1. 序論

1.1 本研究の背景と目的

自動車は人々の暮らしには欠かすことのできない交通手段である．その歴史は蒸気機関車や蒸気船よりも古く，1769 年フランス陸軍の技術大尉ニコラ＝ジョセフ＝キュニョーの作成した蒸気自動車がはじまりであるとされている．自動車の普及は人々に多大なる便益をもたらしたがその反面，生命，健康，安全歩行などの人々の基本的権利の侵害をも同時にもたらした．自動車が自走することに起因する交通事故，自動車が燃料を燃焼させることに起因する有害な排気ガス，騒音などの環境問題がその例として挙げられる．発展途上国の先進化が著しい現代社会において，世界規模でこの自動車による社会的コストが増加しているにもかかわらず，自動車の発明以来 200 年以上の歳月が経過した今もこれら社会的コストを改善するような新たな交通手段としての個人向けの乗り物は普及を見ない．

そこで本研究の対象となる二酸化炭素など有害な物質を排出しない個人向けかつ一人乗りの低速度交通手段である LSPTM(Low Speed Private Travel Mode)をこの次世代交通手段として位置づける．これは自動車にとって替わるものとしてではなく，自動車と共存させることでそのコストの低減に取り組むものとしての新規交通手段である．本研究では都市部，とりわけ娯楽施設が多く点在し，歩行者が多数存在するような都市部での LSPTM の導入を想定する．現在自動車交通が過密状態にある都市部では自動車による交通事故や，交通渋滞に起因する自動車の旅行時間増加に伴う二酸化炭素排出量の増加などの社会的コストが深刻である．そこでそのような都市部に自動車進人禁止区域を設置し，人々にその区域周辺に自動車を駐車させ，そこから別の交通手段で目的地までの交通を行わせるような規制を行う．そしてこの駐車場から目的地までのトリップに LSPTM の導入を検討する．なお LSPTM の所有は個人ではなく，公共交通として個人に貸し出されるようなものを仮定する．

このような利用の想定を踏まえた上で LSPTM の普及を目指すには，まず事前にその需要を予測し，長期にわたって安定したサービス提供ができ得るかどうかを検討することが不可欠である．本研究では徒歩が唯一の LSPTM の競合交通手段となるような目的地が想定されており，その OD 交通量が明確な場合，新規交通手段に対する需要は競合交通手段との分担率を算出することで予測が可能である．その分担率はランダム効用理論に基づく離散選択モデルを作成し，SP 調査でそのモデルのパラメータ推定をすることで算出されることが一般的である．しかし，この離散選択モデルからの分担率だけで需要を予測することは不十分なように思われる．なぜなら人はそれぞれ自らで意思決定を行うだけでなく，他の人との相互作用によって自らの行動選択を日々変化させ得るからである．これら人と人との相互作用に基づく需要の時間変化は静的なこのモデ

ルだけでは再現できない.

そこで本研究の目的は,複雑系からのアプローチとしてこの需要予測をとらえ,MAS 用いて長期的な LSPTM の需要の変化を観察し,安定してサービスを提供できるシステムの構築にむけた施策の提案を行うこととする.

1.2 複雑系と MAS

まず本研究のキーワードとなる複雑系と MAS の概要[1]をそれぞれ記述する.

(1) 複雑系

複雑系とは多数の因子または未知の因子が関係して系全体の振る舞いが決まるシステムにおいて,それぞれの因子が相互に影響を与えるために一般的な解析手法でシステムの将来の挙動を予想することが困難な系のことを言う.複雑系の厳密な定義は未だ明らかではないが,一般的には Casti の定義がよく用いられており,本研究においても Casti によって定義された複雑系を考えることとする.それによると複雑系とは(1)系を構成する主体の数は中程度であること,(2)個々の主体は自己の利益を追求する,利己的または合理的な存在であること,(3)主体は局所的な情報をもとに相互作用すること,の三点によって定義される.

(2) MAS (Multi Agent Simulation)

MAS とは,コンピューター内の多数の主体(エージェント)に一定のルールを同時に実行させ,その結果出現する現象を観察するためのシミュレーション技法である.複雑系のシミュレーションには従来システムダイナミクスの手法が用いられることが主であった.この手法は全体の結果は正確に部分の総和となる線形システムであり,「全体は部分の総和である.」という還元主義に基づくシミュレーション技法である.しかし現実社会においては還元主義では説明ができないような現象がしばしば発生する.その主な理由として次の二つの要因が挙げられる.

一つは因果関係の複雑さである.一見何の繋がりもないような要素がシステム全体の振る舞いを決定していたり,複数の原因が複数の結果を導くといった因果関係のネットワークが存在するからである.つまり,あらかじめ局所的に因果関係を切り出すことで全体像のモデル化を行うシステムダイナミクスの手法では因果関係のネットワークを完璧に記述することができず,因果関係の複雑さに起因する現象を記述することができない.

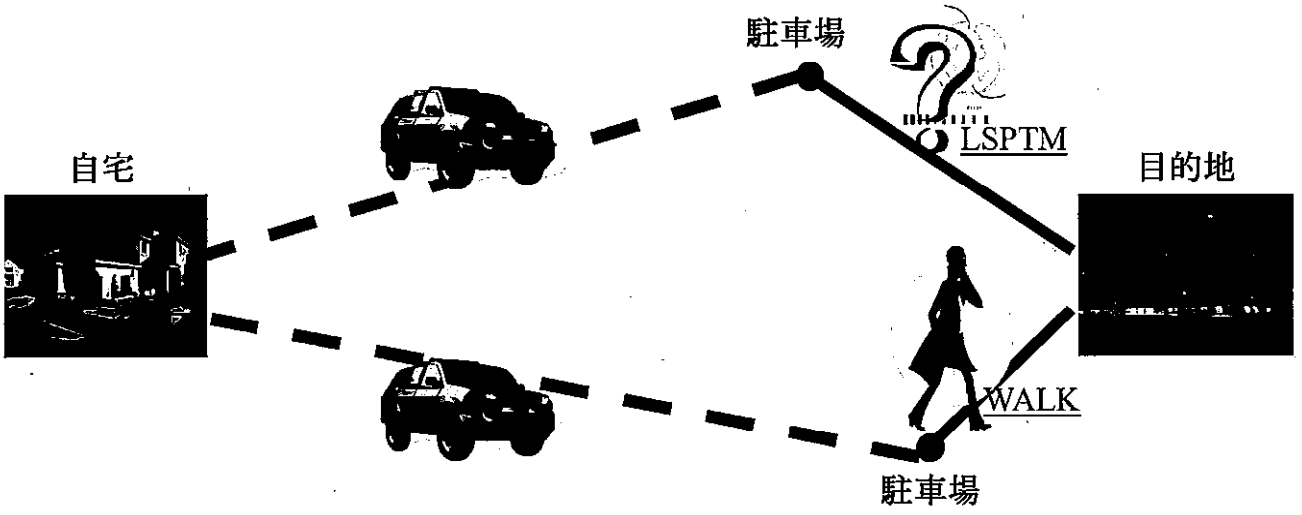
もう一つは要素間の相互作用の存在である.多数の要素がそれぞれ局所的に相互作用を繰り返し複雑に組織化することで,個別の要素の振る舞いかからは予測もできないような性質が全体として現れるからである.このような現象は創発と呼ばれるが,これは「全体が部分の総和以上である」現象であり,還元主義を基にもつシステムダイナミクスの手法では記述することができない.

以上の二点の要因より複雑系のシミュレーションには現象をボトムアップ的に

再現できるMASの技法が現在では主に用いられている．本研究でもLSPTMを導入することにより発生し得る現象があらかじめ想定されているわけではない．その利用形態，個人の振る舞いを想定することで現象をボトムアップ的に発生させ，その現象について分析を行うことが目的である．よってLSPTMの利用者となり得るエージェントの簡単な相互作用ルールを記述することで，そこから発生する複雑な現象をボトムアップ的に再現できるMASの手法を用いている．

1.3 本研究の前提仮定

LSPTMの利用形態に関する前提仮定を以下に記述する．まず自動車で自動車進入禁止区域周辺の駐車場までトリップが行われる．このLSPTMへの乗り換えが行われる駐車場は複数存在するが，今回のシミュレーションではそのうちの一箇所に着目している．そして，そこから目的地までのトリップにおいてエージェントのLSPTMか徒歩（WALK）かの行動選択が行われることとする．本研究においてはエージェントが認識する目的地は全エージェント共通で一箇所とし，駐車場から目的地までの回遊行動は考慮せず，エージェントの行動選択は駐車場から目的地までの片道のそれぞれの行動に対する効用をもとに行われる．以下の図(1.3.1)にその概念図を示す．



図(1.3.1) LSPTMの利用に関する前提仮定の概念図

1.4 本論文の構成

本研究では既存研究[1]のLSPTMとWALKの選択行動に関するロジットモデルを参考にし，そのロジットモデルにおけるLSPTMの効用に含まれる乗換え待ち時間に着目した．この乗り換え待ち時間をエージェントの相互作用の要素とし，これに対してエージェントがなんらかの学習を行いながら，自らの行動選択を変化させていくことを想

定した．これら一連の想定を学習ルールと呼ぶことにする．本論文ではこの学習ルールでのシミュレーション概要を2章で記述し，それに対するシミュレーションの結果，および考察を3章で,そして学習ルールの結果をうけての施策の提案を4章で行っている．

2.学習ルールのシミュレーション概要

各エージェントは想定された効用を用いて行動選択を行うが，その効用の説明変数の一つにLSPTMの乗り換え待ち時間がある．これは毎回の利用時にその値が異なる確率変数であるため，エージェント自らが経験した待ち時間から何らかの学習を行うことにより，その待ち時間，あるいはそれに相当するものがどの程度なのかを認識していき，行動決定を行う必要がある．本研究では学習のルールに機械工学の分野でよく用いられる強化学習の一種であるQ学習を採用した．そして待ち時間を通じて間接的に人と人との相互作用を記述することでシミュレーションを行っている．よって本章では離散選択モデル，Q学習，シミュレーションを行うにあたって想定したシミュレーションパターンについて説明を行う．

2.1 離散選択モデルについて

エージェントの行動決定過程の基礎となる離散選択モデルは大野によって作成されたロジットモデル[2)]を用いる．これはLSPTMか徒歩(WALK)かの選択に関する二項ロジットモデルであり，以下(2.1.1)の式で表される．

$$\begin{cases} P_1 = \dfrac{\exp V_1}{\exp V_1 + \exp V_2} \\ V_1 = B_1 * WT + B_2 * TT_1 + B_3 * CPF_1 + B_4 * F \\ V_2 = B_2 * TT_2 + B_3 * CPF_2 \end{cases} \qquad (2.1.1)$$

P_1 ：LSPTMを選択する確率
V_1 ：LSPTMの効用の確定項
V_2 ：WALKの効用の確定項
WT ：乗換待ち時間変数 (分)
TT_1 ：目的地までの移動時間変数 (分)
TT_2 ：目的地までの徒歩時間変数 (分)
CPF_1 ：LSPTM利用時の駐車料金変数 (円)
CPF_2 ：徒歩時の駐車料金変数 (円)
F ：LSPTM1回利用料金変数 (円)

B_1：乗換待ち時間変数に関するパラメータ
B_2：目的地までの移動(徒歩)時間変数に関するパラメータ

B_3：駐車料金変数に関するパラメータ

B_4：LSPTM1 回利用料金(年会費)変数に関するパラメータ

パラメータの値は同じく既存研究[1]より，

$$\begin{cases} B_1 = -0.09662919 \\ B_2 = -0.35531573 \\ B_3 = -0.0108468 \\ B_4 = -0.01722779 \end{cases} \quad (2.1.2)$$

となっている．
これにより全エージェントのそれぞれの行動に対する選好は共通である．
本研究においては目的地が一箇所に定められおり，駐車料金，LSPTM 一回使用料金に関しても変更されることはないと仮定するため，乗り換え待ち時間変数以外の説明変数は定数として扱う．その値は LSPTM の利用がある程度見込める目的地を想定し，

$$\begin{cases} TT_1 = 10 \\ TT_2 = 25 \\ CPF_1 = 400 \\ CPF_2 = 400 \\ F = 300 \end{cases} \quad (2.1.3)$$

とした．
よって(2.1.1)の効用関数は$V_1 = -0.0966WT + const.$，$V_2 = const.$となり，
$P_1 = 1/1 + \exp(0.0966WT + const.)$となるから，選択確率$P_1$は乗り換え待ち時間$WT$の関数となる．エージェントは確率変数であるこの乗り換え待ち時間に対して何らかの学習を行い，次の行動選択にその結果を反映させていく．

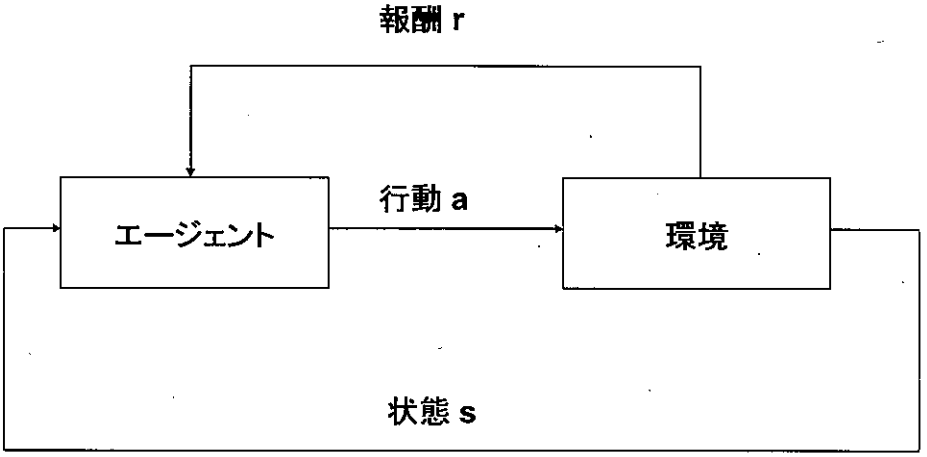
2.2 学習ルールついて

学習ルールには強化学習における Q 学習の理論[3]を用いて記述している．ここではその理論の中で使用されるキーワード別に説明し，それから Q 学習の学習過程を本研究における学習ルールとして概要を説明する．

(1) 強化学習

図(2.3.1)のように，動物，人間，プログラムなどを総称した学習エージェントが環境の状態sをモニターし，それに対して行動aを取り，その結果報酬rを受け取る一連

のプロセスをいい，報酬の期待値を最大化するような行動則$P(a|s)$を獲得することを強化学習と呼ぶ．この枠組みは図(2.2.1)の通りである．



図(2.2.1) 強化学習の枠組み

(2) 累積報酬

強化学習では自らの行動の結果得られる累積報酬を最大化ように学習を行うものを考える．累積報酬は以下の式(2.2.1)で示されるもので，χによって遠い将来に得られる報酬ほど割り引いて評価されている

$$R_t = \sum_{k=0}^{T} \chi^k r_{t+k+1} \quad (2.2.1)$$　T：最終時刻（通常は∞）　r:報酬　$\chi(0 \le \chi \le 1)$：割引率

本研究では$\chi = 0$として計算を行った．よって$R_t = r_{t+1}$となり，時刻tにおける累積報酬は時刻$t+1$で得られる報酬のみで決まる．

(2)マルコフ性

Q学習ではマルコフ性が仮定されており，マルコフ決定過程となっている．「マルコフ性をもつ」とはt+1の応答は一時刻前のtの状態と行動によって決まることを意味し，数式では以下の式(2.2.2)のように定義される．

$$\Pr\{s_{t+1} = s', r_{t+1} = r | s_t, a_t\} \quad (2.2.2)$$

よって，マルコフ決定過程での強化学習では，現在の状態と行動から次の時刻の状態と報酬が予測することができ，式では以下の(2.2.3)，(2.2.4)のように表される．

$$P^a_{ss'} = \Pr\{s_{t+1} = s' | s_t = s; a_t = a\} \quad (2.2.3)$$

$$R^a_{ss'} = E\{r_{t+1} | s_t = s, a_{t=a}, s_{t+1} = s'\} \quad (2.2.4)$$

$P^a_{ss'}$：遷移確率（状態sで行動aを行ったとき，次の状態s'に状態が遷移する確率）

$R^a_{ss'}$：現在の状態sと行動aが与えられたとき次の状態s'での報酬の期待値

(3) 行動価値関数

現在の状態，もしくは行動がどれくらい良いのかを計る関数を行動価値関数と呼ぶ．「どのくらい良いのか」は将来にわたって得られる報酬によって定義される．本研究においては$\chi = 0$であることより，時刻t状態sにおいて行動aをとることの行動価値関数は以下の式(2.2.5)のように定義される．

$$Q(s_t, a_t) = E\{R_t | s_t = s, a_t = a\} = E\{r_{t+1} | s_t = s, a_t = a\} \quad (2.2.5)$$

この行動価値関数を直接近似で求める方法としてQ学習がある．

(4) Q学習

繰り返し計算を行うことで$Q(s,a)$を近似的に求める方法としてQ学習がある．これは各状態における可能な行動の中で，最も行動価値関数の値が高い行動が最も行われやすくなるように学習を行う方法である．Q 値の更新プロセスは以下の式(2.2.6)のように表現される．

$$Q(s_t, a_t) \leftarrow (1-\alpha) Q(s_t, a_t) + \alpha \left[r_{t+1} + \chi \max_a Q(s_{t+1}, a_{t+1}) \right] \quad \alpha(0 \le \alpha \le 1)：学習率 \quad (2.2.6)$$

$\chi = 0$のとき，

$$Q(s_t, a_t) \leftarrow (1-\alpha) Q(s_t, a_t) + \alpha r_{t+1} \quad (2.2.7)$$

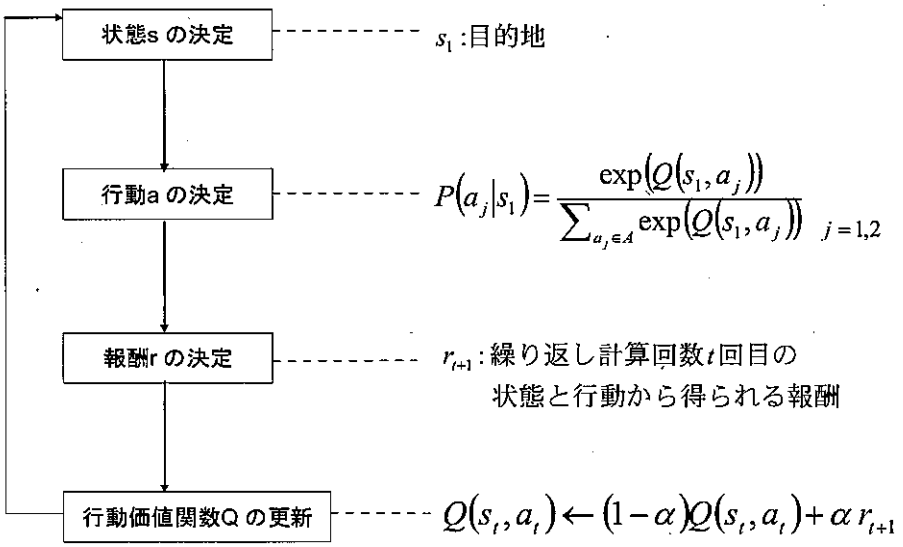
となる．

(4)学習率αとQ値収束性の関係

Q学習では$\sum_{t=0}^{\infty} \alpha \to \infty, \sum_{t=0}^{\infty} \alpha(t)^2 < \infty$　を満たすとき，全てのQ値は確率1でその期待値に収束することが数学的に証明されている．

(5)学習過程

Q学習の学習過程については以下の図(2.3.2)に示す．



図(2.2.2) Q 学習の学習過程

本研究において状態sは目的地を表し，その目的地も唯一であるため，

$s\in S \quad s=\{s_1|TT_1,TT_2\} \quad s_1$:目的地　　(2.2.8)

となる．なおsが複数存在する場合，sは一様な確率でランダムに選ばれる．
sが決まった後，エージェントは自らの行動aを決定する．
行動aとは

$a\in A \quad A=\{a_1,a_2\}\ a_j$：行動パターン

（$j=1$ ；LSPTM で行動　$j=2$ ；WALK で行動）　　(2.2.9)

で定義される．
行動の決定はランダムでも繰り返し計算回数が大きければQ値が収束するので問題はないが，その効率を上げるためにそれぞれの行動に対する行動価値関数を基にして行動の決定が行われる．その決定方法として小さな確率εでランダムに選択を行い，それ以外ではQ値が最大の行動を選択するε-greedy法とボルツマン分布を利用したsoftmax法とのどちらかが一般的に用いられる．本研究ではsoftmax法を用いる．
Softmax 法は任意の状態sにおける任意の行動aの選択確率$P(s,a)$が，

$$P(s,a)=\frac{\exp(Q(s,a)/T)}{\sum_a \exp(Q(s,a)/T)} \qquad T:\text{パラメータ} \quad (2.2.10)$$

と表され，本研究では$T=1$としてシミュレーションを行った．
報酬rに関しては様々な与え方が考えられる．この与え方については次の 2.3 で詳しく説明するが，本研究では機械工学の分野で一般的に行われる行動に関する良し悪しの学習としてのrの与え方と，式(2.1.1)のロジットモデルを行動決定に反映させられるような

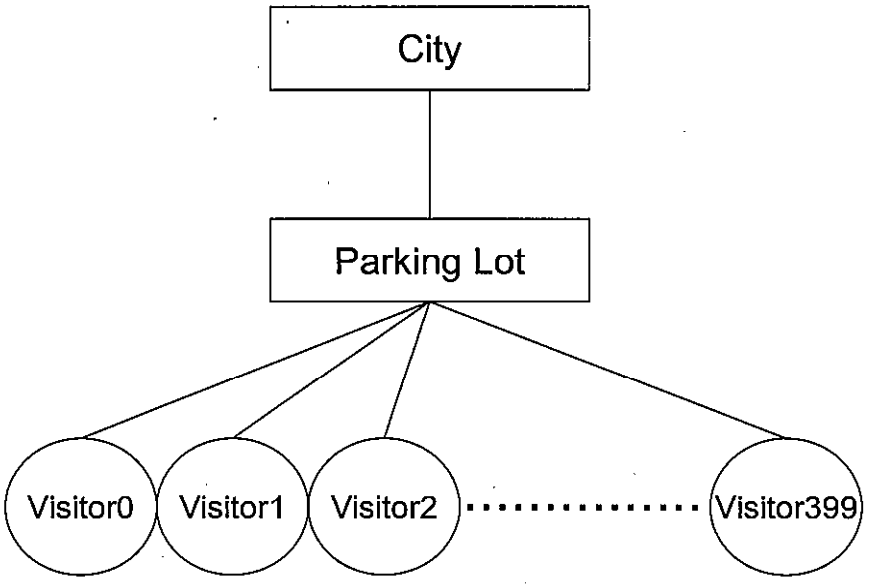
報酬rの与え方で行う二通りの設定を考えた.

2.3 シミュレーションの概要

学習ルールに Q 学習を用いるとしたが，報酬の与え方によって学習内容にバリエーションをもたせることが可能である．本研究では二種類の学習パターンを作成した．加えて待ち時間の設定に関しても様々な想定が可能で，ここでは相互作用の有無を表現するための二種類の待ち時間の設定パターンを作成した．これらそれぞれの組み合わせで四種類のシミュレーションパターンができ，それぞれについてシミュレーションを行うことで，シミュレーション後に結果を比較考察できるようにした．なお本研究では MAS のソフトウェアとして株式会社構造計画研究所の artisoc 1.0 を使用している.

(1)シミュレーションの構造

本研究でのシミュレーションは三段階の階層構造によって構成されている．この階層構造は以下図の(2.3.1)に示す通りである.



図(2.3.1) シミュレーションフィールドの階層構造

最上位の階層に City が存在し，これは LSPTM の利用形態の想定で述べた車の乗り入れができない都市部を指す．そしてその下に Parking Lot という階層を作成した．これは自動車進入禁止区域周辺で LSPTM の貸し出し所がある駐車場を指す．そして Visitor というエージェント 400 名が Parking Lot の階層の下に属しており，これら同一の OD となる Visitor が乗り換え待ち時間を通じて相互作用しながら LSPTM か WALK かの行動選択をステップごとに行っていく．これにより Visitor 個々の選択行動というミクロな動きを Parking Lot の階層からマクロな視点で観察することが可能となる.

(2)エージェントの学習率α

Q学習においては学習率$\alpha(0 \le \alpha \le 1)$がエージェントの特性を表す．式(2.2.7)からわかるようにαが1に近いエージェントほど行動に対して得られた報酬に次の行動決定が依存しやすく，逆にαが0に近いエージェントほど行動価値関数に次の行動決定が依存する．本研究では400名のエージェントそれぞれに対して0から1までの一様乱数を発生させ，それを学習率αとした．

(3)シミュレーションパターン

報酬rの設定，待ち時間の設定をそれぞれ二種類ずつ設定し，その組み合わせとしてシミュレーションパターンを四通り作成した．このそれぞれについてシミュレーションを行うことで学習内容，相互作用の有無の違いによる結果の違いが生じるかどうかが観察できる．それぞれの設定内容については以下（ｉ），（ii）で記述する．

（ｉ）報酬rの設定

①Learning_Q

行動の良し悪しを学習することによりLSPTM, あるいはWALKを選択することがどれくらいよいのかについて学習を行うことを想定したパターンである．報酬としてエージェントが時刻tでの行動が良かったと感じれば$r_{t+1}=1$，悪かったと感じれば$r_{t+1}=-1$が与えられる．報酬の決定基準，つまりその行動の良し悪しの基準は，それぞれの選択に対する効用を計算し，式(2.1.1)に基づいて確率的に計算される．具体的には，エージェントが時刻tでLSPTMを選択した場合（$a_t=a_1$），その利用によって認識した待ち時間から計算されたV_1とV_2から，0から1までの一様乱数lを利用して，

$$\begin{cases} if \quad l \le P_1 \quad then \quad r_{t+1}=1 \\ else \quad r_{t+1}=-1 \end{cases} \qquad (2.3.1)$$

エージェントが時刻tでWALKを選択した場合，時刻tまでで最後にLSPTMを利用した際に認識した待ち時間から計算されたV_1とV_2から，

$$\begin{cases} if \quad l \le (1-P_1) \quad then \quad r_{t+1}=1 \\ else \quad r_{t+1}=-1 \end{cases} \qquad (2.3.2)$$

となる．

②Learning_Q改

エージェントがLSPTM利用時に発生する待ち時間を学習し，利用時にどの程度の待ち時間が発生するのかを予測することを想定した学習パターンである．

報酬としてエージェントが時刻tでLSPTMを選択した場合，その使用によって認識した待ち時間から計算された効用V_1が報酬として与えられ，WALKを選択した場合，そのま

ま WALK の効用 V_2 が与えられる．具体的には

$$r_{t+1} = V_1^{t=t} \qquad (2.3.3)$$

WALK 選択時には

$$r_{t+1} = V_2^{t=t} \qquad (2.3.4)$$

となる．
以上のことより LSPTM 選択時は Q 値の更新が行われるが，WALK 選択時はその値に変化がなく，どの時刻においても一定である．これにより図(2.2.2)の行動選択の構造と式(2.1.1)のロジットモデルが等価となるので，式(2.1.1)のロジットモデルで行動選択が行われることになる．

(ii) 待ち時間の設定

ここで定義される待ち時間はLSPTMの貸し出しの際に生じる手続き上の処理の時間は考慮せず，貸し出し所の保有する LSPTM が全台貸し出し状態になり，LSPTM が返却されるまで次の利用者が待つことを想定した待ち時間である．

①WT_Random

待ち時間を 0(min)から 15(min)までの整数として，一様な確率でランダムに与える．

②WT_Ruled

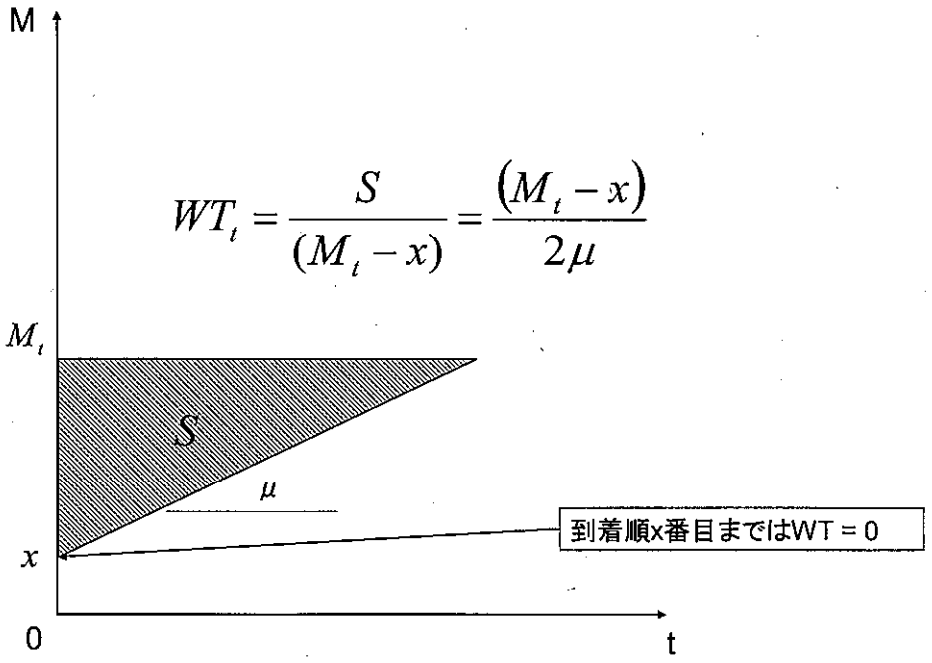
待ち時間をある規則をもって与える．この規則には確定的な待ち時間の設定方法を採用した．LSPTM が貸し出される場所においては，貸し出し所の LSPTM の保有台数 x 台分の利用者は待ち時間が発生することなく LSPTM が利用可能である．つまり LSPTM を選択したエージェントのうち，選択順 x 番目のエージェントまでは $WT = 0$ が与えられる．選択順 x 番目を超えた LSPTM 選択者は待ち行列を為し，その処理は一人あたりの平均 LSPTM 利用時間の逆数の速さでもって処理されていく．この時の一人あたりの平均処理速度を μ(人/min)とすると，
x 番目以降のエージェントには，

$$WT_t = M_t / 2\mu$$

WT_t:繰り返し計算回数 t 回目における一人当たりの平均待ち時間

M_t:繰り返し計算回数 t 回目における累積待ち人数 (2.3.5)

が与えられる．よって x 番目以降のエージェントには全員同じ待ち時間が与えられることになり実際の状況とは少し異なるが，少なくとも $V_1 < V_2$ を成立させるに十分な待ち時間が確実に与えられる．以下の図(2.3.2)がその概念図である．図を見てわかるように WT_t は図の斜線部の三角形の面積 S を x 番目以降の LSPTM 利用者人数で徐したものである．



図(2.3.2) WT_Ruled における待ち時間決定方法の概念図

以上(ⅰ),(ⅱ)それぞれの設定の組み合わせとして四通りのシミュレーションパターンができあがる．これらを Case1，Case2，Case3，Case4 と呼ぶこととし，その組み合わせ内容は以下の表(2.3.4)のようになる．

表(2.3.2) Case 分けの概要

	Q-Learning	Q-Learning 改
WT(Random)	Case1	Case2
WT(Ruled)	Case3	Case4

3.学習ルールでのシミュレーション結果および考察

考察の内容は以下の 4 点になっている．

(1) シミュレーションパターン別の α による Q 値収束性

Case1 から Case4 の全てのシミュレーションパターンに対して 5000 回の繰り返し計算を行い，設定した全ての変数値に対して出力を行った．出力は全エージェントのものが行えればよいが，ソフトウェアの都合上困難であるため，代表的な α 値をもつエージェントに対してのみ出力を行った．代表的な α 値をもつエージェントには，エージ

エント 400 名の中でα値が最大であったもの，400 名のαの算術平均値に最も近かったもの，400 名の中でα値が最小だったものを選んだ．そしてそのそれぞれのα値をもつエージェントをα_max，α_average，α_min と表すことにする．これら 3 名のエージェントそれぞれの全ての変数値を出力し，それぞれのシミュレーションパターンにおけるαの違いによる Q 値収束性の特徴に対して考察を行うことで学習の定義を明確にした．その結果，および考察内容は 3.1 に記述する．

(2) シミュレーションパターン間の Q 値収束性に関する比較

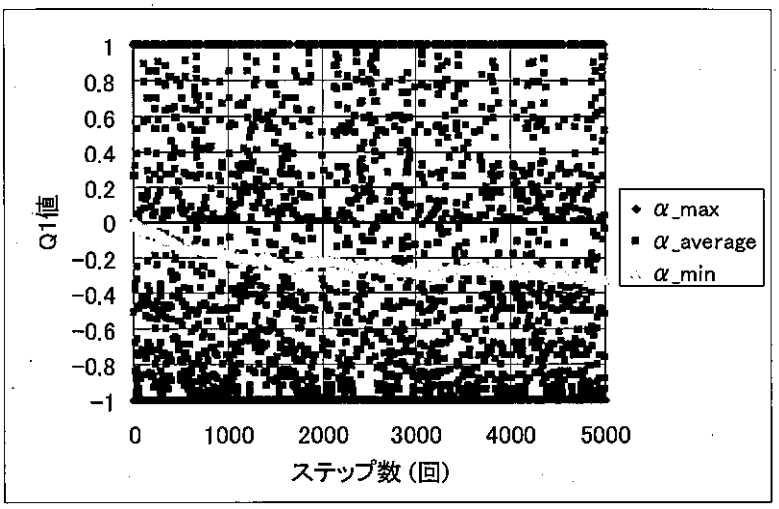
(1)で行ったシミュレーションにおいてQ値の収束が確認されたα_min のエージェントのQ値収束性やその収束値に着目してシミュレーションパターン間で比較を行った．そしてそこで見てとれた特徴的な差異について，その差異を生じさせた要因への考察を行った．これにより Learning_Q と Learning_Q 改の学習の違いを明らかにし，WT_Random と WT_Ruled の待ち時間設定を比較することでわかる相互作用の有無がシミュレーション結果に影響を与えるのかに対して考察を行った．その結果，および考察内容は 3.2 に記述する．

(3) Day to Day の需要変化の特徴

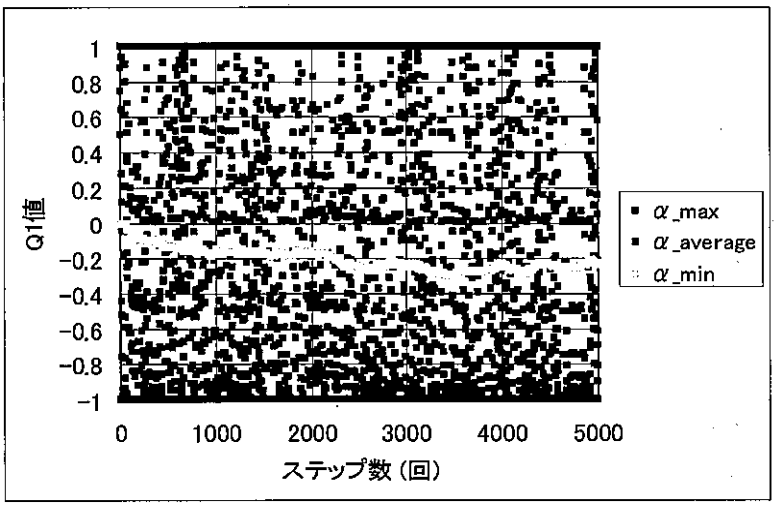
Case4 を用いて 365 回の繰り返し計算を行い，繰り返し計算一回が一日に対応する 1 年間の day to day の需要変化を想定した．このシミュレーション結果について，LSPTM 需要の時間変化を図示することで見受けられた特徴が学習に起因するものであるのか検証を行った．その結果，および考察内容は.3.3 で記述する．

3.1 シミュレーションパターン別のαによる Q 値収束性

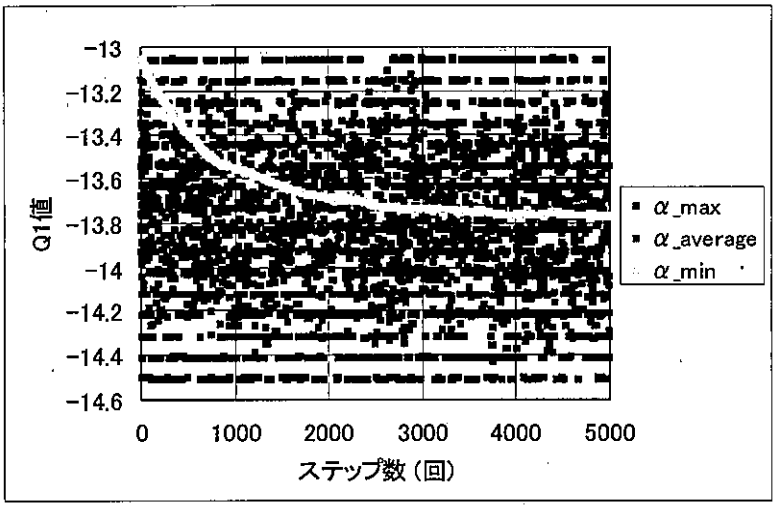
2 章で定義した Case1 から Case4 までの 4 ケースそれぞれについて，代表的なαの値をもつエージェントのステップ毎におけるQ_1値を縦軸に，計算ステップ回数を横軸にとった図を以下の図(3.1.1)から図(3.1.4)で示す．なお簡単のため$Q(s_1,a_1)$をQ_1と表記する．$Q(s_1,a_2)$に関しても同様の分析が可能であるため，ここではQ_1のみを分析の対象とする．



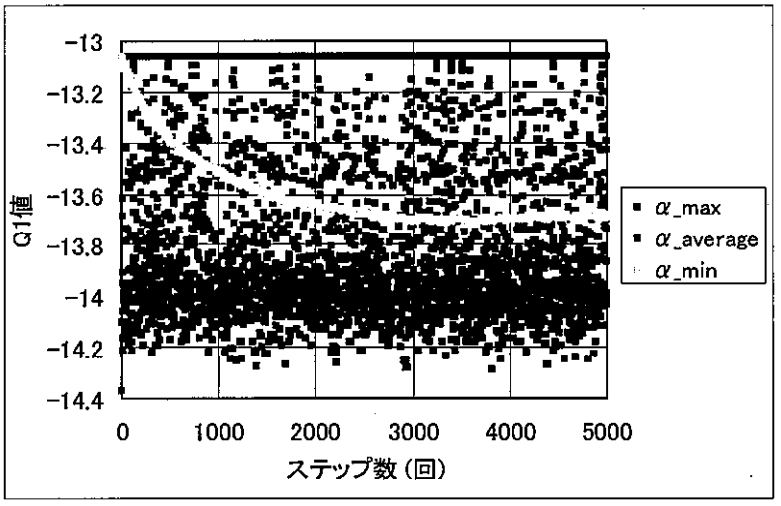
図(3.1.1) Case1 における Q 値収束性



図(3.1.2) Case2 における Q 値収束性



図(3.1.3) Case3 における Q 値収束性



図(3.1.4) Case4 における Q 値収束性

これら代表的な α 値それぞれの学習の特徴であるが，α_max はある時刻において LSPTM 利用時に認識した待ち時間をその次の時刻での行動選択に直接反映させるエージェントであり，α_average はある時刻以前から認識している待ち時間とある時刻においての LSPTM 利用で発生した待ち時間の相加平均を次の時刻での待ち時間として学習を行うエージェント，α_min が学習の結果，待ち時間をある一つの値として予測するようになるエージェントであるということが言える．

図から読み取れる Q 値収束性に関する特徴は，どのケースにおいても α_max のエージェントは Q_1 の値に報酬 r が取り得る値のみを取り続け，α_average のエージェントは報酬 r の取り得る範囲内の値を離散的にとり，α_min のエージェントのみが Q_1 の値をある一定の値に収束させることである．これは 2 章で述べた Q 学習理論における学習率 α と Q 値収束性の関係に起因することである．Q 値が収束し安定することは，そのエ

ージェントがLSPTMで行動するかWALKで行動するかの行動選択確率を安定させることを意味する．よってシミュレーションの結果より，行動選択確率を安定させるエージェントと，時々刻々と行動選択確率を変化させるエージェントが存在することがわかる．ここで本研究における学習の定義を明確にするため，待ち時間に対する学習をエージェントに行わせている Case4 において，一定の値にQ_1を収束させたα_min に着目する．そして学習結果エージェントが認識した待ち時間と，エージェントが得る待ち時間の期待値とを比較することで学習の効果を計る．待ち時間の期待値は，エージェントα_minがLSPTM利用時に得た待ち時間を利用して，α_minのエージェントがシミュレーションにおいてLSPTMを利用した回数をK，α_minの得た待ち時間の算術平均$\overline{WT}$とすると，

$$\overline{WT}=\frac{1}{K}\sum_{k=1}^{K}WT_k \qquad (3.1.1)$$

WT_k：α_minがLSPTMを選択した繰り返し計算回数k回目において得られた待ち時間

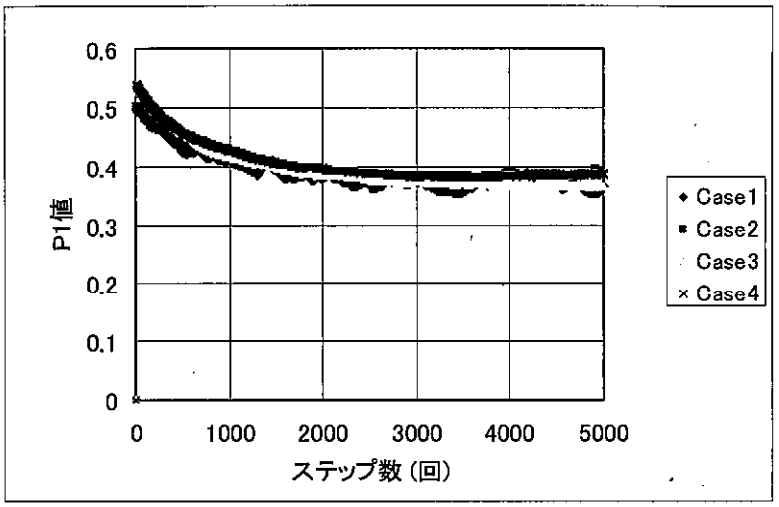
と推定できる．これに対して，あるエージェントが学習を行い，最終繰り返し計算回数T回でのシミュレーションの後，エージェントが予測するようになった待ち時間$WT^{t=T}$は，

$$WT^{t=T}=\left(Q^{t=T}(s_1,a_1)-D\right)/B_1 \quad D=B_2*TT_1+B_3*CPF_1+B_4*F \qquad (3.1.2)$$

で計算される．計算の結果，$\overline{WT}=6.53$，$WT^{t=T}=6.51$となった．よってエージェントが学習を行うことで待ち時間の期待値を近似的に予測していることがわかる．
つまり本研究における学習とはエージェントが報酬の期待値を直接近似により求めていくことであると定義できる．

3.2 シミュレーションパターン別の比較

次にQ値の収束が確認されたエージェントα_minのQ_1値の毎ステップの変化についてケース間比較を行いたい．しかし学習パターン Learning_Q と Learning_Q 改は報酬rのスケールが異なるため直接比較できない．よって，Q値から計算されるLSPTM選択確率P_1の値をもって比較している．その結果は図(3.2.1)のようになった．



図(3.2.1) α_min の P_1 値の Case 間比較

このケース間比較において P_1 の収束性，およびその収束値に着目すると，次の二点の特徴が見受けられた．一つは収束した P_1 の値にはケース間でそれほど大きな差はみられなかったが，Case1,2 と Case3,4 とをそれぞれ比較したとき，Case1，Case3 が Case2，Case4 よりもそれぞれその値が小さいということ，もう一つは，Case1,2 よりも Case3,4 の方が P_1 の値の収束性がよく，Case1,2 には値の若干の振動が見られるという二点である．
この二点の特徴に関して一点目は待ち時間設定の違いに起因するもの，二点目は報酬設定の違いに起因するものであると考えられ，それを以下に詳しく説明する．なおこれら要因の説明にはそれぞれのシミュレーションパターンにおける Q_1 値に焦点をあてて説明を行うこととする．

(1) 待ち時間設定の違い

Case1,2 と Case3,4 とをそれぞれ比較したとき，Case1，Case3 が Case2，Case4 よりもそれぞれ P_1 値が小さくなるということは，Q_1 を定義する報酬 r の期待値が関係している．つまり Case1，Case3 が Case2，Case4 よりもそれぞれその期待値が大きくなるということである．これらは待ち時間設定の違いに起因する．

まず Case1 と Case2 の比較であるが，r は 1 か-1 の二値で与えられることより，$r=1$ となる頻度でもってこのことを説明する．$r=1$ となるためには，報酬の決定の基準より，各行動に対する効用 U_1，U_2 が $U_1 > U_2$ を満たさなければならない．効用の誤差項を除くと，$V_1 > V_2$ を満たす必要がある．$V_1 > V_2$ となるための乗り換え待ち時間 WT の条件は，$WT < 1.67$ であり，このとき $r=1$ が与えられる．WT_Random は WT が 0 から 15 の整数で与えられるため，$WT < 1.67$ を満たし $r=1$ となる確率は，$\frac{2}{16} = \frac{1}{8}$ である．一方 WT_Ruled の場合，$WT = 0$ となる 50 名のエージェント以外のエージェントは，ほぼ確実に $WT \geq 1.67$ となる．よって $r=1$ となる確率は全エージェント数が 400 名であることより，

少なくとも$\frac{50}{400}=\frac{1}{8}$より大きい．つまり全エージェントがLSPTMを使用したとしてもあるエージェントの報酬が$r=1$になる確率はWT_Randomの時のそれと同じであり，多くの場合でWT_Ruledの方が$r=1$が与えられる確率が高い．これらのことより，$r=1$となる頻度はWT_Randomの方がWT_Ruledより少なく，WT_Randomの方が報酬の期待値が小さいため，P_1が小さくなる．

次に，Case3とCase4の比較であるが，これは待ち時間の期待値でもって説明できる．待ち時間WTはどちらの設定においても確率変数とみなすことができ，WT_Randomの待ち時間の期待値を$E[WT_{Random}]$，WT_Ruledのそれを$E[WT_{Ruled}]$とする．

$E[WT_{Random}]$はWT_{Random}が0から15の整数で一様な確率で与えられるので，

$$E[WT_{Random}]=\frac{(0+1+2+\ldots+15)}{16}=7.5 \qquad (3.2.1)$$

となる．

$E[WT_{Ruled}]$は本章3.1での不偏推定量$\overline{WT}$がそのまま適用できるため，

$$E[WT_{Ruled}]=6.53 \qquad (3.2.2)$$

となる．

よって待ち時間の期待値はWT_Randomの方がWT_Ruledよりも大きく，報酬の期待値が小さいため，P_1値が小さくなる．

以上のことより，待ち時間設定の違いによってこの特徴は生じているものと考えられる．つまりシミュレーション前に想定していた待ち時間設定での相互作用の有無に起因した特徴ではないため，今回のシミュレーションでは相互作用が選択行動に影響を与えているような現象は見受けられなかった．

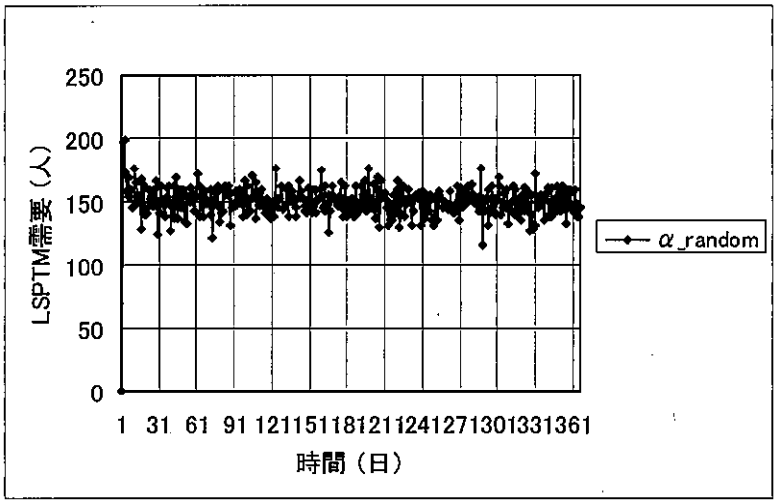
(2)報酬設定の違い

Case1,2とCase3,4の収束性の違いはLearning_QとLearning_Q改の報酬の与え方の違いが原因で生じているものと考えられる．その違いについては二点が挙げられる．まず報酬の取りうる値のバリエーションである．Learning_Qはrに1か-1かの二値しかとらない．一方Learing_Q改は，それぞれの時刻において得られたWTを反映したV_1がrとなる．よってLearning_QではLearning_Q改のようなr値のバリエーションが無い．もう一つはステップ間でのrの差Δrである．Learning_Qではrは1か-1の二値であり，その差の絶対値$|\Delta r|$は$|\Delta r|=2$である．一方Learning_Q改では，想定され得る最大の報酬の差として時刻tで$WT=0$だったものが時刻$t+1$で$WT=15$になった場合を考えても，$|\Delta r|=|V_1^{t+1}-V_1^t|\cong 1.45$とLearning_Qほど大きくない．これら二点の要因が主に合わさって，Q値の収束性の違いが発生していると考えられる．

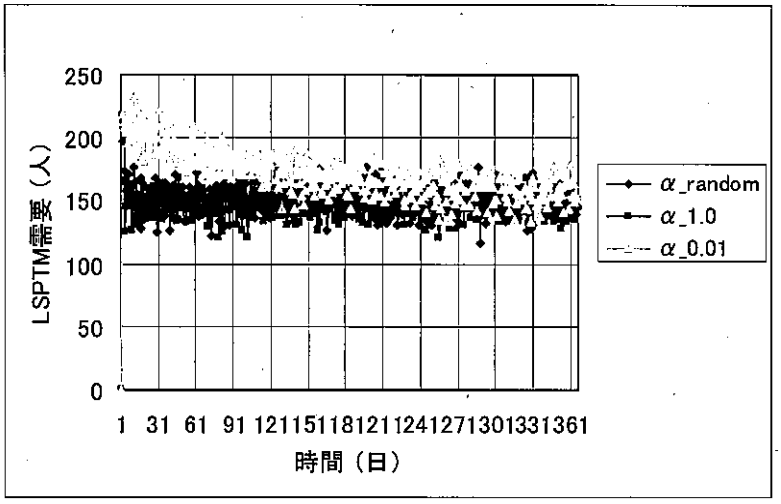
結果として，報酬の設定の違いによるシミュレーション結果の差はその振動性を除けば大きな差は見受けらなかったが，この二つの学習パターンは想定されていることが異なる．まずLearning_Qは行動の良し悪しを表現するポイントからエージェントの行動決定が行われ，そのポイントの決定基準として潜在的にロジットモデルが存在する．つまりLearning_Qにおいてはそれぞれの行動に対する効用が行動決定後に決まるため，エージェントが必ずしも効用の高かった方の行動を選択しているわけではない．一方Learning_Q改はロジットモデルでエージェントの行動決定が行われる．つまりそれぞれの行動に対する効用は行動決定前に決まり，エージェントは常に効用の高かった方の行動を選択している．よってこれら二つの学習パターンはエージェントの行動決定方法，およびロジットモデルの使用方法の点で異なる．

3.3 Day to Dayの需要変化に関する考察

シミュレーションの結果に関して，縦軸にLSPTM利用者数，横軸に時間をとった図を以下図(3.3.1)に示す．この図を読み取ると，シミュレーション開始から30日以内で需要が急激に下がり，その後ある幅をもって安定するという特徴が見受けられる．この現象を（1）シミュレーション30日以内における需要の低下，（2）需要安定期における需要の振動幅，の二点にわけ，これらがエージェントの学習に起因する特徴であるかを検証する．検証には学習を行わなかった場合を想定した全エージェントの学習率をα=1として同様のシミュレーションを行ったもの(凡例α_1.0)，365日以内にQ値を収束させられる学習を行うことを想定した全エージェントの学習率をα=0.01として同様のシミュレーションを行ったもの（凡例α_0.01）と今回のシミュレーション結果(凡例α_random)とを比較することで行う．比較を行った図は以下図(3.3.2)に示す．



図(3.3.1) α_randomにおけるday to dayの需要の時間変化



図(3.3.2) 学習率による day to day の需要の時間変化比較

(1) シミュレーション初期における需要の低下に関する分析

この現象はシミュレーション初期の段階で待ち時間を 0 と予測して利用を開始したエージェントが待ち時間を認識するようになり，それが行動選択に影響を及ぼすことで需要が下がったものと考えられる．全エージェントが学習を行わない α_1.0 は，エージェントが待ち時間を認識するとすぐに次の行動選択にその待ち時間を考慮するため，LSPTM の選択確率が急激に低下する．一方 α_0.01 は全エージェントが待ち時間を認識しても学習率 α を通して徐々に行動選択に反映させていくため急激な需要の低下は見られない．よってこの現象は図(3.3.2)を見てわかるように，エージェントの学習率の分布によりその需要の低下の勾配に差が生じている．

(2)需要安定期における需要の振動幅に関する分析

図(3.3.2)より，需要の振動はどのシミュレーションにおいても概ね同じ幅で存在する．この振動幅はエージェントの行動が確率的に決定されることに起因するが，ここで特筆すべきことはシミュレーションパターン間に学習に関する差があるにも関わらず，その振動幅にそれほど変化がないことである．エージェントには α_0.01 におけるエージェントのようにある一定の LSPTM 選択確率を獲得するものと，α_1.0 におけるエージェントのように学習を行わず時々刻々とその選択確率を変化させるものとの二種類が大きく分けて存在する．α_random においてはこれら両方のエージェントが同時に存在するなかでシミュレーションを行っている．しかしこれらシミュレーションパターン間での学習に関する違いがあるにもかかわらず，個々のエージェント(Visitor)が属するもう一つ上の階層(Parking Lot)から需要の時間変化を観察したときに，シミュレーションパターン間でその振動幅に大きな差がない．よってこのことは下の階層の構造からは予測できなかった現象がその上の階層で生じる創発であると言える．

(1),(2)の分析結果を受けて，シミュレーション 30 日以内における需要の低下は

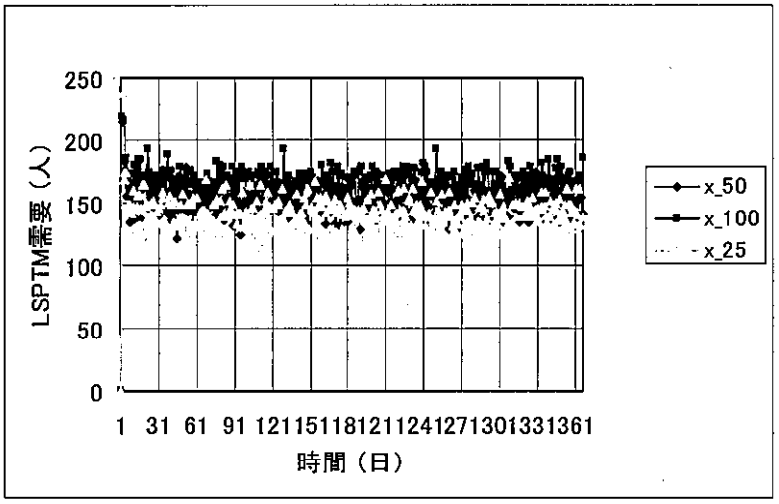
エージェントの学習率の分布により低下の勾配に差が生じるが，需要安定期における需要の振動幅は学習とは無関係に存在することがわかる．

4. 施策の提案

エージェントの行動選択過程，待ち時間の設定方法の構造を考えたとき，人為的に操作が可能な変数は待ち時間の設定における LSPTM の台数 x である．また x と単位時間あたりの処理人数 μ には相関があり，少なくとも x が増加すれば μ も増加するような関係であると考えられる．よってここでは(1) x による需要への影響，(2) μ による需要への影響，と二点に独立してシミュレーションを行い，それぞれを分析することでシステムの安定供給に向けた施策の提案を行う．なお安定供給にむけた施策をここでは LSPTM 需要の拡大につながるような施策と位置付ける．

(1) x による需要への影響

本研究ではこれまで $x=50$ としてシミュレーションを行ってきたが，ここではこれに加えてその 2 倍の $x=100$ と，$\frac{1}{2}$ 倍の $x=25$ についても同様のシミュレーションを行い，Day to Day の需要の変化について比較を行う．それぞれのシミュレーションパターンを x_50，x_100，x_25 と表すこととし，その結果は以下の図(4.1)のようになった．

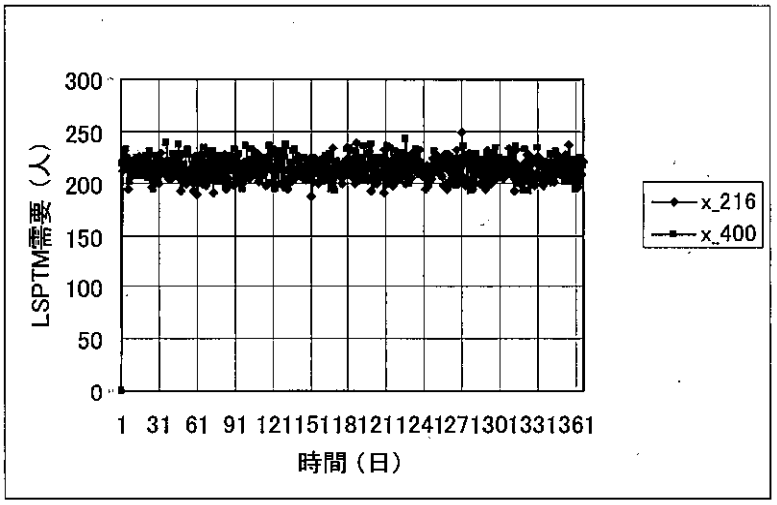


図(4.1) x による day to day の需要の時間変化比較

図(4.1)より，x を増減させることで LSPTM 需要の上下方向へのシフトが見られる．よって x を増加させることは LSPTM 需要の拡大に繋がる有効な施策である．しかしこの施

策は理論上，考えうる最大の LSPTM 選択確率で全エージェントが行動を行ったときの LSPTM 需要の期待値を超えて LSPTM 台数を増やすことは有効でないと考えられる．そのことを検証するため，最大 LSPTM 選択確率である $WT=0$ のときの $P_1 \cong 0.54$ で全エージェントが行動することを考える．このときの需要を n（人），その期待値を $E[n]$，n の最大値を $n_{\max}$ として，$x=E[n]$ でのシミュレーション結果と，$x=n_{\max}$ でのシミュレーションを比較することで施策の有用性の限界を検証する．確率変数である n は全エージェント数を N（人）とすると，二項分布に従うので $E[n]=NP_1 \cong 216$，$n_{\max}=400$ である．

よって $x=E[n]$ でのシミュレーションを x_216，$x=n_{\max}$ でのシミュレーションを x_400 と表し，その結果を以下の図(4.2)にまとめた．

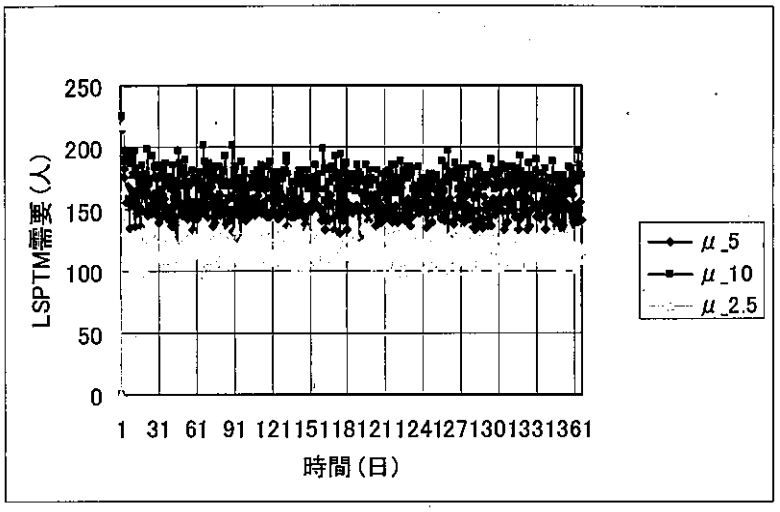


図(4.2) 施策が有効となる x の限界値の検証

図(4.2)より，x_216 と x_400 の Day to Day の需要はほぼ同じであることがわかる．よって $x=E[n]$ を超える LSPTM 台数の導入には需要拡大の効果はない．

(2) μ による需要への影響

本研究におけるこれまでのシミュレーションでは，WT が概ね 0 から 15 の間の値を取ることを想定していたため，μ=5 でシミュレーションを行ってきた．ここではその 2 倍である μ=10 と，その $\frac{1}{2}$ 倍である μ=2.5 で同様のシミュレーションを行い，Day to Day の需要変化について分析する．それぞれの μ の値でのシミュレーションを μ_5，μ_10，μ_2.5 と表すと，その結果は以下図(4.3)のようになった．



図(4.3) μによる day to day の需要の時間変化比較

図(4.3)より，μの値の増減によって LSPTM 需要の上下シフトが見られる．よって単位時間あたりの処理人数が大きくなると LSPTM 需要の拡大に繋がる．

(1)(2)の分析結果をうけて LSPTM の導入台数を(1)で検証した限界値までの範囲内で増やすことは LSPTM 需要の拡大させることがわかった．加えてxとμの相関関係より，xを増加させることでμも増加し，(2)で示した需要の拡大も同時に実現される．よってLSPTMの導入台数を増やすことはLSPTM需要の拡大にむけた有効な施策であるといえる．

5 結論

本研究では主に Case4 のシミュレーションパターンを用いることで，乗り換え待ち時間で相互作用を記述し，その待ち時間についてエージェントが Q 学習に基づいた学習を行うエーシミュレーションを行った．そしてそのシミュレーション結果について Q 値の収束性や LSPTM の需要の時間変化に関して分析を行い，待ち時間に着目した際の需要の拡大にむけた施策の提案を行った．

5.1 需要予測に MAS を使用することの意義

本研究では説明変数をあるルールで記述し，ロジットモデルを人間の行動則として使用した．そしてこの説明変数のルールに乗り換え待ち時間が確率変数となることを考たため，各エージェントの待ち時間の予測方法が重要となった．この予測方法を動的かつ非集計的な視点で捉え，シミュレーションを行ったのが Case4 のシミュレーションである．Case4 では各エージェントの待ち時間の予測方法の違いが学習率αを通して動的に表現できる．例えばα_max のエージェントは LSPTM 利用時に経験した待ち時間

を次の行動選択時に直接反映させて予測し，α_min のエージェントは，経験した待ち時間を少しずつ次の行動に反映させ，最終的にはある一定値（待ち時間の期待値）として待ち時間を予測するようになるというような違いである．今回行ったシミュレーションでは，乗り換え待ち時間によるエージェント間相互作用がエージェントの行動選択にそれほど影響を及ぼさなかったため day-to-day の需要が非常に安定し，MAS を用いて動的に需要を予測することの意義があまり示されなかった．しかしシミュレーション開始 30 日以内での需要の低下など待ち時間の学習過程で生じた需要の変動が観察できたことに少なくとも MAS を使用する意義があったと考えられ，今後エージェントの行動選択に直接影響を及ぼすような相互作用を考慮すれば，よりその意義を高められると考える．

5.2 今後の課題と展望

今回のシミュレーションで何点かの課題が明確となった．複雑系の視点から MAS を捉えると，エージェントは 1 章 1.2 で記述した Casti の定義：(1)系を構成する主体の数は中程度であること，(2)個々の主体は自己の利益を追求する，利己的または合理的な存在であること，(3)主体は局所的な情報をもとに相互作用すること，の三点を満たす必要がある．これを踏まえると(2)と(3)に関して今回のシミュレーションでは不十分な点があった．

まず(2)に関して，予測方法を戦略としてとらえたとき今回のシミュレーションでは各エージェントが一つの戦略しかもたないことである．例えば α_min のエージェントであるが，最終的に待ち時間の期待値を近似的に予測するようにはなるが，その収束途中に関しては大きい待ち時間を経験しているにもかかわらず，次の行動にはあまり反映させないため利己的または合理的な行動をとっているとは言いがたい．よって各エージェントに学習率を複数もたせこれを戦略とし，そして戦略の決定過程に進化型計算における遺伝的アルゴリズムなどの方法論を用いることでより合理的かつ利己的なエージェントを作成し，エージェント個々の特性を出すことが必要であると考える．

次に(3)に関してだが，今回のシミュレーションではエージェント間の相互作用として乗り換え待ち時間を考えたが，これはあくまでロジットモデルにおける説明変数が相互作用するだけであり，エージェントの行動が最終的にはロジットモデルで記述されるので相互作用によるシミュレーション結果への影響が小さいことである．実際 3 章 3.2 で記述したように今回のシミュレーションでは相互作用による創発は起こっておらず，相互作用のない WT_Random の待ち時間設定のシミュレーション結果と比較してもそれほど需要の変動に差異が見受けられなかった．よってエージェントの行動選択に直接的に影響を及ぼす相互作用を記述することで複雑系からのアプローチとしての意義を高める必要がある．現段階ではその相互作用として「流行」というものを想定しており，流行を統計物理学の強磁性モデルである二次元イジングモデルを参考に記述し，それについて分析を行っている．二次元イジングモデルの概要，およびそのシミュレーション方法であるメトロポリス法の概要は付録を参考にされたい．

これらの課題をうけて，今後の研究の展望としてはまず流行ルールにおけるシミュレーションの分析を行い，このルールでのシステムの安定供給に向けた施策を検討

したい．そして戦略を考慮した学習ルールと流行ルールを組み合わせるようなシミュレーションモデルを構築し，独立してシミュレーションを行ったものとを比較しながら最終的な施策の提案を行いたい．

参考文献

1) 株式会社　構造計画研究所；http://mas.kke.co.jp/
2) 大野　稔起；ロジットモデルによる新低速度個別移動手段の需要分析，神戸大学大学院工学研究科，修士論文，2008
3) 日本神経回路学会；http://www.jnns.org/niss/2000/text/koike2.pdf
4) 高山　純一　中山　晶一朗；適応的マルチエージェントを用いた交通規制時の交通シミュレーションモデルに関する研究；適応的マルチエージェントを用いた災害時交通モデルの構築とネットワーク信頼性解析，5-27，2006
5) 生天目　章；マルチエージェントと複雑系，1-22，23-53，1998