報酬系と嫌悪系によるネズミの学習モデル

大阪歯科大学一年　米田正器

１．序

ネズミの脳の多数の神経細胞発火による意思決定システムをマルチエージェントシミュレーターによってモデル化した。

このモデルの特徴は単一のニューロンの発火により意思決定を行っているのではなく、一見ランダムに発火する多数の神経細胞群の比率の変化によって行動が決定されるところにある。

さらに私は報酬系（チーズの報酬度）や嫌悪系（電気ショックの嫌悪度）等の設定を調整することによって、モデル性格、状況をつくり学習能力の比較を行った。

２．マルチエージェントシミュレーター（ＭＡＳ）について

各々の内部属性に関連づけられた独自の意思決定メカニズムと行動計画に基づき、自己の利益を追求する活動主体のことをエージェント、そしてそれらが相関関係をもつ集合体のことをマルチエージェントという。

今回用いたマルチエージェントシミュレーターとは構造計画研究所(http://www.kke.co.jp)でつくられた、これらの動き（複雑系）をシミュレートとするソフトである。

３．モデルの説明

Ⅰ．モデルの概略

モデルはネズミの意思決定、思考を表した空間（脳内空間）とネズミが動き回るＴ字迷路（環境空間）、そして数種類の agent から構成される。

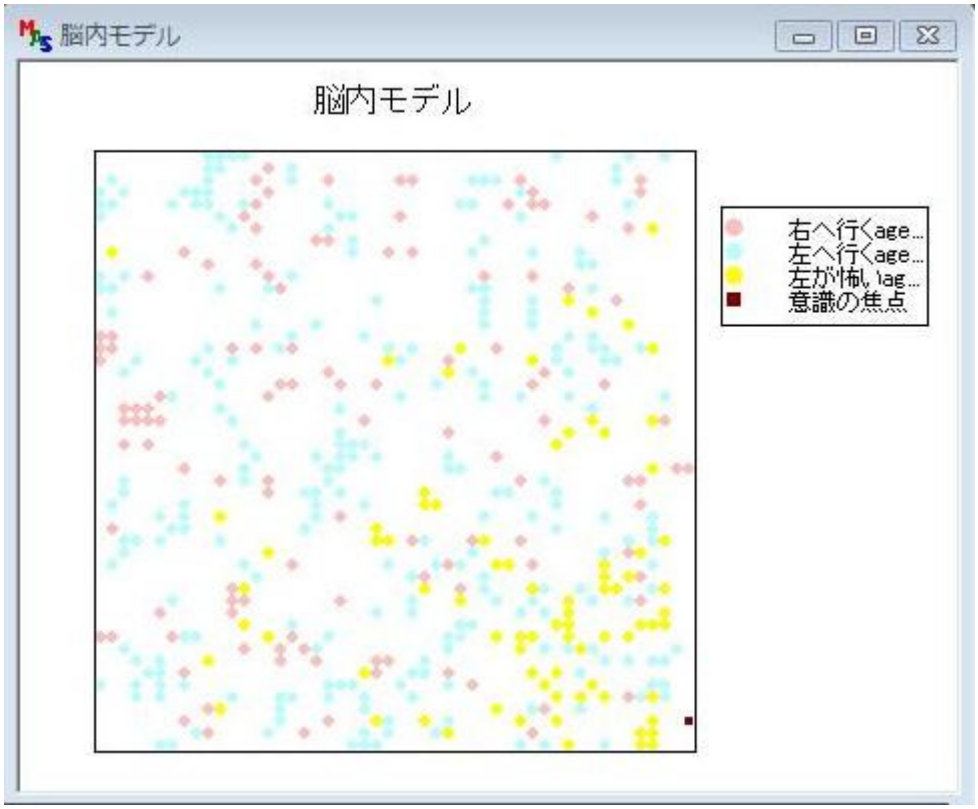
ネズミは脳内空間の agent の動きから意思決定を行い、環境空間を動く。

環境空間での刺激（報酬系＆嫌悪系）は脳内空間における agent の構成に変化を与え、ネズミの意思決定に影響を与える。この際、学習が行われる。

この一連の流れは、私たちが自分の考え方によって行動を起こし、その行動によって得られた結果から反省をし、反省によって変化が与えられた考え方が次なる行動をおこす。そのようなものをイメージしてもらえばわかりやすい。

Ⅱ．脳内空間

脳内空間では**３種類の意思決定 agent** とランダムに動く**意識の焦点**という点によりネズミの意思決定が行われている。



３種類の意思決定 agent

●…**右へ行く agent。**ネズミがチーズを食べるとこの agent が増える。この agent が増えるとネズミは右（チーズ）へ行きやすく、前進しやすくなる。

●…**左へ行く agent。**この agent が増えるとネズミは左（電気ショック）へ行きやすく、前進しやすくなる。

●…**左が怖い agent。**ネズミが電気ショックを受けるとこの agent が増える。この agent が増えるとネズミは右へ行きやすく、左へ行きにくくなると同時に、前進することも恐れだす。

脳内空間におけるルール

◆意思決定 agent は以下のように、配置される。

・プログラム開始時に初期設定の数だけランダムに設置される。（右へ行く、左へ行く agent のみ）

・環境空間においてネズミがチーズか電気ショックに到着すると設定数だけ対応した agent が生成される。このとき生成された agent は意識の焦点を中心にランダムに配置される。（右へ行く、左が怖い agent のみ）

◆意思決定 agent は以下のように消滅する。

・総 agent 数が設定数以上になった時、余分な agent の割合をプログラムが計算する。agent はそれぞれ割合から出された確率によって自殺し均衡を保つ。

◆意思決定 agent は以下のように行動する

・自分の近傍１マスを見渡し自分と同じ agent が多ければその場所を動かない。自分と同じ agent が少なければランダムに移動する。このため **agent は同じ色（種類）同士が固まりをつくる**ことになる。

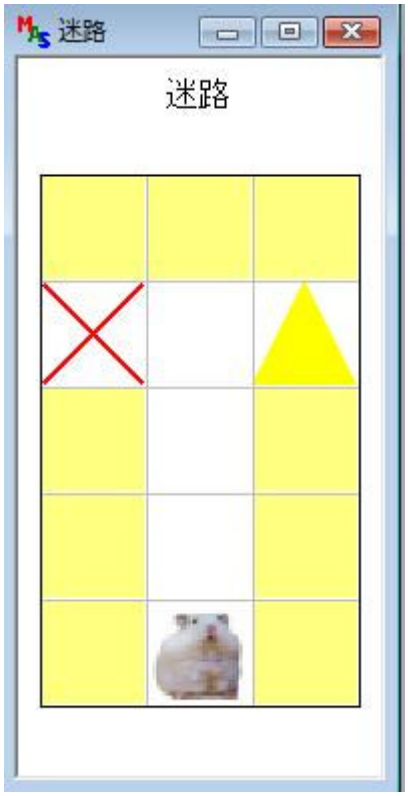
◆意識の焦点とは３種類の意思決定 agent から独立した存在である

・意識の焦点は常に１つのみであり生成、消滅をしない。これ自体はランダムで動き続けるだけの点である。

この点の役割は自分の近傍の意思決定 agent を採取することにある。脳内空間には多数の agent が存在するが、ネズミの意思決定は全ての agent の総合で決められているわけではない。この点の近傍１マス（８マス分）の agent のみにより決められる。

だから、ある種の意思決定 agent が脳内空間においてマイノリティであったとしても、場所が上手くかみ合えば、意見を取り入れられることになる。もちろん、数が多ければ多いほどこの点に意見を取り入れられやすいことは言うまでもない。

Ⅲ．環境空間



環境空間（T字迷路内）にはそこに住むネズミ、そして報酬系であるチーズ（黄色い三角）と嫌悪系である電気ショック（赤いバツ）が存在する。

チーズと電気ショックは固定であり、この中で動的な存在はネズミのみだ。**ネズミは脳内空間での意思決定により動き、**チーズか電気ショックへ向かう。チーズか電気ショックに到達したネズミはそこから受けた刺激を**脳内空間に新規の意思決定 agent を生成する**ことによりフィードバックし、次のステップで再びスタートへと戻される。

◆ネズミの行動ルール

１．ネズミが通路（含スタート地点）にいる場合

脳内空間の意識の焦点近傍で**「右へ行く agent＋左へ行く agent＞左が怖い agent」**なら前へ１マス進む。そうでなければ、すくんで動かない。

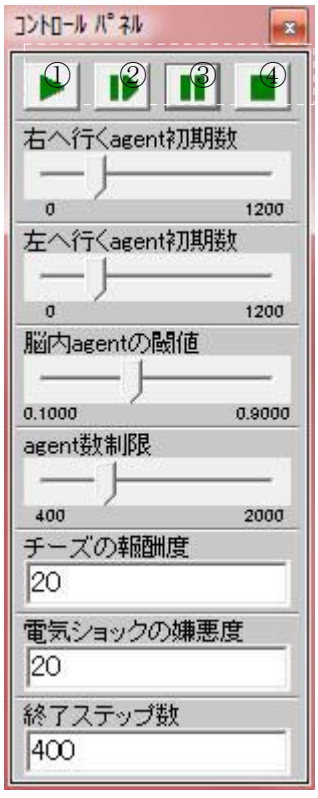
２．ネズミが分岐（チーズと電気ショックの間のマス）にいる場合

脳内空間の意識の焦点近傍で**「右へ行く agent＋左が怖い agent＞左へ行く agent」**なら右のチーズへ進む。

「右へ行く agent＋左が怖い agent＝左へ行く agent」なら迷って動かない。

「右へ行く agent＋左が怖い agent＜左へ行く agent」なら左の電気ショックへ進む。

Ⅳ．コントロールパネル



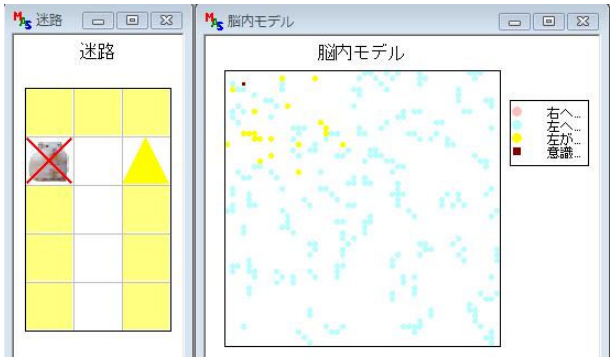
①　プログラム開始　②　プログラムコマ送り　③　一時停止　④　プログラム終了

右へ行く agent 初期数…脳内空間に初期状態から存在する右へ行く agent の数
（初期値 250　これが多いと最初から右へ行きたがる）
左へ行く agent 初期数…脳内空間に初期状態から存在する左へ行く agent の数
（初期値 250　これが多いと最初から左へ行きたがる）
脳内 agent 閾値…脳内空間において同種類の agent との固まりをつくりやすさを表す
（初期値 0.4　数値が高すぎると逆に固まりにくくなるのでこれくらいが適量）
agent 数制限…脳内空間の意思決定 agent の合計数がこの設定以上になるとランダムでの淘汰が始まり数を制限する。
（初期値 800　多すぎてもいいものじゃないのでこれくらいが適量）
チーズの報酬度…ネズミがチーズに到達した時に脳内空間に増える右へ行く agent の数
（初期値 20　報酬系。積極的な学習を促すため効果は強く、数値を大きくしすぎるとデータとしてすぐに飽和し、面白みに欠ける。）
電気ショックの嫌悪度…ネズミが電気ショックに到達した時に脳内空間に増える左への恐れ agent の数
（初期値 20　嫌悪系。消極的な学習を促す。報酬系と上手く組み合わせてやれば面白い結果を観察できる。今回の考察は主にこの数値の調整によって行った）
終了ステップ数…プログラムが終了するまでのステップ数
（初期値 400　ステップ数が多すぎると脳内空間が右へ行く agent だけで飽和する）

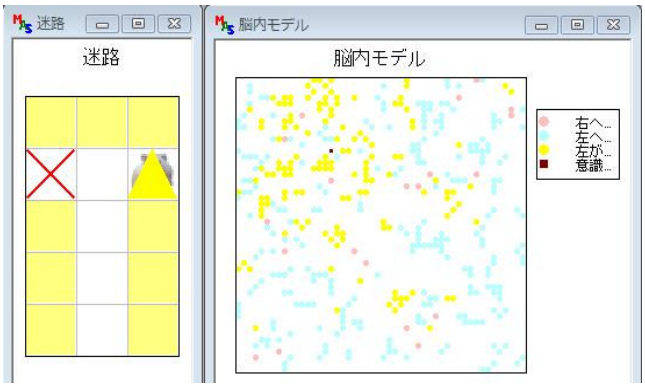
４．学習の流れ

学習の流れは脳内空間における（意思決定 agent の）色の分布に表される。

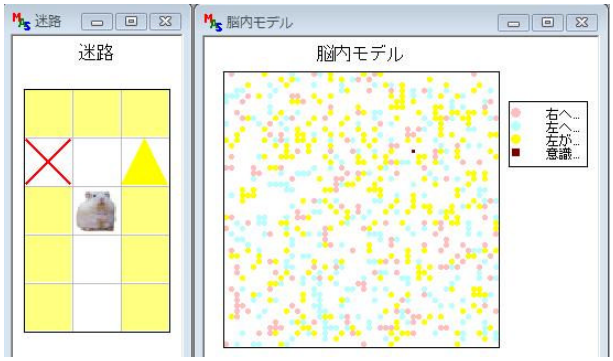
下に示した例は初期設定で左に行く agent しかない場合の例だ。最初は電気ショックにしか向かわないネズミがどのようにしてチーズを学ぶか、順を追って示した。脳内空間で色（意思決定 agent）が変化していく様を見てもらいたい。



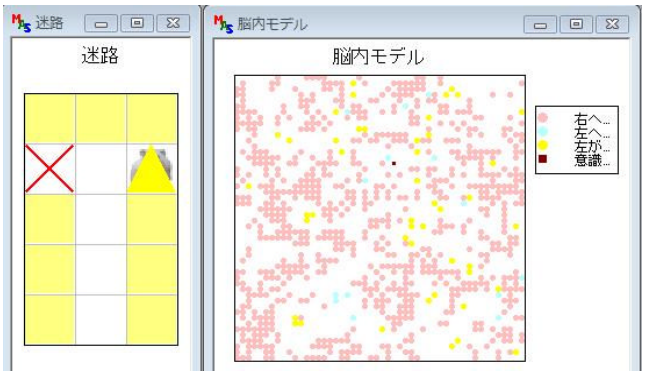
① はじめ、ネズミの脳内には左に行く意思（青色の点）しか存在しないのでネズミは電気ショックへ行く。電気ショックに到達すると、電気ショックは嫌なので、ネズミの脳内には左への恐れ（黄色の点）が生成される。



② ネズミの脳内で左への恐れ（黄）が増えてきた。ネズミは左へ行くのが嫌になり、消去法的選択で右へ向かいだす場面がでてくる。
するとそこにはチーズがあったので右に行く意思（赤）が出現した。



③ 左への恐れ（黄）と右に行く意思（赤）が組み合わさり、ネズミがチーズに行く確率が増えてくる。チーズへ行く度に右へ行く意思が（赤）生成されるので、右へ行く意思（赤）の割合が大きくなる。



④ 右へ行く意思（赤）がほぼ全体を支配してくる。こうなればネズミはほとんどチーズにしか行かない。しかも恐れによる消極的選択（黄）ではなく積極的選択（赤）が主要な動機となるため、ネズミの行動スピードは早くなる。

５．モデル性格を設定しての考察

Ⅰ．考察方法

このプログラムはコントロールパネルの設定により様々なモデル性格や状況をつくりだすことが可能だ。たとえば、電気ショックの嫌悪度は、ネズミへの罰の量を表すだけでなく、ネズミの危機的刺激に対する鋭敏さを示すとも考えられる。危機的刺激（電気ショック）に対する反応である左が怖い agent は消極的選択として正解（チーズ）へ導く反面、ネズミをすくませる作用がある。つまり、この左が怖い agent が少なすぎるとネズミは何も恐れず素早く動くが間違い（電気ショック）が多く、多すぎると正解へ多く進むが、すくんだ動きになるのだ。

この数値（電気ショックの嫌悪度）を変化させながらシミュレートすると、０付近は動きが素早いが間違いが多く、それを改善しようと数値を増やすと間違いこそ少なくなるが動きが鈍くなるという二律背反になる。この中でネズミにとっての**最適**はどこにあるのだろう？

そこでネズミ行動の適切さを測る基準として以下の式を設けた。

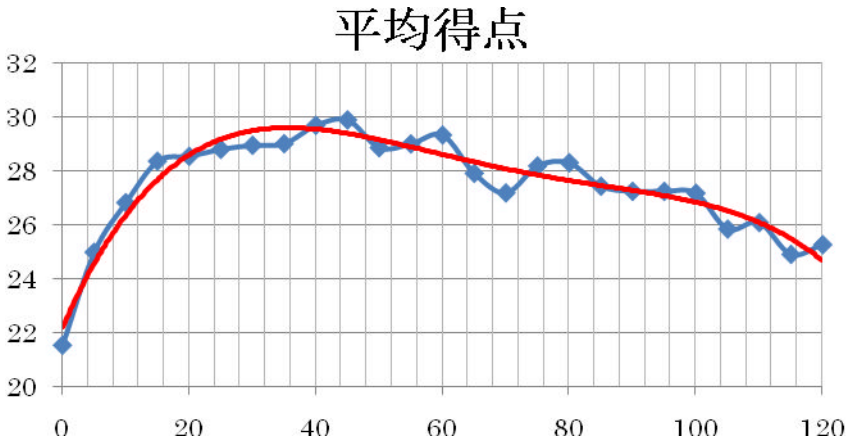
得点　＝　チーズを得られた回数　－　電気ショックを受けた回数

この項では設定（電気ショックの嫌悪度）の変化に伴う得点の変動を調査し、どのような鋭敏度を持ったネズミが優れているのか？平均点、分散、分布などの面から考察した。

Ⅱ．白紙からの学習モデル

まずは基本的な設定のシミュレーションを行った。初期設定でネズミがチーズと電気ショックに到達する確率は五分と五分。つまり白紙の状態にある。この状態からネズミが 300 ステップの間にどれだけ学習をし、適応することができるか。電気ショックの嫌悪度を変化させ、それに伴う平均得点の変動をグラフ化してみた。
電気ショックの嫌悪度は 0~120 まで 5 ずつ変化させ、それぞれにつき 100 回ずつ実行した（合計 2500 回）

（詳細は添付の Excel データ資料１にて）

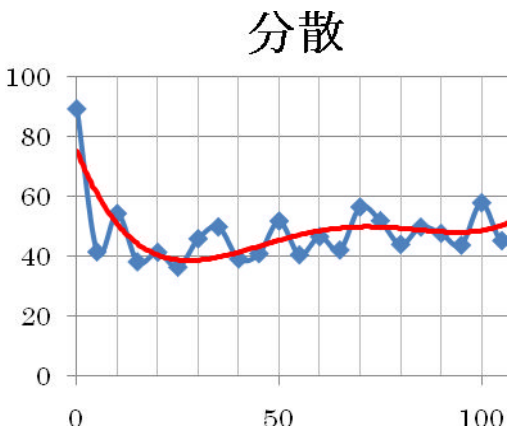


設定
右へ行く agent 初期数　250
左へ行く　agent 初期数　250
脳内 agent の閾値　0.4
agent 数制限　800
チーズの報酬度　20
終了ステップ数　300
電気ショックの嫌悪度　変化
（0 から 120 まで 5 ずつ）

縦軸：得点　横軸：嫌悪度数値

平均得点は電気ショックの嫌悪度が 0 の時点で最低であり、ここから 15 まで急上昇する。報酬系だけでなく、嫌悪系が組み合わさることによって学習の効果が上がる場面だ。グラフはその後、電気ショックの嫌悪度 20~50 付近で飽和を見せ、そこからは緩やかな下り坂を描き全体としては**くじらの背中**のように見える。
この状況におけるネズミは嫌悪系に対する鋭敏度が 20~50 の域でバランスを保つことがわかった。

それは右の分散グラフからも読み取れる。
右のグラフは上と同じデータを嫌悪度ごとの平均得点ではなく、分散で表したグラフだが、20~40 付近で分散が低くなっている。これは平均得点が高くなっている域と重なっている。
つまり、**最も得点が高い域は同時に安定した結果が返ってくる域**でもあった。

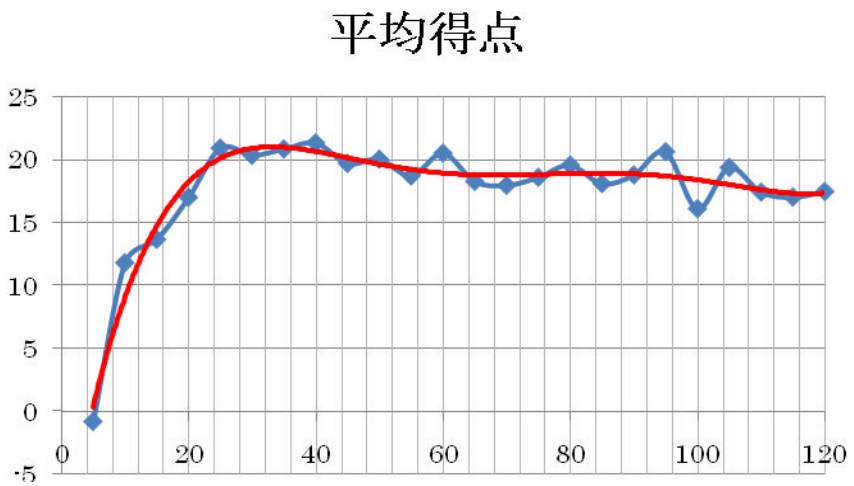


Ⅲ．急激な環境の変化に対する適応学習モデル

今度のモデルでは**急激な環境の変化にどれだけ適応できるか**について注目してみた。これは種が生き残るために重要なことである。このモデルは初期状態において左へ行く agent しか存在していない。つまり、**今までずっと左へチーズがあって慣れていたのに、ある日を境にチーズと電気ショックの場所が入れ替わってしまった。**そんな状況を表すモデルだ。agent 初期数以外の設定は上の「白紙からの学習モデル」とほとんど同じだが、今回は真逆の状況に適応しなければならないのでチーズの報酬度を 20→30 に、終了ステップ数を 300→400 にして若干学習をしやすくしている。

グラフはその設定のもと電気ショックの値を 5～120 まで 5 ずつ変化させて計測したもの。今回もそれぞれにつき 100 回実行し、その平均得点をプロットした。（合計 2400 回。0 を値として加えなかったのは嫌悪度 0 の場合、電気ショックを受けても左への恐れ agent が出現せず学習しようがないから）

（詳細は添付の Excel データ資料 2 にて）



設定
右へ行く agent 初期数 0
左へ行く agent 初期数 250
脳内 agent の閾値 0.4
agent 数制限 800
チーズの報酬度 30
終了ステップ数 400
電気ショックの嫌悪度 変動
（5 から 120 まで 5 ずつ変化）

縦軸：得点　横軸：嫌悪度数値

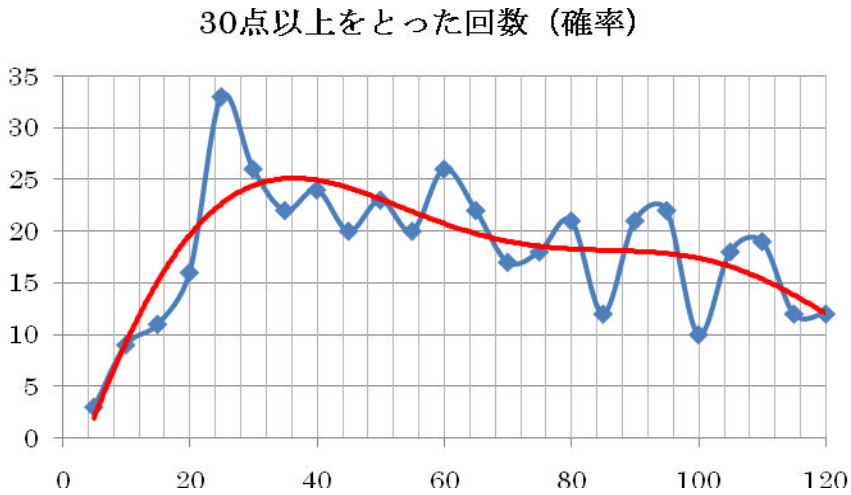
結果は前モデルと似た**くじらの背中型**になった。ただし、今回は新しい環境への適応ということもあり上下の揺れ幅は大きくなった。前回が最低平均点と最高平均点の間に 8 程の差しかなかったのに対し、今回は 22～23 と約 3 倍もの差がある。この差は主に電気ショックの嫌悪度が低い時の落ち込みから来ている。（今回は電気ショックの嫌悪度が 5 のとき平均得点は 0 未満となっている！）

そのためＸ軸の最低値を 20 に設定している前回のグラフと比べると、後半部分（電気ショックの嫌悪度 60 以降）の勾配が緩くなっているような印象を受けるが、実際は似たような勾配だ。今回も電気ショックの嫌悪度が 25~50 あたりで山を迎えているところからすると、この辺の数値が生き残るために持つべき鋭敏度として相応しいのかもしれない。

ところでこのグラフを見て次のことを思うかもしれない。

前半の得点が急速に伸びるところは情報として価値が高いが、後半の緩やかな下り坂に情報としての価値はあるのか？単なる飽和状態ではないのか？大事なことは**どれどれ以上の鋭敏度があれば生き残れるか**（このグラフなら 25 くらいだろう）であり、あとは 40 であれ 60 であれ 100 であれ似たようなものではないか。

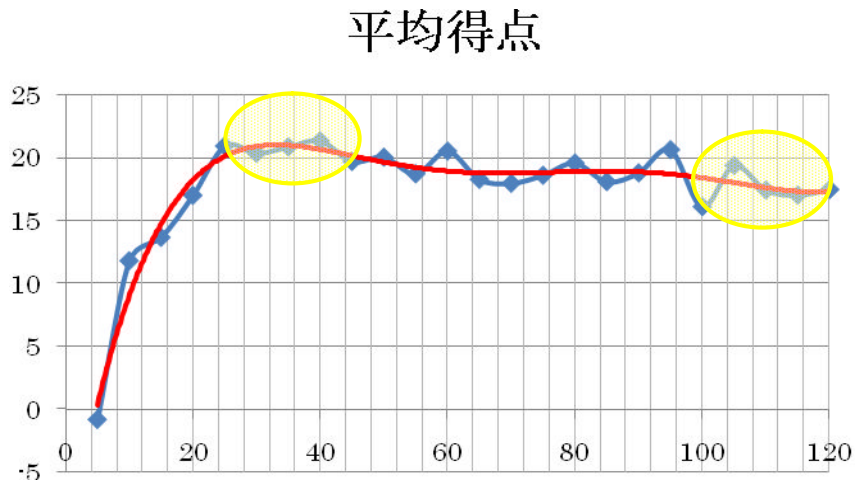
これに関してはデータを別の視点から眺めることによって答えを得られる。



このグラフはネズミが 30 点以上をとった回数（確率）を同じデータから拾ってきたものだ。**前のグラフがクラスあたりの平均点なら、このグラフはクラスあたりの合格者の数といえる。**生き残るということは淘汰をしのぐということであるから、評価としては平均点を見るより適切かもしれない。

この 30 点以上という基準をクリアーしたのは 2400 回のうち 437 回で約 1/6 となる。仮にこの 30 点を淘汰の基準とするなら、このグラフは淘汰に生き残れたものの数を表す。

ここにおいて緩やかだった後半の傾斜が牙をむく。平均得点ではなく合格ラインを決めての淘汰になると、電気ショックの嫌悪度が **25**~40 の場面において、100~120 の場面の**倍近くの生存者、適応者を出す**ことになる。緩やかな傾斜が無意味なものではなかったことがわかる。



実はこの僅差に生き残るための
倍差が潜んでいた？

６．結論と今後の課題

たとえば、ゲームのキャラクター。おもちゃのロボット。本物を似せたニセモノは数知れない。このプログラムのネズミもその一部だ。

これらニセモノ造りのセオリーとしては、本物から具体性のある要素をとりだし真似ることが一般だ。しかし、私はなるべく具体性や意図性のある要素を排除することにした。このプログラムで唯一具体性を持つのはネズミの画像くらいで、これをとってしまえばネズミは点の散らばりに分解される。それはネズミが精巧なニセモノを目指してつくられたのではなく、あやふやなニセモノを目指してつくられたからだ。

ネズミは明確な命令を与えられず、いくつかの簡単なルールを持たされただけでＴ字迷路に放り込まれる。放り込んだ時点で人の手を離れ、自らの意思決定に従い動きだす。ネズミにとって私は造物主だが、造物主はネズミがどう動くかを知らない。彼はあやふやにつくられたからだ。だから私はネズミがどう動くかを眺めた。

「５．モデル性格を設定しての考察」がその結果である。

もちろん、この結果をもって「本物のネズミもこうなんだ！」と結論付ける気はない。遠い将来あったとしても、少なくとも今は、あやふやなニセモノの本分ではない。

大事なことは私の手を離れたネズミがそこに現象を残してくれたことだ。たとえば「白紙からの学習モデル」において平均得点と分散との最適点が重なっていること。または、「急激な環境変化に対する適応学習モデル」では平均得点の僅差が淘汰を意識した選別において倍差を生んでいたことなど。

単純な要素を組み合わせただけのものにも関わらず、生命を感じさせるような複雑な動きを残してくれたところにプログラムはひとまずの成功を収めたと思う。

今後の課題はプログラムをより汎用性と有用性のあるものに改造することだろう。

今のプログラムをたとえて言うならば、十得ナイフからペーパーナイフだけを取り出して、それで切り絵をしているようなものだ。紙を切るだけならハサミを使えばよい。それだけの為に器用貧乏の代名詞である十得ナイフをひっぱり出すのはナンセンスだ。右が正解なら一度右へ行けばあとは右に進むようにすればよい。その方が簡単に学習するし、プログラムもシンプルですむからだ。

このようなシステムを取り入れた限りには、私はその有用性を示さなければならないし、このシステムがその範囲をＴ字迷路の内だけに留めないと考えたからこそ私はこれを取り入れることにした。

つまり現状では非効率的なシステムが、ある種の使用方法では一般的な効率化されたシステムには出来ない働きをしてくれると考えている。

次のモデルではその具体的な形を１つ提案したい。

７．参考文献

[1] 山影　進，服部　正太（2002），「コンピューターのなかの人工社会　マルチエージェントモデルと複雑系」，共立出版
[2]Richard S.Sutton，Andrew G.Barto（1998），*Reinforcement Learning*、MITPress,（三上　貞芳，皆川　雅章　訳（2000），「強化学習」，森北出版株式会社）
[3]星野　力（1994），人工生命の夢と悩み，裳華房