2024 年度卒業研究概要

TCU タワーバトルの対戦 COM における強化学習を用いた難易度設計

大谷 紀子 研究室
2172040　亀井　絢太

1. 背景と目的

どうぶつタワーバトルとは，オンラインで行う 1 対 1 のターン制のゲームである．各プレイヤが交互に，動物のイラストが描かれたピースを 1 つずつステージの上に積み上げていき，ステージからピースを落とした方が負けというルールである．どうぶつタワーバトルをオフラインゲームにした TCU タワーバトルには，プレイヤ vs COM と COM vs COM という 2 種類の対戦形式がある．TCU タワーバトルで使用されている対戦 COM は，事例ベース推論によってピースの積み方を決定する．事例ベース推論は，過去の類似した事例の解法に基づいて新たな事例の解法を導く推論方法である．事例ベース推論に用いる過去事例データベースは，各ピースをランダムに落下させ，積み上げに成功したときの概形と，落下させたピースの角度，位置を種類ごとに保存することで，事前に作成しておく．対戦時には，データベース中の過去事例のうち，現在の事例に最も類似した事例の解法に基づいて，現在の事例の解法を導く．

現在の TCU タワーバトルでは，対戦 COM の選択が過去の事例に大きく依存する問題がある．過去事例データベースに登録されている事例のパターンに偏りがあると，明らかに失敗しそうな位置や角度で落下させることがある．また，安定志向と挑戦志向の 2 通りの対戦 COM が実装されているが，具体的な難易度選択はできない．プレイヤが TCU タワーバトルをより楽しむためには，自分のレベルにあった COM と対戦することが必要である．

本研究では，プレイヤの TCU タワーバトルへの満足度を向上させることを目的として，難易度の異なる COM を構築する．

2. 強化学習による対戦 COM の学習

本研究では，対戦 COM の行動方針，すなわち落下させるピースの角度と位置を導くルールをニューラルネットワーク（以下　NN）で表現する．NN への入力としては，落下させるピースの位置，角度，種類に加え，タワーの概形，ステージの範囲，および現在までに積みあがっているタワーの総ピース数とし，NN のパラメータの学習に強化学習を用いる．

強化学習とは，学習の主体であるエージェントが，環境との相互作用を通じて最適な行動方針を学習する手法である．エージェントが環境の状態を観測し，環境から得た情報に基づいて行動する．エージェントは行動によって変化した新たな環境と報酬を受け取り，受け取った報酬をもとに将来的な累積報酬が最大化されるように行動方針を更新する．本研究におけるエージェントは対戦 COM である．

本研究における強化学習の手順を以下に記す．

1. 対戦 COM が単独でピースを積み上げるプレイを N 回繰り返し，各プレイで落下させたピースとタワーに関する情報，およびプレイによって得られた報酬をプレイデータとして記録する．
2. N 回分のプレイデータを M 回ずつに分割する．
3. 分割されたプレイデータを順に用いて行動指針を更新する．
4. 3 を R 回繰り返す．

過学習や性能の悪化を防ぐために，3 での行動

表 1：報酬付与時のピース数の条件

	積み上げ成功時		ゲームオーバー時	
	報酬 5	報酬-5	報酬 20	報酬-30
easy	7 未満	7 以上	4 以上 7 未満	7 以上
normal	7 以上 10 未満	7 未満，10 以上	7 以上 10 未満	7 未満，10 以上
hard	0 以上	-	-	0 以上

表 2：学習条件の差と落下位置制限の有無

	N	M	R	落下位置制限
easy	256	64	1	あり
normal	1024	256	3	なし
hard	2024	512	4	なし

方針の更新量は低く抑えられている．4 において同じデータで行動方針の更新を繰り返すことで，十分な学習を実現する．

3. 難易度の異なる対戦 COM の構築手法

3 つの対戦 COM，easy，normal，hard において報酬付与時のピース数の条件を表 1 に示すように変化させることで，3 つの対戦 COM の難易度を調整した.hard の対戦 COM に対しては,ピースが 7，10，14 個積み上げられたときや，タワーの高さが 3，6，9 に到達したときにそれぞれ 5，10，15 の報酬を与える．また，同じデータで行動指針の更新を繰り返すため，ゲームオーバー時の報酬も重要であることから，ゲームオーバー時に積み上げたピースの数だけ追加で制の報酬も与える．

報酬付与時の条件に加え，行動方針を更新する際の変数 N, M, R, およびピースの落下位置の制限の有無で難易度を調整する．対戦 COM ごとの設定内容を表 2 に示す．easy の対戦 COM では，意図的に敗北するような行動をとることがあるため，ピースを落下させる位置をステージの上空部分に限定して，弱くなりすぎないように制限する．

4. 評価実験

評価実験では，16 名に対し本研究で開発した TCU タワーバトルをそれぞれの難易度ごとに 3 回ずつプレイさせ，難易度ごとの積み上げたピース数を記録し，3 つの対戦 COM の強さに違いがあるかどうかを評価する．また，ゲーム全体を通しての満足度と，対戦 COM に対する満足度は 5 段階でアンケートを取る.それぞれの満足度については，評価値が高いほど高評価とし，細かい評価を自由記述で回答させる．

表 3：16 名が 3 回ずつゲームをプレイした結果

	最頻値	最小値	最大値	平均値
easy	6	2	14	5.4
normal	8	3	12	7.3
hard	11	3	26	9.6

難易度ごとに記録されたピース数の最頻値，最小値，最大値，平均値を表 3 に示す．また，ゲームに対する満足度について，被験者の 19%が 5，69%が 4，12%が 3 と回答し，対戦 COM に対する満足度については，31%が 5，31%が 4，25%が 3，13%が 2 と回答した．

自由記述には，ゲーム自体には満足している一方で，キー操作とマウス操作によってできることが違い，操作性が悪いという意見が多かった．また，全体的に hard の評価が高く，「難易度が上がるごとに COM が賢くなっていて負けてしまった」という意見がある中，easy，normal に対しては，人によって差が出ていないという評価もあった．

5. 考察

表 3 の最頻値および平均値より，本研究で実装した 3 つの対戦 COM には難易度の差をつけることができたといえる．また，対戦 COM に関するアンケートでは 2 の評価も一定数いるが，4 や 5 の高い評価が 6 割以上を占めており，ゲームに関するアンケートでは 1 と 2 の評価がなく，ほとんどが 4，5 の高評価だったため，本研究で作成した TCU タワーバトルおよび 3 つの対戦 COM は有用であったことがわかる．