

# 20XX 年度人工知能学会全国大会・L<sup>A</sup>T<sub>E</sub>X スタイルファイル

L<sup>A</sup>T<sub>E</sub>X Style file for manuscripts of JSAI 20XX

相良博喜 <sup>\*1</sup>

Hiroki Sagara

屋藤翔麻 <sup>\*2</sup>

Yato Shoma

草場壽一 <sup>\*3</sup>

Shuichi Kusaba

<sup>\*1</sup>九州大学

Kyushu University

<sup>\*2</sup>東京工芸大学

Tokyo Polytechnic University

<sup>\*3</sup>東京大学

University of Tokyo

後で書く Here is an abstract of up to 150 words in English or 300 characters in Japanese. This document describes the format guidelines for Japanese manuscripts in L<sup>A</sup>T<sub>E</sub>X of the annual conference of JSAI. This is also a sample of a formatted manuscripts (see 2.4 for writing the abstract).

## 1. 研究背景・目的

強化学習 (Reinforcement Learning; RL) の分野は過去 10 年間で目覚ましい成功を収め、多様なタスクへの対応、意思決定から複雑なゲームの習得に至るまで、その応用範囲を広げてきた。これらの進展は、計算能力の向上と機械学習手法、特にディープラーニングのブレークスルーによって推進されてきた。しかし、報酬信号が希薄な環境における効率的な探索は依然として大きな課題である。このような環境では、エージェントが複雑な状態空間を効率的にナビゲートし、限られたフィードバックから学習する能力が求められる。この問題に対する有望な解決策として、内発的動機 (Intrinsic Motivation; IM) が挙げられる。IM は、新奇性や驚きなどの内部信号を活用して探索を導き、学習効率を高める。この分野で注目すべき研究の一つが、Random Network Distillation (RND) である。RND は、新奇性検出に基づく手法であり、固定されたランダムなターゲットネットワークと予測モデルを用いることで、エージェントに探索の指標となる内発的報酬を提供する。具体的には、予測モデルがターゲットネットワークの出力を模倣しようとする際に生じる誤差を、新奇性の指標として用いる。この手法は、報酬が希薄な環境で一定の成功を収めたが、固定されたターゲットネットワークに依存するため、内発的報酬の多様性が低く、学習が進むにつれて動機付け信号が消失するという課題がある。こうした課題を克服するために提案されたのが、Self-supervised Network Distillation (SND) である。SND は、ターゲットネットワークを自己教師あり学習によって動的に学習させることで、ターゲットモデルの表現を適応的に進化させ、新奇性検出の精度を向上させる。SND は、RND の一般化された形として位置づけられ、より高い性能と柔軟性を実現した。本研究では、この SND フレームワークをさらに改良するために 2 つの取り組みを行った。最初の取り組みは、内発的動機に基づく行動において、エピソード終了時の挙動を調整するための最適化である。内発的動機に基づく行動は、新たな状態に対して無作為に取り組むため、短期間で敵と衝突するなどしてエピソードが終了してしまい、効率的な学習が妨げられる可能性がある。これに対処するため、高い内発的動機によってエピソードが終了した場合、その内発的動機に基づく報酬にマイナスの補正を加える方法を試みた。しかし、このアプローチでは、学習に要するステップ数

の短縮やパフォーマンスの顕著な向上には繋がらなかった。これにより、次の改良に焦点を移した。2 つ目の取り組みは、内発的動機の学習方法を見直すことである。特に対照学習を行う際、Noisy TV 問題が発生することを考慮した。この問題は、エージェントが一時的な新奇性に過度に引き付けられ、環境の本質的な目標から注意が逸れる現象を指す。たとえば、Pitfall のようなゲームでは、NPC が点減するプリンクやランダムな動作がエージェントの関心を不適切に引きつける可能性がある。他の環境でも、初期状態におけるランダムなキャラクターの無作為な前後移動がノイズとして作用し、エージェントの探索効率を妨げる可能性がある。これらの問題に対処するため、対照学習を行う際に前後フレームを考慮し、連続するフレームを同一の状態として学習させる手法を導入した。この方法により、初期状態でのランダムな動作や視覚的なノイズをエージェントの興味の対象から外し、より本質的な違いに基づく状態に関心を移すことが可能となった。このアプローチにより、より短期間で報酬を得ることができ、探索効率の向上が実現した。我々の実験では、これらの取り組みが探索効率と報酬取得の改善に寄与することを確認した。モンテズマの復讐や ProcGen ゲームなどの困難な環境でテストを行った結果、Noisy TV 問題を効果的に緩和し、ベースラインモデルと比較して優れたパフォーマンスを示した。我々の手法は、特に初期状態でのランダムなノイズの影響を軽減し、エージェントが重要な状態遷移を識別し、効率的に活用する能力を向上させた。本研究は、内発的動機に基づく強化学習の新たな可能性を提示し、報酬が希薄な環境における探索効率を大幅に改善するための道を開く。

## 2. 関連研究

## 3. 提案手法

### 3.1 SND の基本的な手法

SND は、内発的動機付けを活用し、報酬が希薄な環境における探索効率を向上させる手法である。具体的には、強化学習における報酬に、内発的報酬  $r_t^{\text{intr}}$  と外的報酬  $r_t^{\text{ext}}$  の合成を用いて、新奇性の高い行動へエージェントを誘導する。

$$r_t = r_t^{\text{ext}} + \eta \cdot r_t^{\text{intr}}$$

ここで、 $\eta$  は内発的報酬の重み付け係数である。内発的報酬は、予測モデル  $\Phi_P$  とターゲットモデル  $\Phi_T$  の間の誤差を内発的報酬として利用する。この誤差は以下の式で定義される：

$$r_t^{\text{intr}} = \|\Phi_T(s_t) - \Phi_P(s_t)\|_2^2,$$

ここで、 $s_t$  は現在の状態を表し、 $\|\cdot\|_2^2$  は L2 ノルムの 2 乗を意味する。この誤差は、新奇性を評価する指標として利用され、新しい状態ほど大きな値を取る。予測モデルは、以下の損失関数で定義される通り、ターゲットモデルの出力を模倣するよう学習が行われる。

$$L_P = \frac{1}{N} \sum_{s \in S} \|\Phi_T(s) - \Phi_P(s)\|_2^2,$$

つまり、RND のようにターゲットモデルが固定されたアルゴリズムの場合、最終的には内発的報酬が収束してしまい内発的報酬が発生しなくなる。SND では、自己教師あり学習を利用してターゲットモデルを動的に更新し、状態表現の分散性を向上させる手法である。本研究では既往論文にて提案されている SND-V というアルゴリズムを用いた。SND-V におけるロス関数は以下の通りである：

$$L_T = \sum_n (\tau_n - \|Z_n - Z'_n\|_2^2)^2,$$

ただし、 $\tau_n = 0$  の場合、 $Z_n$  および  $Z'_n$  は同じ状態に対してノイズを付与した状態に対する特徴ベクトルを表す。 $\tau_n = 1$  の場合、 $Z_n$  および  $Z'_n$  は全く異なる状態、異なるエピソード、異なるステップに対してノイズを付与した状態に対する特徴ベクトルである。

ノイズの付与方法は既往研究と同様に、以下の手法を用いた。

- 一様ランダムノイズ：ピクセル値に  $[-0.2, 0.2]$  の範囲でノイズを追加。
- ランダムマスキング：画像タイルをランダムにマスクする（タイルサイズは 2、4、8、12、16 ピクセル）。
- ランダム畳み込み：ランダムフィルターを用いた畳み込みを適用。

### 3.2 提案する改良手法

本研究では、SND の基本的な枠組みに加え、探索効率および報酬取得の向上を目指して以下の 2 つの改良手法を提案する。

### 3.3 エピソード終了時の報酬調整

エピソード終了時における新たなルールを導入した。具体的には、エピソード終了時に `level_complete = false` かつゲーム終了までのステップ数に到達していない場合、その状態を”死亡”とみなす。この場合、報酬は以下のように定義される：

$$r_t = \begin{cases} r_{\text{ext},t} - \eta \cdot r_{\text{intr},t}, & \text{死亡の場合,} \\ r_{\text{ext},t} + \eta \cdot r_{\text{intr},t}, & \text{それ以外の場合.} \end{cases}$$

内発的報酬が大きい状況にて死亡が発生した場合、その行動がエージェントの継続的な探索を妨げてしまっていると考えられる。このため、死亡時に内発的報酬にマイナスの補正を加えることで、エージェントがより長期的に探索を行うことができ、効率的な学習が可能となることが期待される。

### 3.4 類似状態ペア選択方法の改善

前章で述べた通り、Noisy TV 問題とはエージェントが時系列的なノイズに不適切に関心を持ち、探索が非効率化する現象を指す。この問題を解決するため、SND-V に下記の改良を加えた。

1. 同一環境および STEP が  $-N$  から  $+N$  ステップの範囲内にある  $2N + 1$  個の状態からランダムに選択する。
2. 選択した状態にノイズを加え、類似状態ペアを構成する。
3. 非類似状態の選択は既往研究と同様にランダムに行う。

このアプローチでは、元の SND-V が従来採用していた、完全に同一の状態ペアからノイズを加える方式よりも多様な状態表現を学習できる。特に周期的に変化する状態を同一状態と捉え、無駄な内発的報酬の発生を抑えることが期待される。これにより、状態空間全体でより精度の高い新奇性検出が可能となり、エージェントが短期的なノイズではなく本質的な環境変化に基づいて行動できるようになる。

## 4. 実験・考察

## 5. まとめ

### 5.1 ファイル形式・サイズ

Adobe(R) PDF (Portable Document Format) 形式のファイルを提出してください。その他の形式での提出は受け付けませんので、ご注意ください。ファイルサイズはファイル受付システムの制限がありますので、3MB 以下にしてください。また、ファイル名の拡張子は .pdf にしてください。

### 5.2 ヘッダー部分

2015 年大会から講演番号およびヘッダの会議名は、原稿提出後に運営側で挿入しますので、著者が作成する原稿には記入しないでください。

### 5.3 タイトルと著者名

原稿におけるタイトル・筆署名等は、発表申し込み時に入力したものに一致させてください。**原稿のタイトルや著者名が発表申し込み時と異ならないように十分注意してください。**

### 5.4 アブストラクト

概要 (Abstract) には (1) 目的と (2) 結果の概要あるいは結論を含めてください。必要に応じ方法を記載してください。**(1)(2) の記述のない場合は不採択となることがあります。**

### 5.5 原稿枚数

下記指定フォーマットで A4 用紙 2 ページです。希望によりさらに 2 ページまで無料で追加できます。

### 5.6 国際セッション

**国際セッションについては英語のみとなります。**国際セッション論文は、会議時には大会ホームページに掲載されますが、特に本会議と関連が深く、優秀とみなされた論文は、その拡張版の New Generation Computing 誌の特集号への投稿を推奨する予定です。

## 6. L<sup>A</sup>T<sub>E</sub>X 原稿のスタイル

論文のスタイルを統一するために、原稿はできるだけ以下のスタイルファイルを使ってください。基本的には 2000 年度までの全国大会論文集用に配布されていた原稿用紙と同じ形に仕上がるようになっています。スタイルファイル自体は昨年度用のものと同一です。

スタイルファイルは以下のように指定してください。

ASCII 版 L<sup>A</sup>T<sub>E</sub>X2.09 なら

```
\documentstyle[twocolumn,jsaiac]{jarticle}
```

NTT 版 L<sup>A</sup>T<sub>E</sub>X2.09 なら

```
\documentstyle[twocolumn,jsaiac]{j-article}
```

ASCII 版 L<sup>A</sup>T<sub>E</sub>X 2<sub>ε</sub> なら

```
\documentclass[twocolumn]{jarticle}
\usepackage{jsaiac}
```

欧文使用の L<sup>A</sup>T<sub>E</sub>X2.09 なら

```
\documentstyle[twocolumn,jsaiac]{article}
```

欧文使用の L<sup>A</sup>T<sub>E</sub>X2<sub>ε</sub> なら

```
\documentclass[twocolumn]{article}
\usepackage{jsaiac}
```

“jsaiac.sty” は以上のように、標準で配布されるパッケージである jarticle.sty, j-article.sty, jarticle.cls (欧文論文の場合は article.sty, article.cls) を主のスタイルファイルとして、それにオプションという形で使うように設定されています。“jsaiac.sty” はタイトル部分、文字組の調整、一部脚注の調整以外は行っていないが、共通版にする関係から、オプションの twocolumn の指定が必須です。以上のことから、“jsaiac.sty” を使う場合は、上記の指定方法を必ず守るようお願いいたします。

“jsaiac.sty” は以上の 3 つの L<sup>A</sup>T<sub>E</sub>X のバージョンに対応しています。NTT 版の L<sup>A</sup>T<sub>E</sub>X 2<sub>ε</sub> は動作確認を行っていません。

## 6.1 “jsaiac.sty” を使うことで指定が不要なもの

“jsaiac.sty” を使えば、次の指定は必要ありません。

- ページ番号の書式
- マージン等の位置
- 用紙 (A4) 用紙
- 本文 (2 段組)

## 6.2 “jsaiac.sty” を使うことで指定が必要なもの

○ タイトル領域: “jsaiac.sty” の書き方のきまりは次のようになります。

- タイトル:

```
\title{
  \jtitle{和文タイトル}
  \etitle{欧文タイトル}
}
```

なお、欧文論文の場合は、単に

```
\title{欧文タイトル}
```

としてください。

- 筆者名 (同一所属の場合):

```
\author{%
  \jname{筆者氏名}
  \ename{Given-name Surname}
\and
  \jname{筆者氏名}
  \ename{Given-name Surname}
\and
  Given-name Surname
}
```

なお、欧文論文の場合は、単に

```
\author{%
  Given-name Surname
\and
  Given-name Surname
}
```

としてください。 \jname{ } や \ename{ } は指定しません。

- 筆者名 (所属が異なる場合):

```
\author{%
  \jname{第 1 筆者氏名\first{}}
  \ename{Given-name Surname}
\and
  \jname{第 2 筆者氏名\second{}}
  \ename{Given-name Surname}
\and
  Given-name Surname\third{ }
}
```

所属が異なる場合、違いを識別するため、

```
\first \second \third ....
```

の指定を加えてください。これは同一の所属は同一のコマンドを与えます。さらに所属の方にも、該当する \first, \second, \third... の指定を加えますが、その順序は自由です。具体的な出力は、\first と指定すると、“\*1” が筆者名の右上 (所属は左上) に表示されます。これは単純なコマンドです。全部で 9 つ用意してあります。以下がその内訳です。

```
\def\first{\hbox{$\m@th^{\ast 1}$\hss}}
\def\second{\hbox{$\m@th^{\ast 2}$\hss}}
\def\third{\hbox{$\m@th^{\ast 3}$\hss}}
\def\fourth{\hbox{$\m@th^{\ast 4}$\hss}}
\def\fifth{\hbox{$\m@th^{\ast 5}$\hss}}
\def\sixth{\hbox{$\m@th^{\ast 6}$\hss}}
\def\seventh{\hbox{$\m@th^{\ast 7}$\hss}}
\def\eighth{\hbox{$\m@th^{\ast 8}$\hss}}
\def\ninth{\hbox{$\m@th^{\ast 9}$\hss}}
```

- 所属: \jname{ } や \ename{ } の指定は筆者名の場合と同じです。次のように指定します。

```
\affiliate{
  \jname{\first{}}所属和文 1}
  \ename{Affiliation #1 in English}
\and
  \jname{\second{}}所属和文 2}
  \ename{Affiliation #2 in English}
\and
  \third{ }Affiliation #3 in English
}
```

なお、欧文論文の場合は、単に

---

```

\affiliate{
  \first{ }Affiliation #1 in English
\and
  \second{ }Affiliation #2 in English
\and
  \third{ }Affiliation #3 in English
}

```

とします。`\jname{ }`や`\ename{ }`は指定しません。ただし、和欧文とも所属が同一の場合は、`\first`の指定は不要です。

- 連絡先： 代表者の氏名，所属，所在地，電話番号，Fax 番号，e-mail アドレスなどをお書き下さい。

```

\jaddress{氏名，所属，住所，電話番号，Fax 番号，電子メールアドレスなど}

```

とすれば，脚注の位置に“連絡先: ”という形で出力されます。なお，欧文論文の場合は，

```

\address{name, affiliation, address,
phone number, Fax number,
e-mail address}

```

とすれば，脚注の位置に“Contact: ”という形で出力されます。

#### ○ その他

- 脚注： 脚注は，下にある例のように <sup>\*1</sup> 通常の  $\text{\LaTeX}$  (`[?]`) の書き方である `\footnote{ }` を使って書きます。
- 参考文献： `j(-)article.cls(sty)` (欧文論文の場合は `article.sty(cls)`) が用意しているものを使うことになります。著者名，文献名，ジャーナル (出版社)，発行年など，イニシャル，略語のスタイル，記載順などは論文誌の規則に従ってください。 $\text{\LaTeX}$  を使う場合は論文誌用の  $\text{\LaTeX}$  スタイルファイルと同時に配布されている“`jsai.bst`”を使うことをお勧めします。参照ラベルの `\cite{ }` も使えます。最後の部分に参考文献のサンプルが添付してあります。
- 他のコマンド 通常の  $\text{\LaTeX}$  の組版と変わりありません。`j(-)article.clssty` (欧文論文の場合は `article.sty(cls)`) で扱えるものはすべて使うことができます。

## 参考文献

[Knuth 84] Knuth, D. E.: *The  $\text{\TeX}$ book*, Addison-Wesley (1984), (邦訳:  $\text{\TeX}$  ブック, 斎藤 信男 監修, 鷲谷 好輝 訳, アスキー出版局 (1992)).

[Lamport 86] Leslie, L:  *$\text{\LaTeX}$ : A Document Preparation System* (Updated for  $\text{\LaTeX}2\epsilon$ ), Addison-Wesley, 2nd edition (1998) (邦訳: 文書処理システム  $\text{\LaTeX}2\epsilon$ , 阿瀬 はる美 訳, ピアソン・エデュケーション, (1999)).

---

\*1 この例が脚注です。