

# 共創型都市開発シミュレーションゲームにおける QS-Learningを用いた協調エージェントの試作

坂田 匠悟<sup>†</sup><sup>†</sup> 静岡大学 情報学部行動情報学科福田 直樹<sup>‡</sup><sup>‡</sup> 静岡大学 学術院情報学領域

## 1 試作ゲームの概要

本研究の背景として、著者らによって過去に行われた共創型に基づいて自身の管理する都市開発シミュレーションゲームの試作 [3][4][5] がある。この試作ゲームでは、プレイヤーや他のエージェントは各々の都市にタイルを設置することができ、そのタイルから収益を得ることができる。文献 [3] では地価というパラメータが各タイルに設定されており、収益は  $t$  と地価  $p$  に基づいて、そのタイルに対応する収益が式 (1) のように決定される。

$$\text{収益} = \max(0, \text{floor}(\min(t, k * \max(j, p)))) * i \quad (1)$$

本試作ゲームでは、 $i = 2$ ,  $j = 0.5$ ,  $k = 4$  としている。

また、各都市は隣接しているという設定であり、各都市の外周に設置したタイルは他のエージェントの都市に地価の影響を与えることがある。エージェントはプレイヤーのプレイ状況や様々な価値観で都市の開発を行う。本試作エージェントでは比較しやすい指標である収益を重視したエージェントを試作した。ヒューリスティックな手法を用いた場合には、プレイヤーの楽しみ方ごとにエージェントの実装を手直しする必要があることから協調的動作を作りこむことが開発コスト的な面から困難である。本研究では、強化学習をエージェントに適用することにより、協調的な動作を行っているように見えたり収益をより多く出せるような強力なエージェントを試作することを目標とする。

A Preliminary Prototype Implementation of Cooperative Agents using QS-Learning for Collaborative City Development Games

<sup>†</sup>Sakata Shogo Faculty of Information, Shizuoka University Email: sakata.shogo.18@inf.shizuoka.ac.jp

<sup>‡</sup>Fukuta Naoki Department of Informatics, Shizuoka University Email: fukuta@inf.shizuoka.ac.jp

## 2 試作エージェント

強化学習で構成するエージェントは、都市の開発をプレイヤーと同じように行うものである。強化学習に必要な現在の状態は都市の開発しているマップから構成でき、アクションはどこの座標に何のタイルを置くかで構成できる。Q学習を用いてこのエージェントの動作を構成しようとする、状態の数が非常に膨大そのために必要な計算を現実的な時間および機材で行うことが困難となる。この条件でQ学習をエージェントに適用すると、最序盤はエージェントとして価値のある動作を行う場面があっても、その後現れる盤面の多くが未知の状態となり、Q学習による学習が行われた範囲外になってしまうという課題に対するアプローチの1つにQS-Learning[1]がある。QS-Learning[1]は、強化学習の1つでQ学習よりも条件次第では収束が速くなるものである。このQS-Learningでは $\sigma$ と呼ばれる関数を対象となる問題ごとに適切に定義することになる。今回は2種類の $\sigma$ 関数について検討を行う。

### 2.1 代表値を用いた手法

代表値を用いた手法は、横  $w$  マス、縦  $h$  マス、合計  $w * h$  マスのタイルから横  $u$ 、縦  $v$  マスのタイルを抽出し、QS-Learning 実行時に保存が必要となる状態空間を少なくする手法である。今回は  $w = 10$ ,  $h = 5$ ,  $u = 5$ ,  $v = 5$  とし、代表値は1つとばしてタイルを取得するという方法をとった。 $u$  と  $v$  の値を  $w$  と  $h$  よりも小さく設定することにより、状態空間を非常に小さくすることができ収束を速くすることが可能となる。

この手法で学習を10000回行ったところ実装者が実際に違和感なくプレイできるような状態のエージェントとなった。エージェントは収益が大きくなるような行動を重ね、マスがすべて埋まる頃に

はプレイヤーと、他のヒューリスティックな動作を行うエージェントよりも収益が高くなるという動作を確認した。

## 2.2 タイル種類削減手法

タイル種類削減手法では、配置されているタイルの種類を同じようなもので分類し、学習に利用するタイルの種類を減らした方法である。エージェントは空白タイルを含めて計6種類の中から設置するが、役割の似ている複数のタイルを1つのタイルとして認識し、学習に用いる。これによって状態空間を減らすことで、学習の収束を速くする狙いがある。

タイル種類削減手法を用いた学習は、代表値を用いた手法よりも各ステップの計算コストが高く、同じステップをシステムタイムを利用し各100000回計測し平均をとったところ、タイル種類削減手法が1.1ミリ秒で、代表値を用いた手法は5.0ミリ秒となり1ステップあたり5倍程度違うことが観測できている。学習を3000回行い、タイル種類削減手法を用いた学習を適用したエージェントに興味深い動きがみられた。エージェントは先に公園等の地価調整タイル設置していき、収益がマイナスになってから地区を配置して収益が大きくなるよう行動を行う。学習回数が少なく、学習結果と一致している状態にならなかった場合にはランダムな行動を行う。しかし公園タイルの上手な配置により、収益を保ったままエージェントが勝利する場面が非常に頻繁に見られた。

二つの手法を用いたエージェント各1体と、ヒューリスティックな手法により構成したエージェントの1体を用いてプレイされたゲームにおける、収益の増減の一例をグラフで示したものが図1である。

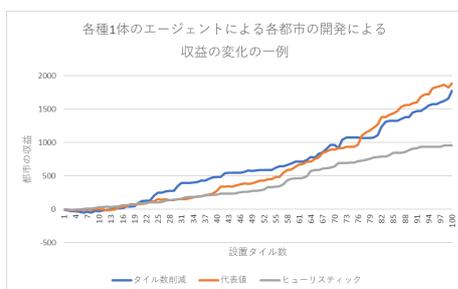


図 1: 各種1体のエージェントによる各都市の開発による収益の変化の一例

## 3 おわりに

本研究では、共創型都市開発シミュレーションゲームに敵エージェントを導入し、強化学習の1つである QS-Learning を用いたエージェントの試作について述べた。強化学習により、収益がヒューリスティックなエージェントよりも高い都市を開発するエージェントの試作を行うことができた。今後の課題としては、学習に順位などの指標学習させることの検討を計画している。また、他のエージェントに対する敵対と協調の選択に関する手法の適用 [2] も今後の課題である。

## 参考文献

- [1] Ariel Rosenfeld, Matthew E. Taylor, and Sarit Kraus. Leveraging human knowledge in tabular reinforcement learning: A study of human subjects. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, pp. 3823–3830, 2017.
- [2] Heechang Ryu, Hayong Shin, and Jinkyoo Park. Cooperative and competitive biases for multi-agent reinforcement learning. In *Proceedings of the 20th International Conference on Autonomous Agents and MultiAgent Systems, AAMAS '21*, p. 1091–1099, Richland, SC, 2021. International Foundation for Autonomous Agents and Multiagent Systems.
- [3] Shogo Sakata and Naoki Fukuta. On a preliminary implementation of enemy agent on a city development simulation game. In *Proc. 10th IIAI International Congress on Advanced Applied Informatics (IIAI AAI2021 / SCAI2021)*, 2021. poster.
- [4] 坂田匠悟, 福田直樹. 都市開発シミュレーションゲームにおけるプレイヤーエージェントの設計. 研究報告知能システム (ICS), Vol. 2021, No. 7, pp. 1–4, 2021.
- [5] 坂田匠悟, 福田直樹. 都市開発シミュレーションゲームにおける敵エージェントの試作. 情報処理学会 第 83 回全国大会, 2021.