シミュレータを使用した実験の検討

1. 実験概要

2019年度実施したSCCFの特性調査実験では、カテゴリ数が状態行動数の削減のために6個であることや、2回前までの発話内容を格納するなど、状態行動数が少ない状態での学習ができることを確認した。

本実験では、状態行動数を増やし、深層強化学習アルゴリズムをSCCFに適用することで学習が可能であるかについて、性能を調査する。

ただし、実験を簡単にすることで、サイクルを早くするため、IDAは実際の利用者ではなく、エージェントを評価するシミュレータを対象として実験を実施する。

対象アルゴリズム

- REINFORCE
- DQN（DQNの派生系も実装する予定です）
- そのほか未定

2. エージェントのパラメータ/実験環境

- エージェントが発話するカテゴリの総数: C
- エージェントが持つ、過去の発話履歴の長さ: T (Tステップ前までの発話を保存)

この環境で、エージェントのもつ状態行動の総数は $C^T * T = C^(T+1)$

3. シミュレータのパラメータ

- シミュレータがエージェントが与える報酬:

$$r \in \{1, 0, -1\}$$

DQNの場合、報酬値をClipping(1,0,-1の値に絞る)することで学習が安定化するため。

2019年度のREINFORCEの実験で性能の良かった実験設定（報酬設定B）

- シミュレータがエージェントに与える報酬がrである確率

$P(r|s,n)$：特定の発話$c \in C$がn回連続で続いた時、さらにcが発話された条件のもと、報酬rを与える確率

具体例:

$$
\begin{aligned}
\leftline{P(r=1|(s,n))=[} \\
&[0.5, 0.9, 0.9\ldots](2\text{回目以降}\alpha=1.0)(c=1) \\
&[0.4, 0.45, 0.1, \ldots](3\text{回目以降}:\alpha=0.2)(c=2) \\
&[0.5, 0.15, \ldots](3\text{回目以降}:\alpha=0.3)(c=3) \\
&[0.35, 0.5, \ldots](3\text{回目以降}:\alpha=0.3)(c=4) \\
&[0.45, \ldots](3\text{回目以降}:\alpha=0.1)(c=5) \\
&[0.4, \ldots](3\text{回目以降}:\alpha=0.1)(c=6) \\
]
\end{aligned}
$$

SCCF特性調査実験において、特徴的だった実験協力者4名のデータから、協力者固有のパラメータを用意する。

各ステップにおける減衰項（連続して発話されたときの評価の減衰に関する値）

協力者固有のパラメータ:

- 特定の状態がn回続いた時後にカテゴリmを発話し、報酬rを受け取る確率: P(r|(m|n))

  三次元配列のようなもの
- 減衰係数 α_(r,m)

  上記の三次元配列の値を決めるもの。

全ての被験者において、特徴的な箇所（1回目の発話よりも2回連続した時の発話の方が評価が高い、など）があるので、

途中まで決められた値 -> 途中からは減衰係数α_(r,m)をかけるのように報酬を与える確率を変化させる。

αはSCCF特性調査実験で、利用者が実際に報酬を与えた確率から求めている。

- データのなかった箇所: α=0.1
- データがある箇所: α: 直前の2回の勾配を減衰係数に指定 (ex 0.5, 0.3 -> α=0.3/0.5)

  ただし、1以上になる場合（確率が増加して終わったもの）については、0.2(仮)で固定
- user1
  Chatbot-REINFORCEでの実験結果:1,2,3,4,5,6位と順に発話確率が高くなった

```
P(r|(m|n)) =
[ (r = 1)
    [0.5, 0.9, 0.9 ... (2回目以降 α=1.0)] (m=1)
    [0.4, 0.45, 0.1, ...] (3回目以降: α=0.2) (m=2)
    [0.5, 0.15, ...] (3回目以降: α=0.3) (m=3)
    [0.35, 0.5, ...] (3回目以降: α=0.3) (m=4)
    [0.45, ... ](3回目以降: α=0.1) (m=5)
    [0.4, ... ](3回目以降: α=0.1) (m=6)
],
[ (r = 0)
    [0.45, 0.1, 0.1, ... (2回目以降α=0.1)] (m=1)
    [0.55, 0.3, ...] (3回目以降: α=0.55) (m=2)
```

```
    [0.5, 0.1, ...](3回目以降: α=0.1) (m=3)
    [0.55, 0.5, ...] (3回目以降: α=0.9) (m=4)
    [0.4, ... ](3回目以降: α=0.1) (m=5)
    [0.5, ... ](3回目以降: α=0.1) (m=6)
],
[ (r = -1)
    [0.001, ...] α=0.1
    [0.05, 0.2, ...] α=0.2
    [0.02, ...] α=0.1
    [0.1, ...] α=0.1
    [0.15, ...] α=0.1
    [0.1, ...] α=0.1
]
```

- user4
  Chatbot-REINFORCEでの実験結果: 1,3位のカテゴリの発話確率が増加

  ```
  調査中
  ```

- user7
  Chatbot-REINFORCEでの実験結果: 2,4位のカテゴリの発話確率が増加

  ```
  調査中
  ```

- user8
  Chatbot-REINFORCEでの実験結果: 3,4位のカテゴリの発話確率が増加

  ```
  調査中
  ```