

# シミュレータ設定の検討

シミュレータのパラメータは以下の通り

- 発話カテゴリ: C
- 発話内容を記録する長さ: T (Tステップ前までの発話を保存)
- 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)
- 報酬設定: B (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。
  - REINFORCEの実験でこちらの方が精度が良かったため

## 協力者固有のパラメータ

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。

協力者固有のパラメータ:

- 特定の状態がn回続いた後にカテゴリ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  $\alpha(r,m)$   
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 $\alpha(r,m)$ をかけるのように報酬を与える確率を変化させる。

$\alpha$ 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:  $\alpha=0.1$
- データがある箇所:  $\alpha$ : 直前の2回の勾配を減衰係数に指定 (ex 0.5, 0.3 ->  $\alpha=0.3/0.5$ )  
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  $\alpha=1.0$ )] (m=1)  
  [0.4, 0.45, 0.1, ...] (3回目以降:  $\alpha=0.2$ ) (m=2)  
  [0.5, 0.15, ...] (3回目以降:  $\alpha=0.3$ ) (m=3)  
  [0.35, 0.5, ...] (3回目以降:  $\alpha=0.3$ ) (m=4)
```

```

[0.45, ... ](3回目以降:  $\alpha=0.1$ ) (m=5)
[0.4, ... ](3回目以降:  $\alpha=0.1$ ) (m=6)
],
[ (r = 0)
  [0.45, 0.1, 0.1, ... (2回目以降 $\alpha=0.1$ )] (m=1)
  [0.55, 0.3, ...] (3回目以降:  $\alpha=0.55$ ) (m=2)
  [0.5, 0.1, ...](3回目以降:  $\alpha=0.1$ ) (m=3)
  [0.55, 0.5, ...] (3回目以降:  $\alpha=0.9$ ) (m=4)
  [0.4, ... ](3回目以降:  $\alpha=0.1$ ) (m=5)
  [0.5, ... ](3回目以降:  $\alpha=0.1$ ) (m=6)
],
[ (r = -1)
  [0.001, ...]  $\alpha=0.1$ 
  [0.05, 0.2, ...]  $\alpha=0.2$ 
  [0.02, ...]  $\alpha=0.1$ 
  [0.1, ...]  $\alpha=0.1$ 
  [0.15, ...]  $\alpha=0.1$ 
  [0.1, ...]  $\alpha=0.1$ 
]

```

- 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