C-06

モジュールの切替えとモジュール自体の同時学習による エージェントの内的表象の意味獲得

Acquisition of Meanings of Symbols by Simultaneous Learning of Functions of Modules and Their Free Combinations

坂戸 達陽†　　岡 夏樹†　　大森 隆司††　　長井 隆行‡
Tatsuya Sakato Natsuki Oka Takashi Omori Takayuki Nagai

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．遊び場面における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，このような場面において知識や行動を学習し，適切なインタラクションを行うことができるエージェントのモデルは，子どもの発達に関する研究に有用な知見を与えることができると考えられる．そこで本研究では，遊び場面における学習エージェントを提案する．本研究では，子どもの遊びの中でも，特に見立てを含む真似遊びに注目し，そのような遊び場面におけるエージェントの知識や行動の獲得，他者とのインタラクションについて検証する．本研究では基本的な学習環境として，子どもと養育者が対面で一緒に遊ぶ場面を想定し，学習環境において，学習エージェントを子ども，インタラクションを行う他者を養育者と見なす.

1.1　見立て遊び

見立ては 1 歳半ごろから 4 歳ごろにかけて獲得され，見立て遊びなどの形で出現する．見立ては行動の対象となる物体をそれ以外のものとして扱うことで成立すると考えられている[1][2][3]．このとき，物体の本来の表象は 1 次表象と呼ばれ，見立てられる表象は 2 次表象と呼ばれる．見立て遊びの例としては，バナナを電話に見立てて遊ぶことが挙げられる．この場合，行動の対象であるバナナの表象を 1 次表象，見立ての対象である電話の表象を 2 次表象として見立て遊びが成立する．このように，見立てには事物を抽象的に扱う高度な認知能力が必要であり，見立てを含む行動の認識，生成を学習するモデルを提案することは，子どもの発達に関する研究に有用な知見を与えることができる．また，人と人との対話場面において，人は事物を直接的に表現するだけでなく，見立てや比喩を用いることも多いため，見立てを含む行動を認識，生成できるモデルは，対話システムの設計においても有用である．

1.2　モジュールの切替えとモジュール自体の同時学習

モジュールの組み合わせによって複雑な問題を解決しようとする研究が多数行われてきた．モジュールの切替え方を限定した場合はモジュール自体の学習と切替え方の学習を同時に行うことができるが，モジュールの切替え方の自由度を上げた場合は，モジュール数に対して組合せ爆発的に切替え方が増え，学習が難しくなる．そこで，モジュール自体はあらかじめ作りこんでおき，切替え方だけを学習する手法や，逆にモジュール自体は学習するが，切替え方は与えておく方法が提案されてきた．こうした中で，岡はモジュールの組み合わせ方を限定しないことを特徴とするモデル[4]を提案し，坂本らはこのモデルを用いて，モジュールの機能の学習と，モジュールの切替え系列の同時学習を行うことができることを示した[5]．坂本らは単純な仮想空間上での迷路探索という，比較的単純なタスクを用いて評価実験を行っているが，神山らは，「何色ですか」「なんという形ですか」といった発話の意図に応じて物体の色や形を答えるエージェントの，モジュールの切替えとモジュール自体の同時学習を行っている[6]．また，岡らは，終助詞「よ」「ね」「か」といった機能語や，「色」「形」「大きさ」といった抽象語の意味獲得モデルを提案している[7].

これまでの研究で，モジュールがエージェントの意図を生成する際に，見立てを含んだ意図を生成することができるモデルを提案した[8]．また，見立てが発生した際の，エージェントの意図の他者への伝わりやすさから，物体の種類間の関係性を学習できることも示した[9]．本研究ではさらに，見立てを含む真似遊びにおけるインタラクションを通して，観測した他者の行動の意図を学習するモデルを提案する．本研究でエージェントが扱う意図　は，予め「積み木遊び」「ミニカー遊び」などのラベル付けはされておらず，モジュール間の関係性によって意味が定義される．例えば，真似遊びにおいて，他者のエージェントの積み木を積む行動を観測したエージェントが，その行動の意図として生成した表象から，そのエージェントの積み木を積む行動が生成されたとき，その意図は，認識した行動および生成した行動の関係性によって，「積み木遊び」と定義される．表象の意味を予め定義しないことによって，未知の環境や複雑な環境において適切な表象を，インタラクションを通して獲得することを期待する．

2. 実験環境および学習エージェントの構成

2.1　実験環境

図 1，図 2 のような，子どもと養育者が対面で遊んでいる場面を想定した仮想的な学習環境で評価を行う．今回の実験では，プログラムで構成されたエージェントを養育者とする．学習エージェントは養育者エージェントとのインタラクションによって学習を行う．環境中には，学習エージェント，養育者エージェント，積み木あるいはミニカーが存在する．また，環境中には物体を積むための固定された積み木も存在する．実験は，

1. 環境を初期化する．

†京都工芸繊維大学，Kyoto Institute of Technology
††玉川大学，Tamagawa University
‡電気通信大学，The University of Electro-Communications

2. 養育者が行動し，学習エージェントがその行動を観測する．
3. 環境を再度初期化する．
4. 学習エージェントがモジュール切換えによって行動を生成，実行する．
5. 学習エージェントが実行した行動を養育者が評価し，学習を行う．

という単位を 1 エピソードとして進行する．学習エージェントの構成は 2.2 節で述べる．

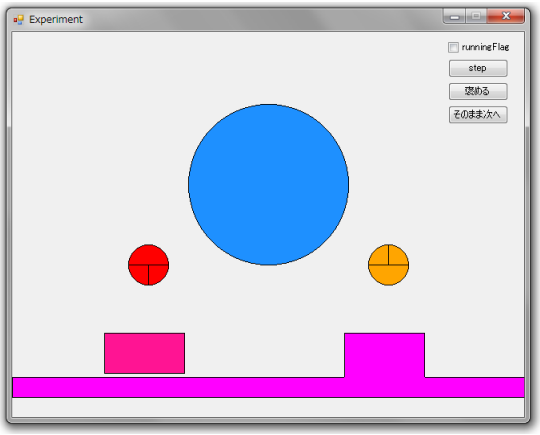

図 1　実験環境（積み木）

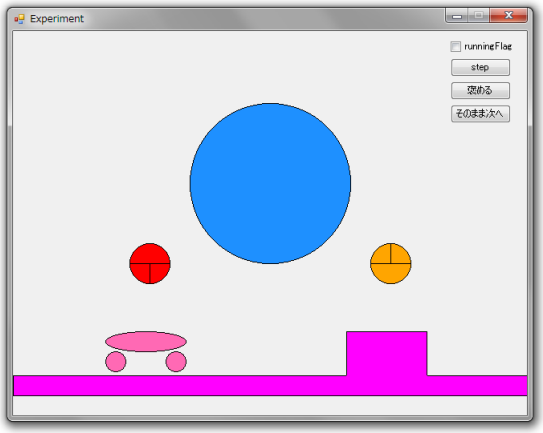

図 2　実験環境（ミニカー）

環境が初期化されると積み木あるいはミニカーがランダムに配置される．養育者エージェントは配置された物体に対して，固定された積み木の上に積む，あるいは左右に動かすという行動を行う．養育者エージェントは積む行動を行っているときには積み木遊びを意図しているとし，同様に，左右に動かす行動を行っているときにはミニカー遊びを意図しているとする．ミニカーを積む，あるいは積み木を左右に動かす行動を行っているとき，養育者エージェントはそれぞれ，ミニカーを積み木に，積み木をミニカーに見立てて行動しているとする．養育者エージェントは学習エージェントが物体を積む，あるいは左右に動かす行動を行っているとき，それぞれ積み木遊び，ミニカー遊びを意図していると認識し，養育者エージェント自身の意図と一致している場合は褒め，一致していない場合は褒めずに次のエピソードに移る．学習エージェントには褒めると 1.0 の報酬が与えられ，褒めずに次のエピソードに移ると報酬は与えられない（報酬0.0で学習する）．

2.2　学習エージェントの構成

本実験における学習エージェントの構成について述べる．学習エージェントの概要を図 3 に示す．本実験では，インタラクションを通して，観測した他者の行動に基づく意図生成，および意図に基づく行動生成が学習できるかどうかを評価するため，他者の行動の意図を生成し，生成した意図に基づいて行動を生成，実行するための最小の構成，すなわち，物体認識モジュール，他者モデルモジュール，行動生成モジュール，実行モジュール，そして制御モジュールおよびワーキングメモリで学習エージェントを構成する．学習エージェントを構成するモジュールのうち，物体認識モジュール，実行モジュールは学習済みとし，学習は，他者モデルモジュール，行動生成モジュールおよび制御モジュールで行う．

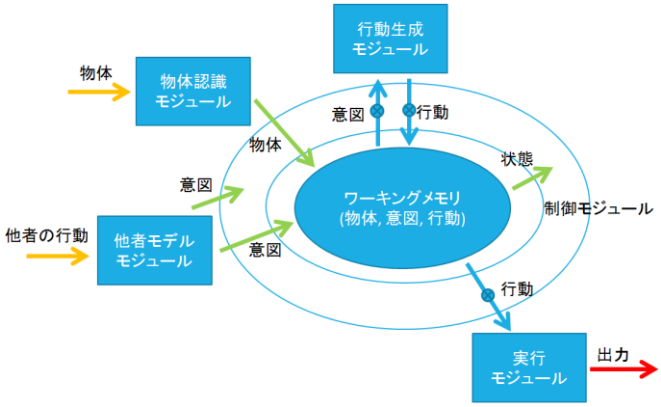

図 3　提案モデル

2.2.1　ワーキングメモリ

ワーキングメモリは，物体，意図，行動の情報をそれぞれ 1 つまで記憶することができる．他のモジュールはゲートを介してワーキングメモリと情報をやり取りする．ゲートは，ボトムアップに，あるいは制御モジュールからの制御を受けて開閉される．

2.2.2　物体認識モジュール

物体認識モジュールは，環境中にある物体を認識し，その種類を特定する．物体認識モジュールは学習エージェントがモジュールの組換えを始める際に，認識した物体の情報をワーキングメモリへ送る．認識できる物体は，積み木，ミニカーの 2 種類とする．

2.2.3　実行モジュール

実行モジュールはワーキングメモリ内に行動の情報が存在すると，その行動を実行する．行動は，実行モジュールの入力ゲートが開かれた際に実行される．実行する行動が積む行動のとき，エージェントは認識している物体を固定された積み木の上に置く．実行する行動が左右に動かす行動のとき，エージェントは認識している物体を左右に動かす．

2.2.4　制御モジュール

制御モジュールは，ワーキングメモリ - 各モジュール間のゲートの開閉を制御する．ただし，物体認識モジュールの出力ゲート，他者モデルモジュールの出力ゲートは制御の対象外とする．これらのゲートは学習エージェントがモジュールの組換えを始める際にボトムアップに開き，物体認識モジュールはワーキングメモリへ，他者モデルモジュールはワーキングメモリと制御モジュールへそれぞれ情報を送る．制御モジュールは，(i)ワーキングメモリ内に物体，意図，行動それぞれの情報が存在するかどうか，(ii)どのゲートが開いているか，(iii)他者モデルモジュールが認識した意図の種類を状態，次にどのゲートを開くかを行動とする Q 学習[10]によって学習する．状態s_tで切替えaを行い，報酬rを獲得した際，行動価値$Q(s_t, a)$は(1)のように更新する．

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

ここで，αは学習率，γは割引率，rは切替えの際に獲得した報酬である．次に開くゲートは，行動価値$Q(s, a)$に基づきソフトマックス法(2)によって決定する．

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

ここで，Tは温度パラメータ，Aは制御モジュールが開くことのできるゲート全体の集合である．

2.2.5　他者モデルモジュール

他者モデルモジュールは，養育者の行動を認識し，その意図を推測する．認識する行動は，(i)積み木を積む，(ii)ミニカーを積む，(iii)積み木を左右に動かす，(iv)ミニカーを左右に動かすの 4 種類とし，そこから推測される意図は，意図 1，意図 2 の 2 種類とする．

他者モデルモジュールでは，推測される各意図の価値が，認識可能な行動ごとに与えられている．各意図の価値はエージェントの獲得報酬によって学習され，認識した行動から養育者の意図を推測するために用いられる．行動aを認識したとき，そこから推測される意図xの価値$Q(a, x)$は，即時報酬のみの Q 学習で，(3)のように更新される．

$$Q(a, x) \leftarrow Q(a, x) + \alpha_s\left(r - Q(a, x)\right) \tag{3}$$

ここで，α_sは学習率，rはモジュールが価値を更新する際に与えられる報酬である．学習は，エージェントの行動が評価された際に，ワーキングメモリ内に他者モデルモジュールが生成した情報が存在する場合に行う．行動aを認識したとき，意図xであると推測するための方策$\pi(a, x)$は，価値$Q(a, x)$に基づき，(2)のTをτに，sをaに，aをxに，そしてAをXに置き換えたものとする．ここで，τは他者モデルモジュールにおける温度パラメータ，Xはモジュールが推測する意図全体の集合である．他者モデルモジュールは学習エージェントがモジュールの組替えを始める際に，認識した行動の情報をワーキングメモリへ送る．

2.2.6　行動生成モジュール

行動生成モジュールは，ワーキングメモリ内に存在する意図の情報に応じて行動を生成し，生成した行動の情報をワーキングメモリへ送る．行動生成モジュールで扱う行動は，(i)物体を積む，(ii)物体を左右に動かすの 2 種類とする．意図と行動の対応付けは，即時報酬のみの Q 学習(3)で行う．行動生成モジュールにおける学習率は，α_aとする．学習は，エージェントの行動が評価された際に，ワーキングメモリ内に行動生成モジュールが生成した情報が存在する場合に行う．行動認識モジュールは，入力ゲートが開くと，入力された意図から生成する行動を，学習結果に基づきソフトマックス法(2)によって決定する．ソフトマックス法における温度パラメータは，τ_aとする．生成した行動の情報はモジュール内部で保持される．出力ゲートが開くと，行動生成モジュールは保持している行動の情報をワーキングメモリへ送る．モジュールが行動の情報を保持していないときはワーキングメモリに対する操作は行わない．ワーキングメモリへ情報を送ると，行動生成モジュールは保持していた情報を破棄する．

2.2.7　各パラメータ

各パラメータの値は，制御モジュールにおける学習率αを0.1，割引率γを0.9，温度パラメータTを0.01，他者モデルモジュールにおける学習率α_sを0.1，温度パラメータτを0.2，行動生成モジュールにおける学習率α_aを0.1，温度パラメータτ_aを0.2とする．

3.　結果および考察

3.1　他者モデルモジュール，制御モジュールの同時学習

行動生成モジュールにおける，意図と行動の対応付けを固定した状態での学習実験を行った．すなわち，意図 1 を積み木遊び，意図 2 をミニカー遊びとし，積み木遊びからは物体を積む行動，ミニカー遊びからは物体を左右に動かす行動を生成することとして学習実験を行った．

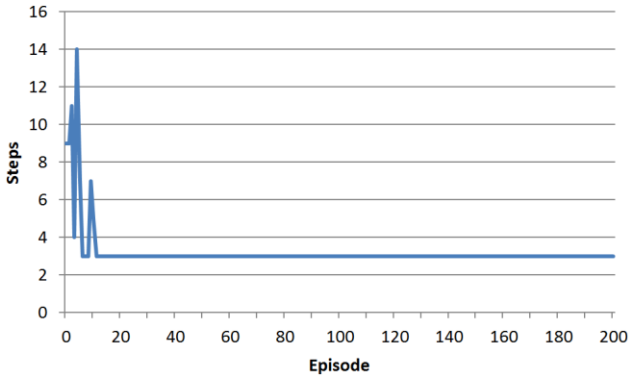

図 4　ゴールまでのステップ数

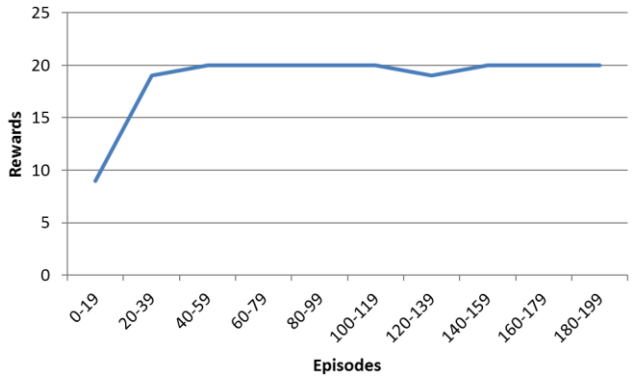

図 5　20 エピソードごとの獲得報酬

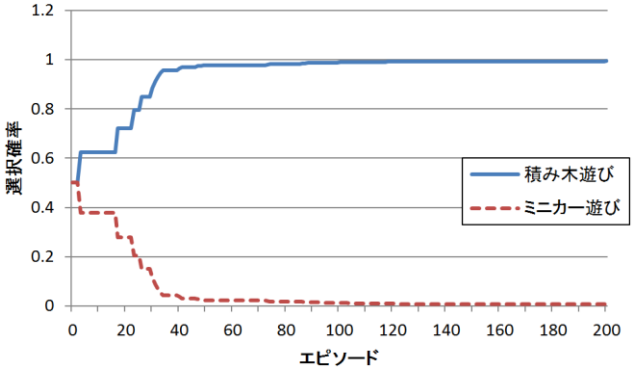

図 6　積み木を積む行動によって推測される意図の変化

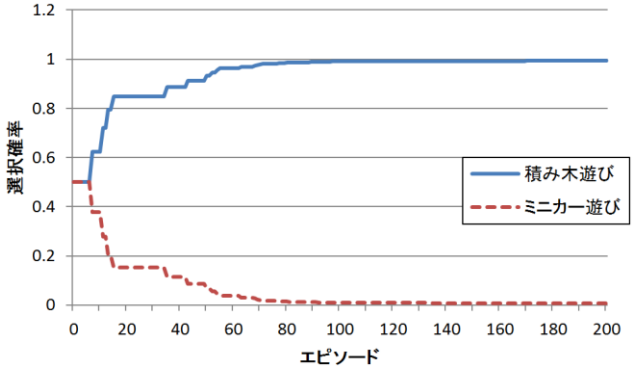

図 7　ミニカーを積む行動によって推測される意図の変化

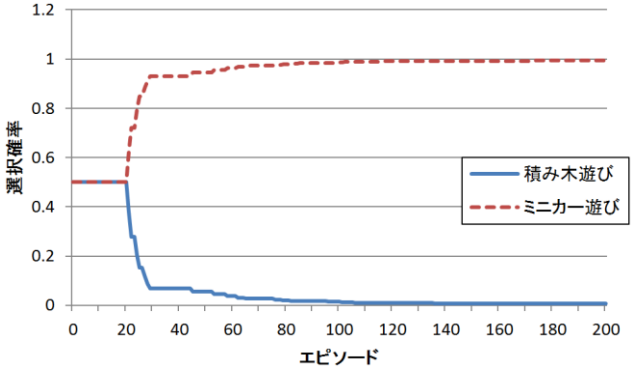

図 8　積み木を左右に動かす行動によって推測される意図の変化

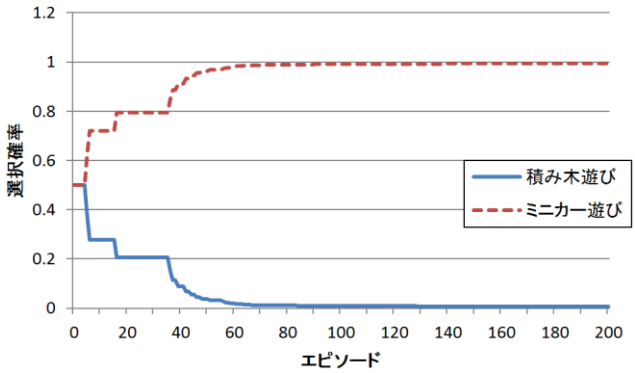

図 9　ミニカーを左右に動かす行動によって推測される意図の変化

図 1 に各エピソードにおいてエージェントが行動を実行するまでにかかったステップ数を示す．図より，エピソードを重ねるにつれてゴールまでのステップ数が減少していることから，モジュールの切替えに関する学習が進んでいることがわかる．図では，切替え数は 3 ステップ以下を示していないが，これは，エージェントが行動を実行するために必要な最低ステップ数が，(i)行動生成モジュールへの入力ゲート，(ii)行動生成モジュールの出力ゲート，(iii)実行モジュールへの入力ゲート，の 3 ステップであるためである．

図 5 に各エピソードにおいてエージェントが獲得した報酬を 20 エピソードごとに合計したものを示す．図より，エピソード初めの方，つまり学習の初期段階に，エージェントが報酬を獲得できないエピソードが多いことがわかる．要因としては，(i)切替え系列の学習が不十分であるため，(ii)意図を理解する能力が不十分であるため，ということが考えられる．モジュールの切替え系列は，エージェントの行動の実行に関わっているが，今回の実験においては，エージェントの行動の実行によってエピソードの終了を判断しており，また，エピソードの終了時以外では報酬を与えられず，与えられる報酬も実行された行動の種類によるもので，行動に何ステップかかったかということは影響していないため，(i)は要因ではないと判断できる．図 6 から図 9 に，観測された行動に対し，その意図として各意図が選択される確率を示す．図 5 と，図 6 から図 9 を比較すると，学習エージェントが報酬を獲得できないエピソードは，他者モデルモジュールにおいて，各行動に対して，特定の意図の選択確率が0.8を超える辺り，つまり，他者モデルモジュールの学習がまだ進んでいない辺りに多いことがわかる．よって，エージェントが報酬を獲得できなかった要因は(ii)意図を理解する能力が不十分であるためであると判断できる．

続いて，それぞれの行動に対する各意図の選択確率の学習結果自体を見てみると，学習が進むにつれて，積み木を積む行動，ミニカーを積む行動に対しては，積み木遊びが，積み木を左右に動かす行動，ミニカーに左右に動かす行動に対しては，ミニカー遊びが対応付けられ，養育者と意図の共有ができたことがわかる．

まとめると，今回の実験設定において，学習エージェントは，制御モジュールによるモジュールの切替えの学習と，他者モデルモジュールによる観測した行動の意図の学習とを，同時に行うことができたと結論付けることができる．

3.2　他者モデルモジュール，行動生成モジュール，制御モジュールの同時学習

続いて，行動生成モジュールにおいても，意図と行動の対応付けを予め固定せずに，インタラクションを通して対応付けを獲得させる学習実験を行った．対応付ける意図の数 N は，N = 2，N = 10，N = 100とした．

図 10 から図 12 に，N = 2，N = 10，N = 100のときそれぞれにおいて，エージェントが各エピソードで行動を実行するまでにかかったステップ数を示す．

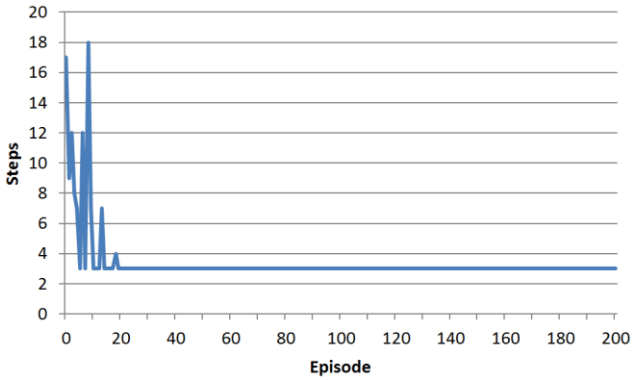

図 10　N = 2におけるゴールまでのステップ数

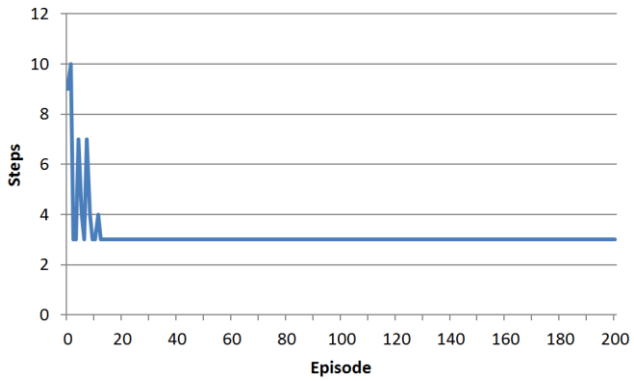

図 11　N = 10におけるゴールまでのステップ数

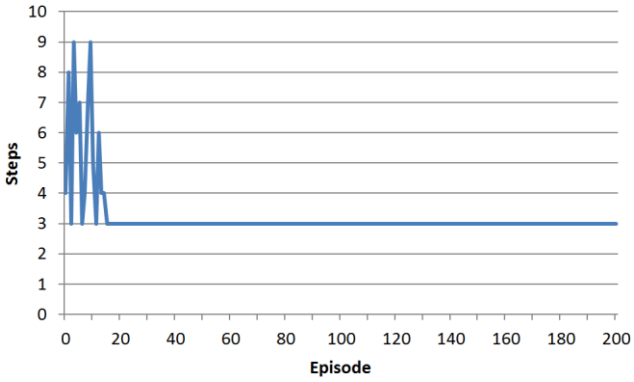

図 12　N = 100におけるゴールまでのステップ数

図よりどのNの値においてもステップ数が収束していることがわかる．収束するまでのエピソード数を見てみると，切替えの収束にかかるエピソード数は，Nの値が増加してもあまり大きな影響を受けていないように見える．

図 13 から図 15 に，N = 2，N = 10，N = 100のときそれぞれにおいて，エージェントが獲得した報酬を 20 エピソードごとに合計したものを示す．また，比較のため，図 16 にそれぞれのグラフを重ねたものを示す．結果を分かりやすく示すため，図 13 から図 16 にはそれぞれ 500，1000，4000，3000 エピソード学習を行った結果を示す．また，図 13 から図 15 で用いたデータと図 16 で用いたデータは同じものである．

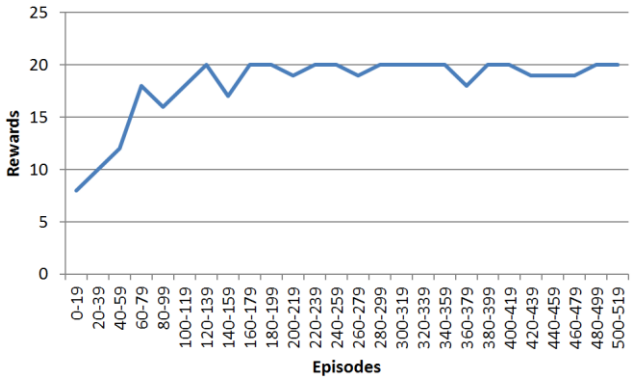

図 13　N = 2における 20 エピソードごとの獲得報酬

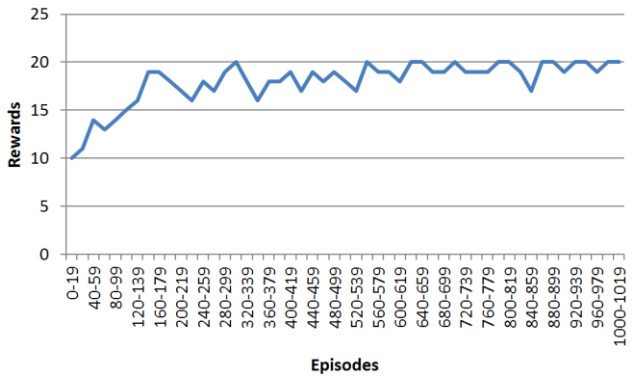

図 14　N = 10における 20 エピソードごとの獲得報酬

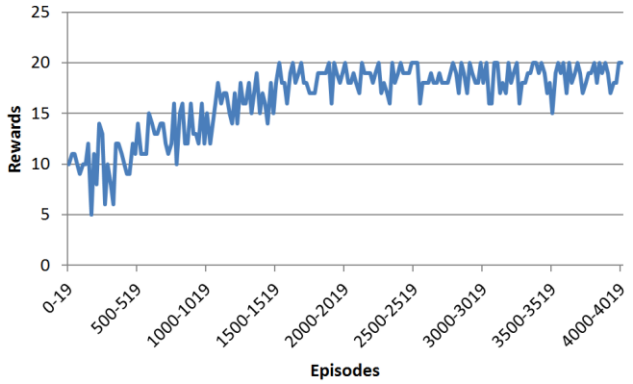

図 15　N = 100における 20 エピソードごとの獲得報酬

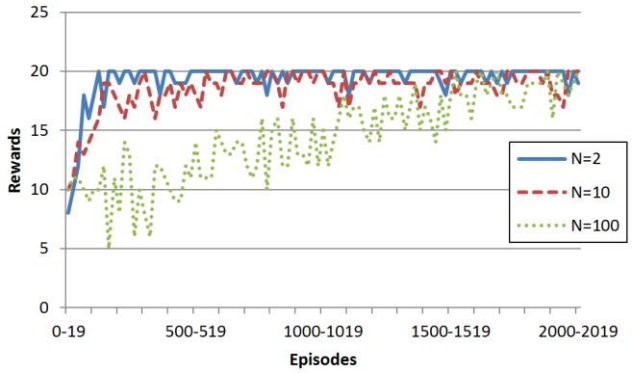

図 16　Nの値による 20 エピソードごとの獲得報酬の違い

図 13 から図 15 より，N = 2，N = 10，N = 100において，それぞれ 80，120，2000 エピソード程度まで学習すれば，20 エピソード中 15 回（75%）以上報酬を獲得できるようになっていることが分かる．また，図 16 より，Nが大きくなるほど学習に必要なエピソード数が増加していることが分かる．

まとめると，今回の実験設定において，学習エージェントは，(i)制御モジュールによるモジュールの切替えの学習，(ii)他者モデルモジュールによる観測した行動の意図の学習，(iii)行動生成モジュールにおける意図と行動の対応付けの学習を，同時に行うことができた．また，切替えの学習は内部で扱う意図の表象の数の影響を受けにくいが，観測した養育者の行動とエージェントの行動との，意図の表象を介した対応付けの学習は，内部で扱う意図の表象の数の影響を大きく受けることがわかった．

4. おわりに

本稿では，見立てを含む真似遊びに注目し，モジュールの切替えとモジュール自体の同時学習によって，観測した他者の行動の意図を学習するモデルを提案した．提案モデルにおいては，エージェント内部で扱う意図は予めラベル付けされておらず，インタラクションによって学習されるモジュール間の関係性によって意味付けられる．本稿では基本的な学習環境として，子どもと養育者が対面で一緒に遊ぶ場面を想定し，学習環境において，学習エージェントを子ども，インタラクションを行う他者を養育者と見なした．評価実験は，(i)他者の行動の意図のみを学習する場合，(ii)他者の行動の意図およびその意図と行動との対応付けの両方を学習する場合で行った．(ii)の場合については，エージェント内部で扱う意図の表象の数Nを$N = 2$，$N = 10$，$N = 100$として実験を行った．

実験の結果，いずれの場合においても，学習エージェントはモジュールの切替えとモジュール自体の同時学習によってエージェント内部の意図の表象を適切に意味付けることができた．また，モジュールの切替えに関しては内部で扱う表象の数に大きな影響は受けず，モジュール自体の学習に関しては内部で扱う表象の数に大きな影響を受けることがわかった．

本稿では，見立てを含む真似遊びに注目し，エージェント自身が主体となる見立てについては取り上げなかったが，それは，自律行動のためのモジュールなど，適切なモジュールを追加することによって可能になると考えている．また，見立て自体に関しても，例えば，本稿や坂戸ら[9]などは他者に見立てが伝わるかどうかでその見立てが適切であるのかを判断していたが，今後は，Zook ら[11]や Magerko ら[12]のように，物体の色や形などの属性を考慮した見立てなども行えるようにしたいと考えている．

その他の課題としては，実世界の情報からの学習，他者の行動の認識とエージェント自身の行動の生成に関するミラーニューロンシステムの実装などが挙げられる．

参考文献

[1] 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.

[2] 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.

[3] 井上 洋平, 幼児期におけるふり行動の発達的研究 ―ふり行動の二重性に関する一考察―, 立命館産業社会論集, Vol. 43, No. 1, pp. 77-93, 2007.

[4] N. Oka, Apparent "free will" caused by representation of module control, No Matter, Never Mind: Proceedings of Toward a Science of Consciousness: Fundamental Approaches, pp. 243-249, 1999.

[5] 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, 2012.

[6] 神山 薫, 深田 智, 尾関 基行, 岡 夏樹, 発話意図に応じたモジュールの切替とモジュール自体の処理の同時学習, HAI シンポジウム 2013, 2013.

[7] 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.

[8] 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, 2015.

[9] 坂戸 達陽, 岡 夏樹, 尾関 基行, 大森 隆司, 長井 隆行, 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, 第 29 回人工知能学会全国大会論文集, 2015.

[10] R.S. Sutton, A.G. Barto, Reinforcement Learning: An Introduction, MIT Press, 1998.

[11] A. Zook, B. Magerko, M. Riedl, Formally Modeling Pretend Object Play, Proceedings of the 8th ACM Conference on Creativity and Cognition, 2011.

[12] B. Magerko, J. Permar, M. Jacob, M. Comerford, J. Smith, An Overview of Computational Co-creative Pretend Play with a Human, Proceedings of the Playful Characters workshop at the Fourteenth Annual Conference on Inteligent Virtual Agents, 2014.