

インタラクションを通じた数の概念の獲得

Acquisition of the Concept of Number Through Human-Agent Interaction

高井 利将 岡 夏樹 早川 博章
Toshimasa TAKAI Natsuki OKA Hirofumi HAYAKAWA

京都工芸繊維大学 大学院工芸科学研究科
Graduate School of Science and Technology, Kyoto Institute of Technology

We developed an agent that voluntarily learns “the stable-order principle” and “the abstraction principle”, two of Gelman and Gallistel’s “five counting principles” which are basic concepts of number. The agent learned them from external rewards given by a teacher, and from internal rewards, motivation to learn and curiosity. The agent successfully learned “the stable-order principle”, i.e. the sequence of natural numbers, and “the abstraction principle” in a limited setting.

1. はじめに

「数の概念」とは数に関する考え方そのものであり、人にとって重要な概念の1つである。これまでになされた研究によって、多くのことが分かってきている。一例として、ゲルマンらの「計数の5原理」があげられる[Gelman 1979]。この原理は幼児期に獲得されるもので、「数の概念」の獲得の初期段階において重要なものである。しかしながら、幼児期の子供の「数の概念」の獲得のモデルはよく分かっていない。

そこで、本研究の目的としては、幼児期における「数の概念」の獲得のモデルを作ることである。そのためにまず、人とのインタラクションを通して学習をすることができるエージェントを開発し、エージェントを子供に見立ててシミュレーションおよび考察をおこなった。またエージェントには内部報酬を定義した。内部報酬とはエージェントがより子供らしく振舞うようにするために導入したものである。これは学習意欲や知的好奇心に相当する機能（子どもがもつ新奇性や親近性への好みに対応）である。

2. 数の概念

「数の概念」とは数に関する考え方そのものである。数に関する機能やその操作は、最も抽象的な思考の次元につながる意味があり、人にとって重要な概念の1つである。人の「数の概念」のレベルは各々の年齢や立場によって異なる。ほとんど獲得していない者もいる一方で、理系大学院を卒業した者ならかなりの高度なものとなっているはずであり、さらにその先まで獲得している者もいる。しかしながら個人の「数の概念」の獲得程度に関わらず、基礎となるものは未就学時に習得する段階である。そしてこの段階は、日本人の成人ならばほぼすべての人が通過している。未就学時に習得する段階の前提として、乳児は生得的に1つと2つを別のものとして認識することができる程度の「数の概念」を獲得していると考えられている[栗山 1998]。この生得的な能力は、その後の概念獲得の基礎となっているとも考えられる。

初期段階とは、Gelman and Gallistel(1978) によって明らかにされた、子供の計数における5つの原理の理解

[Gelman 1979] ができるという段階である。5つの原理とは、1. 一対一対応、2. 安定した順序、3. 順序無関連、4. 基数性、5. 抽象性である。これらの習得は個々が独立したのではなく連動しながら進むものである。一般的にこの原理は言葉の獲得以前の時期から習得し始め、3、4歳の頃までに終わる。特に、1. 一対一対応は保護者の教示に対して子供が習得するものなので、言葉の獲得を必ずしも必要とするわけではない。2. 安定した順序以降は、必ずしもというわけではないが、数唱が重要な役割を果たすので言葉の獲得が大きな支えとなる。

1. 一対一対応：数える対象物に対して数詞を1つ1つ割り当てていくことである。例としては、りんごが3つ並んでいる場合、1つ目のりんごが1、2つ目のりんごが2、3つ目のりんごが3と解釈することである。最も基礎的なものであり、この原理の理解は他の4つの原理の理解に欠かせない。

2. 安定した順序：モノを数える時に使用される数詞が常に一定の順序になっていることである。例としては、りんごが5つ並んでいる場合、「いち、に、さん、し、ご」と数えることができることである。

3. 順序無関連：ものを数える際、どの順序で数えても全体の数に変わりがないことである。同じものを複数回数えたり、数えないで飛ばしたりしないで正確に数えることができるということである。

4. 基数性：集合の最終の項目に与えられた標識は、その集合の項目の数を表す。複数個のりんごを数え場合、最後に数えた数がそのりんごの集合の数を表すと理解できていることである。

5. 抽象性：対象物が変化しても正しく数えられることである。りんごであれ、自動車であれ、消しゴムであれ、5つは5つであるということである。

以降の「数の概念」の獲得は、学校教育およびそれをもとにした家庭学習、自主学習を中心として進む。次の段階としては2つの独立した変数を扱えるようになることである。たとえば、時間と分、ドルとセント等である。その後、天秤の長さとおもりの重さのような異なる次元の数を同時に配慮することが出来るようになり、最終的には異なる次元の数を式で捉えることが出来るようになる程度まで進む[Griffin 2009]。これらは学校教育およびそれをもとにした家庭学習を中心として進む。

本研究では「数の概念」の初期段階の獲得を限られた範囲の記号レベルの「数」を扱って教示学習をおこなうシステムを作成し、その学習モデルのシミュレーションおよび考察をお

連絡先: 〒 606-8585 京都市左京区松ヶ崎橋上町 1

京都工芸繊維大学 大学院工芸科学研究科 情報工学専攻
インタラクティブ知能研究室, takai@ii.is.kit.ac.jp

こなった。具体的には Gelman らの「子供の計数における5つの原理」のうち、抽象性および安定した順序の獲得のシミュレーションを行った。

3. エージェントの内部機能

3.1 学習方法および行動選択方法

今回の研究では学習するエージェントの作成のために、機械学習の方法に Q 学習、行動選択方法にソフトマックス法を用いた。エージェントは学習していく過程でユーザーから報酬を貰い、Q 学習により強化学習を行う。Q 学習の設定は図 1 の通りである。ソフトマックス法は Q 値に応じた確率で行動選択を行う。累積型の更新式であるがゆえに、評価値は目標値を持たず発散し絶対的な値は意味を持たない。人間は絶対的評価よりも相対的評価を行っている [並木 2013] ため、相対的評価を行っているソフトマックス法は人間のメカニズムに近い。よって本研究ではソフトマックス法を採用した。

3.2 報酬

本研究では強化学習の報酬として、ユーザから与えられる報酬 (外部報酬) と内部報酬を導入した。全報酬は外部報酬と内部報酬の和とした。

(1) 外部報酬

外部報酬の値は+1 もしくは-1 である [図 1]。

(2) 内部報酬

内部報酬の定義に関する具体的なアルゴリズムは、以下の通りである。Schmidhuber らの定義 [Schmidhuber 2011] を基にシステム作成を行った。

Schmidhuber らは、エージェントが相手の次行動を予測する行動予測誤差を用いていたが、本研究では自身の行動に対して与えられる外部報酬を予測する、報酬予測誤差を用いる。これは幼児期の子どもは承認要求が高いため、相手の報酬を期待する [櫻井 2009] ためである。

内部報酬とは学習意欲や知的好奇心に相当するものである。これらは入力される情報に対する親近性や新規性によって引き出されるものである。インタラクションをするにつれて相手への親近性は上がるが、予想通りの返答ばかりだと、親近性は感じられない。新規性は学習初期は大きくなるが、学習が完了してくると小さくなっていく。本研究では以下のようにすることで知的好奇心として内部報酬を定義した。

- エージェントは自分の行動に対して、ユーザーが自身に与える報酬を予測する。
 - ユーザの行動を状態 $s(t)$ 、エージェントの行動を行動 $a(t)$ と定義し、状態 $s(t)$ と 行動 $a(t)$ 、観測された外部報酬値 $r(t)$ を $P(s(t); a(t); r(t))$ とする。
 - 新たな $s(t)$ と $a(t)$ が入力された時、その $s(t)$ と $a(t)$ の組み合わせにおいて、入力回数が最大の $r(t)$ を予測値 p とする。
実験の設計上、報酬 $r(t)$ は $-1, 0, +1$ の3つなので $(0, \dots, 2)$ とする。
 - 状態 $s(t)$ と 行動 $a(t)$ よりトライグラムで外部報酬 $r(t)$ を予測する。
- 報酬予測の誤差 err を定義する。報酬予測の誤差とは上述したエージェントの予測とユーザーが与えた外部報酬の誤差である。外部報酬 $r(t)$ と 予測値 p の差の絶対値とした。

$$err = |r(t) - p|$$

- 報酬予測の誤差の短期間の平均値 (*shortaverage_err*) および長期間の平均値 (*longaverage_err*) をとる。

$$shortaverage_err = (1 - \beta) \times shortaverage_err + \beta * err$$

$$longaverage_err = (1 - \eta) \times longaverage_err + \eta * err$$

- 短期間の平均値と長期間の平均値の差の絶対値を内部報酬 *internalreward* と定義する。

$$internalreward = |shortaverage_err - longaverage_err|$$

Schmidhuber らの定義 [Schmidhuber 2011] では絶対値をとらずに、短期平均-長期平均が負の値となった場合は内部報酬は 0 としていた。しかし好奇心は短期的に失われるものではないことを考慮して、本研究では内部報酬を短期平均-長期平均の絶対値として定義した。

短期間と長期間という抽象的な表現になっているが、シミュレーションの際にはパラメータ (β と η) で具体的に設定する。

これらのパラメータを変動させた場合、長期平均のパラメータ (η) を小さくすると長期平均になり、短期平均のパラメータ (β) を大きくすると短期平均になる。いずれにしても予測が観測と合致し続けると興味を失う。また学習が完全に完成した場合も興味を失う。また短期平均と長期平均の初期値は、それぞれ 1 と 0 に設定した。長期平均の初期値を小さく設定すると、報酬予測誤差の変化量の初期値は大きくなるので、最初は興味があるというのを表現可能なためである。

行動	エージェント側の行動
状態	被験者の行動
報酬	被験者が「おっけーい！」ボタンを押したら「正」の報酬 被験者が「それちがうよ！」ボタンを押したら「負」の報酬

図 1: Q 学習の設定

4. シミュレーション

4.1 シミュレーション設定

シミュレーションでは図 2 のような画面を設計した。シミュレーション方法としては、まず「ユーザー選択ボタン」を選択することによって、ユーザーがエージェントに教示する。その後エージェントが返答を返してくる。その返答に対してユーザーは「報酬授与ボタン」を押す。この一連の動作を学習回数 1 回と定義する。ユーザーはエージェントが学習できたと思うまでこの動作を繰り返す。

「報酬授与ボタン」は「おっけーい！」ボタンを押すと+1 の外部報酬が入り、「それちがうよ！」ボタンを押すと-1 の外部報酬が入る。「おっけーい！」ボタンおよび「それちがうよ！」ボタンの設定定義としては、エージェントの正解に対してユーザが褒める行為および、エージェントの不正解に対してユーザが指摘する行為である。

本シミュレーションではまず、ゲルマンの 5 原理にあたる「抽象性」を学習させる。これは本研究の設定では、「1」と「*」や「|」, 「2」と「**」や「||」の対応を学習することにあ

たる。「抽象性」は対象物が何であれ個数を認識できるということなので、「*」や「||」を「1」、「**」や「|||」を「2」と認識できることを意味する。これらの学習が完了すると「安定した順序」の学習に移る。これは「1」「2」「3」「4」の順番を学習することにあたる。逆順や数字以外の順番、例えば「*」「**」「***」は本シミュレーションでは学習させない。後者の学習では12個の「ユーザー選択ボタン」のうち、「1」「2」「3」「4」の4つのみを用いた。学習内容の切替に関しては、正解の項目のQ値の増加が誤差の範囲内である0.001となった時点とした。

本研究ではパソコンがランダム選択を繰り返すシミュレーションをおこなった。この際、外部報酬は正解ならば+1、不正解ならば-1を自動的に与えている。なお前半の学習では「ユーザー選択ボタン」のうち、12個すべてからランダムに選択を行なったが、後半の学習では用いた4つのボタンからランダム選択を行なった。

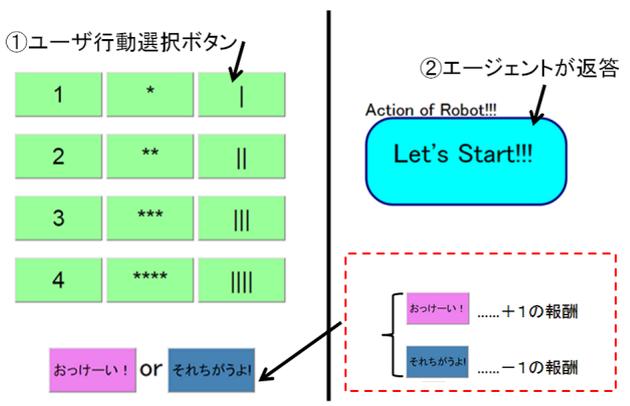


図 2: 設計画面

5. 結果と考察

5.1 シミュレーション結果

学習回数を経るごとの内部報酬の推移を表したグラフが図3である。内部報酬および、それを求めるための報酬予測誤差の短期平均と長期平均の値のグラフである。この時のパラメータは、図5である。

またQ学習の推移の図において、内部報酬を導入した場合とそうでない場合の比較を行う [図4]。これらグラフが生成された際のパラメータは、図5と図6である。内部報酬が導入されていない場合、内部報酬に関するパラメータである、初期値 (*longaverage_err*, *shortaverage_err*) やパラメータ (η , β) の値は影響がなく、 α , γ , τ のみが影響する。横軸は学習回数の定義は、「4.1 節シミュレーション設定」に記載した通りである。縦軸のQ値が2に収束している理由であるが、Q学習の更新式の割引率を0.5、ソフトマックス法の温度定数 τ を0.1にした場合の収束値である。立ち上がりは2か所あるが、最初は「抽象性」の学習であり、後が「安定した順序」の学習を示している。

図4の黒点線より内部報酬を導入した方が、学習初期のQ値の上がり方が急であることが分かる。これは正解の行動を取った場合は、正の外部報酬に加えて内部報酬がプラスされるので学習回数1回あたりのQ値の上昇幅は大きくなるためである。

また最も学習が早く進んだものと、最も学習が遅く進んだものの差が大きいことが分かる (赤色縦破線で挟まれた部分)。これは不正解の行動を取った場合は、負の外部報酬に加えて正の内部報酬が入るのでQ値の減衰量が小さくなる。そのため、その後の行動選択における正解の行動を選択する確率が減少するためである。

図7および図8は、報酬予測誤差の短期平均と長期平均にかかるパラメータ β と η を変化させてシミュレーションをおこなった結果である。 β は0.1から1までの0.1刻みの値を用い、 η は0.01から0.1までの0.01刻みの値を用いた。図7はQ値が時定数の値に到達するまでの、Q値の推移のグラフの近似直線の傾きを示している。図8は最速でQ値が時定数の値に達した項目の学習回数と、最遅でQ値が時定数の値に達した項目の学習回数の差である。

$\beta = \eta = 0.1$ の枠が内部報酬0の時を示している。

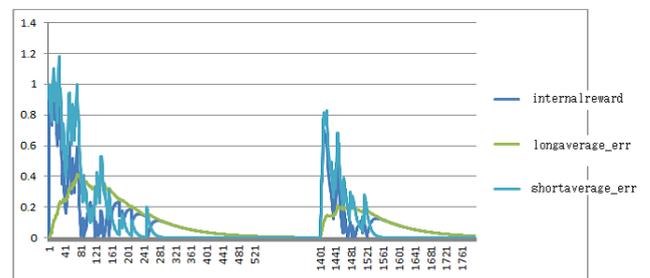


図 3: 内部報酬の推移

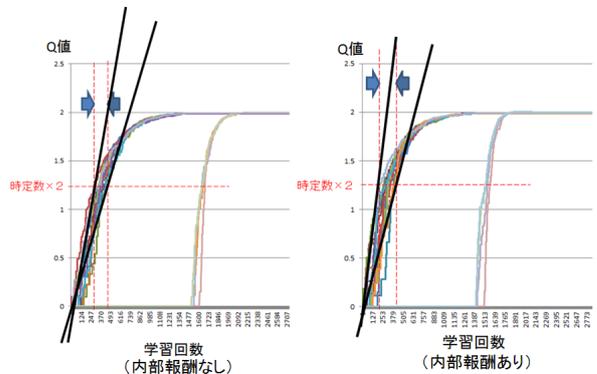


図 4: Q 値の推移

パラメータ名	<i>longaverage_err</i>	<i>shortaverage_err</i>	η	β
意味	初期値(長期平均)	初期値(短期平均)	パラメータ(長期平均)	パラメータ(短期平均)
数値	0	1	0.01	0.1

図 5: 図4のパラメータ (内部報酬)

5.2 考察

学習回数を経るごとの内部報酬の推移より (図3)、内部報酬の値が、学習初期では値が高いが徐々に低くなり、別の学習が始まるとまた値が上昇し、その後減少している。3.2.(2) で示した知的的好奇心に相当する内部報酬の特徴が表れている。

パラメータ名	α	γ	tau
意味	学習率	割引率	温度
数値	0.1	0.5	0.1

図 6: 図 4 のパラメータ (強化学習)

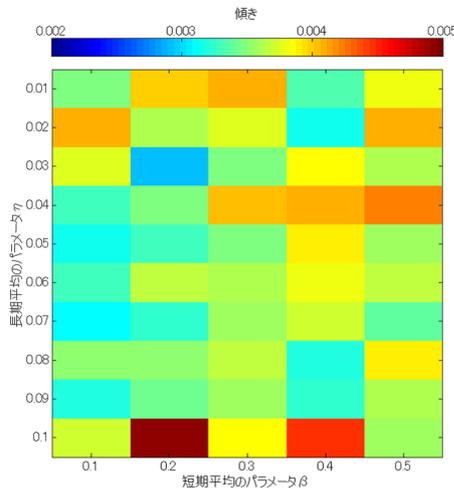


図 7: 図 4 の黒点線の傾きの平均

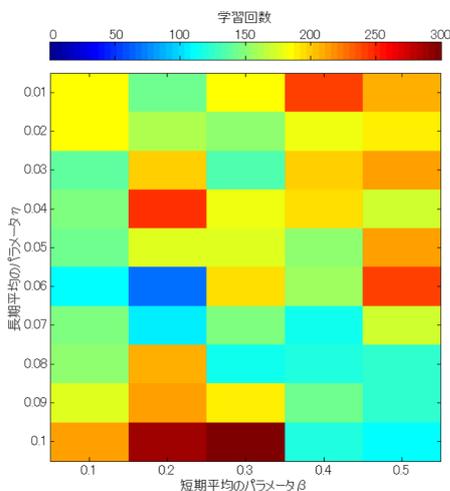


図 8: 図 4 の縦赤点線間の幅

よって学習意欲や知的好奇心に相当するものとしての内部報酬が導入できていると考えられる。

図 7 および図 8 より、本シミュレーションの範囲では短期平均のパラメータ $\beta=0.2$ 、長期平均のパラメータ $\eta=0.1$ とした場合、内部報酬の効果を最も示せることが分かった。また内部報酬を導入した結果、傾き、学習回数の差の両者において、導入以前と比較して、値が減少する場合も増加する場合も見受けられた。

数の概念の中で基本的なものと考えられている、「計数の 5 原理」のうち「安定した順序」と「抽象性」に関して学習シ

ミュレーションをおこなった。

「抽象性」に関しては、本研究のシミュレーションは 1 つの画面内という限られた狭い範囲内でおこなった。対象物の位置がバラバラの状態や変化する場合は今後の課題である。また対象物の種類が増えた場合も今後の課題である。

「安定した順序」に関しては「1」「2」「3」「4」の 4 つのみの場合は完璧に学習できた。同様の方法で学習をおこなった場合、逆順や数が増えた場合も対応できると予想できる。「安定した順序」の学習は、順番が 1 以外から始まる場合や、複数の種類の対象物の場合、数字とものが混同した場合、「2」「4」「6」のように飛び番号になった場合も定義内である。よって今後の課題として、これらの学習シミュレーションを行なうことがあげられる。

さらに、「安定した順序」と「抽象性」の両者を同時に考慮することや、時間差が関係してくる場合も今後の課題である。例えば、宝探しにおいて、異なる種類の宝を順番に探すことである。

6. 今後の方針

2 つの学習の切り替えにおける、タイミングや方法に関して吟味していく予定である。また計数の原理のうち、本研究で扱ってない「一対一対応」、「安定した順序」、「基数性」に関しても新たな画面設計を用いた上でシミュレーションをおこなう予定である。

参考文献

- [櫻井 2009] 櫻井茂男：自ら学ぶ意欲の心理学 (2009).
- [並木 2013] 並木尚也, 大用庫智, 高橋達二：知識利用と探索に対する因果的直感と相対評価の処方箋的効果, 人工知能学会全国大会論文集, vol.27, pp.1-4 (2013).
- [Gelman 1979] Gelman.R and Gallistel.C.R: The child's understanding of number (1979).
- [栗山 1998] 栗山和広：子どもの数概念の発達について,” 宮崎女子短期大学紀要, vol.24, pp.81-96 (1998).
- [Griffin 2009] S. Griffin: Learning Sequences in the Acquisition of Mathematical Knowledge: Using Cognitive Developmental Theory to Inform Curriculum Design for Pre-K-6 Mathematics Education, *Mind, Brain, and Education*. Vol.3. No.2. 96-107. (2009).
- [Schmidhuber 2011] J. Schmidhuber: Artificial curiosity with planning for Autonomous perceptual and cognitive development, *Formal Theory of Creativity, Fun, and Intrinsic Motivation*, pp.1-8, Aug. (2011).