

Q 学習を含むエージェントによって 構成されるエージェントベース シミュレーションにおける割引率と 流行の流布との関係に関する研究

中 野 統 英

1. 緒 言

近年、ファッションの流行状況をはじめとする様々な経済活動や社会活動による結果や影響を望ましい方向にコントロールするメカニズムを求める手法が研究されるようになった^[1, 2]。特に人間同士での情報伝達、いわゆる口コミを何らかの手段でコントロールすることができれば各種のマーケティング活動に利用することができ、また悪い噂を広まらないようにすることもできると考えられる^[3]。

そこで経済・社会活動をミクロ的な視点、つまりそこで活動する人間（エージェント）のレベルからモデルを作成して再現・予測を行うエージェントベースシミュレーションと呼ばれる手法が近年多く用いられるようになってきた^[1, 2, 4-9]。エージェントベースシミュレーションとは、仮想空間内に配置されたエージェントが決められたルールに従って意思決定や行動を行うことにより社会現象などを仮想空間上に再現して調査を行う手法である。著者は人々の間で起こる流行の流布や衰退といった物事に対する採用行動をエージェントベースシミュレーションで再現し動向を考察する

研究を行っている^[5-9]。近年には周囲の環境変化に対応して学習を行うことができるエージェントについて研究を行っており、実際にはQ学習と呼ばれる強化学習の一手法を導入している^[7-9]。これらのエージェントについては、報酬の与え方を最適化することにより採用行動を適切に行わせることができることが示されている^[9]。

本研究ではエージェントが目先の報酬（即時報酬）を重視して行動した場合の採用行動と将来得られるであろう報酬を重視して行動した場合の採用行動とがどのように異なるかという問題を考える。Q学習では割引率というパラメータを操作することによってこの問題について検討することができるので、ここではQ学習における報酬に関するパラメータである割引率と流行などの採用行動との関連についての研究を行った^[10, 11]。

2. Q学習によるアルゴリズム

2.1 エージェントの性格と仮想空間

新しいうわさなどの情報が人々の間に伝わる時、その情報を採用する革新的な人と容易には採用しない保守的な人があるのは論をまたない。過去にRogersは（うわさなどの情報に対する）採用者カテゴリーを5つに分類した^[12]（図1参照）。図1ではInnovatorsがうわさを最初に採用し、時間の経過とともにEarly Adopter, Early Majority, Late Majorityの順で採用行動を行い、Laggardsが最後に採用する（または最後まで採用しない）

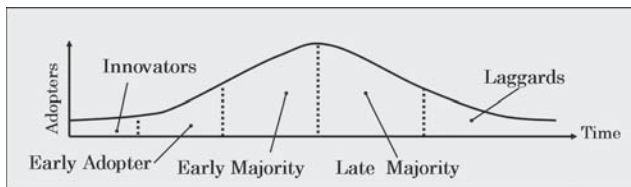


図1 Rogersの提案した採用者カテゴリー

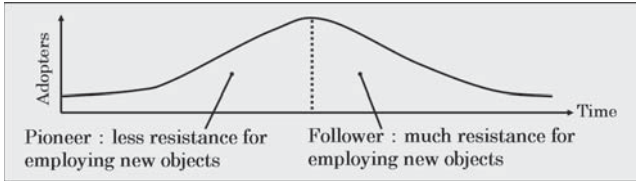


図2 本研究で用いた採用者カテゴリー

ことを示している。その後にこれらをもとにした2カテゴリーへの分類が提案されており^[4] (図2参照), 本研究でもこのカテゴリーを採用する。ここで新しい情報(うわさ)の採用に抵抗感が低い革新的なエージェントを Pioneer, 逆に抵抗感が高い保守的なエージェントを Follower としており, それぞれのエージェントは周囲エージェントの採用態度を参考にして, うわさなどについて採用 (Accept) または非採用 (Reject) のどちらかの態度をとるものとする。

エージェントが配置される仮想空間は 10×10 の格子状空間とし, 空間の端に位置するエージェントが反対側のエージェントとつながるトーラス構造であるとする^[4] (図3参照)。この空間にエージェントを隙間なく配置し予め定めたルールをもとにして自律的に行動させることにより各エージェントのうわさに対する採用行動の変化をみるが, 各エージェントは採

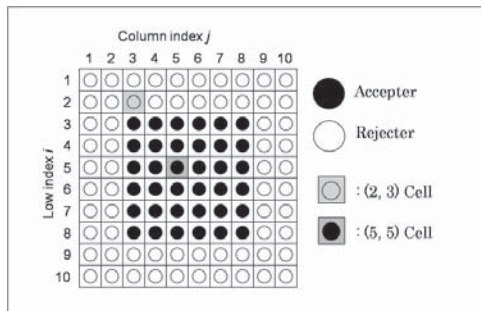


図3 仮想空間とそこに配置されたエージェント

用 (Acceptance), 非採用 (Rejection) または変化無 (No Change) の 3 種類の態度変化のうちのいずれかの一つを必ず選択するものとする。ただし本研究では仮想空間中の 100 エージェントはすべて Follower かすべて Pioneer かのどちらかであるとする。またここでは i 行 j 列目のセルを (i, j) セルと呼ぶことにする (図 3 参照)。

なお 2 種類のエージェントに対するルールや仮想空間の設定については、過去の研究^[7-9]との比較のために同一のものをういたということをここに記しておく。

2.2 意思決定手法

ここでは図 4 の例を用い意思決定プロセスと Q 値の更新手順の説明を行うこととする。図 4 中の左図において実線で描かれた中央のエージェントは Follower であつ非採用者であるとし、このエージェントの採用行動に注目して意思決定および Q 値の更新について説明する。

中央のエージェントは意思決定の際には破線で示された上下左右に位置する 4 エージェント中の採用者数 n_t ($n_t=0, 1, 2, 3$ および 4) をカウントし、その n_t に対応する Q 値の状況をもとにして採用行動を決定する。ここでは黒丸で示されたエージェントは実戦の中央エージェントの意思決定には

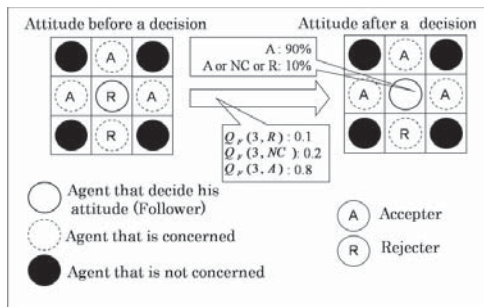


図 4 意思決定プロセスの例

関係しない。採用行動 b_t については採用 ($=: A$)、非採用 ($=: R$) または変化無 ($=: NC$) のうちの 1 つを選択するものとし、Q 値 $Q_e(n_t, b_t)$ は周囲 4 エージェントの採用者数 n_t とその時の採用行動 b_t およびエージェントの種類 (Follower ($=: F$) または Pioneer ($=: P$)) で決まるとする。

本研究における Follower および Pioneer の意思決定の初期ルールと Q 値の初期設定値を表 1 および表 2 のように定義する。また意思決定手法は確率 $1-\varepsilon$ で greedy 手法による行動選択を行い確率 ε ($0 \leq \varepsilon \leq 1$) でランダムに行動を選択する ε -greedy 手法を用いる。

この例の場合、周囲の 4 エージェントのうち 3 エージェントが採用者であるとカウントされるので $n_t=3$ の時の Q 値をチェックするが、ここでは次のようになっている：

$$Q_F(3, R)=0.1, Q_F(3, NC)=0.2 \text{ and } Q_F(3, A)=0.8.$$

この例の場合、 $Q_F(3, A)$ が 3 つのうちで一番大きな値となっている。本研究では行動決定において ε -greedy 手法 ($\varepsilon=0.1$) を用いることにしている

表 1 Follower の意思設定ルールと Q 値 (初期設定値)

n_t	Initial Rule	Initial values of Q values		
0	R	$Q_F(0, A)=0$	$Q_F(0, NC)=0$	$Q_F(0, R)=1$
1	R	$Q_F(1, A)=0$	$Q_F(1, NC)=0$	$Q_F(1, R)=1$
2	NC	$Q_F(2, A)=0$	$Q_F(2, NC)=1$	$Q_F(2, R)=0$
3	A	$Q_F(3, A)=1$	$Q_F(3, NC)=0$	$Q_F(3, R)=0$
4	A	$Q_F(4, A)=1$	$Q_F(4, NC)=0$	$Q_F(4, R)=0$

表 2 Pioneer の意思設定ルールと Q 値 (初期設定値)

n_t	Initial Rule	Initial values of Q values		
0	R	$Q_P(0, A)=0$	$Q_P(0, NC)=0$	$Q_P(0, R)=1$
1	A	$Q_P(1, A)=1$	$Q_P(1, NC)=0$	$Q_P(1, R)=0$
2	NC	$Q_P(2, A)=0$	$Q_P(2, NC)=1$	$Q_P(2, R)=0$
3	R	$Q_P(3, A)=0$	$Q_P(3, NC)=0$	$Q_P(3, R)=1$
4	A	$Q_P(4, A)=1$	$Q_P(4, NC)=0$	$Q_P(4, R)=0$

ので、90% の確率で態度を“採用”に変化させ、残りの 10% の確率で“採用”，“非採用”，“変化無”のうちのどれかの態度をランダムに選択することとなる。

2.3 Q 値の更新式および割引率について

本研究では、Q 値 $Q_e(n_t, b_t)$ の更新は時刻ステップ t における意思決定が行われた後に以下のアルゴリズムによって行われることとする。逐次更新される Q 値を学習成果とみなしこれを用いてエージェントの行動選択を行わせることにより、学習の成果に対応した行動選択を可能にしている^[10, 11]：

$$Q_e(n_t, b_t) \leftarrow Q_e(n_t, b_t) + \alpha [r + \gamma \max_{b_{t+1} \in B(n_{t+1})} Q_e(n_{t+1}, b_{t+1}) - Q_e(n_t, b_t)]$$

ここで $Q_e(n_{t+1}, b_{t+1})$ は、1 時刻先に周囲の状態が n_{t+1} になった場合の採用行動 b_{t+1} に対応する Q 値である。 r は状態 n_t から n_{t+1} へ遷移において得られる報酬、 $B(n_{t+1})$ は状態 n_{t+1} で実行可能な行動全体の集合を表し、 α ($0 < \alpha \leq 1$) は学習率、 γ ($0 \leq \gamma \leq 1$) は割引率を表している。Q 値の更新はエージェントが一つの行動を実行し、次の状態に移るごとに行われる。

割引率は、将来獲得予定の報酬を現時点でどれだけ重要に考えるかの割合と考えることができる。具体的には、割引率 γ が大きいときは将来得られる報酬を重視し、逆に割引率 γ が小さいとその場で得られる即時報酬のほうを重視することになる。本研究ではこの割引率 γ の値の大きさがエージェントの採用行動にどのような影響を与えるかについてコンピュータ・シミュレーションによる調査を行った。

3. シミュレーションの方法

これまでの議論に従い、本研究では以下の方法によってコンピュータ・シミュレーションを行うことにする。なお数値解析についてはすべて JAVA 言語を用いてプログラム作成を行った。

- [1] 表 3 から表 5 に示した 32 パターンのシミュレーションから 1 つを選択して各種のパラメータを設定する。
- [2] 100 個のエージェント（すべて Follower または Pioneer）を格子空間上

表 3 シミュレーションパターン結果（その 1）

No.	Agent	Initial Accepters	Setting of γ	$t=2500$ 時の Acceptor 数
1	Follower	$36(6 \times 6)$	0.5	91 ~ 100
2	Follower	$36(6 \times 6)$	0.8	89 ~ 100
3	Follower	$36(6 \times 6)$	0.9	89 ~ 99
4	Follower	$36(6 \times 6)$	0.95	67 ~ 91
5	Follower	$36(6 \times 6)$	0.99	92 ~ 100
6	Pioneer	$36(6 \times 6)$	0.5	90 ~ 100
7	Pioneer	$36(6 \times 6)$	0.8	89 ~ 99
8	Pioneer	$36(6 \times 6)$	0.9	86 ~ 100
9	Pioneer	$36(6 \times 6)$	0.95	79 ~ 94
10	Pioneer	$36(6 \times 6)$	0.99	89 ~ 99

表 4 シミュレーションパターン結果（ $t=4000$ ）

No.	Agent	Initial Accepters	Setting of γ	$t=4000$ 時の Acceptor 数
11	Follower	$36(6 \times 6)$	0.95	88 ~ 99
12	Follower	$36(6 \times 6)$	0.96	88 ~ 99
13	Follower	$36(6 \times 6)$	0.97	81 ~ 98
14	Follower	$36(6 \times 6)$	0.98	23 ~ 98
15	Follower	$36(6 \times 6)$	0.99	92 ~ 100
16	Pioneer	$36(6 \times 6)$	0.95	90 ~ 99
17	Pioneer	$36(6 \times 6)$	0.96	89 ~ 100
18	Pioneer	$36(6 \times 6)$	0.97	89 ~ 100
19	Pioneer	$36(6 \times 6)$	0.98	88 ~ 98
20	Pioneer	$36(6 \times 6)$	0.99	89 ~ 100

表5 シミュレーションパターン結果 (初期採用者数)

No.	Agent	Initial Accepters	Setting of γ	$t=2500$ 時の Acceptor 数
21	Follower	16(4×4)	0.8	92 ~ 100
22	Follower	16(4×4)	0.9	87 ~ 100
23	Follower	16(4×4)	0.99	92 ~ 100
24	Follower	64(8×8)	0.8	91 ~ 100
25	Follower	64(8×8)	0.9	88 ~ 100
26	Follower	64(8×8)	0.99	91 ~ 99
27	Pioneer	16(4×4)	0.8	89 ~ 100
28	Pioneer	16(4×4)	0.9	88 ~ 100
29	Pioneer	16(4×4)	0.99	89 ~ 100
30	Pioneer	64(8×8)	0.8	92 ~ 100
31	Pioneer	64(8×8)	0.9	89 ~ 100
32	Pioneer	64(8×8)	0.99	88 ~ 100

に配置し、初期採用者数とそれらの配置を与える。なお本研究では格子空間の中央部分に正方形状に初期採用者を配置する (図3参照)。

- [3] Follower と Pioneer の各行動パターンに対する Q 値の初期値を表1 および2のようにそれぞれ設定する。また本研究では $\alpha=0.1$ および $\varepsilon=0.1$ とし、 $\gamma=0.01$ は個々のエージェントが採用という意味を示したときにその都度与えることとする。
- [4] 各エージェントは Q 値に基づいて ε -greedy 手法により 1 時刻ステップにつき 1 回態度を決定し、それから (1) 式を計算して Q 値を更新する (図4参照)。
- [5] ここでの各エージェントの態度をチェックし、採用行動を行っているエージェントに報酬 γ を与える。
- [6] 全てのエージェントに対して時刻ステップ $t=2500$ または $t=4000$ までは採用行動を継続させ、そこでシミュレーションを強制終了させる。
- [7] 上記のプロセス [4] から [6] を 100 回繰り返し、これらのデータをもとにして考察を行う。

4. 結果および考察

4.1 割引率の変化と採用状況の関係について

まず初めに行ったシミュレーションの条件と結果を表 3 に示す。ここでは同条件で 100 回のシミュレーションを行った場合における採用者数の途中での変化やシミュレーションの終了時刻 ($t=2500$ 時点) での採用者数の収束先のばらつきを, 図 5 から図 14 から読み取って比較することにより検証する。

表 3 に記載したシミュレーション No. 1 から No. 10 までの採用者数の時

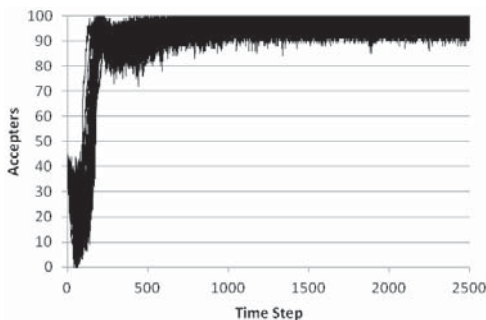


図 5 Simulation No. 1 における採用者数の推移

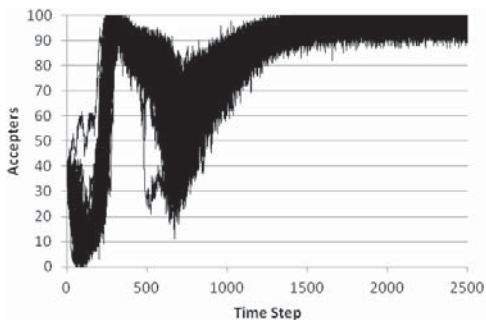


図 6 Simulation No. 2 における採用者数の推移

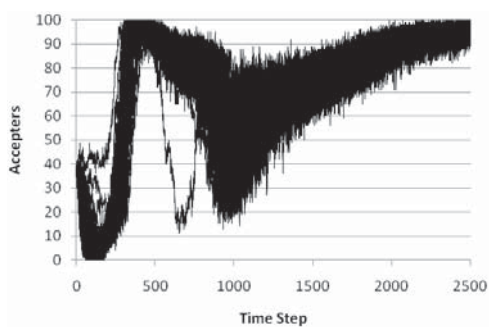


図 7 Simulation No. 3 における採用者数の推移

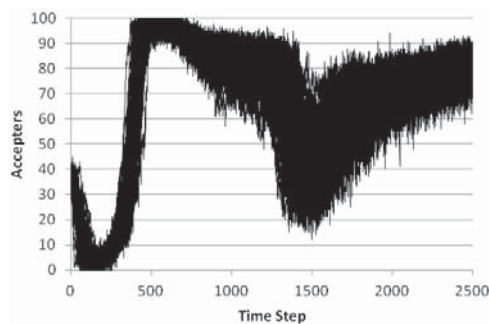


図 8 Simulation No. 4 における採用者数の推移

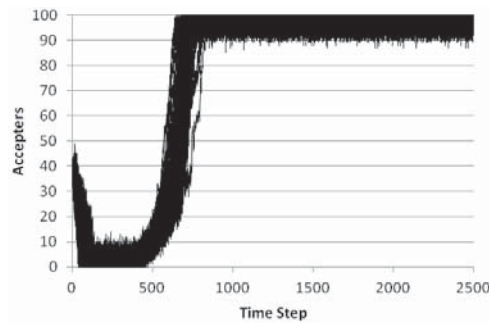


図 9 Simulation No. 5 における採用者数の推移

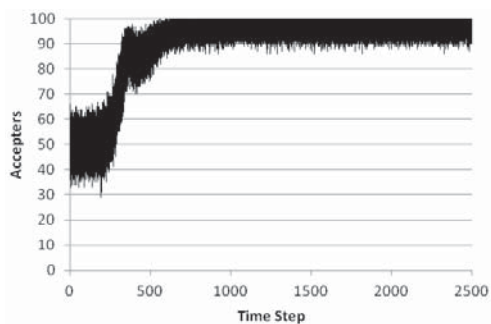


図 10 Simulation No. 6 における採用者数の推移

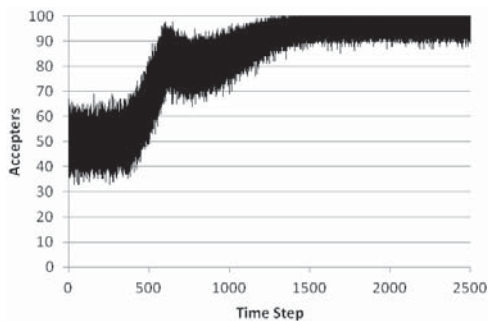


図 11 Simulation No. 7 における採用者数の推移

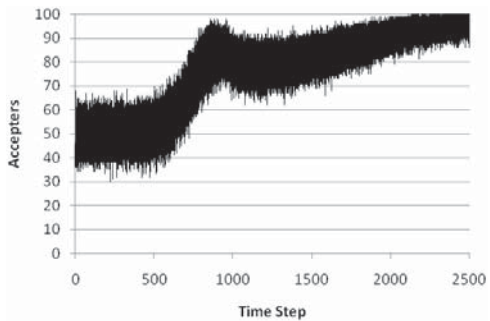


図 12 Simulation No. 8 における採用者数の推移

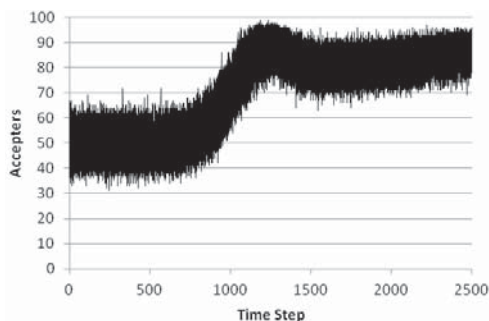


図 13 Simulation No. 9 における採用者数の推移

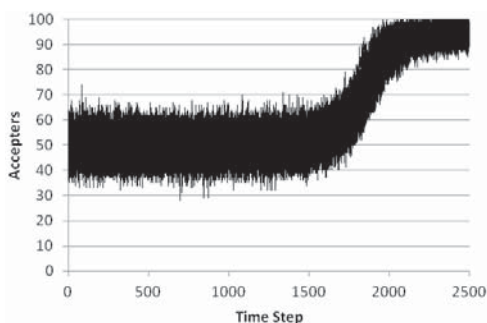


図 14 Simulation No. 10 における採用者数の推移

間変化は図 5 から図 14 に示されているが、これらの図にはそれぞれ 100 回のシミュレーションにおける採用者数の変化が同時に記されている。ここではこれら 100 回のシミュレーション結果のデータの挙動やばらつき具合から考察を行うこととする。

まず Follower からみると、 $\gamma=0.5$ から $\gamma=0.95$ までは一度採用者数が減少したあと増大し、再度減少してからまた増大してその後の挙動が落ち着くこと、そして割引率が増大するにつれて挙動が落ち着くまでの時間が長くなっていることが分かる。ただ $\gamma=0.99$ のとき、すなわち Sim. No. 5 のときは一度採用者数が減少したあと増大しそのまま挙動が安定しており

$\gamma=0.95$ までとは挙動が明らかに異なっていることが分かる。

Pioneer については、 $\gamma=0.5$ から $\gamma=0.95$ までの場合は途中まで採用者数が 50 前後で推移するが、途中から 100 近くまで増大してから少し減少し、そこから少しずつ増加して挙動が安定している。Pioneer のこれらの場合についても割引率が増大するにつれて挙動が落ち着くまでの時間が長くなっていることが分かる。Pioneer においても $\gamma=0.99$ のとき (Sim. No. 10) のときは、途中まで採用者数が 50 前後で推移した後に採用者数が増大してそのまま挙動が安定するというそれまでとは違った挙動を示している。

Follower にしても Pioneer にしても、 $\gamma=0.5$ から $\gamma=0.95$ までの場合と $\gamma=0.99$ の場合では挙動が明らかに異なっていることが分かる。 $\gamma=0.95$ から $\gamma=0.99$ での挙動はシミュレーションの終了時刻を $t=4000$ にして再度シミュレーションを行ったので、これらの結果については 4.2 で後述することとする。

4.2 割引率が高い場合 ($\gamma=0.95$ から 0.99)

表 3 でのシミュレーションだと $t=2500$ までの採用者数しか計算できないので、 $\gamma=0.95$ から 0.99 については 0.01 刻みの間隔で設定し、 $t=4000$ までのシミュレーションを再度行った。表 4 にここで行ったシミュレーションの条件を示しており、このときの採用者数の時間変化は図 15 から図 24 に示されている。

まず Follower の場合だが、 $\gamma=0.95$ から 0.97 までの場合は割引率が大きくなるにつれて採用者数が 80 以上に落ち着く時刻が遅くなるのがここでもはっきりと分かる。問題は $\gamma=0.98$ のときで、 $t=4000$ までだと採用者数が不安定になって落ち込むところで終わっている。今回使用したプログラムとパソコンではメモリ等の関係から $t=4000$ が限界だったので、これ以上のシミュレーションができなかったことを追記しておく。 $\gamma=0.99$

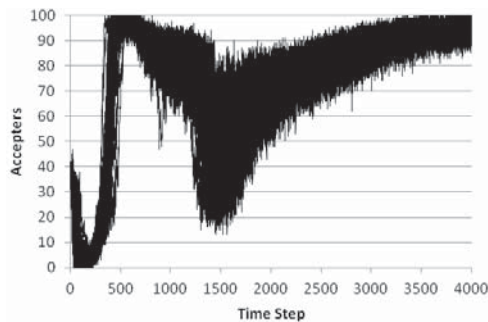


図 15 Simulation No. 11 における採用者数の推移

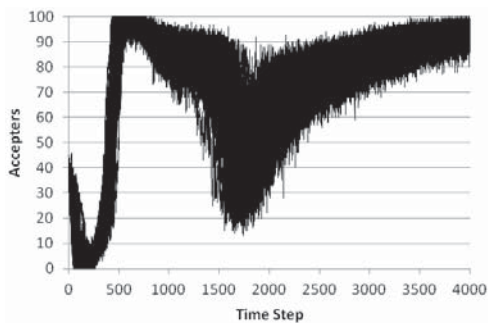


図 16 Simulation No. 12 における採用者数の推移

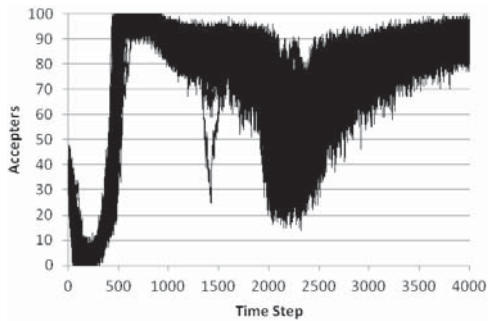


図 17 Simulation No. 13 における採用者数の推移

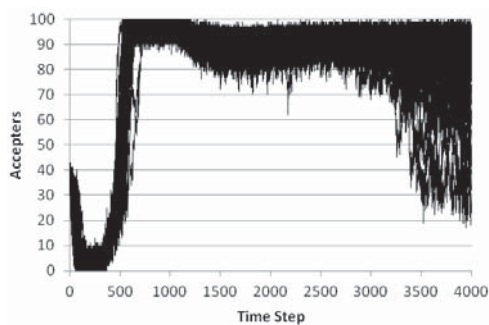


図 18 Simulation No. 14 における採用者数の推移

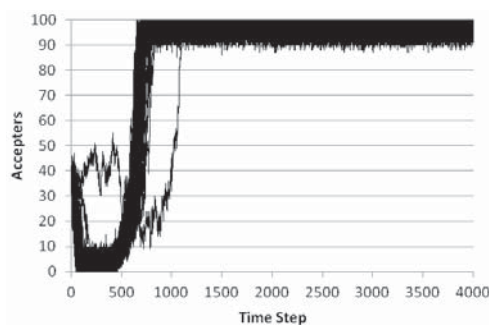


図 19 Simulation No. 15 における採用者数の推移

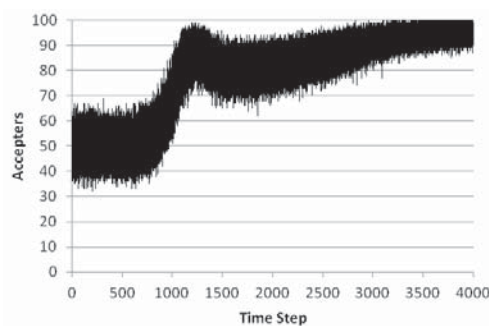


図 20 Simulation No. 16 における採用者数の推移

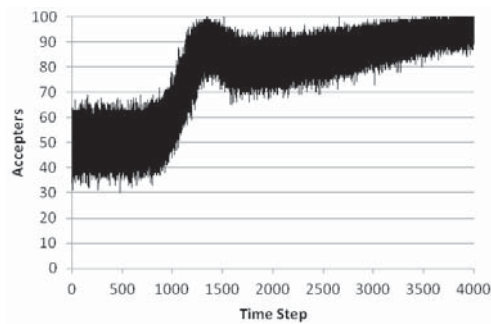


図 21 Simulation No. 17 における採用者数の推移

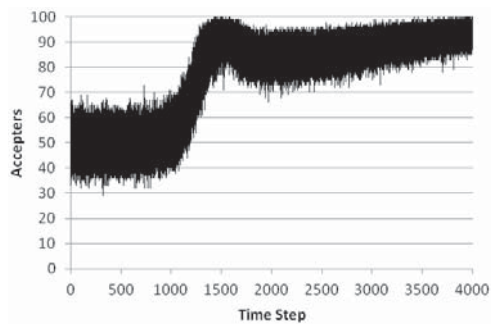


図 22 Simulation No. 18 における採用者数の推移

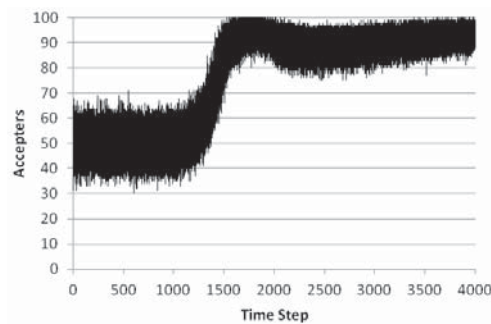


図 23 Simulation No. 19 における採用者数の推移

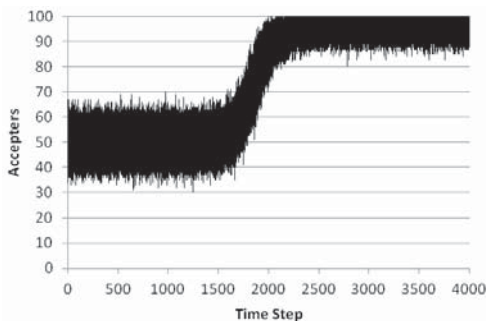


図 24 Simulation No. 20 における採用者数の推移

のとき (Sim. No. 15) の挙動がそれまでとは違ったものになっていることを考えると、今回設定したエージェントベースシミュレーションは $\gamma = 0.98$ から $\gamma = 0.99$ の辺りが挙動の変わり目であると推測される。

一方 Pioneer の場合は、図 20 から図 24 より $\gamma = 0.95$ から $\gamma = 0.99$ までのときのエージェントの採用状況については Follower ほどの大きな違いがみられないことが分かる。

4.3 初期採用者数の違いについて

ここでは初期採用者数がシミュレーション結果に及ぼす影響について考察する。なお本研究では初期配置による違いをなくすため、初期採用者については図 3 に示したように仮想空間内に正方形状に配置することにした。表 5 にここでの条件を示しており、このときの採用者数の時間変化は図 25 から図 36 に示されている。

初めに Follower の場合についての考察を行う。初期採用者が 16 のときは初期採用者が 36 のときとほぼ同じ挙動を示しているが、初期採用者が 64 の時は最終的には採用者がほぼ 9 割以上に収束するものの途中での採用者数の推移がシミュレーションによってまちまちになってしまっていることが分かる。これはランダムな要素を含む行動選択手法 (ϵ -greedy 手法)

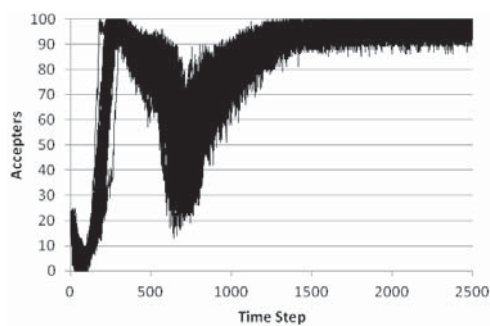


図 25 Simulation No. 21 における採用者数の推移

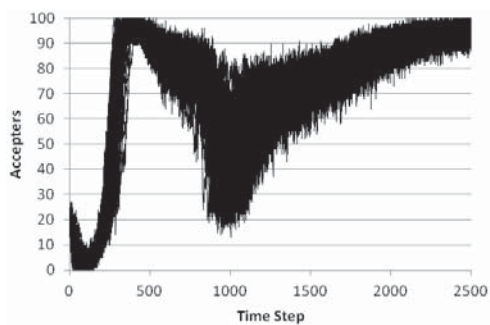


図 26 Simulation No. 22 における採用者数の推移

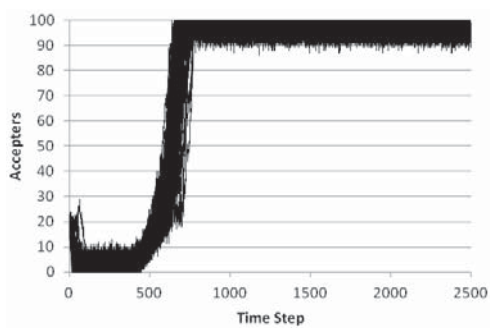


図 27 Simulation No. 23 における採用者数の推移

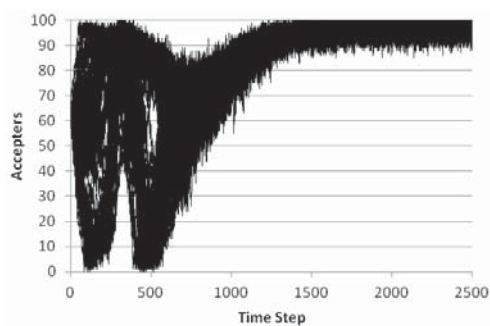


図 28 Simulation No. 24 における採用者数の推移

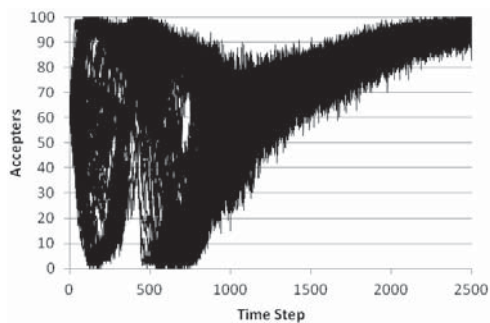


図 29 Simulation No. 25 における採用者数の推移

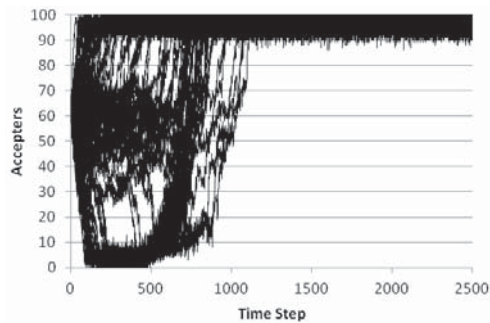


図 30 Simulation No. 26 における採用者数の推移

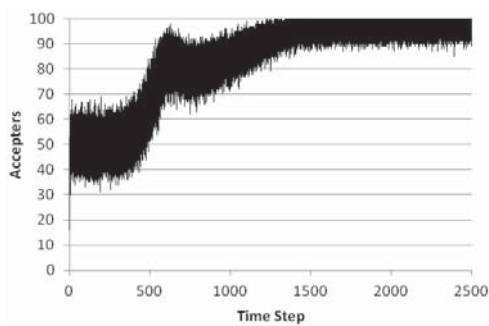


図 31 Simulation No. 27 における採用者数の推移

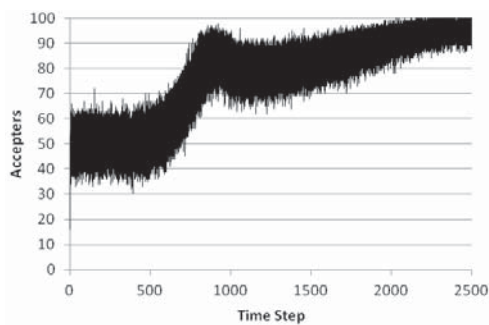


図 32 Simulation No. 28 における採用者数の推移

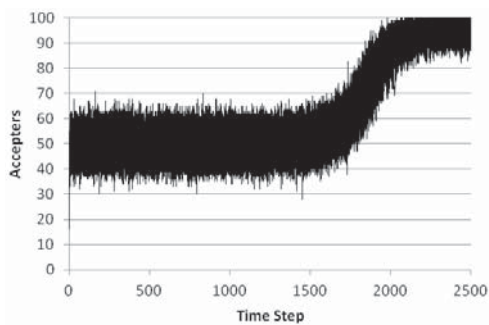


図 33 Simulation No. 29 における採用者数の推移

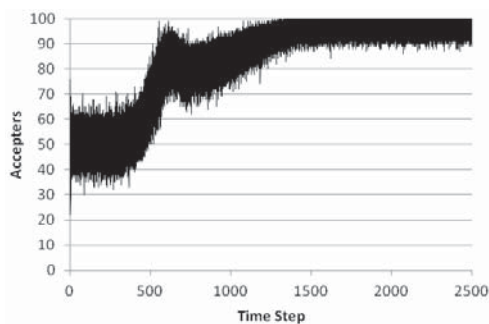


図 34 Simulation No. 30 における採用者数の推移

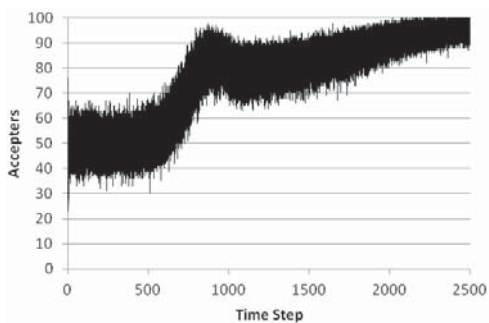


図 35 Simulation No. 31 における採用者数の推移

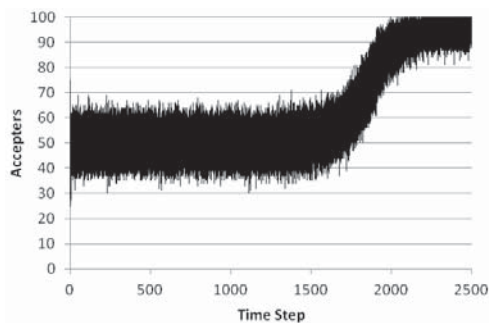


図 36 Simulation No. 32 における採用者数の推移

の影響がこの場合に限っては強く出ているものと思われる。

一方、Pioneer の場合は初期採用者が 16 であっても 64 であっても採用者数の変化は初期採用者が 36 のときとほぼ同じ挙動を示していることが分かり、初期採用者が 64 であっても Follower の場合ほど ε -greedy 手法の影響が強く出ない傾向であることが分かった。

5. 結 言

本研究では、著者が研究している Q 学習によるエージェントベースシミュレーションにおける割引率とエージェントの採用行動との関係にある程度明らかにすることができた。

まず今回のエージェントベースシミュレーションでは、初期採用者数が 36 に固定して検討した場合、 $\gamma=0.5$ から $\gamma=0.98$ までについては採用者数が 80 以上で安定するまでの時間が長くなり $\gamma=0.99$ になると途中の採用行動の経過が変わるだけでなく安定するまでの時間が $\gamma=0.98$ のときよりも短くなることが分かった。これは Follower, Pioneer とともに同じ現象がみられた。

初期採用者数が 36 と異なる場合についても今回シミュレーションを用いて考察を行ったが、初期採用者数が 36 と 16 の場合については Follower, Pioneer とともに途中経過および多くのエージェント (80 以上) が採用行動をとって安定するまでの時間について大きな差は見られなかった。初期採用者数が 64 の時は、Pioneer については初期採用者数が 16 や 36 のときとあまり違いがみられないものの、Follower のときは初期採用者数が 16 や 36 の場合と比べて最終的には同じように安定するものの、途中経過がシミュレーションによってまちまちになり全体的にみるとランダムな挙動が出現して傾向をひとくくりに議論できない状況になることが分かった。

今回のシミュレーション結果より、多くの結果についてはこれまでの研究結果や Q 学習の理論に沿った結果を得ることができたが、一部（初期採用者が 64 のときの Follower）については予想外の結果を得ることとなった。これまでの研究で Q 学習を用いた学習エージェントのシミュレーションを構築し、このシミュレーションについての様々な考察を行ってきた^[7-9]。エージェントのルール設定や報酬の与え方、行動選択手法や初期採用者の配置など検討する余地のある事項は多々あるものの、今回の研究を含むこれまでの研究結果により感覚的ではあるがシミュレーションの傾向や挙動をつかむことができたのは大きな成果である。今後これらの結果をもとにして再度シミュレーションのルールや報酬、外部からの影響（Global Influence^[5]）や各種パラメータの設定などを行う予定である。

参考文献

- [1] 山影進，服部正太編：コンピュータのなかの人工社会 マルチエージェントシミュレーションモデルと複雑系，共立出版（2002）.
- [2] J. M. Epstein and R. Axtell（服部正太，木村香代子訳）：人工社会 —— 複雑系とマルチエージェント・シミュレーション ——，共立出版（1999）.
- [3] 二瓶善博：うわさとくちコミマーケティング（第二版），創成社（2003）.
- [4] 藤井慎太郎：エージェントベースシミュレーションを用いた社会的拡散現象の解析，北陸先端科学技術大学院大学知識科学研究科修士論文（2003）.
- [5] N. Nakano: An Agent-Based Approach to Simulations of Global Influence with Random Effect, Trans. International Journal of Innovative Computing, Information and Control (IJICIC), Vol. 2 No. 5, pp. 1153-1163 (2006).
- [6] 中野統英：社会的拡散現象における重要度を考慮した影響力の解析 ～エージェントベースシミュレーションを用いて～，日本繊維機械学会誌 Vol. 61, No. 9, pp. 615-620 (2008).
- [7] 中野統英：社会的拡散現象における学習効果を考慮したエージェントが引き起こす影響の解析～Q 学習を含むエージェントベースシミュレーション

- を用いて～, 日本繊維機械学会誌 Vol. 63, No. 9, pp. 499-506 (2010).
- [8] N. Nakano : A Study on Modeling and Analysis of Agent-Based Simulations with Q-Learning, Trans. International Journal of Innovative Computing, Information and Control (IJICIC), Vol. 7 No. 1, pp. 51-60 (2011).
- [9] 中野統英：エージェントベースシミュレーションにおける Q 学習を用いたエージェントによる報酬と流行の流布との関係に関する研究, 合同エージェントワークショップ & シンポジウム 2011 (JAWS 2011), 8 pages (2011).
- [10] 高玉圭樹：マルチエージェント学習 —— 相互作用の謎に迫る ——, コロナ社 (2003).
- [11] R. S. Sutton and A. G. Barto (三上貞芳 皆川雅章 共訳)：強化学習, 森下出版 (2000).
- [12] E. M. Rogers (青池慎一, 宇野善康訳)：イノベーション普及学, 産能大学出版社 (1990).

Studies on Relations between Discount Rate and Circulation of Rumors for Agent-Based Simulations Comprising Agents with Q-learning

Nobuhide NAKANO

Abstract

Many researchers study estimation and simulation problems for economic and social activities. In a few decades, agent-based approaches have been attracting much attention as one of simulating method for not only various social and economic phenomena but also marketing researches. Agent-based simulation is the method for analyzing various phenomena in virtual spaces which researchers set up. At first, agents are given some initial conditions and performance rules. Then, these agents are arranged in a virtual space and they perform autonomously according to their performance rules.

Among marketing researchers, agent-based simulations are one of the best methods for analyzing diffusion processes of rumors and fashions, for example selections of goods which consumers purchase. The author researches agent-based approaches for diffusion processes of rumors, their characters are simply changed randomly in some fixed patterns. As many agents, especially trendsetters, often change their action rules by attitudes of neighbor agents in real society, we cannot disregard the effect of learning for agents in the circulation of rumors and fashions. Hence, the author introduces the reinforcement learning for the research of diffusion processes.

Though there are some methods in reinforcement learning, the author consider that Q-learning is one of good methods for analyzing diffusion of rumors and fashions because Q value, which is action-value function at Q-learning, is suitable for expressing characters of agents and convergence to optimal states is proved at Q-learning.

In this study, the author treats the relation between discount rate and circulation of rumors for agents with Q-learning in agent-based simulations because discount rate is one of the important parameters which influence reward given agents. At first, the action rules for two

types of agents, called pioneers and followers, are modeled by using Q values. Then, the updating algorithm of Q values is designed with optimal various parameters. In this study, ϵ -greedy method is used as choosing the method of action rules. After some simulation studies are performed, simulation results and analysis are shown. As the results, we know that the relation between discount rate and acceptance actions of agents are very sensitive in limited areas.

Keywords : agent-based simulation, Q-learning , discount rate, ϵ -greedy method and categorization of agents