

モジュールの学習とモジュール組換え計算による見立て遊びの成立過程のモデル化

Modeling the Development of Pretend Play as a Computational Process of Module Recombination and Learning

坂戸 達陽^{*1} 岡 夏樹^{*1} 尾関 基行^{*1} 大森 隆司^{*2} 長井 隆行^{*3}
Tatsuya Sakato Natsuki Oka Motoyuki Ozeki Takashi Omori Takayuki Nagai

^{*1}京都工芸繊維大学 大学院工芸科学研究科 ^{*2}玉川大学 工学部
Graduate School of Science and Technology, Kyoto Institute of Technology School of Engineering, Tamagawa University
^{*3}電気通信大学 大学院情報理工学研究科
Graduate School of Informatics and Engineering, The University of Electro-Communications

Interaction occurred in playing scene is important for child development. Complex interactions that occur in play scene are important in the development of children. Therefore, the proposal of the model to explain the child's behavior in the play scene is significant. In this paper, we focus on the pretend play, and propose a model to reproduce the development process by the learnable module recombination. We discuss pretend play in the proposed model, and describe the future challenges.

1. はじめに

子どもと養育者が一緒に遊ぶ場面では、言語コミュニケーションや共同注意、協調作業などの複雑なインタラクションが発生する。このようなインタラクションは子どもの発達において重要であると考えられており、遊び場面における子どもの振る舞いを説明するモデルは、(i) 遊びを通した子どもの発達を説明する、(ii) 大人-子どもインタラクションにおける子どもモデルを提案することで、子どもとインタラクションを行うロボットが内部に持つ子どもモデルの設計に寄与する、などの意義がある。

我々は、子どもの遊びの中でも、特に見立て遊びに注目し、その成立過程を再現することを目指している。見立て遊びには、環境を認識するだけでなく、認識した情報を高度に処理する能力が必要であるため、見立て遊びの成立を説明するモデルを提案することは、子どもの発達を説明する上でも重要である。

見立て遊びは1歳半ごろから出現する。見立て遊びは、行動の対象となる物体の表象を1次的表象、見立ての対象となる表象を2次的表象とすることによって成立すると考えられている[久崎 03][志波 09]。これまでの研究で我々は、エージェントのワーキングメモリ内で行動の対象になっている物体の表象を1次的表象、その行動を生成する方策モジュールが対象としている表象を2次的表象として見立て遊びが成立するモデルを提案し、見立てが成立するためのモジュール組換え型モデルによる基本的な枠組みを示した[坂戸 15]。

子どもと養育者が一緒に遊ぶ場面の一例として、互いが互いの意図を読み取りながらインタラクションを行う場面が考えられる。これは見立てを含む場面でも同様である。見立てを含む場面では、見立てを含まない場面におけるインタラクションに必要な能力に加え、物体が何に見立てられているかということを理解する能力が必要である。子どもが見立てを行ったときそれが何を意味しているのかは必ず伝わるわけではない。本来の物体とあまりにもかけ離れた物体で見立てた場合などは当然伝わらない。先行研究における提案モデル[坂戸 15]では、子

もエージェントはその見立てが相手に伝わるかどうかは考慮していなかった。本稿では、子どもエージェントが見立て遊びを成立させるだけでなく、相手に伝わる見立てを養育者とのインタラクションによって獲得するモデルを提案する。相手に伝わる見立てを獲得するために、我々は子どもエージェントに物体のクラス間の見立てに関する類似度をもたせ、子どもエージェントの行動に対する養育者の反応によってこれを学習させる。評価実験の結果を基に提案モデルにおける見立て遊びについて考察し、今後の課題について述べる。

2. エージェントの構成

本研究では、子どもと養育者が積み木やミニカー、ボールなどで遊ぶ場面を想定し、モジュール組換え型モデル[坂本 12][岡 14][坂戸 15]を用いて、子どもモデルをエージェントとして実装する。エージェントは図1のように構成される。各モジュールの動作を以下に示す。

2.1 状態メモリ

状態メモリは、各モジュールから出力される情報の内、状態価値の計算に用いられる情報を格納する。状態メモリには4つの情報を格納することができる。状態メモリ内で参照された情報はタイムスタンプが更新され、容量以上の情報を格納しようとした場合はタイムスタンプの古いものから上書きされる。

2.2 行動メモリ

行動メモリは、各モジュールから出力される情報のうち、行動選択に用いられる情報を格納する。具体的には、各方策モジュールからのゲートが開くと、(intention, 意図, 行動価値, パラメータ,...)の単位で情報を格納する。行動メモリに新しい情報が入ると、既にメモリ内に存在する情報を含めて行動価値の高い順に4つの情報が格納され、残りの情報は破棄される。行動モジュールへのゲートが開くと、行動モジュールに情報が送られ、モジュール内の情報は破棄される。

2.3 物体認識モジュール

物体認識モジュールは、(belief, クラス, オブジェクト, 状態)の単位で外部からの情報を取得する。取得する順番はラン

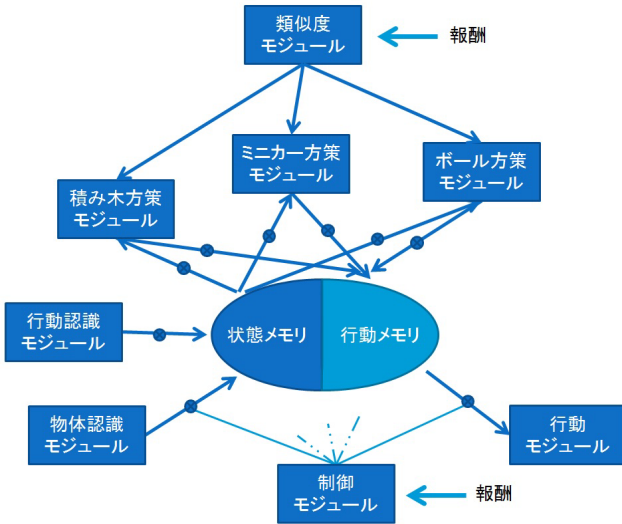


図 1: モジュール組換え型モデルによるエージェントの構成

ダムに決定する。状態メモリへのゲートが開くと、取得した情報が状態メモリに送られ、モジュール内の情報は破棄される。

2.4 行動認識モジュール

行動認識モジュールは、(intention, 意図, パラメータ,...) の単位で養育者の行動を認識する。認識した行動は、モジュールの切換えの各ステップにおいて、切換えのための状態価値を評価する前に状態メモリに送られ、モジュール内の情報は破棄される。

2.5 積み木方策モジュール

積み木方策モジュールは、状態メモリからのゲートが開くと、状態メモリ内に存在する (belief, クラス, オブジェクト, 状態) に対して、積み木遊び (intention, moveBlock, 行動価値, オブジェクト) を生成する。行動価値 Q は、状態メモリ内における情報の新しさに基づき (1) のように決定する。

$$Q := \gamma_p^{x-1} \quad (1)$$

x は行動の基となる情報の新しさの状態メモリ内での順位、 γ_p は行動価値の割引率である。生成される各行動の行動価値は、行動の対象となるオブジェクトのクラスの積み木との類似度に応じて割引される。行動メモリへのゲートが開くと、生成された各行動が行動メモリに送られ、モジュール内の情報は破棄される。

2.6 ミニカー方策モジュール

ミニカー方策モジュールは、状態メモリからのゲートが開くと、状態メモリ内に存在する (belief, クラス, オブジェクト, 状態) に対して、ミニカー遊び (intention, moveCar, 行動価値, オブジェクト) を生成する。行動価値は、状態メモリ内における情報の新しさに基づき決定し (1)、オブジェクトのクラスのミニカーとの類似度に応じて割引される。行動メモリへのゲートが開くと、生成された各行動が行動メモリに送られ、モジュール内の情報は破棄される。

2.7 ボール方策モジュール

ボール方策モジュールは、状態メモリからのゲートが開くと、状態メモリ内に存在する (belief, クラス, オブジェクト, 状態)

に対して、ボール遊び (intention, moveBall, 行動価値, オブジェクト) を生成する。行動価値は、状態メモリ内における情報の新しさに基づき決定し (1)、オブジェクトのクラスのボールとの類似度に応じて割引される。行動メモリへのゲートが開くと、生成された各行動が行動メモリに送られ、モジュール内の情報は破棄される。

2.8 行動モジュール

行動モジュールは、行動メモリからのゲートが開くと、行動メモリ内に存在する (intention, 意図, 行動価値, パラメータ,...) の 1 つを実行する。実行する行動は、行動価値に基づきソフトマックス法 (2) を用いて決定する。

$$\pi(s, a) = \frac{\exp(Q(s, a)/T)}{\sum_{p \in A} \exp(Q(s, p)/T)} \quad (2)$$

2.9 制御モジュール

制御モジュールは、ゲートの開閉を制御する。制御モジュールは、状態メモリ内のタイムスタンプの最も新しい 2 つの情報を状態、どのゲートを開くかを行動とする Q 学習 [Sutton 98] によって学習する (3)。開くゲートの選択は、ソフトマックス法 (2) によって行う。区別のため、制御モジュールの温度パラメータは τ とする。

$$Q(s_t, a) \leftarrow Q(s_t, a) + \alpha(r_{t+1} + \gamma \max_p Q(s_{t+1}, p) - Q(s_t, a)) \quad (3)$$

2.10 類似度モジュール

本稿の提案モデルでは、子どもエージェントは物体のクラス間に見立てに関する類似度も持っている。類似度は各方策モジュールにおいて行動が計画される際の行動価値の割引に用いられる。類似度は、子どもエージェントが見立てを行ったときに養育者が本来の物体で真似した場合、子どもエージェントが本来の物体で行動を行ったときに養育者が見立てで真似した場合に増加し、子どもエージェントが見立てを行ったときにそのクラスの物体で異なる行動をとった場合、見立てで行動することに失敗した場合に減少する。クラス x と y の間の類似度 $\text{sim}(x, y)$ の更新式は、

$$\text{sim}(x, y) \leftarrow \text{sim}(x, y) + \alpha_s(r - \text{sim}(x, y)) \quad (4)$$

とする。ここで、 α_s は学習率、 r は報酬である。報酬 r は値を増加させるときは 1.0、減少させるときは 0.0、値を変化させない場合は $\text{sim}(x, y)$ とする。

3. 見立て遊びの成立

見立て遊びは、行動を生成する方策モジュールが対象とする表象 (e.g. ミニカー方策モジュールに対するミニカー) と、生成された行動が対象とする表象 (e.g. 積み木を左右に動かす行動における積み木) が異なるときに、後者を 1 次的表象、前者を 2 次的表象として成立する。本稿の提案モデルでは、本来の物体での行動と見立て行動を、子どもエージェント自身が区別することはない。

4. 実験設定

図 2 のような、子どもと養育者が対面で遊ぶ場面を想定している。環境中には子どもエージェント、と操作可能な養育者エージェント、2 つの積み木、1 つのミニカー、そして 1 つの

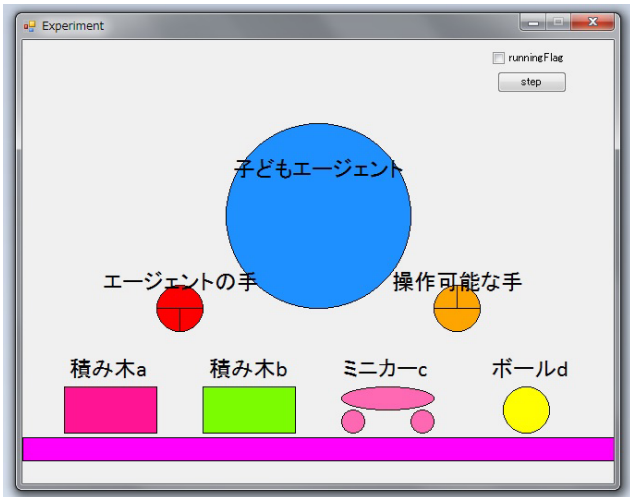


図 2: 実験環境の説明

ボールが存在する。子どもエージェントは、積み木を持ち上げランダムな位置に積む積み木遊び、ミニカーを持ち上げて左右に動かして元の位置に置くミニカー遊び、ボールを真上に放り投げて元の位置に置くボール遊びを行うことができる。積み木、ミニカーの上には積み木あるいはミニカーを積むことができる。これに対し、ボールの上には何も積むことはできない。また、ボールを何かの上に積むこともできない。養育者エージェントは、これらの遊びに加えて、子どもエージェントを褒めることができる。

実験が始まると、まず養育者が、遊ぶ、褒める、そのまま進めるのうちのいずれかを選択する。遊ぶ場合は現在の状態で可能な遊びのみを選択できる。いずれかを選択すると、子どもエージェントがモジュール組換えを行う。子どもエージェントのモジュール組換えは、子どもエージェントが何か行動するまで続く。子どもエージェントが何か行動を行うと、養育者はまた上記のいずれかを選択する。養育者の行動選択と子どもエージェントのモジュール組換えを繰り返すことによって実験は進行する。

今回の実験では、養育者はプログラムで構成される。養育者はモジュール組換えの結果として子どもエージェントが行動を成功させると、褒めるまたは見立てを含む形で子どもエージェントの行動を真似る。ただし、積み木またはミニカーでボール遊びを見立てで行った場合、ボールで積み木またはミニカー遊びを見立てで行った場合には、養育者は子どもエージェントの意図を理解できず、何もせずに子どもエージェントの行動を観察し続ける。

養育者が褒めたとき、あるいは養育者が子どもエージェントの行動を真似たときには 1.0 の報酬が与えられる。行動は、対象となる物体の上に何か載っていた場合は失敗する。行動を行わなかった場合、行動が失敗した場合には報酬が与えられない。

なお、Q 学習の学習率 α は 0.1、割引率 γ は 0.9、ゲート選択の際のソフトマックス法における温度パラメータ τ は 0.2、行動選択の際のソフトマックス法における温度パラメータ T は 0.2 とする、行動価値の割引率 γ_p は 0.9、見立てに関する類似度の学習率 α_s は 0.1 とする。

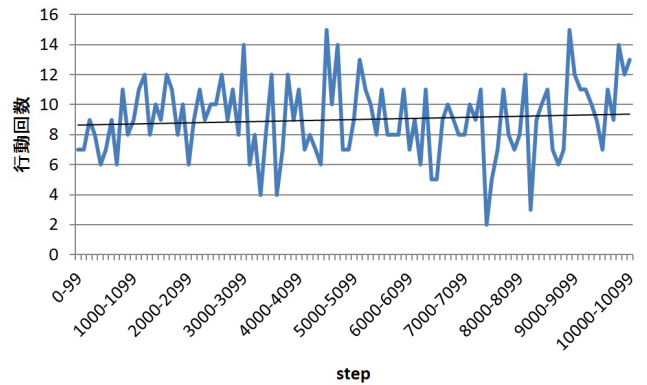


図 3: 100 回の切換えごとの行動回数

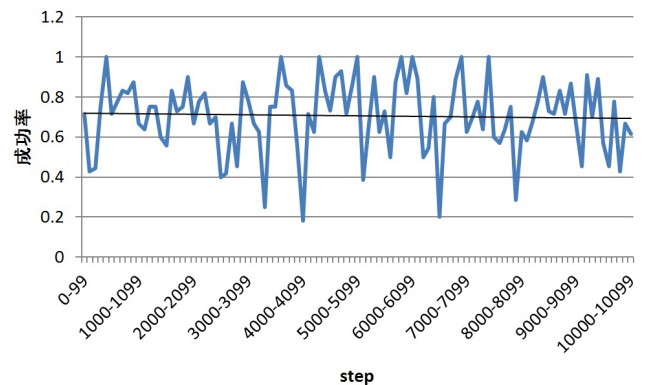


図 4: 100 回の切換えごとの行動の成功率

5. 結果および考察

図 3 に 100 回の切換えごとの行動回数、図 4 に 100 回の切換えごとの行動の成功率、図 5 に 100 回の切換えごとの見立てに関する類似度の変化を示す。図 3 および図 4 から、適切なモジュール切換えの学習については進んでいないことが読み取れる。一方、図 5 より、見立てに関する類似度については、積み木-ミニカー間の類似度が高くなっており、学習が進んでいることがわかる。積み木-ミニカー間の見立ては行動自体が失敗しない限りは養育者に正しく伝わるので、このような結果になったと考えられる。一方、積み木-ボール間、ミニカー-ボール間の見立ては、行動自体が失敗しなくても、養育者に正しく意図が伝わらないため、類似度が高くならなかったと考えられる。

今回の実験では適切なモジュール切換えの学習は進まなかったが、これは恐らく、方策モジュールでの行動生成の際に、各行動の行動価値を、状態メモリ内の行動の対象となる物体に関する情報の新しさのみで決定しており、行動の対象となる物体の状態を考慮していなかったことが一因ではないかと考える。各行動の行動価値の決定には、情報メモリ内の情報の新しさだけでなく、行動の対象となる物体の状態を考慮した計算が必要であると考えられる。

6. おわりに

本稿では、モジュール組換え型モデルを用いて見立て遊びの成立およびインタラクションによる伝わる見立ての獲得を試み

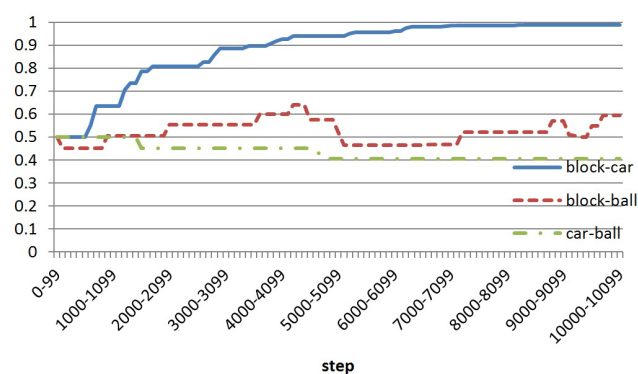


図 5: 100 回の切換えごとの見立てに関する類似度の変化

た。実験の結果、モジュール組換えの学習はうまく進まなかったが、伝わる見立てを行うための、物体のクラス間の見立てに関する類似度については学習することができた。モジュール組換えの学習がうまくいかなかった要因としては、方策モジュールにおける各行動の行動価値を、状態メモリ内の行動の対象となる物体に関する情報の新しさのみで決定しており、行動の対象となる物体の状態を考慮していなかったことが挙げられる。今後の課題としては、行動の対象となる物体の状態を考慮した行動価値の決定、各モジュールの同時学習、注意モデルの実装、飽きモデルの導入、メタ認知のモデル化、より自然なインタラクションの実現などが挙げられる。

参考文献

- [Sutton 98] Sutton, R. and Barto, A.: *Reinforcement Learning: An Introduction*, MIT Press, Cambridge, MA (1998)
- [岡 14] 岡 夏樹, 呉 霞, 神山 薫, 深田 智, 尾関 基行: 機能語や抽象語の意味表現とその獲得—モジュール組換え演算に基づくモデル化の試み—, 信学技報, Vol. 113, No. 426, pp. 101–106 (2014)
- [久崎 03] 久崎 孝浩: 生後 2 年目における認知発達—表象機能という視点からの考察—, 九州大学心理学研究, Vol. 4, pp. 37–55 (2003)
- [坂戸 15] 坂戸 達陽, 尾関 基行, 大森 隆司, 長井 隆行, 岡 夏樹: 見立て遊びの成立過程のモジュール組換え計算によるモデル化, 第 77 回情報処理学会全国大会論文集, 3D-04 (2015)
- [坂本 12] 坂本 裕太, 坂戸 達陽, 尾関 基行, 岡 夏樹: モジュール組換え型モデルにおけるモジュールの学習とモジュール組換え系列の学習, 第 26 回人工知能学会全国大会論文集, 3B2-R-2-6 (2012)
- [志波 09] 志波 泰子: 2 歳児は誤信念を理解するだろうか: Perner と Leslie の論争を再考する, 京都大学大学院教育学研究科紀要, Vol. 55, pp. 75–87 (2009)