人　工　知　能　99−20
（１９９５．３．７）

# 状況に応じた戦略選択による実時間プランニング

月岡　陽一　　鈴木　英之進　　志村　正道

{tsukky, eino, shimura}@cs.titech.ac.jp
東京工業大学　大学院
情報理工学研究科　計算工学専攻

## 概要

動的世界における資源の有限性や不確定性に対応した実時間問題解決として、リアクティブプランニングに関する研究が注目されている。しかし、計算能力や状況に応じた適切な問題解決戦略の選択は困難である。本研究では動的に変化する領域の状況に応じて戦略を適切に使い分ける基準を、学習によって自動的に獲得する手法を提案する。本学習では組合せ最適化の近似解法を用いて戦略の選択基準を洗練する。実験では、対象領域として二次元平面上で複数のタクシーが平均待ち時間を小さくするように客を迎えに行く Taxi World を提案する。そして本学習手法を用いて複数の戦略を状況に応じて使い分ける選択基準を生成し、本手法の有効性を検証した。

# Real-time Planning by Selecting the Most Appropriate Strategy according to Situations

Yoichi Tsukioka,　Einoshin Suzuki,　and　Masamichi Shimura
Department of Computer Science,　Tokyo Institute of Technology
2-12-1 Ohokayama, Meguro, Tokyo 152, Japan

## Abstract

Due to its high capability in coping with the finiteness of resources and the uncertainty in the dynamic world, reactive planning has attracted much attention of the researchers in the real-time problem solving community. Different strategies can be considered in reactive planning, however, it is difficult to select the best strategy according to the situations. In this paper, we propose an automatic learning method for acquiring the criteria for selecting the best strategy. The effectiveness of the learning approach has been validated in an experimental domain "Taxi World", in which several taxis transport clients with minimum waiting time.

## 1 はじめに

近年、動的世界における実時間問題解決システムとして、リアクティブプランニングに関する研究が注目されている [4]。リアクティブプランニングにはプランの作成方式、リプランニングの度合など特定の状況で有効となる複数の戦略が存在する。例えばプランの作成方式には、プランの再利用や並列化を導入することによりプランを求めるコストを軽減、合理化しようとする熟考性を重視したアプローチと、あらかじめ用意した部分的な動作パターンを利用してエージェントを暫定的に行動させ、目標状態までのプランを無理に生成しない即応性を重視したアプローチの二種類が存在する。一般に前者の場合、計算コストが大きい反面高品質のプランを生成できるため、対象領域の変化の度合が小さく複雑な動作が要求される状況において効果的である。一方、後者は動作の最適性は低いが計算コストが小さいのが特徴で、対象領域の変化の度合が大きく単純な動作パターンだけでも容易に目標が達成可能な状況に向いている。

そこで Pollack らは、エージェントの行動基準となる意図構造のプランを維持するに当たり、プラン候補の数をフィルタリングで調節することによって、様々な状況において計算時間が少ない即応戦略と良質のプランを生成する熟考戦略の効率を評価している [2]。Kinny らは、エージェントが一度決めたプランを持続する度合 (degree of commitment) に注目して、プランをあまり変更しない大胆な (bold) 戦略と、逆にプランを変更し易い用心深い (cautious) 戦略の効率を状況毎に比較している [1]。しかし、これらの研究は領域の状況に応じた各戦略の効率の変化を確認したに過ぎず、積極的に状況に応じた戦略を選択するまでには至っていない。大沢 [3] は、分散協調問題解決の対象領域である追跡問題 において、状況に適した組織スキーマをあらかじめ求めておくことによって、変化する環境に応じて動的に組織スキーマを再編する手法を提案している。これは状況に応じた戦略選択の一種と見なすことができる。しかし様々な領域においてここで示されたような手法を適用する場合、領域の解析に手間がかかり様々な状況に適した戦略を求めることは困難である。

本研究では動的領域において特性の異なる戦略を複数個用意して、**状況に応じて戦略を適切に使い分ける基準を学習によって自動的に獲得**する手法を提案する。まず、動的世界の具体例として、二次元平面において複数のタクシーが中央局の指示に従って平均待ち時間を小さくするように客を迎えに行く「Taxi World」を提案する。中央局は暫定戦略や熟考戦略など特性の異なる三種類の基本的な戦略の中から一つを選択してプランを生成する。領域の状況は「客数」、「客発生率」、「直前のプラン」の組合せに基づいて区別し、この三つの要素を学習で用いる属性とする。そして客の発生についてのデータを利用して、刻々と変化する領域の状況に対して適切な戦略の選択基準を、貪欲法 (greedy method) を応用した学習手法によって獲得する。その結果、状況に応じた効率の良い戦略を選択するシステムが自動的に構築される。

本研究で提案する学習手法は以下の性質を持つ領域に適用可能である。

- 動的領域の非決定的な変化がエージェントの動作と独立である。
- エージェントが複数存在する場合は、それらはマネージャに当たる唯一のエージェントによって集中的に制御される。

このような条件を満たす対象領域は Taxi World に限らず、ネットワークを通じた複数の計算機への仕事の割り振りや、Factory Automation など数多く存在し、本研究で提案する手法の様々な領域への応用が考えられる。

学習手法の有効性を検証する実験では、以下の場合において単一の戦略では実現できないほどの大きな効率の向上が実証された。

- 領域の状況の変化が明示的で激しい場合
- 基本的な戦略が複数用意されていて、それらの相対的な能力が状況に応じて大きく変動する場合

以下、2節では Taxi World を説明する。3節では中央局のプランニング手法について議論し、4節では戦略の選択基準を学習する手法を提案する。5 節では学習手法の有効性を実験によって検証する。6節は結論である。

## 2 Taxi World

Taxi World とは、100 × 100 の二次元格子世界において客の要求 (客の所在地と目的地の座標) がランダムに発生し、その要求を満たすために複数のタクシーが

中央客の指示に従って動く領域世界である。タクシーは1台当たり一度に高々一人の客を乗せることが可能で、単位時間当たり上下左右に1マス移動もしくは静止することが可能である。客を乗せていないタクシーが客の待っている場所に到着すると自動的に(瞬時に)客が乗車し、その客の目的地に到着すると同様に客は降車する。領域内に発生する客は一定の周期性をもって変化するものとする。客の座標は所在地、目的地共にランダムに決定する。度発生した客はタクシーが迎えに来るまで消滅しない。

プランとは全てのタクシーに対してそれぞれ迎えに行く客とその順番を指定したものである。以後、任意の時点においてタクシーの行動基準となっているプランを、**有効プラン**と呼ぶことにする。有効プランは、新しいプランが生成されるまで有効プランであり続ける。有効プランにおいて迎えに行くべき客が割り当てられていないタクシーは静止しているものとする。従来の研究では、プランの生成とエージェントの動作は同時に実行できないことが前提とされていたが、Taxi World におけるエージェント(タクシー)は**プランナがプラン生成中でも有効プランに従って動作する**ことができる。その違いは、エージェントの形態や能力、つまりプラン生成モジュールと動作モジュールが独立に存在するか否かに依存する。

プラン生成の要求は、以下の二つの場合に限って発生する。

- 新たな客が発生した場合。
- まだ運ばれていない客が存在していて、かつ客が割り当てられていないタクシーが生じた場合。

客が待つ場所及び目的地はすべて正確に把握されているものと仮定し、プラン生成の要求は以下の情報から構成される。

- タクシーの現在位置、及び全ての待ち状態の客の位置と目的地
- その時点で有効なプラン

プラン生成の要求が生じた場合、プランナはが休止中であれば即刻、実行中であればそれが終り次第プランニングを実行し新しいプランを生成し、生成されてきたプランを有効プランとする。

Taxi World における目標は、**発生した客全ての平均待ち時間が出来る限り小さくなるようにタクシーに客を迎えに行かせる**ことである。客の待ち時間は、その客が発生してから任意のタクシーが迎えに来て乗車するまでの時間とする。図1は、Taxi World において4台のタクシーが与えられたプランに従って4人の待ち状態の客を迎えに行こうとしている状態を示したものである。ただし、直線は客の要求、破線はプランに基づくタクシーの予定経路を表す。

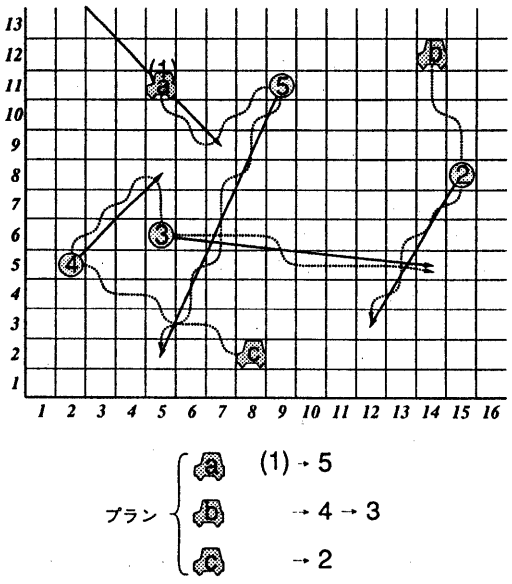


図 1: Taxi World

## 3 プランニング

### 3.1 基本戦略

本研究におけるプランナは、状況に応じて1つのプランニング戦略を選択してプランを作成する。戦略としては、暫定戦略、ビーム戦略、修正戦略の三種類を用意した。暫定戦略とビーム戦略はそれぞれ典型的な即応戦略と熟考戦略に相当する。修正戦略は現在の有効プランを再利用する戦略である。以下、それぞれの戦略について説明する。

**暫定戦略**

1台のタクシーに対して最寄りの客を高々一人だけ割り振ったプランを生成する戦略である。必要最低限の大きさのプランしか生成しないために計算量は小さ

い。反面、目先のことしか考えていないために後から効率の悪い客の運搬を強いられる可能性が高い。よって客がたくさん発生していて、かつ客発生率が高い時に効果的な戦略である。客数$N$に関する計算量のオーダーは$O(N\log N)$である。

**ビーム戦略**

全ての客を運ぶプランをビーム探索を用いて生成する戦略である。ビーム幅を適切に設定すれば質の良いプランを生成する。よって、客数が少ない場合や、客発生率が低い時に効果的な戦略といえる。待ち状態の客数$N$に関する計算量のオーダーは$O(N^2\log N)$である。

ビーム戦略は特に計算量が大きく、プラン生成が要求されてから実際にプランを生成するまでの遅延時間が比較的大きくなってしまう。この遅延時間が大きくかつ有効プランが大きく変更された場合、タクシーの移動が生成されるプランの実行に悪影響を及ぼすこともある。そこでプランの生成を開始する前に、プラン生成にかかる時間を推定し、プラン生成後その状態からプランニングを行なうことにした。このことによって応答の遅延によるタクシーの非効率的な動作を大幅に減らすことが可能となった。

**修正戦略**

有効プランを修正することによって新しくプランを生成する戦略である。ここでの修正とは、有効プラン中に含まれていない待ち状態の客を、一人ずつ有効プラン中の客の割り振り行列に割り込ませることである。新しく客を割り込ませた際に、全体の客の平均待ち時間の増分が最も小さいプランを選ぶ。修正戦略の計算時間は、有効プラン中に含まれていない客の数に大きく依存する。待ち状態の客数が多い上有効プランが暫定戦略で作られた場合は、有効プラン中に含まれていない客が多くなり、その分修正の手続きが繰り返されて計算時間は大きくなる。その時生成されるプランの質も一般的に悪い。それ以外の場合は計算時間は小さく、生成されるプランの質も程々に良い傾向がある。待ち状態の客数$N$と有効プランに含まれていない客数$n$に関する計算量のオーダーは$O(N^2n)$となる。つまり、有効プランの生成元が暫定戦略の場合は$O(N^3)$、ビーム戦略や修正戦略の場合は$O(N^2)$となる。

## 3.2 状況の区別

状況は、適切な戦略を選択できるように詳細に状況を区別することが望ましい。しかし区別する状況の数が増加すると、学習にかかるコストは大きくなる。本研究では状況を区別する基準として**客数、客発生率、有効プランの質**の三つの属性に注目する。また各属性はそれぞれ3個の属性値をとるものとした。これによってTaxi Worldにおいて変化する状況が明示的に表現される上、学習の際に先に紹介した3種類の戦略の特性を引き出した選択を行なうために必要最低限と考えられる。

客数はプラン生成の計算量に大きな影響を及ぼす。閾値を二つ設定して'少ない'、'中間'、'多い'の三つの状況に区別し、それらを属性値とする。客発生率は100単位時間当たりに発生する客の数を表し、プラン生成要求の発生する頻度、ついては生成したプランの有効期限の長さの指標となる。ここでは客の発生パターンが周期性を持っていることを前提としており、更に客の発生パターンに関する情報は既知であるものと仮定する。従ってある時点での客発生率は時刻のみによって獲得することができる。客発生率も客数と同様に閾値を2つ設定して'低い'、'中間'、'高い'の三つの状況に区別し、それらを属性値とする。客数と客発生率に対する閾値は、学習の際に変更される。有効プランの質は基本的に有効プランが直前にどの戦略によって生成されたかを示し、修正戦略を使用するのが効果的か否かの大きな判断基準となる。有効プラン中に含まれていない待ち状態の客の数が多い場合は'不完全'、有効プランの質が良い場合は'新しい'、悪い場合は'古い'としてこの三つの状況に区別する。プランの質のみを基準にして状況を区別した場合、使用された戦略による状況遷移図を図2に示す。暫定戦略を使用した場合は常に'不完全'、ビーム戦略を使用した場合は常に'新しい'となる。修正戦略を使用した場合は、直前の状況が'不完全'もしくは'古い'なら'古い'、直前の状況が'新しい'でビーム戦略が実行されてからしばらくの間なら'新しい'、直前の状況が'新しい'でも修正戦略がある回数以上繰り返された場合なら'古い'となる。ここで「しばらくの間」とは、客数:Nと最後にビーム戦略が実行されてから修正戦略が繰り返された回数$i$に依存し、$i$/Nが一定値$k:(0\leq k\leq 1)$を越えなければ状況は'新しい'のまま、それを越えたら'古い'に遷移する。これは修正戦略が繰り返されるとプランの質が

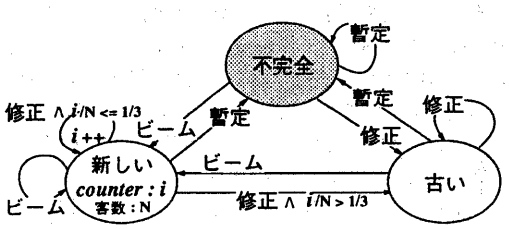


図 2: プランの質に対する状況遷移図

徐々に劣化していくことが予想されるためである。$k$の値が大きいほど劣化したプランを新しいと見なすことになる。本研究では$k$の値を 1/3 とした。

## 4 学習手法

本節では戦略の選択基準を獲得するための学習手法について述べる。

まず3.2節で説明した状況を区別する 3 種類の基準を用いて、組合せ的に 27 通りの状況を区別する。本研究では各状況で選択する戦略と、客数や客発生率を区分する閾値の値を示したものを**戦略テーブル**と呼ぶ。また、学習用のデータとして一定時間分の客発生に関するデータを**客発生サンプル**と呼ぶ。客発生サンプルが小さ過ぎる場合は十分な学習結果が得られないが、逆に大き過ぎると学習にかかるコストが大きくなる。

本研究で構築した学習アルゴリズムは、**Preparation Phase** と **Main Phase** の二種類のフェーズから構成される。Main phase は組合せ最適化の近似解法である貪欲法 (Greedy Method) のアルゴリズムを応用した手法により戦略テーブルを洗練する。つまり戦略テーブルと客発生サンプルを基に実世界環境に関するシュミレータを実行し、戦略テーブルを変更する手続きを繰り返すことにより、できる限り評価値 (客の平均待ち時間) が小さくなるような戦略テーブルの生成を目指す。シミュレータが実行されると、戦略テーブルに対する評価値と状況の**出現頻度順リスト**の 2 つが生成される。出現頻度順リストとは、シミュレーション中に各状況でプラン生成要求が生じた数を求めておき、各状況を降べきの順にソートして生成したリストである。

Main Phase では以下の一連の手続きを実行する (図3参照)。図中の各部分に記された番号は、以下に示す手続き番号に対応する。

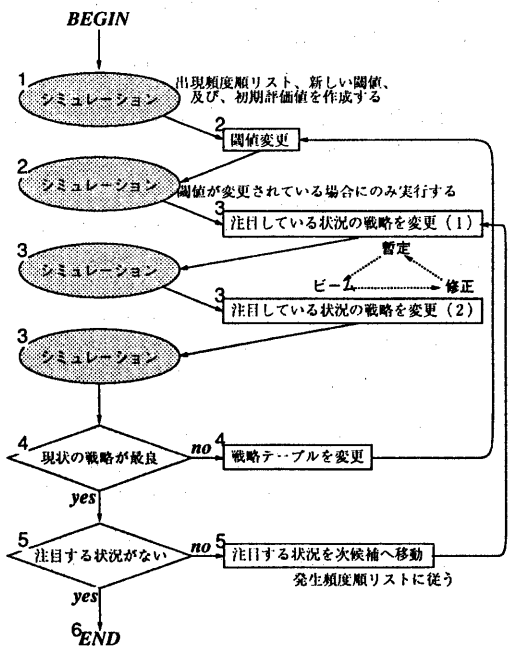


図 3: Main Phase の流れ図

1. 初期設定の戦略テーブルを基にシミュレーションを行ない、出現頻度順リストを作成する。
2. 出現頻度順リストを基に閾値を変更してシミュレーションを行ない、新たに出現頻度順リストを作成する。
3. 出現頻度順リストにおいて最も出現頻度の高い状況に注目して現状の戦略以外の戦略全てと交換してそれぞれシミュレーションを行ない、それぞれの評価値、出現頻度順リストを生成する。
4. もし現状の戦略の評価値以外のものが最良の場合は、注目している状況で選択される戦略を評価値が最良の戦略に変更した上で 2 へ戻る。
5. 出現頻度順リストにおいてまだ注目する状況がある場合は、出現頻度順リストから注目していた状況を削除した上で 3 の手続きを行なう。
6. 終了

以上の手続きを実行する際に閾値の設定は、

「各状況の出現頻度が均等であるほど適切である」

という仮説に基いて、各状況で選択される戦略を変更すると同時に閾値も変更する。

本アルゴリズムの特徴を以下にまとめる。

a. 最も出現頻度の高い状況から順に適切な戦略を設定していくため収束が早い。

b. 戦略テーブルが変更された場合、戦略テーブルのチェックを最初からやり直すために異なる状況同士の間に依存関係などがある場合に対応できる。

c. 出現する全ての状況において全ての戦略が試されることが保証されているので、生成されるプランの評価値は常に極小値をとることが保証される。

Main Phase における戦略テーブルの洗練においては、たとえ他に優れた極小値が存在しても最初に陥った極小値から抜け出すことができない。この問題は近似解法を実行する際には避けられない問題である。しかし、対象領域においてある程度の連続性が存在する場合には、探索の初期状態の設定次第では陥る極小値がより小さくなる可能性を大きくすることは可能である。Taxi World においては初期の戦略テーブルの設定方法がそれに当たる。そこで本学習手法では、1個の状況毎に選択すべき戦略の吟味をする学習手続き (Main Phase) を実行する前に、9個の状況を1個の状況集合と見なして戦略を吟味する大まかな学習手続きを実行し、初期設定の戦略テーブルを生成する手続きとして Preparation Phase を導入した。9個の状況、すなわち状況集合は、3つの属性のうちのいずれか1個に注目し、その属性がとる値が同一の状況を集めてくることによって得られる。このような状況集合は9種類存在する。ただ1つの状況に注目するのではなく、以上のようにして得られる状況集合に注目しながら Main Phase のアルゴリズムと同様に戦略選択の学習を行ない、大まかな学習を行なうことによって初期設定の戦略テーブルを生成する。Preparation Phase を導入することにより、先に述べた初期設定の戦略テーブルに関する問題点がほぼ解消され、より効率の良い戦略テーブルの生成が可能になると考えられる。

## 5 実験

本節では4節で説明した学習手法の有効性を検証するため、その実験結果と考察について述べる。

本実験で選択可能な戦略は暫定戦略、ビーム戦略、修正戦略の3種類で、単位時間長 0.5sec、ビーム戦略のビーム幅は 10 とした。客数、客発生率をそれぞれ3つの属性値に区別する閾値の初期設定は、出現し得る状況を3等分する設定として ([客数小, 客数大], [客発生率小, 客発生率大]) = ([10, 20], [4, 7]) とした。タクシー数については予備実験の結果から以下の事実が知られている。

- タクシー数が多いほど単位時間あたりに消費できる客数は単なる比例の割合分以上に増加するが、反面プランニングにおける計算量が増加する。
- タクシー数が多いほど互いの非合理的な客の運搬を補い合い易くなるため、プランの質による客の待ち時間への影響が小さくなる。

以上のことを考慮して本実験におけるタクシー数を5台に設定した。客発生サンプルは、1周期が2500単位時間で正規分布に準じるものとしその分布の様子が緩慢なものから急激なものまで5種類を用意し、更に予備実験の結果から学習用データの客発生サンプルの大きさはそれぞれ10000単位時間 (4周期) 分と決定した。

以上のような実験条件のもとで、4節で説明した学習手法を用いて5種類の客発生サンプルに対してそれぞれ戦略テーブルを獲得した。そして学習時に用いた客発生サンプルとは別に検証用の客発生サンプル10個をそれぞれ用意して、学習前後の戦略テーブルを評価した結果を表1に示す。表中の'限界' とは、参考のためプラン生成にかかる計算時間を0とした理想的なビーム戦略を実行した結果であり、客の平均待ち時間をどれだけ小さくすることができるかの目安である限界値として示した。学習前の2つの項目はそれぞれ、

暫：全て暫定戦略を選択する戦略テーブル
ビ：全てビーム戦略を選択する戦略テーブル

を評価した結果であり、各客発生サンプル毎の限界値と比べて客の平均待ち時間がどれだけ増加するかについて示した。同様に学習後の3つはそれぞれ、

暫：全て暫定戦略を選択する戦略テーブルを初期設定として Main Phase のみの学習を実行した結果の戦略テーブル
ビ：全てビーム戦略を選択する戦略テーブルを初期設定として Main Phase のみの学習を実行した結果の戦略テーブル

表 1: 学習による客の平均待ち時間の変化

| | | 学習前 | | | 学習後 | | |
|---|---|---|---|---|---|---|---|
| | 限界 | 暫 | ビ | 修 | 暫 | ビ | 準備 |
| 緩慢 2 | 144 | 20 | 7 | 33 | 13 | 8 | 7 |
| 緩慢 1 | 189 | 22 | 12 | 39 | 21 | 11 | 15 |
| 普通 | 225 | 25 | 20 | 36 | 22 | 16 | **16** |
| 急激 1 | 245 | 29 | 28 | 40 | 29 | **22** | **20** |
| 急激 2 | 280 | 33 | 37 | 36 | **21** | **26** | **19** |
| | | ※ 限界値で引いた値 | | | | | |

準備 : Preparation Phase → Main Phase の順に学習した結果の戦略テーブル

を評価した結果、限界値よりどれだけ増加するかについて示した。

表 1 に示した実験結果から確認できた事実を以下に示す。

1. 客の発生が突発的になるに従って客の平均待ち時間が長くなっていく。

2. 学習前の戦略テーブルとしては、客の発生が緩慢 (loose) な場合は ビーム戦略、突発的 (severe) な場合は暫定戦略のみからなる方が効率が良い。また、修正戦略のみから成るテーブルはあまり効率が良くない。

3. 客の発生が緩慢な時は、学習前の方が効率が良い場合がある。逆に、客の発生が突発的であるほど学習の効果が大きくなる。

4. Preparation Phase → Main Phase の順に学習を行なう手法はあらゆる客の発生パターンにおいて有効である。

1 の原因は、客の発生が突発的になるに従って客がタクシーの客消費能力を上回る発生頻度で発生し、それに応じて客が待ち状態で存在する時間が増加するためである。2 の結果についても、ごく自然な結果といえる。客の発生が緩慢な場合、待ち状態の客はそれほど増加し過ぎることはない。よって質の高いプランを作る ビーム戦略で所要する時間もそれほど大きくならないため、ビーム戦略の方が効率が良くなる。逆に客の発生が突発的な場合は、一時的に多くの待ち状態の客が出現する。よってビーム戦略で所要する時間が大

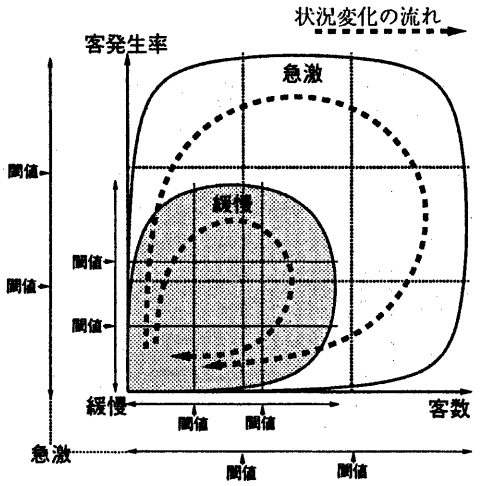


図 4: 状況 (客数, 客発生率) の変域に関する概念図

きくなり過ぎるため、単純に近い客から消費していく計算計算コストの小さい暫定戦略の効率が上回るようになる。また 3 については、まず学習用データとして用いた客発生サンプルが十分な大きさではない可能性を考慮しなければならない。一般性を欠いた客発生サンプルを用いることによって、そのサンプル固有の特徴に従った学習が行なわれてしまう。以降これを不十分な数の客発生サンプルによる雑音の学習と呼ぶことにする。この時 '緩 2' や '緩 1' など客の発生が緩慢な場合は、(客数や客発生率の) 属性の変域が狭く状況の変化が小さいために (図 4濃網掛部)、雑音が生じる可能性のある学習を行なうよりも、全ての状況において単一戦略 (ここではビーム戦略) を用いる方が効率が良くなるものと思われる。一方 '急 1' や '急 2' など客の発生が突発的になってくると、属性の変域が広くなってきて状況の変化が大きく激しくなるため (図 4淡網掛部)、多少の雑音を学習する可能性が生じるリスクよりも、学習で得られる各状況に応じたほぼ適切な戦略選択基準によって全体的に効率の良いプランニングを実現するメリットの方が上回るために、学習を行なった方が良い結果が得られるものと考えられる。

次に 5 種類の客発生サンプルのうち、最も客の発生が急激なもので Preparation Phase → Main Phase の順に学習した結果、獲得された戦略テーブルの内容と各状況で選択された戦略の回数 (状況の出現頻度) についてまとめた結果を表 2に示す。

表 2: 学習で得られた戦略テーブル

| 閾値(客数, 発生率) | | ( [8.5, 27.0], [6.5, 10.0] ) | | |
|---|---|---|---|---|
| 高い | 不完全 | ビー 7 | ビー 2 | 修正 |
| 高い | 新しい | 暫定 9 | 修正 89 | 修正 28 |
| 高い | 古い | ビー 3 | ビー 21 | 修正 3 |
| 中間 | 不完全 | ビー 4 | ビー | 修正 10 |
| 中間 | 新しい | 修正 16 | 修正 | 修正 9 |
| 中間 | 古い | ビー 12 | ビー | 暫定 11 |
| 低い | 不完全 | 暫定 57 | ビー | 修正 1 |
| 低い | 新しい | 暫定 4 | 修正 5 | 修正 4 |
| 低い | 古い | ビー | ビー 4 | 修正 42 |
| 発生率 | 質 ／ 客数 | 少ない | 中間 | 多い |

まず閾値に注目すると、初期設定が、([客数小, 客数大], [客発生率小, 客発生率大]) = ([10, 20], [4, 7]) であったのに対して、学習後の値が大幅に変更されていることがわかる。これは様々な種類の客発生サンプルに基づいて学習するに当たり、それぞれ客数や客発生率に関する状況の散らばりが変化するためである。実験結果では客数や客発生率で区別した時の状況における戦略選択の出現頻度が均等化されており、閾値に関する仮説が有効に生かされていることが確認できる。

選択された戦略について非常に興味深い点は、客数と客発生率が同一の三つの状況をひとまとまりと考えた場合、何箇所かで、

(不完全:ビーム, 新しい:修正, 古い:ビーム)

のようなパターンが生成されていることである。このパターンに従うと、ビーム戦略を使用してからしばらく修正戦略を使用して、プランの質が劣化してくると再びビーム戦略を使用することになる。即ち、修正戦略において有効プランの質が良ければ、計算時間が短くかつ割合質の良い新しいプランを生成することができる特性を生かした戦略が、本学習手法で自動的に獲得されている。その他客数が多く客発生率が高い時において、

(不完全:修正, 新しい:修正, 古い:修正)

のようなパターンが生成されている、プランの質が劣化していても客が発生した場合は常に有効プランを生かしてその客を追加し、プランの質の劣化を避けるよりも計算時間を短くする戦略が生成されていることも興味深い。

## 6 おわりに

本論文では、動的に変化する領域の状況に応じて戦略を適切に使い分ける基準を、学習によって自動的に獲得する手法を提案した。この手法は従来手作業で求められていた戦略の選択基準を自動的に獲得するためのものであり、より知的なエージェントを構築するために必要不可欠の技術である。また対象領域として Taxi World を提案し、実験によって本研究で提案した手法の有効性を検証した。

実験では、領域の状況の変化が明示的で激しい場合、また複数の戦略同士の相対的な能力が状況に応じて大きく変動する場合において、本手法を用いることによって単一の戦略では実現できないほどの大きな効率の向上が実証された。1でもあげたように本手法には対象領域についていくつかの制約が存在するが、ネットワークを通じた複数の計算機への仕事の割り振りや、Factory Automation など様々な領域への応用が考えられる。

今後の課題としては、雑音を学習しない最低限のデータ量を厳密に吟味する必要がある。さらに、待ち状態の客の消滅やメッセージ通信における遅延時間など、動的世界において考えられるより多くの不確定性を考慮することも興味深いと考えられる。

## 参考文献


[1] D.N. Kinny and M.P. Georgeff. Commitment and effectiveness of situated agents. *Proceedings of the Twelfth International Joint Conference on Artificial Intelligence(IJCAI-91)*, pp. 82–88, 1991.

[2] M.E. Pollack and M. Ringuette. Introducing the tileworld: Experimentally evaluating agent architectures. *Proceedings of the Eighth National Conference on Artificial Intelligence(AAAI-90)*, pp. 183–189, 1990.

[3] 大沢英一. 協調プランニングにおける動的組織再編とメタレベル整合戦略——追跡ゲームにおける考察. コンピュータソフトウェア, Vol. 12, No. 1, pp. 43–51, 1994.

[4] 山田誠二. リアクティブプランニング. 人工知能学会誌, Vol. 8, No. 6, pp. 35–41, 1993.