操作する対話システムを試作し、対話の制御、応答文の生成を行うためのプラットフォームを構築した。

2. 対話システム

2.1. 対話システムの構成

対話システムの構成は図 1 に示す通りである。

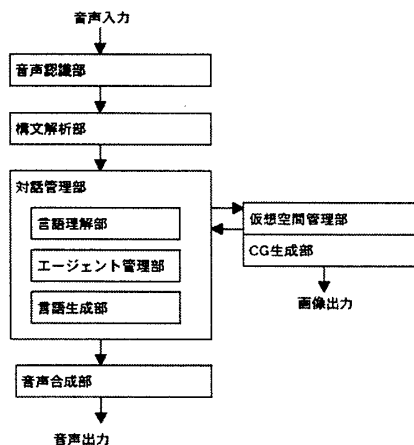


図 1 システム構成

ここでは対話管理部と仮想空間管理部について説明する。対話管理部は対話システムの根幹をなすものであり、言語理解部、エージェント管理部、言語生成部からなる。言語理解部は音声入力の曖昧性を排除する役割を担う。エー

ジェント管理部はユーザの指示に従ってエージェントを動かす。言語生成部はユーザとの対話を行う際に韻律情報を含む文を生成する。仮想空間管理部は仮想空間の物やエージェントの位置や動作などを管理し CG 生成部により仮想空間の状態をリアルタイムに表示する。

2.2. 処理の流れ

対話の処理の流れは次の様になる。まず、音声認識部と構文解析部により構文解析されたユーザの発話が対話管理部に入力される。対話管理部は言語理解部を用いてユーザの入力の中の曖昧な表現を解決する。曖昧性が排除された入力により、エージェント管理部がエージェントの動作を行う。曖昧性の解決の過程やエージェント動作の決定の過程においてシステム自身では問題を解決できない場合、言語生成部を用いて韻律情報を含んだ文を生成し、音声合成部を用いて音声を出し、ユーザとの対話を行ない問題を解決する。

3. 曖昧な入力の解決

3.1. 扱う概念

本システムが扱う主な概念は以下に示す通りである。

アイテム 仮想空間中に配置されている物のことであり、色や種類といった属性を持つ

場所 仮想空間中の位置を表す

本システムでは言語情報は構文木の形で扱う。それぞれの概念に対応する構文木を以下の様に定義する。

アイテムの構文木 種類またはものを表す単語を

root とする構文木 (例: 赤い机, 机の左にある黒いもの)

場所の構文木 位置を表す単語を root とする構文木

(例: 机の左, 白いものの左)

場所がアイテムを修飾する場合は「にある」, 「の」を用い、アイテムが場所を修飾する場合は「の」を用いるとする。

†東京大学大学院工学系研究科

‡東京大学大学院新領域創成科学研究科

§東京大学大学院情報理工学系研究科

3.2. アイテムと場所の決定

システムがユーザの発話内容を理解するためには発話文中のアイテム、場所を一意に特定できなくてはならない。本システムでは以下のようにしてこれを行う。

発話内にアイテムを表す単語があった場合、それを root とするのがアイテムを表す構文木であり、その木の表すものを一意に決定する必要がある。まず、その木の枝を走査し、そのなかにアイテムを表すものや場所を表すものがあれば、先にそれが表すものを一意に決定する。たとえば「赤い机の左にある椅子」というアイテムを表す構文木が存在した場合、「赤い机の左」という場所を表す構文木が「椅子」を修飾しているのだから、まずこの場所を表す構文木の表すものを一意に決定する。

場所もアイテムの場合と同様に構文木の枝要素から曖昧性を解決していく。すなわち、「赤い机の左」という場所の決定は、root となる「左」を修飾している「赤い机」が表すものをまず一意に決定する。空間の位置と言語の関係は[2]に従う。

このような曖昧性の解決の過程で解が一意に求まらない場合はユーザとの対話を行い、解を一つに絞る。

4. エージェントの動作

4.1. 動作と命令

動作とは「空間がある前提となる状態にある時、その状態を終了状態にすること」であり、それぞれの動作には「前提となる状態」と「終了状態」が定義されている。動作には「[場所]に移動する」、「[アイテム]の前に移動する」、「[アイテム]を持つ」、「[アイテム]を[場所]に置く」などがある。

命令とは「空間を目標状態にするように促すこと」であり、それぞれの命令には「目標状態」が定義されている。命令には「持つ」、「置く」、「移動する」などがある。

4.2. 動作の決定

エージェントはユーザから命令を受けると動作を開始する。その際のエージェントの動作は次の様に決定される。まず、命令から目標状態を設定し、目標状態を終了状態とする動作を探す。その動作の前提条件が満たされているかを判断し、満たされていればその動作を実行する。満たされていないならばその状態を新たな目標状態として同じことを再帰的に繰り返す。

この繰り返しの過程で、エージェント自身では解決できない目標状態に達した場合、エージェントはユーザとの対話を開始し、新たな命令を受け付ける。

5. 応答文

文生成は構文木の状態で処理される。構文木を適宜組み合わせることで文を生成する。状態や命令は文構成の情報

を含んでおりそこにアイテムや場所を表す構文を接続することで文を生成する。

応答文の例を以下に示す。

黒色のテレビを持つことを命令されたのですが
黒色のテレビの前にいるという状態ではないので
黒色のテレビの前に移動することをしようとおもったのですが
黒色のテレビの所に移動できるという状態ではないのですが、どうすればいいでしょう？

この様に文生成の過程で、文の内部表現を構文木の状態で扱うことで、音声合成の際の韻律制御との親和性が高くなる。音声合成の過程では韻律制御に文の構造を用いるからである。

6. 試作システム

対話システムを試作した。そのシステムの画像出力を図2に示す。空間には同じ色、同じ種類のアイテムを複数配置する。

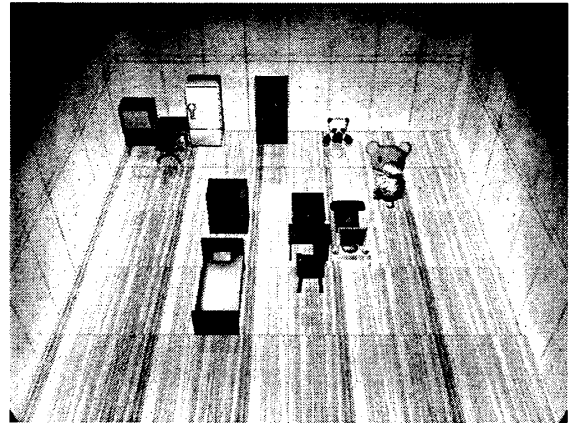


図2 システムの画像出力

7. まとめ

対話システムを構築するにあたり、曖昧な入力の解決法、エージェントの動作決定の方法、応答文生成について説明し、試作したエージェント対話システムを示した。

今後は、入力の曖昧性の解決の際に対話履歴を利用することで照応・省略を扱えるようにする。また、応答文に関して、現段階では非常に冗長な文であるので、照応・省略を利用し、さらに音声の韻律制御も行うことで、ユーザにとって聞きやすい応答を実現する。

参考文献

- [1]. Yusuke Shinyama, Takenobu Tokunaga and Hozumi Tanaka, "Kairai - Software Robots Understanding Natural Language", Third International Workshop on Human-Computer Conversation (2000)
- [2]. 八嶋米美子, 齋藤豪, 奥村学, 中嶋正之, "仮想空間における位置の関係", グラフィックスとCAD研究会 (2001)