

Title	確率文脈自由文法を持つエージェント間での言語共有
Author(s)	武田, 稔也
Citation	
Issue Date	2000-03
Type	Thesis or Dissertation
Text version	author
URL	<a href="http://hdl.handle.net/10119/1358">http://hdl.handle.net/10119/1358</a>
Rights	
Description	Supervisor:東条 敏, 情報科学研究科, 修士

# 確率文脈自由文法を用いたエージェント間での 言語共有

武田 稔也

北陸先端科学技術大学院大学 情報科学研究科

平成 12 年 2 月 15 日

キーワード: マルチエージェント, 言語共有, 確率文脈自由文法, 帰納推論, 分散プログラミング.

言語獲得のモデル化は, 認知科学における大きなテーマであり, 原型言語から, 進化, 適応により, 普段我々が使用する真の言語を獲得する過程においては自動化というスキルが存在する. 本論文ではマルチエージェント環境での確率文脈自由文法 (Probabilistic Context-Free Grammar, PCFG) の文法規則の共有化を自動化のプロセスの一部と考える. ここでは, 各文法規則に確率を割り振り, 会話における有用な規則を重視することにより, エージェント間での文法の共有化を目指す.

会話の相互理解について考察した場合, エージェント間で有用な会話が行われたと主張するためには, 語の概念や発話目的などの設定を必要とする. つまり, 与えられた問題に対して, 各エージェントが目的達成のための会話による協調動作の発現により主張されるべきである. しかし, 文脈自由文法のような階層構造において, 意味の表示や意思・目的の表現を十分に準備することは現実的に困難であり, 本研究では語に対する 意味や目的などは付加していない.

しかし, 最低限のコミュニケーションの枠組で, ある種の規則性の発現は十分な意義を持つものと考えられる. 本研究では, 語と語の対応関係に着目した. 一般的な確率文脈自由文法では, 各文法規則の選択確率は, 任意の導出過程において一定である.

本研究では, 導出過程に依存した選択確率を各文法規則に与えるようなモデルを提案する. これにより, 同じ文法規則でも, その使われ方に応じた確率のチューニングを行うことにより, 例えば日常言語における主語として使用される語 (名詞) と目的語として使用される語 (動詞) の対応関係のような構造的意味を持たせることができる.

実験上のシステムとして、本研究では分散処理環境でのモデルのシミュレーションを行う。分散処理環境とは各エージェントが独立して処理を行い、各エージェント間で同期をとることにより、並行して複数のプロセスを実行できる処理環境である。

ここで、“並列”“分散”は処理の形態である。これに対し、“協調”は処理の目的または方式をさす。つまり、マルチエージェント環境での並列/分散処理実験は、並列/分散環境において動作するエージェントの協調方式という位置付けにある。また、本研究では各プロセッサにそれぞれエージェントを割り当てた大規模シミュレーションを分散処理環境で行い、マルチエージェント環境における大規模システムを提案する。

このような環境のもと、各エージェントが2進数のビット列を入力データとして、与えられたデータから学習を行う対話的学習モデルを作成した。エージェントは入力されたビット列を形成する文法構造を、各エージェントが持つ確率文脈自由文法の文法規則から解析し、得られた解析結果から文法規則の選択確率のチューニングを行う。そして、文法の選択確率を自己組織化し、これをもとに発話を行い他のエージェントにフィードバックする繰り返し最適化モデルを実装した。

選択確率の自己組織化において、次の2つアプローチにより各エージェント間での一般化を行った。

- 文法規則の選択確率に閾値を設定し、閾値以下の文法規則を淘汰する。
- 帰納推論により正例と負例を与えることにより一般化を行う。  
また正例の与え方としては、教師エージェントによる一般の教師つき学習を採用した。

エージェント間のコミュニケーションモデルにおいて、正例と負例を的確に判断することは難しい。これは、先に述べた会話の相互理解の問題も関係しているが、本質的に会話における絶対的な判断基準が存在しないという理由が考えられる。現実の社会においても、文法的にふさわしくない表現が一般に使用されている例は数多く存在する。集団社会においてある事柄が一般化する背景には、その事柄を支持する集団の規模や、エージェントのもつ影響力に左右される。この観点において本研究では以下の結果を得た。

- 実験では、それぞれのパラメータの設定を変えることにより、得られる結果にアルゴリズムの性質を反映した変化が確認された。
- 各エージェントは、割り当てられたプロセッサの処理速度に依存した発話を行うため、発話数に個体差が出現した。これにより現実社会の現象を反映したモデルになった。

今後の課題として、本研究で提案したモデルを認知科学的に説明可能なモデルに改善していくこと、そして、モデルに対して実際の言語データから統計的検証を考えていく必要がある。