エピソード記憶編集による学習アルゴリズムの経験累積効果と課題構造の関係

青田佳士[†] 山口陽子[††]

以前に我々が提案したエピソード記憶編集による学習アルゴリズムは，部分観測マルコフ決定過程下でもとくに時間軸上の不確定さが存在する課題に対して適切なエピソードを選択した．さらに，過去に学習した課題も新規な別の課題に対してより良い解の探索に役立っていることが示された．しかし，過去の経験がどのように役立つのか，あるいは新規な課題でもその課題の構造と過去の経験との関連性は明確ではなかった．そのため本研究では，過去の経験が新規課題の解の改善に役立つ条件を調べた．その結果，課題を解くプロセスがラットの系列学習においてみられるチャンク化とよく似たものであり，課題構造に応じたチャンク化を行うことにより学習を促進させたことを示す．

Relationship between Structure of Tasks and Past Experiences on the Learning Algorithm with Episodic Memory Integration

Yoshito Aota[†] and Yoko Yamaguchi[††]

The learning algorithm, which we proposed previously, showed proper selection of episodes in tasks under the Partially Observable Markov Decision Process, POMDPs, especially including uncertainness along the time axis. Moreover, it was shown that the agents utilized past experiences for solving new tasks. But the relationship between structure of tasks and past experiences was not clear. Here, we investigated the requirements for utilizing past experiences to new tasks. It is also discussed that the learning process is highly similar to "chunking", which is observed in rat's serial learning, and the agent's chunking according to the structure of each task could promote learning.

1. はじめに

これまで我々は，環境変化の規則も学習できるようにするために，エピソード記憶編集に基づいた新しい学習アルゴリズムを提案してきた[1)]．このアルゴリズムを実装したエージェントをラットの系列学習を模したT字型迷路および十字型迷路に適用した結果，部分観測マルコフ決定過程（POMDPs）下でもとくに時間軸上の不確定さが存在する課題に対して，エージェントは適切なエピソードを選択できた．

ここで時間軸上の不確定さとは，報酬やゴールを継続的に得るための，一連の行動の長さやその行動系列の開始と終了のタイミングが不確定な状況を指す．たとえば「Aを2回実行してからBとCを実行して終了」というような行動計画を，試行錯誤を通じてエージェント自ら決めなければ解けないような課題である．時間軸上の不確定さが存在する課題は，Profit sharing（たとえば文献2)）など報酬獲得後に行動の履歴がリセットされるような従来のアプローチでは解決が困難であった[1)]．また前回の結果では，過去に学習した状況も新規な別の状況に対してより良い解の探索に役立つ場合があることも示された．

しかしながら，どのようなときに過去に学習した状況が役立つのか，その条件を明確にすることはなかった．さらに，時間軸上の不確定さの形式として何種類かのタイプが考えられるが，本アルゴリズムが異なるタイプの課題でも通用するかは確かめなかった．

一方，認知科学においては一般に，我々は物事をパターンの集まりとして認識すると考える．多様な事象を既知のパターンの組合せとして認識できれば，学習も効率的になされると期待できる．たとえば漢字を第2外国語として学ぶ場合では，パターン認識能力と学習能力に相関があるという報告もある[3)]．空間的なパターンのみでなく，時間的な変化もパターンの集まりとして認識することで学習を促進させる可能性を考慮

† 横浜国立大学大学院国際社会科学研究科
International Graduate School of Social Sciences, Yokohama National University
†† 理化学研究所脳科学総合研究センター
Brain Science Institute, RIKEN

することは興味深い．

以上をふまえてここでは，Keeping trial タイプと Resetting trial タイプという時間軸上の不確定さの形式が異なる 2 種類の課題について，ゴールの位置変化の時系列に高い周期性がある場合とそうでない場合を用意し，事前学習の有無の効果を比較することによって過去に学習した状況の影響の違いを調べる．

Keeping trial タイプは前回の報告[1] で扱ったタイプである一方，Resetting trial タイプは今回新たに試みる課題構造であり，エージェントのゴール到達失敗時の扱いが異なる．両課題とも，蓄積したエピソードを組み合わせないと継続的な報酬獲得がままならない課題であるが，その組合せの仕方は課題構造に応じて柔軟に対応できなければならない．本アルゴリズムが，異なる 2 つの課題構造に対応できるかを確認する．

またゴールの位置変化の時系列に高い周期性がある場合について，エージェントの学習過程とラットの系列学習との類似性を検討し，本アルゴリズムの解の導き方が認知科学的に妥当であるかも考察する．

2 章では本研究で用いる十字型迷路課題の特徴と 2 つのタイプの課題について述べ，3 章において本アルゴリズムを簡単に説明する．4 章でシミュレーション結果を述べて過去に学習した状況が新規の状況に役立つ条件を検討し，5 章でラットの系列学習との類似性を考察する．最後にまとめを行う．

2. 十字型迷路課題

2.1 課 題 設 定

課題設定は前回の報告[1] と同じであり，既知の読者はこの節を飛ばし，2.2 節から読み進めていただきたい．

本論文でエージェントが解くのは**図 1** のような十字型迷路課題であり，ゴールの位置がある規則で変化する．ここでは，ゴールの位置が ABCABCABC··· といった周期的な変化を仮定する．エージェントはこの隠れた規則を学習しなければならない．観測状態はエージェントの位置と報酬の有無であり，隠れ状態は各位置で報酬が観測されるタイミングである．

本研究では POMDPs 下でもとくにゴールの位置変化の規則をいかに学習するかを考えたいので，ここでは非決定的な状態遷移については取り扱わない．

環境は，マス目に分割された枝分かれを含む通路とし，特定の場所（ゴール）に一定の規則で報酬がおかれる．エージェントの行動は時間，空間とも離散値で表される．単位時間を 1 ステップとして，1 マスを 1 ステップで移動する．行動は各ステップにおけるマス目固有の知覚入力と，移動の方向とで記述される．

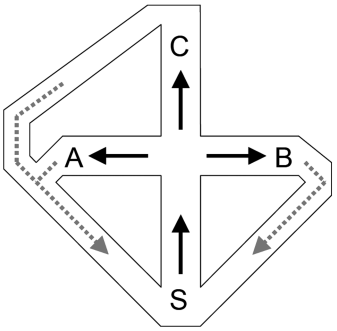

図 1 本研究で用いる十字型迷路課題
エージェントは道端（A，B，C）に着くと S に戻る．ゴールで報酬を受けるが，ゴールの位置は A，B，C の間で規則的に変化する．エージェントはゴールの位置だけでなく，そのような隠れた規則を見つけなければならない

Fig. 1 Cross maze task.
Agent goes back to the position S after the Agent reach A, B, or C. Goal position changes between A, B, and C under some rules. Agent must find out not only Goal position but also such hidden rules.

この課題の難しさは，報酬を継続的に得るために，エージェントはゴールの位置の連続的な変化の中から行動の始まりと終わりのタイミングを決定し，隠れた規則を学習しなければならない点にある．

たとえば図 1 で，ゴールの位置が AABCAABCAABC··· と「AABC」の繰返しとして変化した場合，エージェントは 1 つ目の A と 2 つ目の A を区別し，ゴールの位置変化の周期が 4 試行（の倍数）からなることを学習する必要がある．周期を知るには，行動を自律的に区切ることで試行錯誤を行わなければならない．

また，本研究ではゴールの位置変化の規則そのものが途中から変化するダイナミックな環境を取り扱う．そのため過去の規則から新しい規則へと変化した際に，学習の干渉を極力抑え，逆にエージェントの過去の知識を活かせるような場面では積極的に利用できることが望ましい．

2.2 課題の 2 つのタイプ

エージェントが誤った道を選択したときに，課題を継続するかどうかに応じて 2 つの異なるタイプを考える．

Keeping trial タイプは，課された規則をエージェントが完遂するまで課題を継続するタイプである．たとえばゴールの位置が「AAB」の繰返しとして変化する場合，**図 2** のようにエージェントが 2 回目の A を探索中に誤った道を選択しても課題はそのまま続行し，2 回目の A に到達後は B の探索に移る．

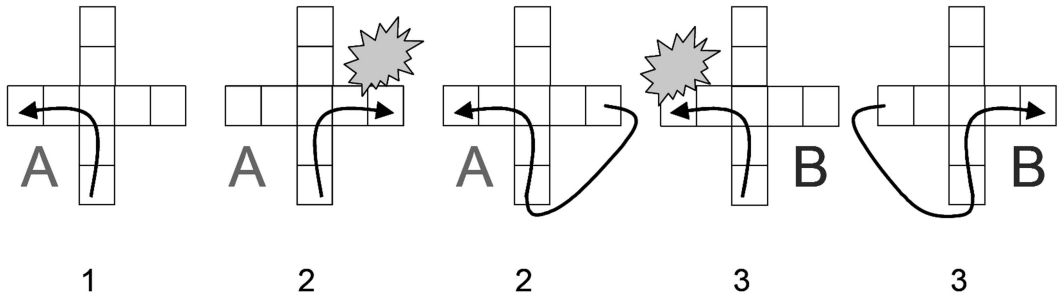

図 2 Keeping trial タイプの例
ゴールの位置は「AAB」の繰返しとして変化する．エージェントは 2 回目の A および B（AAB 中の 3 番目）の探索中に誤った道を選択しても課題はそのまま続行し，2 回目の A および B の探索を引き続き行う．番号は AAB 中の何番目かを示す

Fig. 2 Example of the “Keeping trial” type.
Goal position changes as the repetition of AAB. Even the agent selects a wrong pass at searching the second A, the task keeps its current order as the second A. The agent must continue to search the second A.

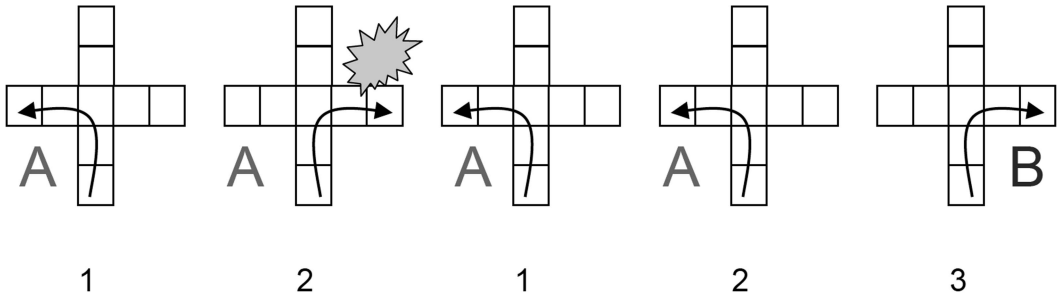

図 3 Resetting trial タイプの例
ゴールの位置は「AAB」の繰返しとして変化する．エージェントは 2 回目の A の探索中に誤った道を選択した場合，エージェントは 1 回目の A を探索するところから再び始めなければならない．番号は AAB 中のどの試行かを示す

Fig. 3 Example of the “Resetting trial” type.
Goal position changes as the repetition of AAB. When the agent selects a wrong pass at searching the second A, the task reset its order to the first A. The agent must search the first A again.

2 つめは Resetting trial タイプで，課された規則の途中でエージェントが誤った道を選択したときに課題はリセットされ，はじめからやり直すタイプである．たとえば先ほどと同様にゴールの位置が「AAB」の繰返しとして変化する場合，**図 3** のようにエージェントが 2 回目の A を探索中に誤った道を選択したら，エージェントは 1 回目の A を探索するところから再び始めなければならない．

なお，Resetting trial タイプでリセットされるのはゴールの位置変化の規則のみであり，エージェントはどちらのタイプの課題が与えられているのか分からない．本研究では，タイプの違いがエージェントの蓄積した知識を役立てるのにどう影響するかを調べる．

2.3 実験条件

本研究では，各タイプに対して，ゴールの位置変化に階層性（高い周期性）の見られる規則とそうでない規則について事前学習の有無を比較する．そのため次のように合計 8 つの実験条件を用意した．なお，エージェントが図 1 の S から出発してゴールに辿り着くまでを 1「試行」と数えることにし，ゴールの位置変化の繰返しを「周期」で数えることにする．たとえばゴールの位置変化が「AAB」の繰返しの場合，1 周期は 3 試行からなる．

① **Keeping trial タイプ**

(1) **階層性のある規則：本課題のみ**
AAB を 7 回のあと BBA を 7 回行うことを繰り返す．そのためこの規則は全体として 1 周期が 42 試行からなる．これを本課題として 200 周期行う．

(2) **階層性のある規則：事前学習あり**
上記の本課題を行う前に，事前学習として A を 200 試行，B を 200 試行，C を 200 試行，AAB を 200 周期，BBA を 200 周期行う．その後で (1) の 1 周期 42 試行の本課題を 200 周期行う．

(3) **階層性のない規則：本課題のみ**
「ABBABBBBBBAABABBBBAAAAAAABBBAAAAAAAAAAABAABA」を繰り返す．1 周期が 42 試行からなる．これを本課題として 200 周期行う．

(4) **階層性のない規則：事前学習あり**
上記の本課題を行う前に，事前学習として A を 200 試行，B を 200 試行，C を 200 試行，「BABBBBAAAAAAABB」を 200 周期，「ABBABBBBBBAA」を 200 周期，「BAAAAAAAAABAABA」を 200 周期行う．その後で (3) の 1 周期 42 試行の本課題を 200 周期行う．

② **Resetting trial タイプ**

(1) **階層性のある規則：本課題のみ**
AAB を 7 回のあと BBA を 7 回行うことを繰り返す．ただし報酬は各回の 3 番目のゴールでのみ与えられるものとし，AAB の途中の AA および BBA の途中の BB では，エージェントは AA および BB の道筋を通るよう強いられるが報酬は与えられない．これを本課題として 10 周期行う．

(2) **階層性のある規則：事前学習あり**
上記の本課題を行う前に，事前学習として AAB を 10 周期，BBA を 10 周期行う．その後で ② の (1) の 1 周期 42 試行の本課題を 10 周期行う．

(3) **階層性のない規則：本課題のみ**
「ABBABBBBBBAABABBBBAAAAAAABBBAAAAAAAAAAABAABA」を繰り返す．ただし報酬は 3 試行ごとに与えられ，1, 2, 4, 5, 7, 8, ···, 40, 41 試行目はそれぞれのゴールへの道筋を強いられるが報酬は与えられない．これを本課題として 10 周期行う．

(4) **階層性のない規則：事前学習あり**
上記の本課題を行う前に，事前学習として「BABBBBAAAAAAABB」を 10 周期，「ABBABBBBBBAA」を 10 周期，「BAAAAAAAAABAABA」を 10 周期行う．その後で ② の (3) の 1 周期 42 試行の本課題を 10 周期行う．

上記の ① および ② の (4) の事前学習は，本課題の規則を 3 分割し 1-2-3 の順番を 2-1-3 の順番にしてそれぞれ Keeping trial で 200 周期，Resetting trial で 10 周期行ったものである．

タイプの違い，階層性のある規則とない規則，および事前学習の有無を比べることで過去に学習した状況が新規な状況に与える影響の度合いを調べ，課題構造と学習の関係を明確にすることを目的とする．

3. 学習アルゴリズム

本論文では前回と同じ学習アルゴリズムを用いるため，ここでは結果の理解に必要な各種定義と，学習アルゴリズムのネットワーク構造の説明のみを行う．前回の報告[1)] と異なる点は，1 つの行動計画ユニット（3.1 節参照）がコードできる成功エピソードの最大数を 30 から 20 に減らした点と，Resetting trial でエージェントが誤った道端に着いた際に強制的にスタート地点に戻される場合も，予期しない移動として失敗とした点である．既知の読者はこれらの点をふまえて 4 章のシミュレーション結果へ進むことができる．

また，本アルゴリズムの情報の流れや数式表現など詳細については文献 1) を参照されたい．

3.1 成功失敗とエピソードの定義

図 1 の S から出発した移動の結果，ゴールに到着して報酬を得る場合を成功と呼ぶ．行動計画の中に報酬の予測が含まれていたにもかかわらず，実際にそのマス目で報酬が得られなかった場合を失敗と呼ぶ．ただし Resetting trial タイプでは，分岐路において誤った道を選択した場合も，強制的にスタート地点に戻されるため，予期しない移動としてその道端で失敗とする．実際のラットでいえば，ラットが誤った道端に辿り着いた時点で実験者がピックアップしてスタート地点に戻すのに似ている．

一連の行動を成功か失敗の直後で区切ったものをエピソードと呼び，さらにエピソードの終わりが成功で区切られたもの（成功を含むエピソード）を成功エピソード，失敗で区切られたものを失敗エピソードと呼ぶ．たとえば各マス目を数字で表したとして，「$1 \to 2 \to 3 \to 4 \to 6 \to 7 \to 1 \to 2 \to 3 \to 4 \to 5 +$ 報酬」で 1 つの成功エピソードといった具合である．つまり Profit sharing のエピソードの定義に失敗も含めたものである．

3.2 行動計画ユニット

1 つまたは複数のエピソードから構成される時系列を行動計画ユニットと呼ぶ．行動計画ユニットは行動

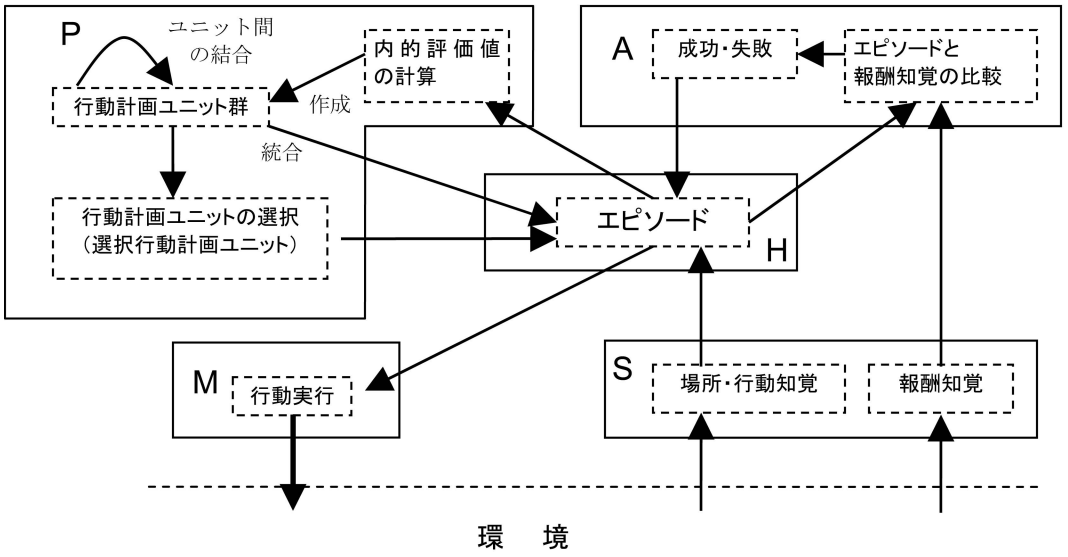

図 4 ネットワーク構造
S は知覚入力，H はエピソードの形成・貯蔵，A は成功・失敗の判断，P は行動計画ユニットの作成・行動計画とエピソードの編集，M はエピソードを通じて選択行動計画ユニットの実行を行う
Fig. 4 Network structure.
S: sensory input part, H: store of episods, A: decision of success and fail, P: episodic memory integration, and M: motion part.

の選択肢として利用されるが，行動計画ユニットとして構成されない単に複数のエピソードが続いたものを連続エピソードと呼ぶ．そのため以降では，1 つのエピソードと複数の連続したエピソードの両方を指して「（連続）エピソード」と呼ぶことがある．さらに各行動計画ユニットはその中に含まれる成功と失敗の回数，および経過ステップ数より決定される内的評価値を 1 つの値として持つ．内的評価値は，その行動計画ユニットがコードする成功エピソードの数が多いほど高い値をとり，経過ステップ数が長いほど低い値となる（詳細は文献 1) を参照）．行動計画ユニットがどのように作成されるかは，次節で述べる．

3.3 学習アルゴリズムのネットワーク構造

提案モデルの全体をネットワーク構造として表すと**図 4** のようになる．図 4 で S 層は知覚入力層で，エージェントの現在位置や移動方向，報酬の有無を H 層および A 層に送る．A 層は成功・失敗を判断する層で，前節で定義したように報酬を得た場合を成功，H 層から入力される予測がはずれた場合を失敗として H 層に出力する．H 層では A 層の成功・失敗の情報に基づいてエピソードが形成・貯蔵される．1 つのエピソードは 1 つの成功か失敗を含む．

P 層は，H 層で貯蔵されたエピソードに基づいて行動計画ユニットを作成・編集し，次の行動計画を立てる．エージェントの行動計画は，行動計画ユニット単位で決まる．行動計画ユニットは，以下の場合でかつ該当する（連続）エピソードが初めて経験したものであるとき，それを 1 つの行動計画ユニットとして P 層に保存する．

＜行動計画ユニット作成対象のエピソード＞

- **報酬を得たとき**の下記のエピソード

1. 探索行動で報酬を得たエピソード
2. 選択行動計画ユニットがコードしている報酬の獲得タイミングと異なるタイミングで報酬を得た場合で，行動開始から最後の報酬までの（連続）成功エピソード

- **直前が失敗のとき**の下記のエピソード

3. 行動開始から直前の失敗までの（（連続）成功エピソード+）失敗エピソード
4. その前の失敗から直前の失敗までの連続成功エピソード

これにより行動計画ユニットは複数のエピソードとして随時編集され，開始と終了のタイミングが様々な行動計画ユニットが作成される．こうして行動の開始と終了は自律的に決定されることになる．行動計画ユニットはエピソード単位で行動を開始・終了するため，ゴールの位置の変化に対応しやすい．

M 層は実行層で，選択行動計画ユニットに基づいた

行動をとる．選択行動計画ユニット中の成功エピソードと同じ行動をとり，失敗エピソードでは探索行動を実行するか次の行動計画に移る．

基本方略としては，大きく以下のような流れで学習を進める．

① 開始時の行動決定

状況（前回の行動計画ユニットと知覚入力）に応じて行動計画ユニットを想起し，より内的評価値の高い行動計画ユニットを選択．これを選択行動計画ユニットとする．

② 選択行動計画ユニットの実行

(1) 選択行動計画ユニットの中に，成功エピソードを含むとき，それを実行し，失敗エピソードもあれば，その時点で行動終了として ① に戻る．

(2) 選択行動計画ユニットが失敗エピソードのみから構成されるとき，探索行動を行う．

③ 選択行動計画ユニット実行後の記憶の編集

選択行動計画ユニットがコードするエピソードと実際に経験したエピソードを比較し，報酬と経験の有無に応じて新しい行動計画ユニットを作成，さらに行動計画ユニット間の統合や順序関係の学習を行う．

4. シミュレーション結果

4.1 Keeping trial タイプの結果

2.3 節の Keeping trial タイプにおける各実験条件（① の (1) から (4)）について，初期値をランダムに変えた 500 体のエージェントを用いた結果を示す．実験条件 ① — (2)「階層性のある規則：事前学習あり」について，学習曲線を**図 5** に示す．また，Keeping trial タイプの 4 つの実験条件すべてについて，本課題の結果（最後の 200 周期）を重ね合わせたものを**図 6** に示す．

図 5 を見ると，ゴールの位置変化の規則自体が次の規則に移行した直後は学習の干渉を示すが，周期を重ねることでエージェントは学習を収束させていることが分かる．また図 6 によれば，本課題へと規則が移行した直後における学習の干渉は本課題のみを始めたエージェントが示す失敗頻度とほとんど変わらない．しかしながら事前学習が本課題の学習にあまり役立っておらず，「階層性のない規則」ではむしろ学習の速度を遅らせていた．

図 7 に「階層性のない規則：事前学習あり」の場合を，**図 8** に「階層性のない規則：本課題のみ」の場合の典型的な例を示した．横軸は何試行目かを示し縦軸は選択行動計画ユニットの番号を示す．「階層性のない規則：事前学習あり」に見られる収束の遅れは，事前学習で獲得した行動計画ユニットがコードする報酬の数が比較的多く内的評価値が高いために，それを本課題において必要以上に頻繁に選択してしまうことに起因すると考えられる．たとえば図 7 において，3 分割された 1 つめの規則の学習（600 から 3,600 試行目）で獲得された行動計画ユニットは，本課題（9,000 試行目以降）においても頻繁に選択され，首尾良く解の一部となっている．しかし，3 つめの規則の学習（6,000 から 9,000 試行目）で獲得された行動計画ユニットは，本課題のはじめの方で頻繁に選択されたにもかかわらず解の一部とはならず，結果として解を導

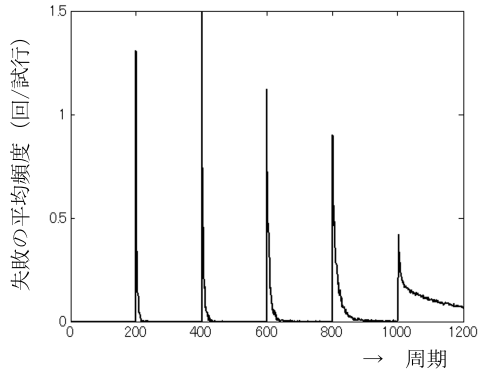

図 5 Keeping trial タイプ「階層性のある規則：事前学習あり」の学習曲線

Fig. 5 Learning curve of Keeping trial type, the periodic rule with pre-learning.

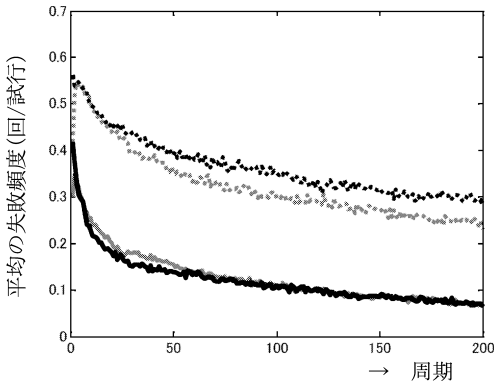

図 6 Keeping trial タイプのすべての実験条件の学習曲線
実線は「階層性のある規則」を，点線は「階層性のない規則」を示す．黒の線は「事前学習あり」を，グレーの線は「本課題のみ」を示す

Fig. 6 Learning curves of Keeping trial type.
Solid lines show the periodic rule, and dotted lines show the non-periodic rule. Black lines show tasks with pre-learnig, and Gray lines show without pre-learning.

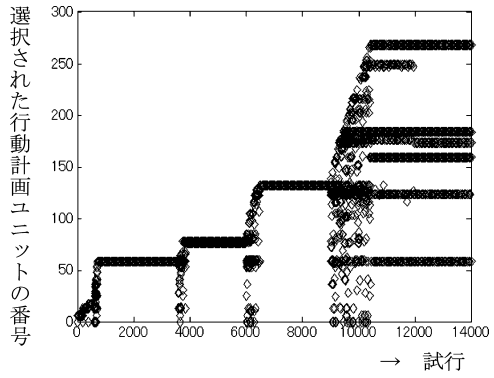

図 7 Keeping trial タイプの「階層性のない規則：事前学習あり」の例
各試行の行動計画で選択された行動計画ユニットを菱形で示す．3 分割された 1 つめの規則の学習（600 から 3,600 試行目）で獲得された行動計画ユニットが，本課題（9,000 試行目以降）においても頻繁に選択されている

Fig. 7 Example of Keeping trial type, Non-periodic rule with pre-learning.
Lozenges indicate Action plan units selected at each trial. The action plan unit, which the agent aquired at the pre-learning of the first rule from 600 to 3,600 trial, was frequently selected also at the final task after 9,000 trial.

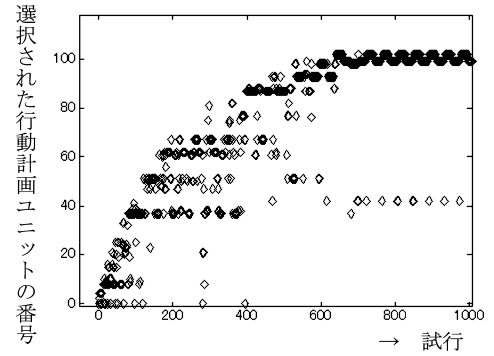

図 8 Keeping trial タイプの「階層性のない規則：本課題のみ」の例

Fig. 8 Example of Keeping trial type, Non-periodic rule without pre-learning.

く妨げとなっているように見える．

また，Keeping trial タイプの場合ゴールまでの道筋自体は成功・失敗の対象にしないため，ゴールまでどのくらいのステップ数を要したかを各実験条件で比べることができる．**図 9** に各試行でゴールまでにかかった平均のステップ数を示す．図 9 によると，どの実験条件にも事前学習の効果が見られるが，学習が進むにつれてその差は縮んでいくのが分かる．

Keeping trial タイプの結果についてまとめると，本課題の規則に階層性のある場合とない場合を比べた場合，階層性のある場合の方が明らかに収束が速く，エージェントは規則の階層性を利用した学習を行っていることが示唆された．この点についてはラットの系列学習との関連とともに 5 章で考察を行う．

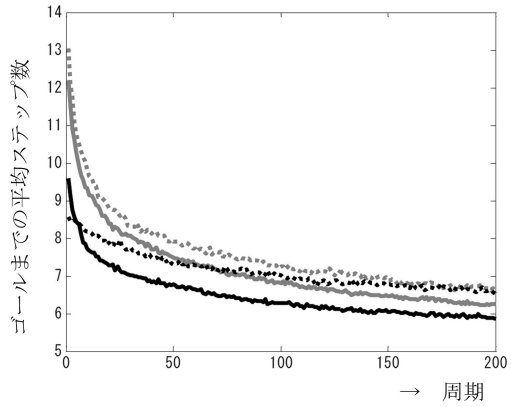

図 9 Keeping trial タイプにおける各周期の 1 試行あたりの平均ステップ数

Fig. 9 Average steps per one trial in Keeping trial type.

さらに，本課題のみの場合と事前学習ありとを比べた場合，事前学習はショートカットのようにゴールまでより近い道のりを選択するのに役立つが（図 9），学習の収束にはあまり寄与しないか，むしろ収束を遅らせる場合もあることが分かった（図 6）．

4.2 Resetting trial タイプの結果

2.3 節の Resetting trial タイプにおける各実験条件（② の (1) から (4)）について，初期値をランダムに変えた 400 体のエージェントを用いた結果を**図 10** に示す．図 10 は本課題のはじめの 3 周期について，各実験条件の周期ごとの失敗数をプロットしたものである．

図 10 によると，階層性のある規則，ない規則ともに事前学習の効果が見受けられ，とくに階層性のない規則の方が事前学習の効果が顕著であった．

事前学習を行うということは，それだけ解くべき課題全体の量が増すということである．そのため事前学習による課題量の増加に比例してエージェントの試行錯誤の量が増えるとしたら，事前学習を行うメリットがやや薄れてしまう．そこで各実験条件について，課題全体を通じて十字型迷路の分岐点を通過した回数をカウントしてみたところ，400 体のエージェントの平均値と標準偏差は次のような結果となった．

PW：14,704 ± 5,778 passes
PN：16,481 ± 5,562 passes
NW：16,597 ± 7,985 passes
NN：25,098 ± 6,185 passes

ここで，PW は「階層性のある規則：事前学習あり」，

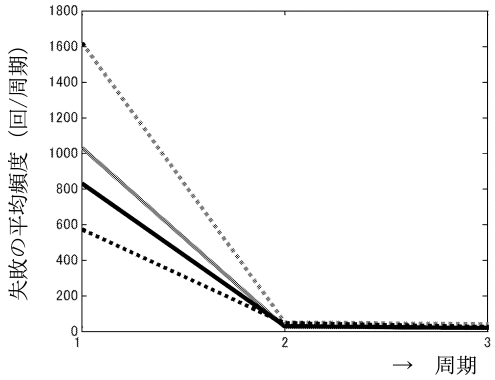

図 10 Resetting trial タイプのすべての実験条件の学習曲線
実線は「階層性のある規則」を，点線は「階層性のない規則」を示す．黒の線は「事前学習あり」を，グレーの線は「本課題のみ」を示す．本課題のはじめの 3 周期を示す

Fig. 10 Learning curves of Resetting trial type.
Solid lines show the periodic rule, and dotted lines show the non-periodic rule. Black lines show tasks with pre-learnig, and Gray lines show without pre-learning. First three period in the final task are shown.

PN は「階層性のある規則：本課題のみ」，NW は「階層性のない規則：事前学習あり」，NN は「階層性のない規則：本課題のみ」を指す．PW と PN，および NW と NN の組合せについて T 検定を行ったところ，T の絶対値がそれぞれ 4.43 と 16.83 であり，いずれも有意（$p < 0.01$）に差があった．

事前学習ありは本課題のみを解くより課題の量が多いにもかかわらず，事前学習を行った方が学習の収束に必要な運動量をかえって減らせることが分かった．そのため，エージェントは過去に学習した状況を，新規な状況を効率良く解くのに役立てているといえる．

以上の，4.1 節と 4.2 節のシミュレーション結果をまとめると，課題構造としては Resetting trial タイプの方が規則に階層性があるかないかにかかわらず，過去に学習した状況を新規な状況に利用しやすいことが分かった．また，Keeping trial タイプであっても，規則に階層性があるような課題なら学習の干渉を示さず，過去に学習した状況をより良い解の探索に役立てうることが分かった．さらに両タイプとも，規則に階層性がある方が学習が速く進む傾向にあった．

5. ラットの系列学習やチャンク化との関連

この章では，考察としてラットの系列学習の，とくにチャンク化との関連性について議論する．ラットの系列学習とチャンク化について簡単に説明しつつ，本研究との関連性や今後の課題について述べる．

5.1 ラットの系列学習

ラットは系列学習を遂行できることが知られている．系列学習でよくなされるのは，直線走路や T 字型迷路の端に設置された目標箱で与えられる報酬量（45 mg の餌ペレットの数）をラットに学習させるものである[4]．実際の実験では，たとえば 14-1-3-7-0 のように報酬量を項目として構成された報酬系列をラットに学習させ，各項目に対する走行速度が測定される．とくに 0 ペレット項目はラットが学習を達成したかどうかの判断の対象となり，0 ペレット項目に対して遅く走行するほどラットは予期をうまく行ったと判断される．Fountain ら[5] による T 字型迷路の実験では，ラットは 1 系列が 25 項目に及ぶ長い系列でも 0 ペレットを正確に予測できることが示されている．ラットの系列学習でなされる課題においてとくに，同じ経路でもどの項目を走行しているかに応じて行動が変化し，かつ 1 系列が何項目で構成されるかをラットは初め分からないといった点は本研究と同様である．

ラットの系列学習を説明する有力な仮説として考えられているのは，記憶弁別理論[6] である．この理論は，主として隣接する項目間の連合記憶形成と報酬強度の類似性による刺激般化の 2 つの原理に基づいて系列学習を説明しようとする．とくに最近 Wallace ら[7),8)] は，これらの原理を仮定した数理モデルによる予測と，実際のラットの行動がよく一致することを報告した．しかし Wallace ら[8)] も指摘しているように，彼らのモデルは複数の系列を分けて記憶することはできない．したがって，本アルゴリズムのようにある成功エピソードと別の成功エピソードをつなげて 1 つの行動計画ユニットにするといった，系列間の編集による問題解決を行えない．さらに系列の始まりを “Start” として実験者がモデルに与えているため，たとえば系列間に明瞭な区切りがないような状況では，その系列が何項目で構成されるかをモデル自身で決めることができないといった問題が生じる．

ただし本アルゴリズムの場合，項目間の連合記憶形成によってエピソードが構成されるが，刺激般化の機能はない．このような刺激般化の機能を本アルゴリズムにも取り入れることができれば，系列間の正の転移[9)] といった実際のラットが示す，より柔軟な学習が期待される．

5.2 チャンク化

Wallace and Fountain のモデルでは，チャンク化を実現するには実験者が適切なタイミングで Start と同等の信号をモデルに入力する必要があった．チャンク化とは，ある系列を適当な小分節に区切り，小分節

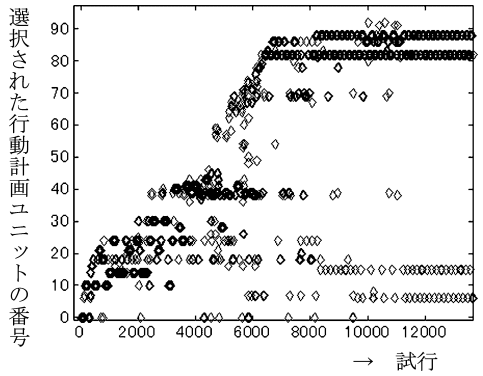

図 11 Keeping trial タイプの「階層性のある規則：本課題のみ」の例
各試行の行動計画で選択された行動計画ユニットを菱形で示す．この例において 88 番目の行動計画ユニットは ABBABBABBABBABBABBA を，82 番目は ABAABAABAABAABAABAAB を，15 番目と 6 番目はそれぞれ BB と A のゴールの位置をコードしていた

Fig. 11 Example of Keeping trial type, Periodic rule without pre-learning.
Lozenges indicate Action plan units selected at each trial. In this example, the 88th action plan unit codes goal positions as "ABBABBABBABBABBABBA", the 82nd action plan unit "ABAABAABAABAABAABAAB", and the 15th and 6th are "BB" and "A" each.

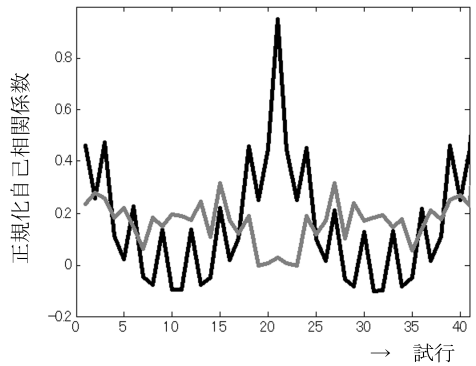

図 12 失敗頻度の周期性の分析
黒の線は「階層性のある規則：本課題のみ」を，グレーの線は「階層性のない規則：本課題のみ」を示す

Fig. 12 Analysis for the periodicity of failed trials.
Black line shows Periodic rule without pre-learnig, and Gray line shows Non-periodic rule without pre-learning.

の連なりとしてその系列を記憶することをいう．

チャンク化によってラットの学習が促進する場合があることが知られている．たとえば 14-7-3-1-0-14-7-3-1-0-14-7-3-1-0-14-7-3-1-0 という系列で，通常は項目間の間隔を 15 秒ほどあけるところを（"-" が 15 秒を示すとする），適当な 4 カ所で間隔を 10–15 分に延ばし，14-7-3-1-0／14-7-3-1-0／14-7-3-1-0／14-7-3-1-0／14-7-3-1-0（"／" が 10–15 分を示す）として系列を分節化することでチャンク化が促される[5]．Fountain ら[5] は，14-7-3-1-0／14-7-3-1-0／14-7-3-1-0··· のように分節化したとき（A）と，14-7-3-1-0-14-7-3-1-0-14-7-3-1-0 のように無分節のとき（B），および 14-7／3-1-0-14-7／3-1-0-14-7／3-1-0 のように誤分節したとき（C）を比べ，A-B-C の順に 0 ペレット項目の予期が優れていたことを示した．

本研究における階層性のある課題でも，適切なチャンク化がなされれば課題の解決が大いに早まると期待される．実際，Keeping trial タイプでは過去の経験の有無よりも階層性の有無による違いの方が大きかった（図 6）．

たとえば**図 11** の Keeping trial タイプの「階層性のある規則：本課題のみ」の例を見ると，エージェントは 6，15，82，88 番目の行動計画ユニットを組み合わせることで，ゴールの位置の予測を達成している．長い 1 つの系列を 4 つの小分節の連なりとして記憶しているという意味では，本アルゴリズムはチャンク化を行ったといえる．

さらに 88 番目の行動計画ユニットは ABBABBABBABBABBABBA を，82 番目の行動計画ユニットは ABAABAABAABAABAABAAB を，15 番目と 6 番目はそれぞれ BB と A のゴールの位置をコードしていた．すなわち「BBA」からなるパターンと「AAB」からなるパターンの大きく 2 つに分類していた．エージェントは，課題構造に応じたチャンク化を実現することで問題の解決を早めたととらえられる．

また，Keeping trial タイプの階層性のある規則とない規則の本課題のみにおいて，失敗頻度について結果を平均が 0 で標準偏差が 1 に正規化したデータについて自己相関関数分析を行ったところ（**図 12**），階層性のある規則は顕著な周期性を示した．図 12 では 3 周期ごとにピークが現れ，21 周期目により高いピークを示している．そのためエージェントは，階層性のある規則においては，初めに AAB や BBA の 3 試行を単位に課題の解決を試みるため，AAB と BBA の切り替わりのタイミングを間違えたり，21 周期目および 42 周期目において AAB と BBA が入れ替わったのに気づかないために失敗の頻度が最も高くなると考えられる．これも，エージェントは 3 試行を単位としたチャンク化を行い，それによって学習が促進されたと理解できる．

ただしチャンク化といっても実際のラットの系列学習でなされたように項目間の間隔を長くするといった

外的な操作がなされたわけではなく，既存の行動計画ユニットに基づいて系列が区切られた，いわば内的な要因によるものである．そのため，外的な操作の影響をいかにアルゴリズムに取り入れるかが課題である．

6. ま と め

本研究では，我々が以前に提案した学習アルゴリズムが，どのような課題構造において過去に蓄積された経験を役立てるかを検討した．そのために，Keeping trial タイプとともに Resetting trial タイプという失敗時の振舞いが前回の我々の報告とは異なるタイプの課題も用意し，階層性のある規則とそうでない規則とで過去の経験の有無の影響を比べた．その結果，エージェントが誤った道端に辿り着くたびにゴールの位置変化の規則がリセットされる Resetting trial タイプの方が，規則に階層性があるかないかにかかわらず，エージェントは過去の経験を新規な状況に利用しやすいことが分かった．また，Keeping trial タイプであっても，規則に階層性があるような課題なら学習の干渉を示さず，過去に蓄積した経験をより良い解の探索に役立てうることが分かった．

さらに，規則に階層性がある方が学習が速く進んだ理由として，ラットの系列学習において見られるチャンク化との関連性について考察することで，チャンク化による学習の促進が示唆された．

今後の課題として，刺激般化の導入，外的要因を取り入れることによるチャンク化の促進，などがあげられる．

謝辞 本論文の査読者に深く感謝します．査読者の指摘により，本論文を大きく改善することができました．また，有益なコメントをくださり，あるいは面倒な編集などの手続を担当してくださった MPS 研究会の諸先生方にも感謝の意を表します．

参 考 文 献

1) 青田佳士，山口陽子：エピソード記憶編集による迷路課題の学習アルゴリズムの提案，情報処理学会論文誌：数理モデル化と応用，Vol.47, No.SIG14(TOM15), pp.125–138 (2006).
2) 宮崎和光，小林重信：Profit Sharing の不完全知覚環境下への拡張：PS-r*の提案と評価，人工知能学会論文誌，Vol.18, pp.286–296 (2003).
3) 高木裕子：非漢字系日本語学習者における漢字パターン認識能力と漢字習得に関する研究，世界の日本語教育，Vol.5, pp.125–138 (1995).
4) Hulse, S.H. and Dorsky, N.P.: Structual complexity as a determinant of serial pattern learning, *Learning and Motivation*, Vol.8, pp.288–506 (1977).
5) Fountain, S.B., Henne, D.R. and Hulse, S.H.: Phrasing cues and hierarchical organization in serial pattern learning by rats, *Animal Learning and Behavior*, Vol.11, pp.193–198 (1979).
6) Capaldi, E.J. and Molina, D.J.: Element discriminability as a determinant of serial-pattern learning, *Journal of Experimental Psychology: Animal Behavior Processes*, Vol.10, pp.30–45 (1984).
7) Wallace, D.G. and Fountain, S.B.: What is learned in sequential learning? An associative model of reward magnitude serial-pattern learning, *Journal of Experimental Psychology: Animal Behavior Processes*, Vol.28, pp.43–63 (2002).
8) Wallace, D.G. and Fountain, S.B.: An associative model of rat serial pattern learning in three-element sequences, *Quarterly Journal of Experimental Psychology*, Vol.56B, pp.301–320 (2003).
9) Hulse, S.H. and Dorsky, N.P.: Serial pattern learning by rats: Transfer of a formally defined stimulus relationship and the significance of nonreinforcement, *Animal Learning and Behavior*, Vol.7, pp.211–220 (1979).

(平成 18 年 4 月 27 日受付)
(平成 18 年 6 月 16 日再受付)
(平成 18 年 7 月 14 日採録)

青田 佳士（正会員）

昭和 45 年生．平成 12 年東京工業大学大学院情報理工学研究科数理・計算科学専攻博士後期課程単位取得退学．同年科学技術振興機構戦略的基礎研究推進事業『脳を創る』山口チーム研究員．理化学研究所脳科学総合研究センター創発知能ダイナミクス研究チーム非常勤研究員．平成 16 年横浜国立大学大学院国際社会科学研究科助手．エピソード記憶による学習理論の研究に従事．情報処理学会数理モデル化と問題解決研究会，日本神経回路学会各会員．

山口　陽子

昭和 30 年生．昭和 56 年東京大学大学院薬学系研究科製薬化学専攻博士課程中退．昭和 57 年東京大学薬学部教務職員．昭和 59 年東京大学薬学部助手．平成 5 年東京大学薬学部講師．同年東京電機大学理工学部情報科学科助教授．平成 8 年東京電機大学理工学部情報科学科教授．平成 12 年より理化学研究所脳科学総合研究センター脳型知能システム研究グループ創発知能ダイナミクス研究チームチームリーダー．また，兼任として平成 11 年科学技術振興機構戦略的基礎研究推進事業『脳を創る』「海馬の動的神経機構を基礎とする状況依存的知能の設計原理」研究代表者．平成 12 年東京電機大学大学院理工学研究科客員教授（生物情報論）．同年東京電機大学理工学研究科講師（生物情報論）．平成 15 年東京大学理学部講師（生物情報プログラム：数理神経生物学）．同年東京大学工学部講師（脳科学入門）．計算論的神経科学の研究に従事．薬学博士．平成 6 年度日本神経回路学会論文賞受賞．日本生物物理学会，日本物理学会，日本神経回路学会，日本神経科学会，電子情報通信学会，計測自動制御学会，ニューロエソロジー学会，数理生物学会，Society for Neuroscience 各会員．