

1 N-7 自然言語を理解するアニメテッドエージェント

鈴木泰山 神野潤 新山祐介 秋山英久 徳永健伸 田中穂積

東京工業大学大学院 情報理工学研究科

1 はじめに

自然言語理解に関するこれまでの研究には機械翻訳や対話システムなどのように自然言語同士の間の関係を探る研究が多い。一方、人間の自然言語理解能力は単独で存在しうるものではなく、自然言語を発した人間の周囲の状況や行動などと密接な関係にある。したがって自然言語と行動や状況などとの間の関係を探る研究もまた自然言語理解の研究を行なう上で非常に有意義である [3]。しかしながら自然言語の理解とそれに伴う人間やロボットの行動との関係を解明しようとする試みは十分に行なわれていない。

そこで我々は計算機上に作られた仮想空間の中で自然言語を理解して行動するソフトウェアロボット（アニメテッドエージェント）の作成を通じて、自然言語理解と人間の行動との間の関係を探る研究を行なっている。この研究ではアニメテッドエージェントの行動を観測し我々人間の行動と比較することによって言語理解などのメカニズムを解明するとともに、その結果を自然言語理解モデルや知識表現に反映させることを最終的な目的としている。また本研究をで得られた技術をアニメーションの作成支援やゲームなどに応用することも検討する。

本論文では我々が実装を行なっているアニメテッドエージェントシステムプロジェクトの概要について述べる。またこのシステムの構成や必要となる要素技術について述べる。

2 概要

計算機上にアニメテッドエージェントを構築する試みは数多く行なわれている。しかしながら、その多くはCGの作成やヒューマンインタフェースの構築を主たる目的としており、人工知能技術の検証に主眼を置いたシステムの研究はあまり多いとはいえない。特に自然言語理解のメカニズムの探求までも目的としている研究は [1, 2] などごくわずかである。そこで、我々はプランニングなどの推論技術だけでなく、自然言語理解までも機能として取り込んだアニメテッドエージェントシステムの構築を目指している。

An Animated Agent Who Understands Natural Language Instructions,
Taizan Suzuki, Jun Kamino, Yusuke Shinyama, Hidehisa Akiyama, Takenobu Tokunaga, Hozumi Tanaka
Department of Computer Science, Graduate School of Information Science and Engineering, Tokyo Institute of Technology,
2-12-1 Ookayama Meguro-ku, Tokyo, 152 JAPAN

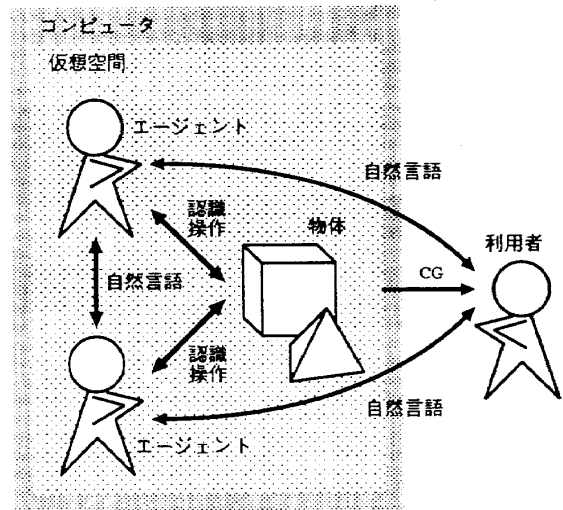


図 1: アニメテッドエージェントシステムの概観

我々が目標とするアニメテッドエージェントシステムの概観は図 1 に示す通りである。本システムでは利用者が自然言語で指令を与えることを前提としている。個々のエージェントは利用者が与えた指令に従って仮想空間上でリアルタイムに行動し、その結果は 3 次元 CG として描画される。また利用者の指令の内容によっては自然言語を利用して利用者との対話を行なう。

3 エージェントシステム

3.1 構成

本研究で目標としているエージェントシステムの構成は図 2 に示す通りである。個々のエージェントは音声認識、言語処理、意味理解を経て利用者の指令の意味を推定する。そして仮想空間の状況などを考慮した上で利用者の指令を実行するために必要なプランを作成する。エージェントはプランに従って仮想空間上で行動し、その結果を 3 次元 CG による動画や自然言語での発話を通じて利用者にフィードバックする。また行動のパターンなどを学習することで新たな言語知識や概念、プランニングの知識を獲得する。エージェントにはロボットだけではなく表示画面を撮影しているカメラなども含まれる。

3.2 要素技術

このようなシステムを実現するには非常に多くの要素技術が必要となる。本節ではこのプロジェクトで必要となる個々の要素技術について述べる。

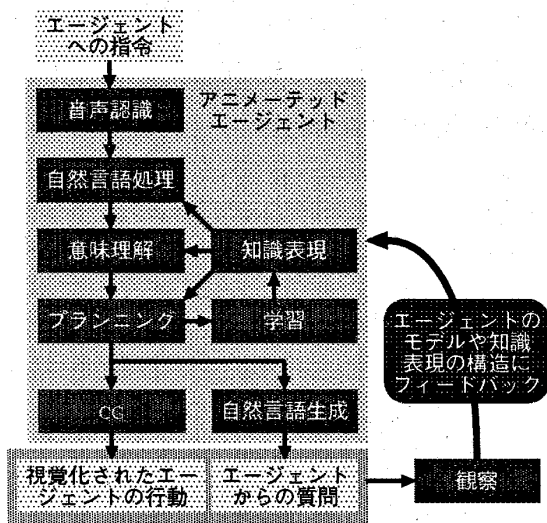


図 2: アニメテッドエージェントの構成

音声認識 本研究ではエージェントが人間の発話に応じてリアルタイムに行動することを想定している。したがって応答時間がある程度以上短く実時間性が高い技術が必要である。このためには入力音声に対して逐次的な音声認識が可能な手法が必要である。

自然言語処理 自然言語処理モジュールでは主に形態素解析、構文解析、意味理解を行なうことになる。エージェントに対する人間の指令は非文や断片的な発話で与えられる可能性が非常に高いので、自然言語処理モジュール、とりわけ構文解析部は非文や断片的な文の入力に対して頑健であることが望まれる。またこのモジュールでは解決できない曖昧性は解消せずに全て出力する。

意味理解部では自然言語処理で得られた構文情報に対して、文脈に依存した曖昧性の解消や補完などを行なう。具体的には係り受けの曖昧性の解消、語義の曖昧性の解消、代名詞の照応、省略表現の補完、断片的な発話の補完などである。さらに個々の詞句からなる表層的な情報とエージェントが保有する概念的な知識との対応付けを行なう。モジュールでは全ての曖昧性を解消し意味を一意に特定する必要があるので、スコアなどで個々の曖昧性の間のもっともらしさを相対的に評価できる手法が必要である。また本研究では主としてエージェントの行動の可視化を行なうので、これまで十分な研究対象とされてこなかったヴェイグネスの問題も扱う必要がある。

プランニング まず、意味理解で得られた利用者の発話の意味解釈からゴールを推測する。そして推測されたゴールを展開しゴールを実現するためのプランを獲得する。そして現在進行しているプランと新たに獲得されたプランとの依存関係を調べ、それぞれのプランのスケジュールを組み直す。利用者の発話の中にプランのスケジュールを左右する指令が含まれている場合はその指令を優先

的に考慮する必要がある。また記号的な処理以外にそれぞれの実際の動作の物理量などの量的なプランニングも行なえる必要がある。

自然言語生成 曖昧性が解消できなかった場合や発話の意味の解釈に失敗した場合などに、利用者に対する質問の生成を行なう。また利用者がエージェントに対して質問をした際には質問に対する返答を生成する。エージェントの発話の内容はプランニングで既に決定されているので、言語生成モジュールでは表層の生成が主になる。

学習 学習は様々なモジュールで必要となる技術である。物体や動作の別称や単語間の概念関係、ゴールを実現するためのプラン、「ちょっと」といった質的表現と具体的な量的表現との関係など学習を必要とする部分は非常に多い。学習の動機の多くは利用者の指示やシステムの質問などによって明示的に得られるので、必要に応じてプランニングの段階で学習を行なうというプランを生成しそのプランを実行することで知識への反映を行なう。**コンピュータグラフィックス** コンピュータグラフィックスモジュールでは仮想空間上のエージェントや物体の動作などを3次元CGで描画する。またエージェントや物体の視覚情報だけではなく仮想空間を支配する物理的な制約の記述やその解消も行なう必要がある。

知識表現 このアニメテッドエージェントシステムでは物体の文章表現と機能記述などのように、個々のモジュールで必要となる知識情報の間に密接な関係がある場合が多い。モジュール間での知識表現の共有をいかにして実現するかが重要な問題となる。

4 今後の予定

今後は、まず意味理解やプランニングなどのモジュールで必要となる知識情報を分析し、共通して利用できる知識の有無や相互の知識間の関係の分析を行なう。その結果に基づいて個々の要素技術で必要となる知識を融合し知識の共有を可能とするような知識表現の研究を行なう。またそれと並行して、現存する技術を応用して個々のモジュールを作成すると共に、モジュール間に流れる情報の検討を行なった上でそれぞれのモジュールを有機的に結合し、第一段階のアニメテッドエージェントシステムの構築を行なう。

参考文献

- [1] N. Badler et al. Intentions and expectations in animating instructions: the AnimNL project. In *Intentions in Animation and Action*. Univ. of Pennsylvania, 1993.
- [2] O. Hasegawa et al. Active agent oriented multimodal interface system. In *Proc. of the 14th IJCAI*, pp. 82-87, 1995.
- [3] T. Winograd. *Understanding Natural Language*. New York: Academic Press, 1972.