

令和 5（2023）年度
データサイエンス学部 卒業論文

才能と運：マルチエージェントシミュレーションと強化学習を用
いた考察

学籍番号 204055
氏 名 森田 健斗
指導教員 越仲 孝文

令和 6 年 1 月 2 5 日 提出

目次

第1章 はじめに.....	3
1.1 研究の背景と目的.....	3
1.2 従来研究.....	3
1.3 本論文の構成.....	4
第2章 関連研究.....	4
2.1 マルチエージェントシミュレーション.....	4
2.2 先行研究.....	5
2.3 強化学習.....	12
第3章 提案手法.....	14
第4章 実験.....	15
4.1 シミュレーションの設定.....	15
4.2 実験 1：シングルシミュレーション.....	18
4.3 実験 2：マルチシミュレーション.....	23
第5章 おわりに.....	26
5.1 まとめ.....	26
5.2 今後の課題.....	26

第1章 はじめに

1.1 研究の背景と目的

人間誰しもが、運を都合よく解釈している。“成功したのは運が良かっただけだ”、“運も実力のうちだ”。これらの言葉に見られるように人間は成功という事象を運というあやふやな言葉でなんとなく理解して済ましている。また、運を高める方法や運を手にする心構えなどが Web や自己啓発本等で見受けられるが、これらのほとんどは事後の結果より運を判断しているので後知恵バイアスがかかっており、運と成功の因果関係は明確になっていない。人の成功が才能や努力のような人間の内在的性質に依存しているのか、それとも運次第なのかということ客観的かつ定量的に評価することは、個々人が成功に近づくことに役立つと考える。

運が成功に対して多大な影響を与えていることは世間一般的に広く言われている。しかし、成功に対して運のみが支配的なのだろうか。個々人の努力によって成功に導くことも可能なのではないかと考えられる。

本研究では“運が成功や失敗にどの程度影響し、人の努力とどちらの方が支配的であるか”を計算機上のマルチエージェントシミュレーションを通して考察する。例えば、本当は運など存在せず、努力や才能のみで成功を手にすることができるのではないか。努力次第で幸運に出会う確率を上げたり、不運を避けたりすることも可能なのではないか。このような成功と運や努力の関係に対して一つの答えを提示することを本研究の目的とする。とりわけ、失敗や成功から学ぶという人間が自然に有する学習の能力をエージェントに持たせ、そのような学習能力が人生の成功や失敗にどの程度影響するのかを明らかにする。

1.2 従来研究

昨今では運の役割が成功に対して影響を与えることを示唆する研究がたくさんなされてきた。

Liran ら[1]は経済学者の専門的成果に対して姓(family name)の初めのアルファベットが若い人ほど終身在職権を得られ, ノーベル賞やクラーク賞を受賞する可能性が高くなることを示唆した. Du ら[2]は生まれた月が年度末の人は他の人と比べて不利になるため, 生まれた月はキャリアの成功に長期的な影響を与えることを述べた. これらの運の要素に対して Taleb [3]は人生における運の役割は想像以上に大きいことを示した. 人間は自分自身や他人の成功を才能や能力によるものとみなす傾向にある一方で, 失敗や不運を偶然の副産物と片付ける傾向にある. 彼は結果の良し悪しを完全に決定論的に説明するのは不可能であると主張し, 確率やランダム性の認識の重要性を強調している. また, 成功を得るタイミングで適切に機会に遭遇することも必要であると結論づけている.

2022年にイグノーベル賞を受賞した Pluchino[4]らの論文では, シンプルなモデルによるマルチエージェントシミュレーションにより, 成功に対する運と才能の役割を現実的に定量化した. 成功は単に才能だけでは説明不可能で, むしろ運の要素が支配的であることを示唆した.

1.3 本論文の構成

先述の通り本研究では成功に対して運・努力の支配度を, マルチエージェントシミュレーションをもとに定量的に評価し, 議論する. 第2章では本研究で使用する技術を示し, 第3章でマルチエージェントシミュレーションによる手法を提案する. 第4章, 第5章で実験概要とその結果を述べ, 第6章で結ぶ.

第2章 関連技術

本研究の関連技術を述べる.

2.1 マルチエージェントシミュレーション

マルチエージェントシミュレーションとは, 複雑な実社会のマクロな現象の性質を調べるために, 何らかの環境の中で複数のエージェントを用いて行われる仮想実験(シミュレーション)を指す[5]. 社会にはいくつもの要素が絡み合う複雑な現象が溢れており, これらの複雑な現象について, 人間をエージェントと見立て, 行動ルールとエージェント同士の相互作用をモデル化してシミュレーションする[6]. これにより複雑な社会現象を説明するモデルを構築したり, そのメカニズム

や性質を理解して振る舞いを予測したりする。

マルチエージェントシミュレーションでは現実社会に見立てたある環境下でのルール・状況に沿ってエージェントが行動する(図1)。今回の研究ではマルチエージェントシミュレーションを用いて実験を行う。

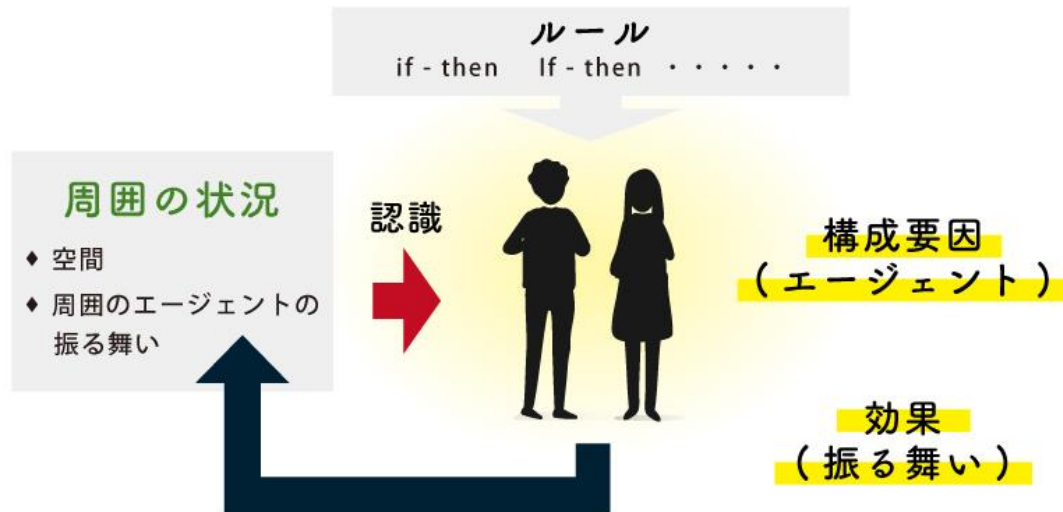


図1：マルチエージェントシミュレーションとは[5]より引用

2.2 先行研究

本研究では先述の Pluchino らの研究[4]をベースとして、ここで使用されているシミュレーションモデルを参照して実験を行なった。Pluchino らは様々な分野において、報酬等を与える場合に才能のある人々に多くの報酬を与える”実力主義パラダイム”に対して、その実力の評価はその人のパフォーマンスを見るだけで事後的に行われており、単純にその時にたまたま高いパフォーマンスを発揮できた人、すなわち幸運な人を実力のある人として評価しているものと捉えている。この実力主義パラダイムに対して警鐘を唱えるのが筆者らの立ち位置で、エージェントベースの統計的アプローチにより、成功したキャリアにおける運と才能の役割を現実的に定量化することを試みた。以下、Pluchino らのシミュレーションモデルについて詳しく説明する。

筆者らが用いたシミュレーションモデルは Talent versus Luck, 略して TvL モデルと呼ばれ、幸運または不運なランダムイベントに影響される人々の資本の増減を分析することを目的としている。非常にシンプルな仮定に基づいており、一定の初期資本、個々人間で異なった才能を持った多数のエージェントが2次元格子の世

界をランダムウォークし, それに伴って幸運・不運に遭遇して資本を増減させる (図 2).



図 2 : Talent versus Luck (TvL) モデルのイメージ[4]. 多数のエージェントが 2 次元の世界を動き回り, 幸運(緑点)や不運(赤点)に遭遇する.

TvL モデルでは N 人のエージェントを考える. 彼らには範囲が $[0, 1]$ に収まるような平均 m_T , 分散 σ_T の正規分布に従う才能 T_i ($i = 1, \dots, N$) が割り当てられる. これらのエージェントは初期状態では 2 次元格子内にランダムに配置される. さらにこの格子内には幸運と不運のイベントが合わせて N_E 個ランダムに配置されていて, 幸運のイベントの割合は p_L とする. 図 2 上では一辺が 201 の正方格子上に $N = 1000$ 人のエージェントが配置され, $N_E = 1000$, $p_L = 0.5$, すなわちそれぞれ 500 個の幸運・不運のイベントがそれぞれ緑色・赤色でランダムに配置されている.

一回のシミュレーションの試行で, エージェントの一生の時間に相当するワーキングライフ P は 20 歳から 60 歳までの 40 年とし, 時間ステップ δ_t は 6 ヶ月とする. シミュレーション開始時点で全てのエージェントは同じ資産 $C_i = C(0)$, $\forall i = 1, \dots, N$ を所有しているとする. ここでいう資産は, それぞれのエージェントの成

功を図る指標として定量的に計りやすく, 比較がしやすいため, 現実世界の通貨とは特に関係のない仮想的なものを用いている. 各エージェントが保有している才能 T_i はシミュレーション上で不変な定数であるのに対し, 資産は変数である. 資産は幸運・不運イベントに遭遇すると変動するが, それらのルールを以下のように定義する.

各エージェントを A_k とすると, それぞれの時間ステップ t , すなわち6ヶ月ごとに, 3つの異なった事象が起こり得る.

(1) 幸運・不運のイベントがエージェント A_k と遭遇しなかった場合:

6ヶ月の期間は何も起こらない.

(2) 幸運のイベントがエージェント A_k と遭遇した場合:

次式のようにエージェント A_k の才能に応じて資産が倍増する.

$$C_k(t) = 2C_k(t-1), \text{ if } \text{rand}[0, 1] < T_k$$

すなわち才能が高い者がより高い確率で利益を得られる.

(3) 不運のイベントがエージェント A_k と遭遇した場合:

次式のようにエージェント A_k の才能に関わらず資産が半減する.

$$C_k(t) = C_k(t-1)/2$$

次に, N 人のエージェントについて各々1回のワーキングライフを試行するシミュレーションの設定について説明する. エージェント数 $N = 1000$ とし, 各エージェントは一定の資産 $C(0) = 10$ と正規分布に従った才能 $T_i \sim N(0.6, 0.1)$ を所有する. T_i は $[0, 1]$ の範囲に収まるものとし, $[0, 1]$ の範囲を超えた場合は範囲内にクリップする. ワーキングライフ40年で, 6ヶ月でのイベントのタイムスパンを考えるので今回のシミュレーションのイテレーションは80回となる. イベントの総数は $N_E = 500$ で幸運イベントの割合は $p_L = 0.5$ である. エージェント, 幸運・不運のイベントが格子内をランダムウォークで80ステップ移動する過程で, 各エージェントの資産の変化を観察する.

上述のシミュレーションを実行した結果, 全てのエージェントの資産分布には現実の社会でも見られるべき乗分布が観察された(図3 (a)). すなわち, 多数の貧しいエージェントと少数の非常に豊かなエージェントで貧富の差が顕著に現れた. 4人のみが500以上もの資産を所有し, 最も成功した20人は総資産の44%をも所有している一方, エージェントの約半分は10資産以下にとどまっている. これらの分布を両対数スケールでプロットすると, 全エージェントの80%が総資産の20%し

か所有しておらず, 残りの20%のエージェントが総資産の80%を所有しているため, パレートの80 - 20則が観察される. (図 3 (b))

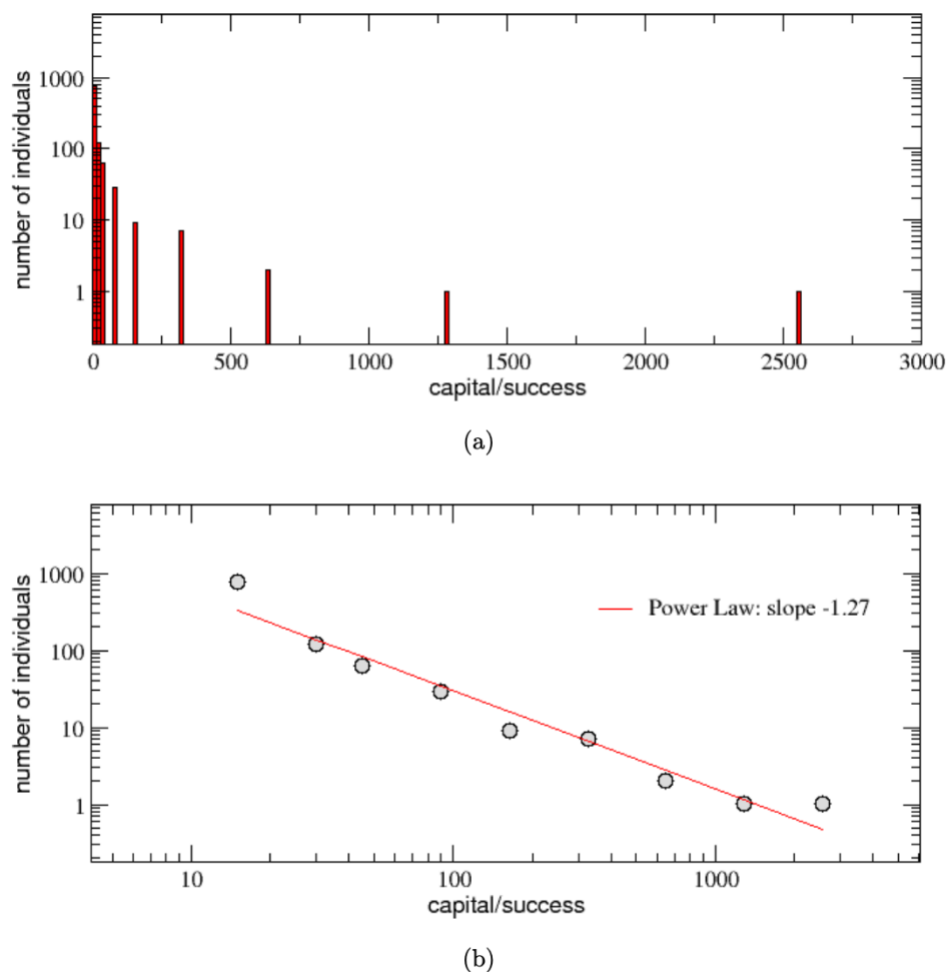


図 3 : シミュレーションの結果([4]より引用)

資産の分布を(a)縦軸のみ対数スケール, および(b)縦横軸どちらも対数スケールで表示.

さらに著者らは才能と資産の関係を調べ, 図 4 の結果を得た. 着目したい点は, 最も資産を所有している(成功した)エージェントは最も才能が高いエージェントではなく, 最も才能の高いエージェントは最も成功したエージェントではないことである. このことから, 先行研究では, 並外れた才能がなければ, 別の要因が強く働いているのではないかと述べている. また, この TvL モデルは現実社会の富の分布をシンプルに表現し, 応用性が高いことにも言及している.

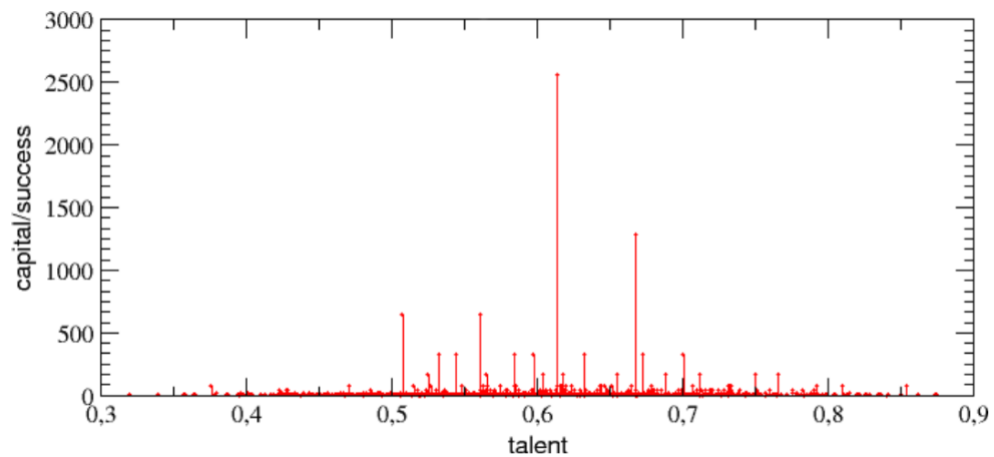


図4：才能と資産の関係([4]より引用)

さらに著者らは最も成功したエージェントと、最も成功しなかったエージェントの両方の資産の時間ステップごとの推移を比較した(図5). 成功したエージェントの初期資産はあまり多いものではなかったが, 時間ステップが30と40の間に上昇を始め, そこから上昇を続けている. 最も成功しなかったエージェントが最も成功したエージェントよりも優れた才能 $T = 0.74$ を持っていたのにも関わらず, 後者が資本を増やすことができたのは, 幸運が原因であることは明らかであると筆者らは述べた.

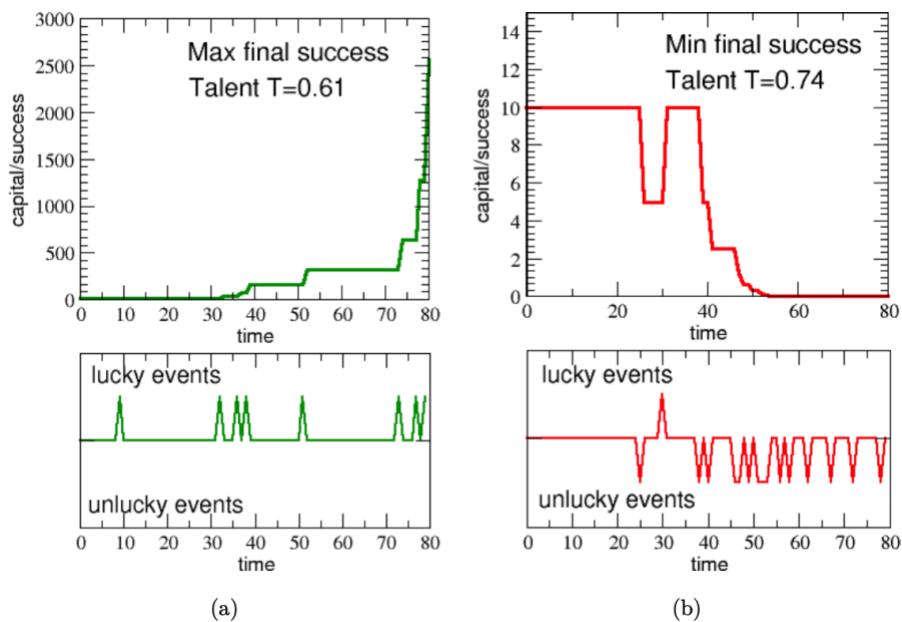
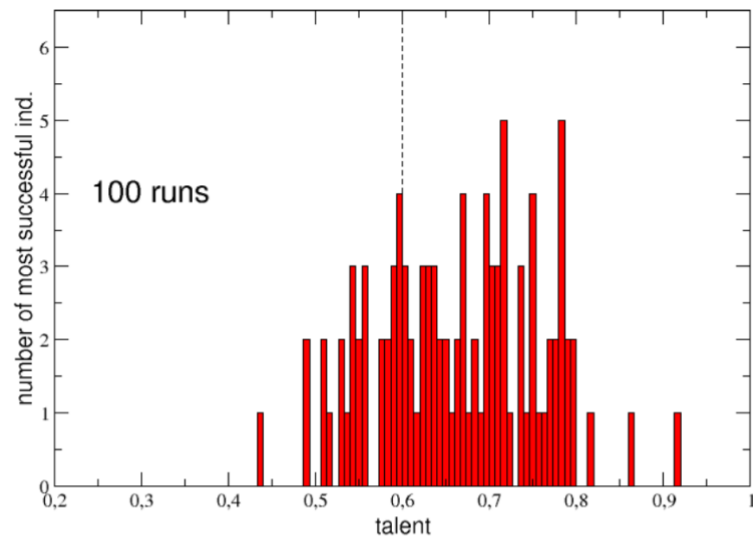
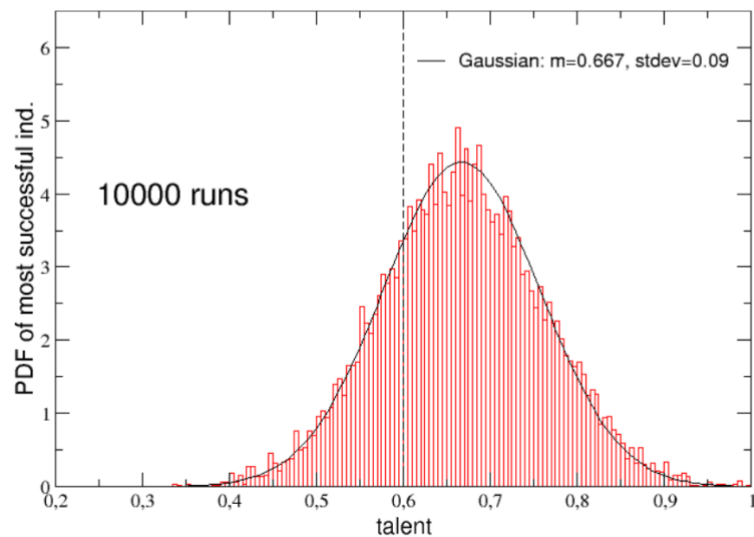


図5：最も成功したエージェントと最も成功しなかったエージェントの時間ステップによる比較([4]より引用)

また, 著者らは, この TvL モデルによるシミュレーションを複数回行った. まず は繰り返し回数 $I = 100$ でシミュレーションを行った. まず全てのエージェントの資産分布はべき乗分布に従い, こちらも同様にパレートの80-20則が観察された. ここでは富裕層と貧困層の差が顕著に確認された. 次に著者らは才能に対して焦点を当てて分析を行ない, 最も成功したエージェントの才能のみを各シミュレーションから抽出して可視化した. 図6(a)は100回のシミュレーション結果である. 分布は才能の平均 $m_T = 0.6$ より少し右にシフトしており, この分布の平均値は $T_{av} = 0.66 > m_T$ であった. これは, 大成功を収めるには, ある程度の才能がしばしば必要であることを裏付けている. しかしながら, 高い才能 ($T > 0.8$) を持っているエージェント達が最も成功する確率は3%であり, それらの最終資産が100回のシミュレーションでの資産の最高値の13%を超えることはなかった. これは才能だけでは成功するのは不十分であることを示している. 筆者らは, より正確に近似をするために10000回シミュレーションを実行し, 各シミュレーションで最も成功したエージェントの才能の分布を計算した(図6(b)). この分布もシミュレーション100回の時と同様に, 才能の分布の軸は平均の $m_T = 0.6$ に比べて少し右にシフトしていることがわかる. さらに, この分布は平均 $T_{av} = 0.667 > m_T$, 標準偏差 $\sigma = 0.09$ の正規分布に従っていると考えられる. 各シミュレーションにおいて才能 T を有するエージェントが最大の資産を獲得する確率は $T = 0.667$ 付近で最大値を取り, その後の確率は才能 T の増加とともに減少していくことを意味する. 才能が成功の主因であるならば, 才能 T を有するエージェントが最大の資産を獲得する確率は T が増加するほど増加する関数であるべきだが, 結果ではそうではないため, 資産が最大となるレベルの大成功を収めるためには, 才能よりも幸運が重要であると結論づけることができる.



(a)



(b)

図6：各回のシミュレーションで最も高い資産を獲得した(成功した)エージェントの才能の分布([4]より引用)

結論として, TvL モデルはその単純さにも関わらず, 我々の社会における富と成功の不平等な分布を再現できた. シミュレーションの結果は, 素晴らしい才能は富の形成や成功を保証するには不十分であり, 平凡な人間が成功の頂点に達することが多いことを示している. このような結果がなぜ起こるかを説明するキーポイントは, 隠れて過小評価されている幸運の役割であると主張した. しかし, この発

見を理解するためにはミクロとマクロの視点を区別することが重要である。ミクロな視点から見ると、才能のある人はどのような機会が来ても自分の利益にする能力が高いので、適度に才能のある人よりも、高いレベルの成功に達成する確率が高い。しかし運を支配することは不可能なため、成功を高めるための最善な選択は、個人で活動の質やアイデアの生成などを積極的に行うことであると著者は結論づけた。一方でマクロな視点から見ると、普通の適度な才能を所有している人間を見つける確率は非常に高い才能を保持している人間を見つける確率よりも高く、すなわち、適度な才能を保持している人間が才能の高い人間よりも普遍的に存在数が多いことを意味すると主張した。

2.3 強化学習

機械学習は大まかに“教師付き学習 (Supervised Learning)”, “教師なし学習 (Unsupervised Learning)”, “強化学習 (Reinforcement Learning)”という三つのカテゴリに分類される。教師付き学習は入力と出力の関係を明らかにすることを目的とし、教師なし学習ではデータに内在する構造的要因を発見することを目的とする。一方強化学習では、エージェントがあるタスクの報酬を最大化すべく行動し、方策と呼ばれる行動規則を獲得することを目的とする。昨今の将棋や囲碁 AI に使用され、計算機上のシミュレーションを駆使して盛んに研究されている分野の一つである。先ほどあげた将棋や囲碁などのように動きが限られていて、自他の駒や石の位置を含むすべての情報が観測できるボードゲーム(完全情報ゲーム)や、ポーカーなどの不確定要素を含んだゲーム(不完全情報ゲーム)など、最近ではゲームの分野に広く応用され、人間の想像を超えた結果をもたらす事例も出ている。強化学習の処理の基本となるのは、環境の状態を観測して次の行動を選択・決定するアルゴリズムや手法である。強化学習のエージェントは、初めは情報が少ないため思うように報酬が得られないが、行動をくり返すにつれて環境の理解が進み、的確な行動を選択して報酬を獲得できるようになる。近年は大規模なニューラルネットワークモデルを用いた深層学習がさまざまな分野で有効性を示している[7]。

強化学習の詳細を述べる。強化学習では、与えられた“環境”における価値、あるいは利益を最大化するように“エージェント”を学習させる。この価値や利益を強化学習では累積報酬又はリターンと呼ぶ。強化学習の要素として、状態・行動・報酬の3つがある。状態は環境が今どうなっているかを表し、行動はエージェントが環境に対して働きかける行為を表す。そして報酬は、ある状態においてエージェン

トの取った行動がどの程度良かった/悪かったのかを表す。報酬に関しては注意が必要で、強化学習ではその時点で獲得できる短期的な報酬ではなく、最終的に将来を鑑みた累積報酬, すなわちリターンを最大化する[7]. (図7 参照)

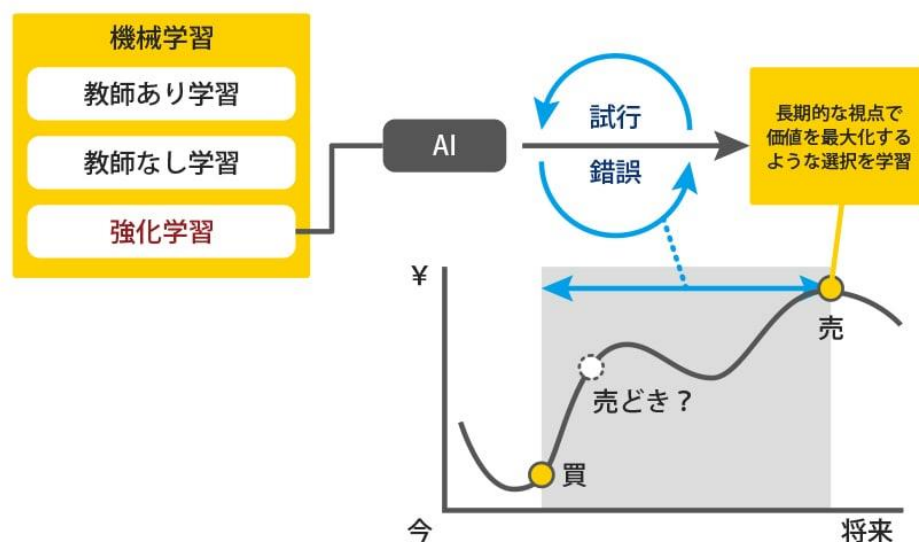


図7：強化学習の仕組み[8]

では強化学習ではエージェントはどのように行動の良し悪しを見極めればいいのか. ある状態 s である行動 a を取った時のリターンがわかれば, 現在の状態から最善の行動を取ることができる. 行動を決定する際に参照する判断基準を行動価値, あるいは Q 値, とよび, それらを全ての状態・行動に対して定めたものを行動価値関数と呼ぶ. ある状態 s に対して, 行動 a を取った時に行動価値関数は $Q(s, a)$ と書く. 行動価値関数はその時点でのリターンの予測値に相当し, 行動価値関数を更新していくことによって強化学習が遂行されていくわけである.

本研究では, 強化学習の代表的なアルゴリズムである Q 学習(Q Learning)を用いて TvL モデルのエージェントを訓練する. Q 学習ではステップ t でエージェントが状態 s_t にあるとして, 行動 a_t を選択した時の行動価値関数 $Q(s_t, a_t)$ を次式に従って更新する.

$$Q(s_t, a_t) = Q(s_t, a_t) + \alpha \left(r_t + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t) \right)$$

ここで α は学習率 (learning rate), r_t は現時点でエージェントが獲得する報酬 (reward), γ は割引率 (discount factor) である. 学習率 α は行動価値関数に対して新たな情報をどのくらい加味するかを制御するパラメタである. α が大きいほど学

習の効率は高まるが、大きすぎると学習が安定しない。本研究のシミュレーションでは、エージェントごとに学習率を変えることによって、個々人の学習能力の差を再現する。 $\max_a Q(s_{t+1}, a)$ は次にある状態 s_{t+1} に至った時に、取るべき行動 a の価値の最大値を表している。割引率 γ は通常 1 よりもやや小さい値を取る定数であり、1 ステップ進んだ未来の行動価値 $Q(s_{t+1}, a_t)$ がやや不確定である(未来のことなので本当にそうなるかどうかわからない)ことに対応している。また、強化学習をするにあたって重要な要素が一つある。それは「探索と活用(exploration and exploitation)」である。学習の途中段階の行動価値関数 $Q(s, a)$ は最適化が不十分であり、正しい行動を導くものではない。したがってこれに忠実に従って学習を進めても最適解に到達できるとは限らず、学習効率も悪い。そこで、学習中はある一定の確率 ε で行動をランダムに選択する。これを ε 貪欲(ε -greedy)探索と呼ぶ。強化学習の目標は、長期にわたる累積報酬(リターン)を最大化させることである。この時、遠い未来の報酬は不確定なので、割引率 γ によって重み付けした和をとっている。 $\alpha, \gamma, \varepsilon$ などを現実的な範囲で適切に定めれば、Q 学習により最適な行動価値関数が得られることが知られている。

強化学習の応用分野はゲームや自動運転など広範である。ボードゲームの囲碁において人間を超越する強さを持つ AlphaGo [10] は DeepMind 社によって開発された強化学習の深層学習モデルである。また、近年の産業用ロボットはさまざまなシチュエーションに合わせた行動を取るように強化学習を用いたプログラムが幅広く実装されている。人のランダムな動きに対して、お互いの作業を干渉しない上、ロボットが部品を押さえたり工具を手渡ししたりしてくれるなど人間の動きに柔軟に合わせた行動が取れるようになっている[6]。

第3章 提案手法

本研究では、先行研究である Pluchino らの論文[3]で用いられている TvL モデルに強化学習を導入したシミュレーションを行い、人間に学習能力があれば、運の要素を上回ることができるのではないかという仮説を検証する。人間の将来のキャリア成功や資産形成に対して、運の要素が支配的であることは Pluchino らが主張した。しかし、人間は高校時代に勉強を頑張り、大学に入り、巨万の富を築く場合もある一方、そのような努力を怠って成功を手にしていない人間も一定数存在して

いるだろう. 努力というのはスポーツ選手にもまさに当てはまる. 彼らは努力をして高度な技術を身につけて, その結果, 報酬として資産を形成している. このように”成功”というものは運のみでは測ることができないのではないか. 自ら学習する努力を積み重ねた末に成功を手にすることはできるのではないかと考える.

本研究で提案する強化学習を加えた TvL シミュレーションモデルによって得られるであろう結果の仮説を図で示す(図8).

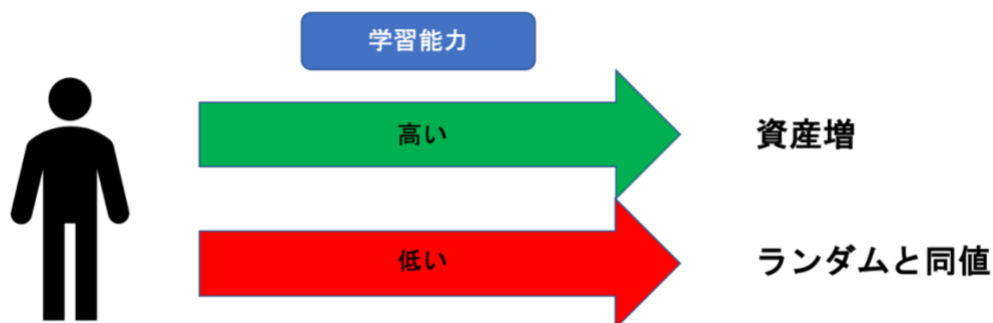


図8：提案するシミュレーションモデルの仮説

提案するモデルでは, エージェントにそれぞれ学習能力を表す学習率という指標をランダムに割り当てる(前章の強化学習での α). この学習率という指標が高ければエージェントは報酬が多く獲得できるような, 価値の高い行動を選択することが期待される. 一方, 学習能力があまり高くない場合, エージェントの行動はランダムに近いものとなり, 資産の増減も従来の TvL モデルと似たような振舞いとなるだろう. 努力が成功に対して支配的であるのではないかとする仮定のもと, 努力をこの学習能力と見立てたときに, 努力をすればするほどシミュレーション後に得られる資本が多くなっていくのではないか. 強化学習を用いた TvL モデルにより, 成功に対しやはり運が支配的であるのか, はたまた努力を続けることができるのか. これらの結論をシミュレーションの結果より導く.

第4章 実験

4.1 シミュレーションの設定

実験にあたって, まずシミュレーションの基本的な動作を説明する. 今回の

シミュレーションでは $N = 1000$ 人のエージェントが 25×25 の格子内を動き回る. 幸運と不運のイベントは合計で $N_E = 100$ 個存在し, そのうち幸運のイベントの割合は $p_L = 0.5$ である. すなわち幸運と不運のイベントがそれぞれ 50 個ずつ存在する. 各エージェントは才能 $T_i, i = 1, \dots, N$ を有し, この才能は平均 0.6, 分散 0.1 の正規分布に従うように設定する. エージェントおよび幸運・不運のイベントはランダムに格子内に配置される. エージェントのワーキングライフは 20 歳から 60 歳までの 40 年間とし, 半年に一回格子内を移動する. つまり 1 回のシミュレーションを通して各エージェントは 80 回の移動を行う. 毎回の移動先に幸運・不運のイベントがあるかどうかによって, 幸運または不運のイベントが起こる. 移動先に何もなければ何も起こらない. また, シミュレーション開始時点で全てのエージェントは同じ資産 $C_i = C(0) = 10, \forall i = 1, \dots, N$ を所有しているとする. ここで従来の (Pluchino らの) TvL モデルと異なる点として, エージェントはランダムに動くのではなく, それぞれが持つ行動価値関数に従って行動する. 一方, 幸運・不運のイベントはランダムな配置を取るのみで, 移動しない. エージェント間で幸運・不運のイベントまでの距離は均一ではないが, そのばらつきが”運”であると解釈する¹. 各エージェントが保有する才能 T_i はシミュレーション中で不変であるのに対して, 資産は幸運・不運イベントに遭遇すると変動する (図 9 参照). 資産変動のルールを以下に示す.

各エージェントを A_k とすると, それぞれの時間ステップ t , すなわち 6 ヶ月ごとに, 3 つの異なった事象が起こり得る.

(1) 幸運・不運のイベントがエージェント A_k と遭遇しなかった場合:

6 ヶ月の期間は何も起こらない.

(2) 幸運のイベントがエージェント A_k と遭遇した場合:

次式のようにエージェント A_k の才能に応じて資産を 1.2 倍とする.

$$C_k(t) = 1.2C_k(t-1), \text{ if } \text{rand}[0,1] < T_k$$

すなわち才能が高いものがより高い確率で利益を得られる.

(3) 不運のイベントがエージェント A_k と遭遇した場合:

次式のようにエージェント A_k の才能に関わらず資産を $1/1.2$ 倍とする.

¹ そもそもエージェントとイベントの両方を動かす必要はない. 一方を固定しても一般性は失われない上に, シミュレーションを簡略化できる.

$$C_k(t) = C_k(t - 1)/1.2$$

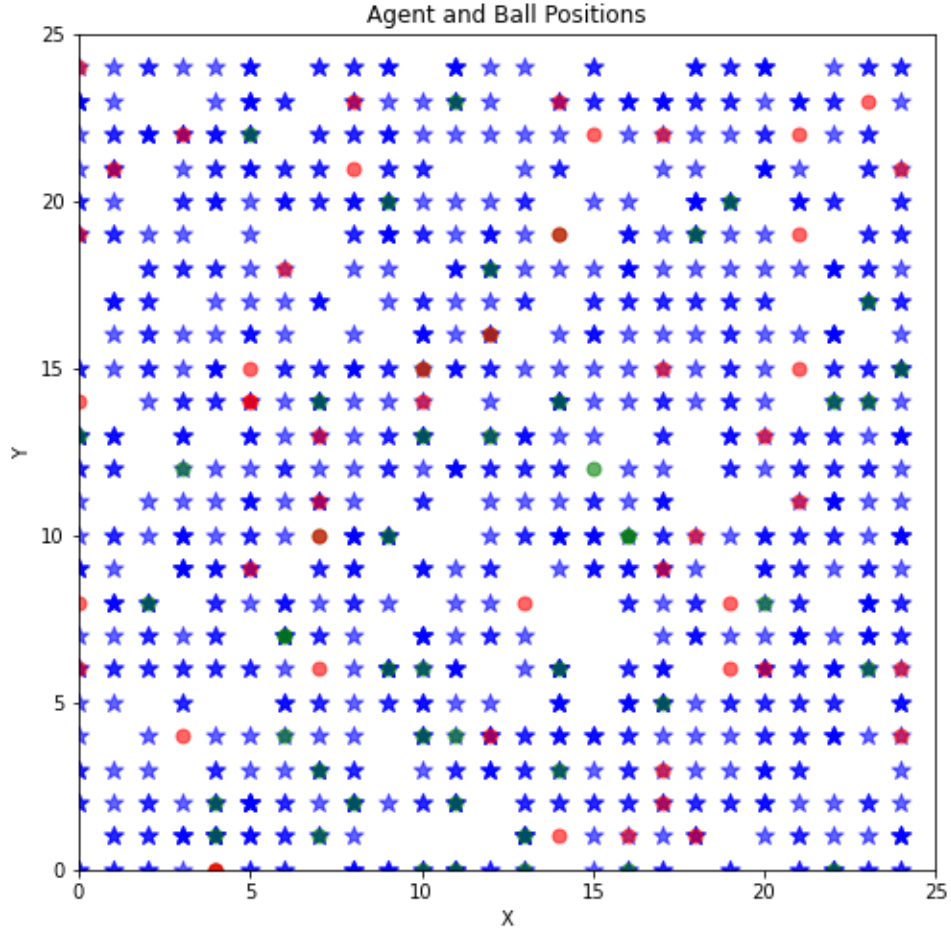


図9：研究で用いるシミュレーションモデル

青い星：エージェント, 緑玉：幸運イベント, 赤玉：不運イベント

さらに, 今回新たに加える強化学習の部分詳しく説明する. 行動価値関数 Q は2章で挙げたように以下で計算される.

$$Q(s_t, a_t) = Q(s_t, a_t) + \alpha \left(r_t + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t) \right)$$

エージェントの取りうる行動 a は上下左右であり, ここでは

{up right, left, down} と表す. また, 状態 s_t はステップ t におけるエージェントの位置座標である. 学習率 α はエージェントによって異なる. このシミュレーションでは学習率はそれぞれのエージェントに対して $[0, 0.02]$ の一様乱数で与えられるとする. また比較のため, 学習率が平均 0.01, 標準偏差 0.005の正規分布に従う場合についても検証する. 報酬である r_t には前タイムステップでの資産からの増減を 10^4 で除したものが入っている. これは, 資産が幸運のイ

ベントと接触するごとに指数関数的に上昇して極めて大きな値になることを防ぐためである。割引率 γ は0.9で定数である。ここで、 ϵ 貪欲探索でランダムな行動をとる確率を $\epsilon = 0.1$ と設定する。これにより、それぞれのエージェントは探索(exploration)の機会を得て、より良い行動価値関数を発見する可能性を高める。

シミュレーションの設定でもう一つ、従来の TvL モデルと大きく異なる点がある。それはこのシミュレーションはトレーニングとテストに分けられているという点である。トレーニングフェーズでは、各エージェントは自身の行動を通して行動価値関数を最適化する。ここでは40回のエピソード(つまり40ワーキングライフ)を実行した。テストフェーズではトレーニングフェーズで取得した行動価値関数是不変のままでエージェントの位置を再度リセットし、ランダムに配置する。ここで全てのエージェントの資産も元の値 $C_i = C(0) = 10$ にリセットしてからテストを実行する。

以上が今回の研究で使用するエージェントシミュレーションモデルの概要である。これらは全て Python を用いて、オブジェクト指向プログラミングの記法により実装した。

4.2 実験 1：シングルシミュレーション

まずは先ほど紹介したシミュレーションを一回のみ試行する。このシミュレーションで得られるエージェントの才能の分布と資産の分布をプロットしたものがそれぞれ図 10 と図 11 である。才能の分布は設定したように平均0.6, 標準偏差 0.1 の正規分布に従っている。シミュレーション後の資産分布では、先行研究同様にべき乗分布が観察された。このヒストグラムで度数最大となったのは資産が $[0, 1000]$ の範囲であり、この範囲に属するエージェントの人数は全体の 98.0% であった。さらに、最も成功したエージェント 20 名の総資産は全てのエージェントの総資産の約 83% を占めている。先行研究と比べて資産の増減倍率を減らしたのにも関わらず、貧富の差が顕著に開いていることが確認された。Pluchino らのシミュレーションとは資産の増減倍率など一部の設定が異なるので一概には言えないが、強化学習の導入によって貧富の差が顕著に開いている可能性がある。

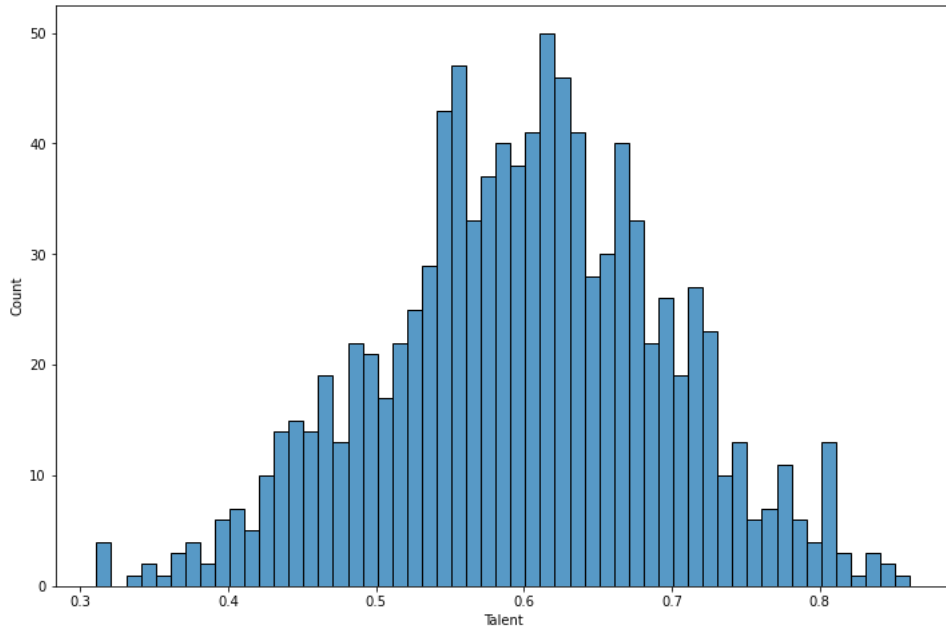


図 1 0 : エージェントの才能の分布

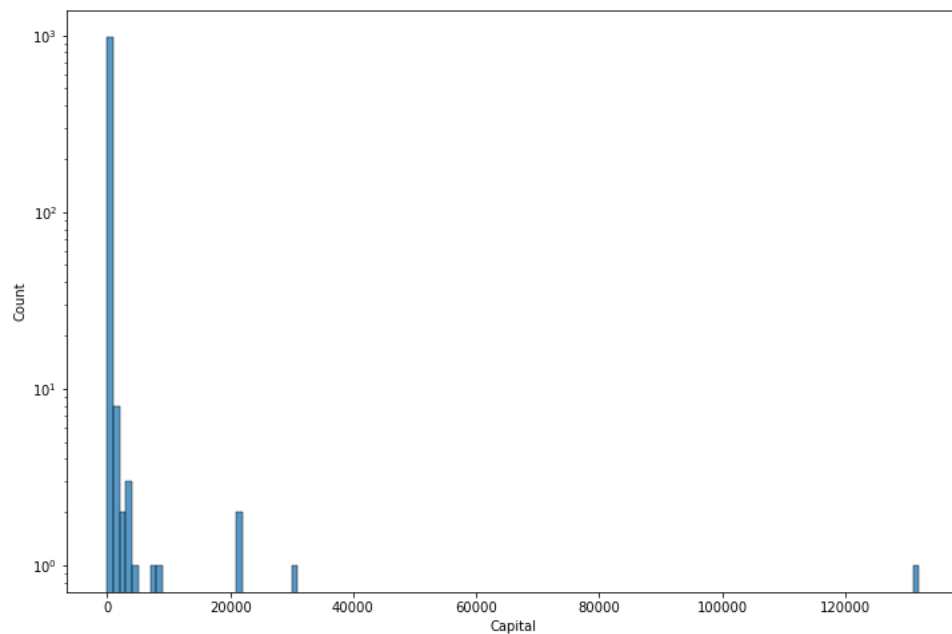


図 1 1 : エージェントの資本分布 (縦軸は対数をとっている)

次に, エージェントの行動源となっている行動価値関数 Q を幸運イベントの座標とともにヒートマップ可視化した(図 1 2). ここに示した行動価値関数は, あるイベント配置においてエージェントの上下左右の各行動の価値を全てのエージェントにわたり総和したものであり, すなわち幸運イベントの位置に

対しての行動の良否を一目で理解することができる。概ね幸運イベントとその周辺の行動価値が他より高くなっており、エージェントが幸運イベントに向かって移動するように学習が行われていることが確認される。この現象の理由としてエージェントの密度が高い、幸運のイベントの密度が高いなど本シミュレーションの設定上の理由が考えられる。いずれにせよエージェントがトレーニングの段階で幸運のイベントを認識していると考えられ、強化学習が機能していることがわかる。上下左右どの方向も似たようなパターンが表れている。例えば座標 (16, 19) 付近などで幸運イベントの密度があまり高くはないにも関わらず高い行動価値が出ているのは、エージェントの密度が高く、密度の高い幸運イベントの方への誘導がなされているので、行動価値が強く表示されていると予想される。

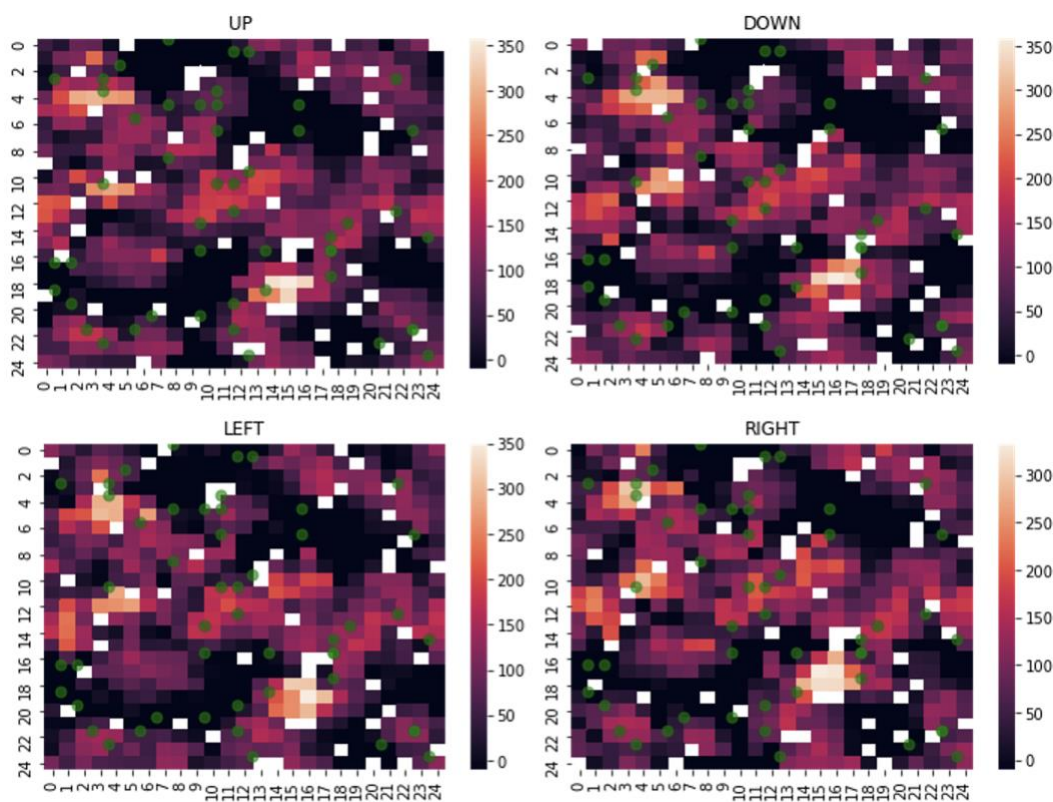


図 1 2：各行動における行動価値関数のヒートマップ。緑玉は幸運イベント。

学習能力が高ければ高いほど成功を手に入れるのではないかという観点で、学習能力を表す学習率と資産の関係、生まれ持った才能と資産の関係を表した散布図を最も成功した 20 人のエージェントと最も成功しなかったエージェン

ト20人をプロットした(図1 3). 上述の観点が成立する場合, 図1 3 (a) では学習率が低いとされる左側にプロットが集まり, 図1 3 (c) では右側にプロットが集まると考えられる. しかしその仮定とは違い, 学習率の分布に特段の偏りがないことから, 学習率は資産に対して影響しないと考えられる. すなわち, 人間の学習能力は成功に対して重要な要素ではなく, やはり運の要素が支配的である. 同様に図1 3 (b), (d) でプロットされている才能と資産の散布図より, 才能がある一定の範囲内, 具体的には 0.6 付近に集中していることが見て取れる. これは才能の分布(本シミュレーションでは平均 0.6 の正規分布)で一番人口の多い層から成功・失敗者が多く出るということであり, 自然なことである. 結果として, 必ずしも学習能力が高いからといって, あるいは豊かな才能を有しているからといって, 多大な資産を手にしたたり, 人生の成功が確約されたりするわけではない. 逆もまた然りで, 学習能力も才能も低い人間が成功を手に入れる可能性がある.

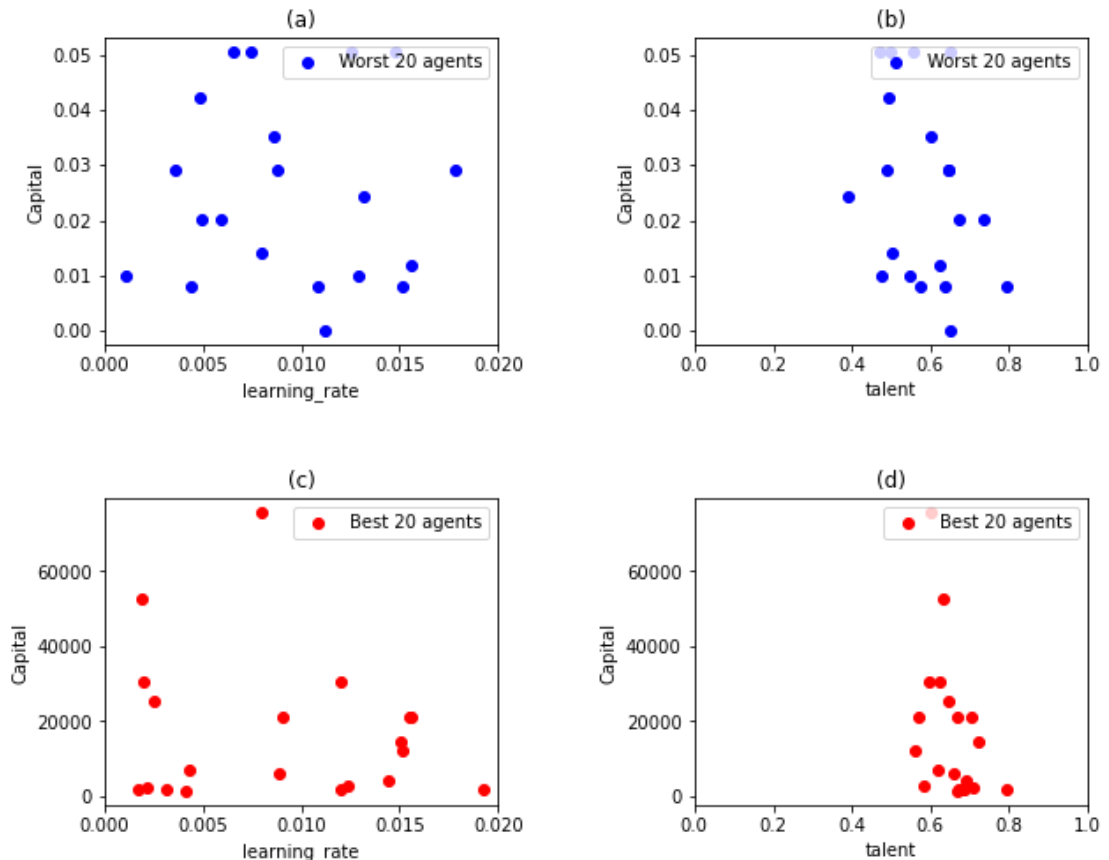


図1 3：学習率, 才能と資産の関係

(a) 最も成功しなかったエージェント20人の学習率と資産の散布図

- (b) 最も成功しなかったエージェント20人の才能と資産の散布図
- (c) 最も成功したエージェント20人の学習率と資産の散布図
- (d) 最も成功したエージェント20人の才能と資産の散布図

ただ、このシミュレーションモデルでの学習率は一様分布で与えられているが、実際の努力している人々の分布は一様分布ではないであろう。そこで追加実験として、学習率を正規分布 $N(0.01, 0.005^2)$ に従うように変えてシミュレーションを再度実行する。結果、資産の分布や行動価値関数の分布に大きな変動はなかったが、学習率と資産の関係に変化があった (図 1 4)。図 1 3 (a), (c) と図 1 4 の (a), (b) をそれぞれ比べると、傾向の変化が観測された。図 1 4 (a) での学習率と資産の相関係数は $r = -0.44$ となり負の相関を持つ。ただ、ここで取り上げているのは最も成功しなかったエージェント20人なので、どの学習率を持っているエージェントでも大きく失敗しうることを示唆している。図 1 4 (b) では学習率の平均 0.01 よりも学習率が高いエージェントは20人中12人いるが、彼らは明らかに学習率が平均よりも低いグループと比べて多くの富を保有している。しかし、シミュレーションを数回行ってみると、実際はこのような分布には従わず、学習率が全ての範囲に均等に分布していることがわかった。すなわち、学習能力が高い人間がより多い資産にたどり着くことはなく、努力は成功に対して影響しないと予想され、どの学習能力を持った人間でも成功に辿り着くことができる。これはつまり、どの学習能力を持っている人間でも他の人間と比べて多くの資産を保有し得るが、多くの資産を持っている人間の中では、学習能力が高い人間がさらに多くの資産を保有できる可能性があることを示唆している。努力はブーストのような役割を果たしており、結果を出せばさらなる高みを目指せるのかもしれない。

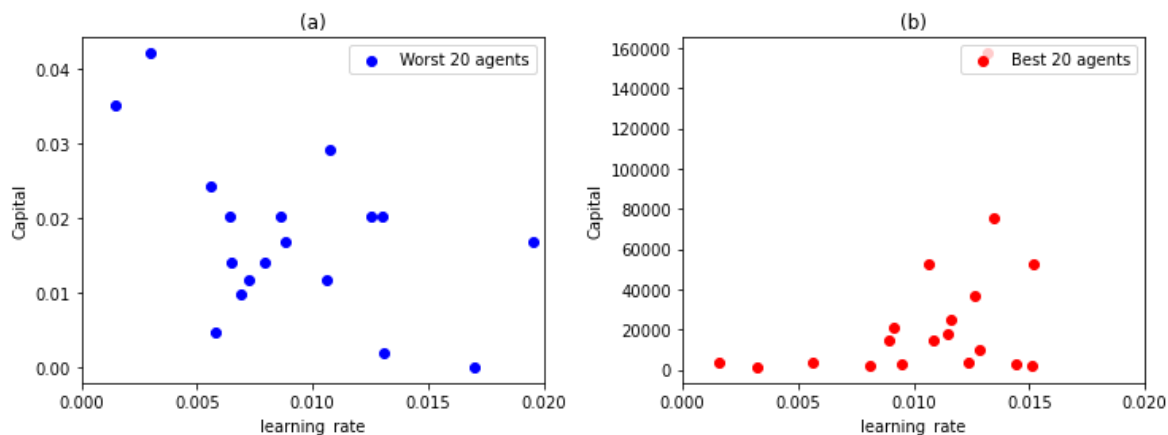


図 1 4 : 学習率を正規乱数で設定した場合の学習率と資産の関係
 (a) 最も成功しなかったエージェント20人の学習率と資産の散布図
 (b) 最も成功したエージェント20人の学習率と資産の散布図

以上述べたようにシングルシミュレーションでは, エージェント全体, 資産を多く保有しているエージェントと保有していないエージェントに分けて考察を行った. 次節ではシミュレーション回数を増やし, 成功者に着目した実験を行う.

4.3 実験 2 : マルチシミュレーション

前節で行ったシミュレーションを100回, 10000回行い, 各回で最も成功したエージェントのみを抽出する. そこで成功するエージェントの特徴を探り, 成功と運, 成功と努力の関係について考察する.

まず各回で最も成功したエージェントの才能の分布をヒストグラムで示す(図 1 5). (a)の分布では荒くはあるが正規分布に従っていると予想され, (b)の分布では正規分布が確認された. ここでの平均は100回の時

$mean(T_{iteration=100}) = 0.703$, 10000回の時 $mean(T_{iteration=10000}) = 0.7247$ である. 先行研究と同じように才能が正規分布に従い, 平均がもともと設定していた $m_T = 0.6$ よりもやや大きいという点で同じ傾向が確認された. しかし元の平均との差が少々異なって観測された. 繰り返し回数 10000回における先行研究との才能の差は 0.065であり, 才能の平均値がわずかに上昇した. これは学習を進めていくにつれて, 幸運イベントと遭遇する回数が増え, その分才能が高いエージェントがより多くの資産を手にしたことが原因と考えられる. 資産の増減倍率や, 幸運・不運イベントが位置を変えないなど, Pluchino らの実

験とは設定が幾分異なるので一概には言えないが, 強化学習によって学習能力を付与した現実により近い環境では才能がさらに高い人が成功する可能性が高まっていると言えるだろう.

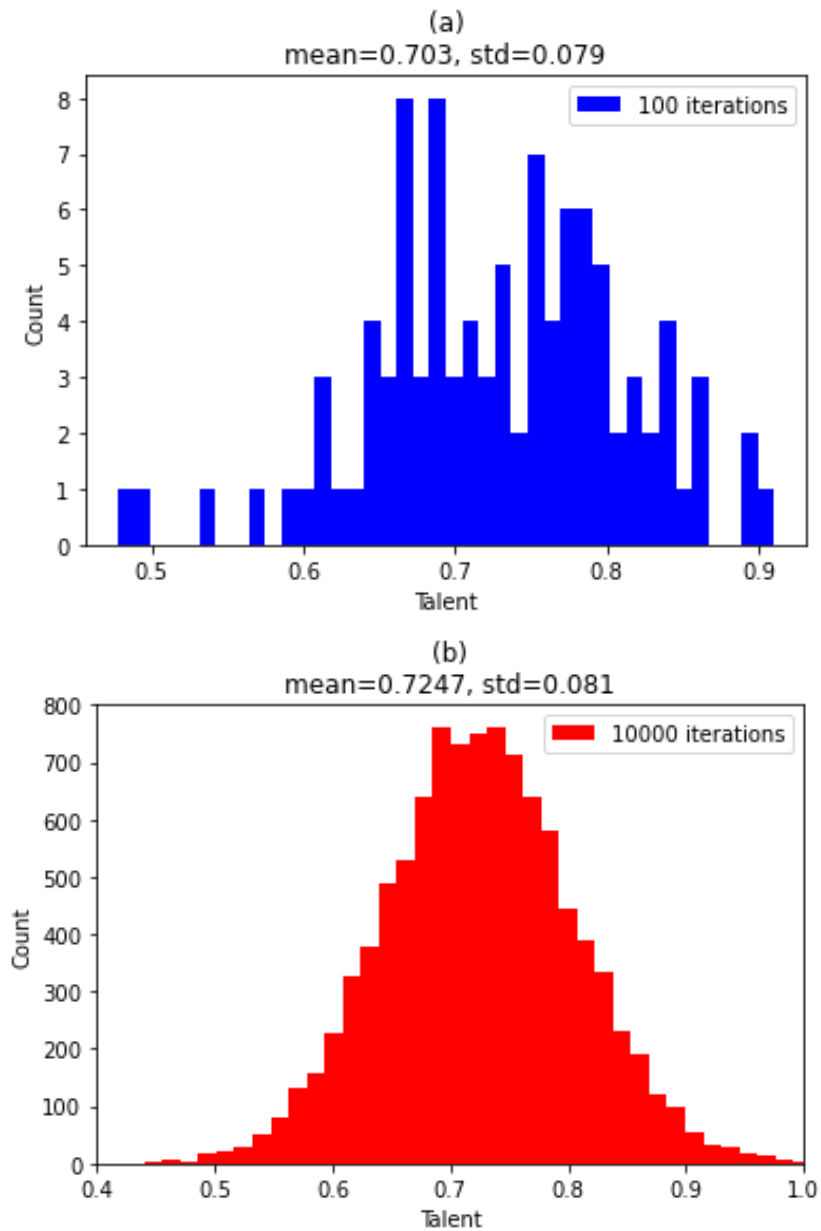


図 1 5 : 最も成功したエージェントの才能の分布,
(a)繰り返し回数100回 (b)繰り返し回数10000回

学習率に関しても各回で最も成功したエージェントのヒストグラムを示す(図 1 6). ここでは学習率は一様分布としている. エージェントの学習率に偏

りは認められず, 学習率の生成に用いた一様分布とほぼ同じ形状が見られる. すなわち, 成功を得るためには学習などの努力よりも運の方がより支配的であると考えられる. 今後より精密なシミュレーションが必要ではあるが, 強化学習の導入により才能が成功に寄与する度合いがわずかに上昇したことから, 学習の要素が才能を通して間接的に人生の成功に影響した可能性がある.

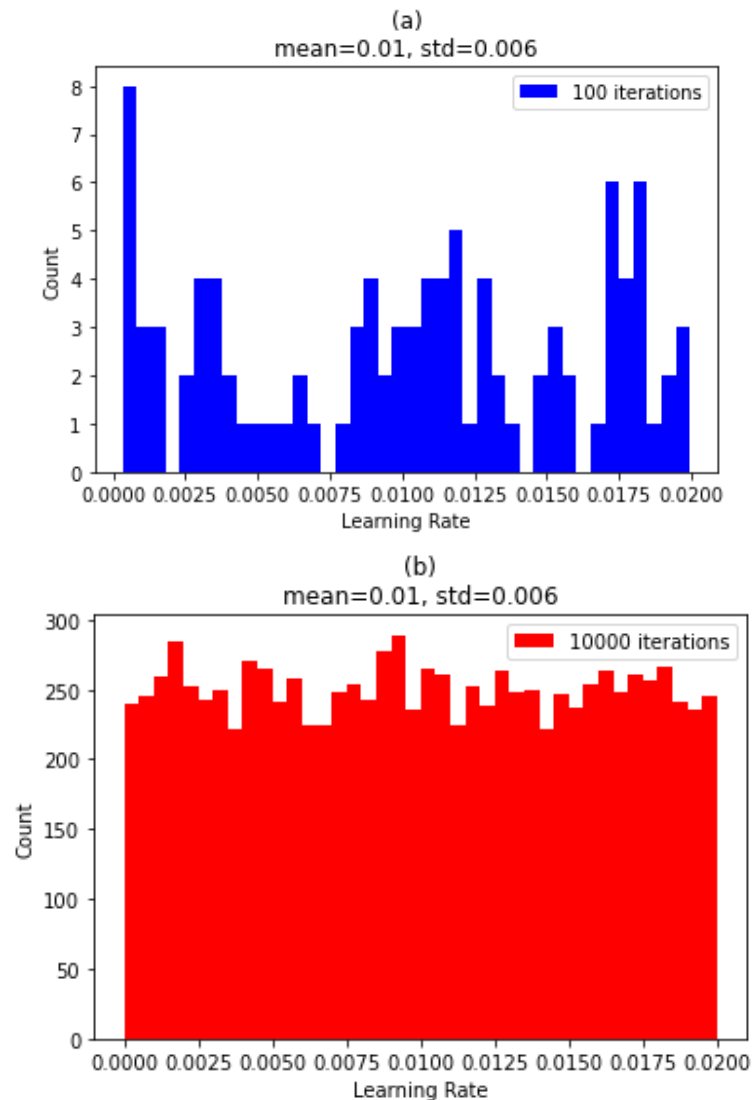


図 1 6 : 最も成功したエージェントの学習率の分布,
(a) 繰り返し回数100回 (b) 繰り返し回数10000回, 学習率は一様分布.

比較として, 学習率が正規分布に従っている場合の分布もプロットした(図 1 7). こちらも一様分布の時と同様に元来の分布に従っていて, 変化が見られ

なかった. すなわち先ほども述べたが, 学習などの努力に成功は依存しない. 運が成功に対して支配的であることがわかった.

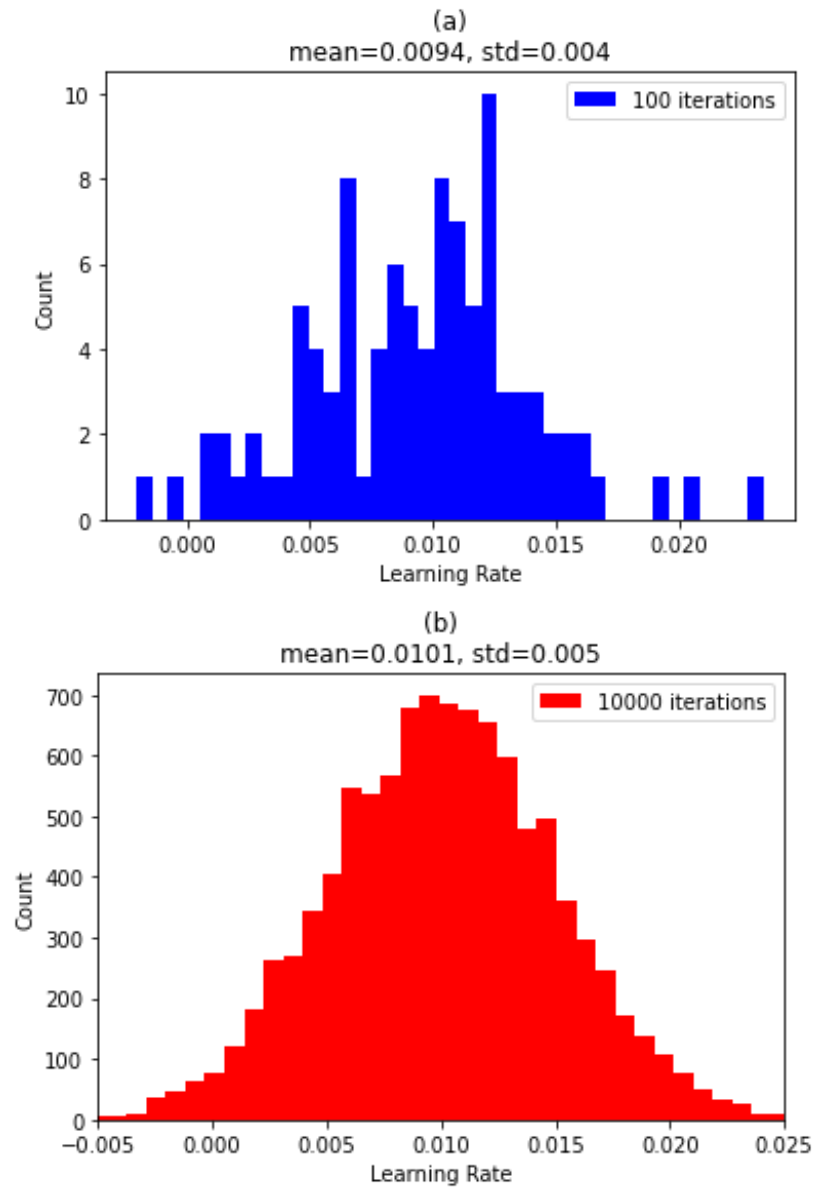


図 1 7 : 最も成功したエージェントの学習率の分布,
(a) 繰り返し回数100回 (b) 繰り返し回数10000回,ただし学習率は正規分布 $N(0.01, 0.005^2)$.

第5章 おわりに

5.1 まとめ

本研究では、成功に対して努力が影響しているのではないかという点に着目し、マルチエージェントシミュレーションに強化学習の手法を組み合わせ考察した。運が成功に対し支配的であると述べた Talent vs. Luck (TvL)モデルを実装し、それに強化学習を適用して新たなモデルを提案した。このモデルでは、強化学習における学習係数の大小が学習能力、ひいては努力の度合いを表す。実験の結果、まず成功は運に支配されるという TvL モデルの結果を再現できた。加えて本研究では、努力は成功にほとんど寄与しないという結果も得た。現実世界で学習能力が正規分布に従っていると仮定すると、平均的な学習能力を持った人間が成功することがわかったので、ある程度の学習能力は必要であると考え。しかし、運の成功への支配性から鑑みて、幸運・不運のイベントに関わらず、積極的に行動を行い、イベントに遭遇する回数を増やした方が成功に近づけると考える。

5.2 今後の課題

今回は成功と運・努力についての研究を行ったが、運と努力の関係性についても調べてみたい。成功を収めている人間は努力をして運を掴んでいる。つまり努力を行ったが故に運を掴むことができたのかもしれない(と多くの人は信じている)ので、運と努力の因果関係をさらに詳細に調べることが可能であると考え。また、今回はシンプルなシミュレーションを行ったが、他のエージェントと遭遇したら他のイベントが起きるなど、マルチエージェント強化学習の枠組みを取り入れて様々な現実的要素を含めたシミュレーションモデルを構築してみたい。

参考文献

- [1], Einav, Liran, and Leeat Yariv, “What’s in a Surname? The Effect of Surname Initials on Academic Success.”, Journal of Economic Perspectives, 2006, 20 (1): 175-187.
- [2], Du, Qianqian., Gao Huasheng, Levi Maurice D., “The relative-age effect and career success: Evidence from corporate CEOs.”, Economic Letters, 2012, Vol. 117(3): 660-662.
- [3]N. N. Taleb, “Fooled by Randomness: The Hidden Role of Chance in Life and in the Markets”, Random House, 2001.
- [4], A. Pluchino, A. E. Biondo, A. Rapisarda, “Talent versus Luck: The Role of Randomness in Success and Failure”, Advanced in Complex Systems, 2018, Vol. 21, No. 3-4.
- [5], 構造計画研究所, “マルチエージェントシミュレーションとは”

[6], NTT データ数理システム, “マルチエージェントシミュレーションによる「人流」研究事例”

[7], 株式会社 doda, 強化学習とは？機械学習との違いなどの用語解説や活用事例などをご紹介します

[8], 株式会社ブレインパッド, 強化学習入門～これから強化学習を学びたい人のための基礎知識～

[9], NTT Communications, 強化学習とは

[10] Silver, D., Huang, A., Maddison, C. *et al.* Mastering the game of Go with deep neural networks and tree search. *Nature* **529**, 484–489 (2016).