階層型強化学習を用いた鋼板ストックヤードにおける鋼板搬出作業のプランニング

正員　白 石 耕 一 郎 [*]　正員　木 村　元 [**]

Planning for Delivering Steel Plates in a Stockyard Using Hierarchical Reinforcement Learning

by　Koichiro Shiraishi, *Member*　Hajime Kimura, *Member*

Summary

In many steel plate stockyards, steel plates are piled up on the yard for a restriction of the yard's site area, and the steel plates must be moved from the top op the pile one by one. Therefore, we must turn over the many steel plates to pick up objective plates for the sorting. Workloads of turning over the steel plates are relevant to an order of pick up the steel plates. So, planning to determine the order can reduce costs for delivering the steel plates. This planning problem is a large scale multi-decision problem including hierarchical decisions. For this reasons, it is difficult to optimize the problem using usual optimization techniques such as Taboo search and Genetic algorithm. In this paper, the planning problems are modeled as Markov decision processes, and optimum planning is achieved by Hierarchical reinforcement learning. The proposed method is demonstrated through several numerical experiments to compare a method using a simple heuristics. The results off the numerical experiments show that the proposed method can save the costs about more than 10% compared to the method using the heuristics.

1. 序　論

地球温暖化などの環境問題の解決のために，船舶海洋分野においても省エネルギー化や温室効果ガス削減のための技術開発が求められている．鋼板の保管や仕分けを行う鋼板ストックヤードにおいても，温室効果ガス削減はもちろんのこと，韓国や中国との国際的競争で勝ち残るためには作業の効率化やコスト削減が求められている．国内の多くの鋼板ストックヤードでは，鋼板の大きさや重量および敷地面積の都合上，鋼板を積み上げて保管している．また，鋼板は工程計画で定められた順番どおりには納入されないため，搬出作業の前段階において仕分け作業を要する．さらに，鋼板ストックヤードの置場スペースが仕分け作業に対して不十分な場合は，鋼板を 1 枚ずつ積み替えて必要な鋼板を取り出す“板繰り” と呼ばれる膨大な作業を繰返す必要がある．この作業が鋼板搬出作業を停滞させる一因となっている．板繰りの作業量は，取り出す鋼板の順序によって大きく変動するため，板繰りを削減する適切な鋼板搬出順序を見つけ出すことによって，クレーンを用いた作業を減らすことができ，作業時間やエネルギー消費などのコスト削減が可能である．しかし，板繰りの作業量を最小にするような最適な鋼板搬出順序を見つけ出すことは，人間の直感で行うことは非常に困難である．熟練したオペレータでも，50 枚以上の鋼板が積み重なった場合の最適な鋼板搬出順序を見つけることは，おそらく不可能であると考えられる．そのため，最適な鋼板搬出順序を求めることが可能な最適化手法が必要とされている．本研究で取り扱う最適な鋼板搬出順序を求める問題は，一般的にはプランニング問題と呼ばれている．プランニング問題は，初期状態から目標状態に到達するまでの意思決定をコスト最小，もしくは報酬が最大になるように定める問題である．プランニング問題として，カーナビゲーションにおける自動車の経路計画やロボットを効率よく動かすための動作計画問題などがあり，様々な分野において研究が行われている．鋼板搬出作業のプランニング問題もこれまでに数多くの研究が行われている．それらの研究の代表的な研究を次に述べる．

Guangmin ら [1] は，鋼板搬出作業のプランニング問題 (以後，プランニング問題と略記) を搬出する種類の鋼板をどのヤードから取り出すのかを定める最適化問題と，目的の鋼板を取り出すのに邪魔となる鋼板をどのヤードに置くのかを定める最適化問題の二段階の最適化問題として定式化を行い，遺伝的アルゴリズムを用いて解を探索する方法を提案している．しかし，この方法は鋼板を種類別に積み上げて保管するような敷地面積の広い鋼板ストックヤードを対象としているため，クレーンの移動距離を最小化することを目的としたプランニングとなっており，板繰り作業の低減を重視していない．一方，本研究で対象とする敷地が狭い

*　海上技術安全研究所 (研究当時 九州大学大学院)
** 九州大学大学院

原稿受理　平成 年 月 日

鋼板ストックヤードでは，板繰り作業の低減が求められているため，Guangmin らの方法を用いても有用なプランニングを得ることができない．

また，荒井ら[2]は，プランニング問題を組み合わせ最適化問題として定式化し，遺伝的アルゴリズムを用いて解を探索する方法を提案している．この方法は鋼板ストックヤードだけでなくコンテナヤードへの適用も考慮した幅広い方法である．しかしながら，遺伝的アルゴリズムの遺伝子の表現に鋼板配置と鋼板搬出順序を遺伝子として表現する方法を採用しているため，何十枚もの鋼板を取り扱うような大規模問題に適用する場合は，状態空間が膨大になり初期解の生成が難しく，大規模問題に適用できないという問題点が存在する．

阿部ら[3]は，プランニング問題を組み合わせ最適化問題として定式化し，最適な鋼板搬出順序をタブーサーチを用いて探索する方法を提案している．しかしながら，この方法も荒井らの方法と同様に，大規模問題になると状態空間が膨大になり初期解の生成が困難になり，有用な解を求めることができないという問題が存在する．

以上のように、これまでに提案されている手法は実問題において必ずしも有用とは限らない．この理由として，鋼板搬出作業のプランニング問題は階層的な意思決定を含む多段決定問題となるため，遺伝的アルゴリズムやタブーサーチなどの一般的な最適化手法を用いても有用な解を得ることが非常に困難なためと考えられる．

また，実際の鋼板ストックヤードでは，鋼板が遅れて入荷されるといった確率的要素が存在するため，優れたプランニングを行うためには，このような確率的要素の考慮も重要である．しかし，先行研究のようにプランニング問題を組み合わせ最適化問題として定式化した場合は，確率的要素を考慮することは不可能であるので，この問題を解決するための新しいアプローチが必要である．

本研究では階層的な意思決定や確率的要素の考慮といった問題点を解決するアプローチとして，プランニング問題をマルコフ決定過程としてモデル化し，階層型強化学習を用いて最適解を探索する方法を提案する．プランニング問題をマルコフ決定過程としてモデル化することによって，最適な鋼板搬出順序を動的計画法によって求めることができる．そして，鋼板の入荷の遅れを不確実性が存在する状態遷移として考慮することも可能となる．しかし，動的計画法は大規模問題に適用する場合，状態数が爆発的に増加し計算不可能になるという問題が存在する．本研究では，この問題を解決するために関数近似を用いた強化学習を用いる．強化学習はシミュレーションを通して学習を行うことで，最適解を求める手法である．また，最適な鋼板搬出順序を探索することは，階層的な意思決定を取り扱う複雑なタスクとなるため，一般的な強化学習を用いても学習に非常に時間がかかり，最悪の場合に有望な解を探索できないといった問題が存在する．この問題を解決するために，複雑なタスクを小さなサブタスクに分割し，学習を容易にすることが可能な階層型強化学習を用いる．

そして，数値実験において，簡易なヒューリスティクスを用いた方法と提案手法を比較することによって，提案手法の有用性を検証する．また，板繰り作業量は搬出する各グループの鋼板の枚数の組み合わせによって大きく変動する．そこで，鋼板の組み合わせによって，両手法の性能がどのように変化するかを調べるために，各グループの鋼板枚数のばらつきを表す指標として，“エントロピー (Entropy)”という概念を新しく導入する．そして，このエントロピーを基準とした数値実験を行うことによって提案手法の優位性を検証する．また，造船所の鋼板水切り・仕分け場における実際の鋼板データを用いた数値実験を行うことによって，実問題における提案手法の有用性も検証する．

2. 鋼板搬出作業のプランニング問題

2.1　鋼板ストックヤードの概要

鋼板はグループに関係なく，無秩序な順序で搬入される．そして，搬入された鋼板を出荷日ごとに 1ヶ所に積み重ねて保管する作業が鋼板搬入作業となる．鋼板搬入作業の流れを Fig.1に示す．

鋼板ストックヤードでは，鋼板はブロックの工程や加工日に応じてグループ化される．本研究では，ブロックの工程や加工日に応じてグループ化される鋼板を “グループ (Group)” と呼ぶ．そして，鋼板ストックヤードでは，搬入された鋼板を必要に応じて次の工程に搬出していく鋼板搬出作業が行われる．鋼板搬出作業の流れを Fig.2に示す．しかし，各鋼板は工程計画で定められた順序ではなく，グループに関係なくばらばらに入庫されるため，次の工程への搬出時には仕分け作業が必要となる．また，仕分け作業において，鋼板のグループ数に対して “ヤード (Yard)” の数が不足している場合は，必要な鋼板を取り出すための板繰り作業が発生する．本研究では，鋼板ストックヤードにおいて鋼板を積み重ねて保管する場所をヤードと呼ぶことにする．この板繰り作業が鋼板搬出作業に要する作業量の増減に大きく影響するので，鋼板搬出作業のコスト削減のためには，板繰り作業量を最小にする鋼板搬出順序を見つけ出すことが可能なプランニング手法が必要である．

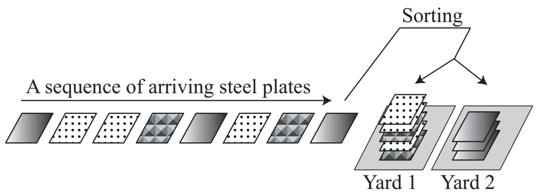

Fig. 1　Putting plates the stockyard.

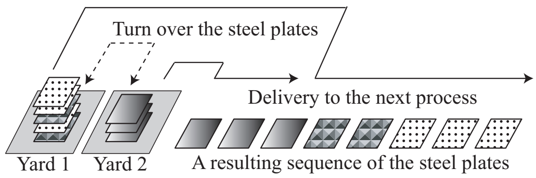

Fig. 2　Taking plates out the stockyard.

2.2　鋼板搬出作業における意思決定

鋼板を次の工程へ搬出する場合，板繰りをできる限り少なくなるように鋼板搬出順序を決定する必要がある．その場合に，どのグループの鋼板から搬出するのか，どのヤードから取り出すのか，邪魔な鋼板をどのヤードに置くのかという意思決定が必要となる．これらの意思決定は三段階の階層構造となっており，各階層における意思決定が複雑に関係している．階層的な意思決定の構成は，最上層が搬出する鋼板のグループを定める搬出グループ (Delivery group) の決定，中間層が搬出グループの鋼板を取り出すヤードを定める取出しヤード (Pickup yard) の決

定，最下層が邪魔な鋼板を一時的に置くヤードを定める一時ヤード (Temporary yard) の決定となる．鋼板搬出作業における各意思決定を Fig.3-Fig.5に示す．

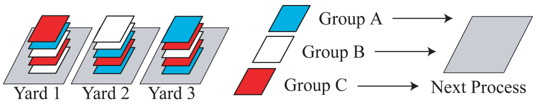

Fig. 3　Decision of the delivery group.

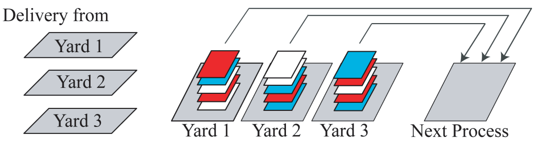

Fig. 4　Decision of the pickup yard.

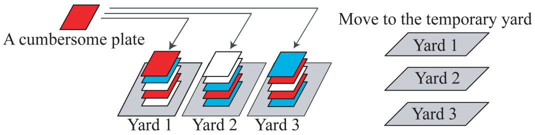

Fig. 5　Decision of the temporary yard.

2.3　プランニング問題の定式化

本研究で取り扱う鋼板搬出作業のプランニング問題は，全ての鋼板が搬入された状態 (スタート) から，全ての鋼板を次の工程へ搬出した状態 (ゴール) へ至るまでの階層的な意思決定をコストの総和が最小となるように定める問題となる．本研究では，プランニング問題を多段決定問題として次のような流れで定式化する．

プランニング問題の定式化のために，鋼板搬出作業の階層的な意思決定を次のように定める．搬出グループの順序を $G = \{g_1, g_2, \cdots, g_l\}$ とする．l は対象とする鋼板搬出作業における鋼板のグループ数である．搬出グループが g_i の場合の取出しヤードの順序を $Y_p(g_i) = \{y_1, y_2, \cdots, y_m\}$ とする．m は鋼板搬出作業で使用できるヤードの数である．そして，搬出グループが g_i，取出しヤードが y_j の場合の一時ヤードの順序を $Y_t(g_i, y_j) = \{y_1, y_2, \cdots, y_n\}$ とする．n は取出しヤードから搬出グループの鋼板を取り出すのに邪魔になる鋼板の枚数である．

プランニング問題で考慮するコスト関数 $Cost(g_i, y_j, y_k)$ を搬出グループが g_i の場合に，クレーンを用いて取出しヤード y_j の最上段の鋼板を一時ヤード y_k に動かすのに要するコストとして定義する．本研究では，鋼板 1 枚を取出しヤードから次の置場まで動かす作業時間を 1step として考え，この作業時間をコスト関数とする．この設定によってステップ数の総和を最小化することがプランニングの目的となる．また，コスト関数は目的に応じて自由に変更することができるので，エネルギー消費をコスト関数とすることでエネルギー最小化のプランニングを行うことも可能である．以上の定義によって，鋼板搬出作業のプランニング問題は次のように定式化される．

Minimize

$$F(G, Y_p, Y_t) = \sum_{i}^{l} \sum_{j}^{m} \sum_{k}^{n} Cost(g_i, y_j, y_k) \qquad (1)$$

Subject to

$$g_i \in G,\ y_j \in Y_p(g_i),\ y_t \in Y_t(g_i, y_j) \qquad (2)$$

プランニング問題は，コスト関数の総和を表す (1) 式を最小にするような，搬出グループの順序 G，取出しヤードの順序 Y_p，一時ヤードの順序 Y_t を決定する最適化問題である．本問題は階層的な意思決定を含む多段決定問題となるため，一般的な最適化手法を用いても有用な解が得ることができない困難な問題である．

2.4　鋼板搬出規則

本研究では，鋼板の搬出方法を決定するために，次のような鋼板搬出規則を定める．

鋼板搬出規則

1. 搬出グループの全ての鋼板の搬出が完了するまで，搬出グループ以外の鋼板は搬出しない．
2. 取出しヤードに存在する搬出グループの鋼板の搬出作業が完了するまで，取出しヤード以外のヤードの鋼板は搬出しない．
3. 取出しヤードの最上段の鋼板が搬出グループだった場合は必ず次の工程へ搬出する．そうでない場合は一時ヤードに積み重ねる．

鋼板搬出規則を満たすプランニングは，必ず鋼板搬出作業が完了することが保障される．本研究では，この鋼板搬出規則を満たすプランニングからコストの総和を最小にする鋼板搬出順序を探索する．その理由は，鋼板搬出規則を満たすことを解の条件に入れることによって，解候補として鋼板搬出作業が完了しない無駄な解を取り除きことができ，解の探索に要する時間を大幅に短縮することできるからである．

3. プランニング問題へのアプローチ

本研究では，プランニング問題をマルコフ決定過程としてモデル化する．これによって，プランニング問題の最適解は動的計画法を用いることによって求めることができる．しかし，動的計画法は大規模問題に適用する場合，計算時間とデータの記憶量の増大といった問題が生じる．本研究では，動的計画法を拡張した手法である強化学習を用いることによって計算時間の問題を解決する．また，関数近似を用いることによって記憶量の問題を解決する．一般的な強化学習では解くことが困難な階層的な意思決定を含む問題において，効率良く最適解を求めるために階層型強化学習を適用する．

3.1　マルコフ決定過程

マルコフ決定過程 [5)6)] とは，時間とともに状態が遷移しながら変化するような動的システムにを表現することが可能な数学モデルである．そして，マルコフ決定過程としてモデル化できる問題では，動的計画法や強化学習を用いることによって解を求めることが可能である．

プランニング問題をマルコフ決定過程としてモデル化するために，次のような定義を行う．まず，鋼板ストックヤードにおける鋼板配置を状態 s とし，本問題で取り得る状態の集合を状態集合 S とする．クレーンを用いた鋼板の移動を行動 a とし，各状態で選択可能な行動の集合を A とする．状態 s において行動 a が取られたときに，次の状態 s' へ遷移する確率を状態遷移確率 $P(s'|s,a)$ とし，そのときに得られる報酬を $R(s'|s,a)$ とする．ただし，鋼板の移動は確実に行われると仮定しているので，実現可能な全ての状態遷移は確率 1 で行われる．また，報酬はプランニング問題の定式化で定めたコスト関数の値とする．そして，状態 s で行動 a を実行した場合に得られる期待報酬和として，価値関数 $Q(s,a)$ を定義する．

マルコフ決定過程としてモデル化するためには，次の状態への状態遷移が，そのときの状態と行動にのみ依存し，それ以前の状態や行動に関係しないというマルコフ性が成立する必要がある．プランニング問題では，現在の鋼板ストックヤードの状態を観測し，その状態で最も良いと考えられるヤードへ鋼板を確実に移動させ，鋼板の移動において不確実性は存在しないとしている．．したがって，マルコフ性が成立するので，プランニング問題はマルコフ決定過程としてモデル化できる．鋼板ストックヤードのある状態における状態遷移をマルコフ決定過程としてもモデル化した場合の例を Fig.6に示す．

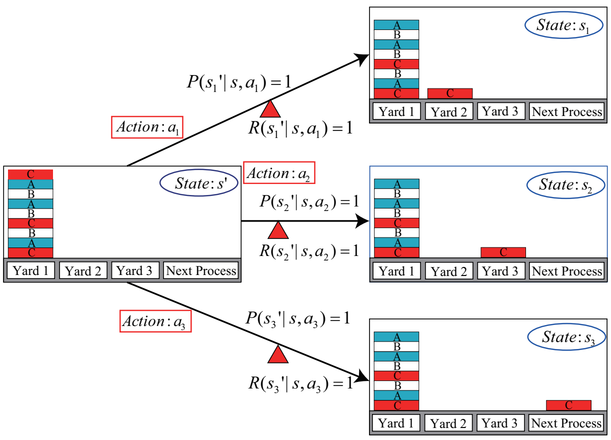

Fig. 6 An illustration of Markov decision processes in the steel plate stockyard.

3.2 強化学習

マルコフ決定過程では，最適価値関数 $Q^*(s,a)$ において次式の Bellman 最適方程式が成立する．

$$Q^*(s,a) = \sum_{s' \in S} P(s'|s,a) \left[R(s'|s,a) + \max_{a' \in A} Q^*(s',a') \right] \tag{3}$$

Bellman 最適方程式を解くことによって最適価値関数が求まり，全ての状態における最適な行動を決定することができる．しかし，Bellman 最適方程式は方程式の中に最大値計算が含まれるため，一般的な行列演算によって解くことは非常に困難である．そこで，Bellman 最適方程式を効率良く解くことが可能な動的計画法[5,6]が開発されている．動的計画法は，全状態における最適価値関数を反復計算によって求めるというものである．全状態の最適価値関数を記憶できるような問題ならば，動的計画法を用いることによって最適解を求めることが可能である．しかし，動的計画法は，プランニング問題のような大規模問題に用いる場合，状態数が指数関数的に増大するため，計算時間も長時間になり，実用の範囲内の時間では解を求めることが難しくなる．そのため，動的計画法は一般的に適用範囲が限られ，現実問題への適用が困難な手法であると考えられている．

そこで，本研究ではこの問題を解決するために，動的計画法を拡張した理論である強化学習[5,6]を適用する．強化学習は，シミュレーションを通して学習を行うことによって，最適価値関数を近似的に求めるという手法である．最適状態価値関数を求めるために，状態遷移確率行列や報酬行列を必要とする動的計画法とは異なり，動的システムのシミュレーションを通して状態と報酬を観測することによって最適状態価値関数を推定するのが特徴である．そのため，状態遷移確率や報酬行列が必要でないので，状態空間が膨大になっても反復計算が可能である．本研究では強化学習の手法として，学習の収束が証明されている Q-Learning[5,6] を適用する．Q-Learning は学習率というパラメータを適切に調整することによって，最適解へ収束することが保証されている．

3.3 関数近似

強化学習を用いる場合も Bellman 最適方程式を解くためには，価値関数の更新を行う状態の価値関数を記憶しておかなければならない．状態空間が膨大な場合，価値関数が更新される状態数も増大し，それら全ての状態の価値関数の記憶は非常に困難である．この問題を解決する方法として，関数近似[5,6]を用いることによって価値関数を表現する方法が提案されており，様々な大規模問題において関数近似の有用性が確認されている．そこで，本研究では関数近似の方法として線形アーキテクチャ(Linear architecture)[5,6,7] を用いる．この方法は，状態と行動の入力を基底関数によって特徴ベクトルに変換し，このベクトルの線形重み付け和で価値関数を表現するというものである．本研究では，特徴量として鋼板配置を抽出した特徴ベクトルを生成し，この特徴ベクトルを用いた線形アーキテクチャによって価値関数を表現する．

3.4 階層型強化学習

鋼板搬出作業のプランニング問題は階層的な意思決定を含む複雑なタスクとなるため，一般的な強化学習を用いた場合，学習に多大な時間を要する．この問題を解決するために，与えられたタスクを構造的に分割し，分割されたサブタスクの処理を受け持つモジュールを階層的に構成することによって，大規模なタスクの学習を可能にする階層型強化学習[8]が提案されている．本研究では，プランニング問題へのアプローチとして階層型強化学習を用いる．階層型強化学習において，タスクの分割作業はシステム設計者の手に委ねられる場合が多く，実際に用いる場合には対象の問題に応じたタスクの分割が重要となる．本研究では，プランニング問題の定式化において，意思決定が階層的な構造をしていることを確認している．この階層構造を用いてタスクを分割することによって，プランニング問題に階層型強化学習を適用することが可能である．具体的には，メインタスクがコストの総和を最小とする鋼板出順序を見つけ出すこととなり，サブタスクは搬出グループの順序の決定，取出しヤードの順序の決定，一時ヤードの順序の決定となる．これらのサブタスクにおける最適な動作を学習によって獲得し，各サブタスクの最適な動作を組み合わせる

ことでメインタスクにおける最適な動作を生成することが階層型強化学習の流れとなる．

本研究では，階層型強化学習アルゴリズムとして，Dietterich が提案する MAXQ Q-learning[9)10)] という手法を用いる．MAXQ Q-Learning は最適解への収束が証明されており，アルゴリズムも Q-Learning に似ており，比較的単純であるので実装が容易であるという特徴をもつ．鋼板搬出作業のプランニング問題に MAXQ Q-Learning を適用した場合，最上層のタスクがグループの搬出順序の決定，中間層のタスクが取出しヤードの決定，そして，最下層が一時ヤードの決定となる．また，最下層の一時ヤードの決定がプリミティブな動作となる．そして，各階層において MAXQ アルゴリズムを適用することによって，最適な行動を求めることが可能である．

3.5 提案手法の流れ

本研究で提案するプランニング手法の流れは次のようになる．まず，状態を鋼板配置，行動を鋼板の移動，報酬をコスト関数とすることによってプランニング問題をマルコフ決定過程としてモデル化する．そして，階層型強化学習のアルゴリズムのひとつである MAXQ を用いて，プランニング問題の最適解を求める．この最適解が最適な鋼板搬出順序となる．また，MAXQ の価値関数の表現方法として，鋼板配置を特徴量とした特量ベクトルを用いた線形アーキテクチャを用いる．以上の流れによって，鋼板搬出作業に要するコストを最小にする鋼板搬出順序を求めることが可能である．

4. 数値実験

本研究では，階層型強化学習を用いたプランニング手法である提案手法 (MAXQ) と簡易なヒューリスティクスを用いた方法 (Heuristics) を比較することによって提案手法の有用性を検証する．プランニング問題の難易度は鋼板の枚数と各グループの鋼板の枚数のばらつきに依存している．そこで，各グループの枚数のばらつきを表す指標としてエントロピーという概念を導入する．そして，エントロピーを基準とした数値実験を行うことによって，問題の難易度による提案手法の性能の変化を調べる．この章では，まず，本研究で用いたヒューリスティクスと各グループにおける鋼板の枚数のばらつきを表すエントロピーについて説明し，その後，数値実験の概要とその結果について述べる．

4.1 ヒューリスティクス

数値実験において，提案手法と比較する手法として鋼板ストックヤードで実際に用いられていると考えられるヒューリスティクス (Heuristics) を用いる．本研究で用いるヒューリスティクスは，鋼板搬出作業において，板繰り回数が少なくなるように鋼板を取り出すことを目的にしたものである．まず，ヒューリスティクスにおいて搬出グループを決定する基準となる板繰りコスト関数について述べる．板繰りコスト関数 $C_t(g_i, y_j)$ は，ヤード y_j におけるグループ g_i の鋼板を全て取り出すのに必要な板繰りの回数として次のように定義される．

$$C_t(g_i, y_j) = H(y_i) - I(g_i, y_j) - N(g_i, y_j) + 1 \qquad (4)$$

ただし，$H(y_j)$ はヤード y_j の鋼板の枚数，$I(g_i, y_j)$ はヤード y_j におけるグループ g_i の鋼板の最下端の位置，そして，$N(g_i, y_j)$ はヤード y_j におけるグループ g_i の鋼板の枚数と定める．板繰りコスト関数を用いることによって，現在の鋼板配置において，搬出時に板繰り回数が最も少ないグループを定めることができる．ヒューリスティクスでは，板繰りコスト関数を用いた次式で表される規則に従って，搬出グループ g_d，取出しヤード y_p，一時ヤード y_t を決定する．

搬出グループの決定

$$g_d = \arg\min_i \sum_j C_t(g_i, y_j) \qquad (5)$$

取出しヤードの決定

$$y_p = \arg\max_j N(g_d, y_j) \qquad (6)$$

一時ヤードの決定

$$y_t = \arg\min_k H(y_k) \qquad (7)$$

この規則では，まず板繰りコスト関数が小さいグループから搬出する．次に，搬出グループの鋼板の枚数が多いヤードから鋼板を取り出す．そして，一時ヤードは積み重なっている鋼板の枚数が最も少ないヤードとする．ヒューリスティクスでは，このような流れで，搬出グループ，取出しヤード，一時ヤードを決定し，これら以外の部分は鋼板搬出規則に従う．また，このヒューリスティクスは積み上がった鋼板の最上段からグループごとに鋼板を取り出していく方法よりも優れていることを数値実験において確認している．

4.2 エントロピー

本研究では，各グループにおける鋼板の枚数のばらつきを，情報科学分野におけるエントロピー [11)] という概念を用いて表現する．次式で表されるエントロピー関数は，ある事象についての選択の自由さ，不確定さ，多様性などを表す尺度として用いられる．

$$H = -\sum_{k=1}^{n} P_k \log P_k \qquad (8)$$

H はエントロピー関数，P_k は事象の発生確率，k は各事象を表すインデックスである．ここで，事象の発生確率 P_k を鋼板の総枚数において各グループの鋼板の枚数が占める割合とする．これによって，エントロピー関数はグループの鋼板の枚数のばらつきを表す尺度として考えられる．また，このエントロピー関数は，あるグループの鋼板の枚数だけ偏っており，それ以外が少ない場合にはエントロピーは小さく，逆にどのグループの鋼板の枚数が均等である場合にはエントロピーは最大となるという特徴をもつ．例題として，鋼板の総枚数を 30，グループ数を 3 とし，グループ A,B,C の鋼板の枚数を変化させた場合のエントロピーの値を Table 1に示す．Table 1から分かるように，エントロピーの値は，1 つのグループに偏っている場合が最小で，どのグループも均等である場合が最大になっている．

また，エントロピーは鋼板搬出作業の難しさを示すという特徴も有している．なぜなら，搬出する鋼板群において，あるグループの鋼板の枚数だけが多く，エントロピーの小さい場合は，枚数の多いグループから取り出していけば良いという自明な解が存在する．そのため，エントロピーの小さい問題は比較的，解を求ることが簡単な問題であるといえる．一方，どのグループの鋼

板の枚数が均等であるようなエントロピーの大きい問題は，どのグループの鋼板も均一に分布する傾向があるため，どのグループから鋼板を取り出しても板繰り作業量は大きく変わらない．したがって，エントロピーの大きい問題は解を求めることが困難な問題だと考えられる．本研究では，各グループの鋼板の枚数のばらつきに応じて，ヒューリスティクスと提案手法の性能にどのような差が現れるかを検証する．

Table 1 The number of plates per group for the examples of entropy.

	A	B	C	Sum	Entropy
Example 1	28	1	1	30	0.420
Example 2	24	5	1	30	0.852
Example 3	19	10	1	30	1.109
Example 4	15	14	1	30	1.177
Example 5	15	10	5	30	1.459
Example 6	10	10	10	30	1.585

4.3 実験概要

本研究では，階層型強化学習を用いたプランニング手法である提案手法 (MAXQ) の有用性を検証するために，比較手法として簡易なヒューリスティクスを用いた方法 (Heuristics) を用いる．また，プランニング問題はグループ数や鋼板の枚数によって問題の難易度が変わってくる．あらゆる難易度における提案手法の有用性を検証するために，次のような 3 つの数値実験を行った．1 つ目は，1 日に処理する鋼板データを用意し，提案手法がヒューリスティクスに比べ優れた解を探索できるかどうか検証するために行う．2 つ目は，各グループの鋼板の枚数が変化することによって，提案手法の優位性がどのように変化するかを調査する．3 つ目は，実際の造船所の鋼板ストックヤードにおける 1 日当たりに処理する鋼板のデータを用いることによって，実問題においても提案手法が有用であるかを検証する．

全ての数値実験において，搬入される全ての鋼板は 1 つのヤードに積み上げられると仮定している．この仮定によって鋼板搬出作業に要する板繰り回数が最大となり，解くのが困難な問題になるようにしている．ただし，初期配置のヤード数が 2ヶ所，3ヶ所，それ以上になっても提案手法を適用することは可能である．また，鋼板搬出作業で使用できるヤードの数は 3ヶ所と設定した．使用できるヤード数が 3ヶ所以上になっても提案手法によってプランニングを行うことが可能である．

提案手法には階層型強化学習の行動選択において確率的要素が含まれるため，探索の度に解が異なる可能性がある．その影響を考慮するために提案手法では，同じ鋼板配置において解の探索を 5 回行い，その平均を結果としている．ただし，ヒューリスティクスには確率的要素が無いのでこの設定は適用しない．また，鋼板の初期配置による問題の難易度の変化を考慮するために，グループ数と鋼板の枚数を同じにし，鋼板の並びだけをシャッフルすることによって 10 パターンの鋼板の初期配置を生成し，それら全ての初期配置において解の探索を行い，その平均を結果としている．この設定は，提案手法とヒューリスティクスの両方に適用する．本研究で用いた強化学習パラメータは，学習率を $\alpha = 0.1$，割引率を $\gamma = 1.0$ とし，行動選択手法としてボルツマン行動選択 [5] を用いる．計算機環境は，CPU: Pentium 4 3.0GHz, Memory: 2038MB, OS: Windows XP とした．鋼板ストックヤードのシミュレーションは，プログラム言語: JAVA を利用して作成した．

4.4 数値実験結果

(1) 鋼板の枚数とグループ数を変化させた場合の比較

提案手法の有用性を検証するために，4 種類の鋼板の搬入データを用意し，それぞれのデータにおいて数値実験を行った．各数値実験で用いた鋼板の総枚数，グループ数，各グループの鋼板の枚数，そしてエントロピーの値を Table 2に示す．Case1 と Case3 は各グループにおいて鋼板の枚数にばらつきが無い場合，Case2 と Case4 はばらつきがある場合である．そして，Case1 と Case2 は総枚数が 50 枚程度，Case3 と Case4 は 100 枚程度となっている．それぞれの数値計算結果を Table 3と 4に示し，4 種類の鋼板データにおける計算結果をまとめたグラフが Fig.7 である．Table 3において，提案手法がヒューリスティクスに比べ作業時間がどれだけ短縮したのかを示す解の差 (Difference) を示している．Fig.7の x 軸は Case 番号，y 軸が鋼板搬出に要した作業時間のステップ数となっている．また，この実験では，学習の最大エピソード数を Max Episode = 30000 とした．

Table 3 The results of the costs for Simulation (1): Changing the number of plates and group of steel plates.

	Heuristics	MAXQ	Difference
Case 1	145.4 ± 13.69	131.64 ± 5.8	9.46 %
Case 2	173.3 ± 15.32	140.58 ± 4.42	18.88 %
Case 3	406.8 ± 21.57	326.08 ± 8.78	19.84 %
Case 4	386.7 ± 16.28	295.86 ± 6.95	23.49 %

Table 4 The results of the calculation time for Simulation (1): Changing the number of plates and group of steel plates.

	Heuristics	MAXQ
Case 1	0.06 ± 0.01 [sec]	8.25 ± 1.76 [min]
Case 2	0.07 ± 0.01 [sec]	9.77 ± 2.18 [min]
Case 3	0.10 ± 0.03 [sec]	37.68 ± 7.91 [min]
Case 4	0.09 ± 0.02 [sec]	33.06 ± 6.68 [min]

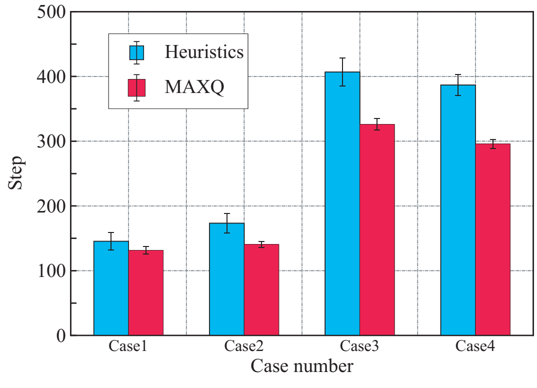

Fig. 7 The results of the costs for Simulation (1): Changing the number of plates and group of steel plates.

Table 2 The number of plates per group for Simulation (1): Changing the number of plates and group of steel plates.

	A	B	C	D	E	F	G	H	I	J	Sum	Entropy
Case 1	5	5	5	5	5	5	5	5	5	5	50	3.3219
Case 2	1	2	3	4	5	6	7	8	9	10	55	3.1036
Case 3	10	10	10	10	10	10	10	10	10	10	100	3.3219
Case 4	5	6	7	8	9	10	11	12	13	14	95	3.2541

Tabel 3より，全てのケースにおいて提案手法が作業時間が短い優れた解を探索できていることが分かる．提案手法は，どのケースにおいても解のばらつきが小さく，最適解に近い解が探索できていることが確認できる．一方，ヒューリスティクスは鋼板の初期配置に依存した方法であるため，解のばらつきが大きい．また，鋼板の枚数が増えて，プランニング問題の難易度が上がるにつれて，提案手法とヒューリスティクスの性能の差が広がっていることも確認できる．ただし，Tabel 4から分かるように，ヒューリスティクスは解の探索に 1 秒もかからないのに対して，提案手法の計算時間は，鋼板の総枚数が約 50 枚のとき約 10 分，約 100 枚のときは約 30 分となっている．そのため，計算時間の点ではヒューリスティクスの方が優れている．提案手法は，シミュレーションを通して学習を行っているため，シミュレーション時間によって解の探索時間が大きく変化する．鋼板の枚数が増えるにつれて鋼板搬出作業で必要となる時間も長くなるため，シミュレーション時間も長くなってしまう．しかしながら，今回設定した最大のエピソード数 Max Episode= 30000 は学習を収束させるために十分に大きい値に設定してあるので，最大エピソード数を短くすることによって計算時間を短くすることが可能である．最大エピソード数の減らすことで若干の解の劣化が考えられるが，Max Episode= 10000 でもヒューリスティクスより優れた解を探索できることを確認している．また，Max Episode= 20000 と Max Episode= 30000 の解の比較を行ったところ，作業時間の差は 3% 程度であったので，Max Episode= 20000 でも十分に優れた解が探索できる．

(2) エントロピーを基準とした比較

この数値実験は，プランニング問題の難易度を変えることによって，提案手法の性能がどのように変化するか調べることが目的である．そこで，鋼板のグループ数と総枚数を変えることによって，3 パターンの数値実験を行った．プランニング問題の難易度の指標としてエントロピーを導入し，エントロピーを基準にして解の性能の比較を行った．また，この実験における学習の最大エピソード数は Max Episode = 20000 とした．

(2-1) グループ数 5，鋼板の総枚数 30 の場合

この設定は，プランニング問題が比較的容易な場合を想定してる．このような簡単な問題においても，提案手法が優れたプランニングを探索できるかを確認することを目的としている．数値実験 (2-1) の結果を Fig.8に示す．x 軸がエントロピーの値，y 軸が両手法によって探索されたプランニングの作業時間のステップ数を示している．

Fig.8より，全ての鋼板の組み合わせにおいて提案手法が優れた解が探索できていることが分かる．ただし，エントロピーが小さい場合は，両手法の解に大きな差は生じていない．これは，エントロピーが小さい場合には，枚数が最も多いグループから取り出せば良いという自明な解が存在し，ヒューリスティクスと提案手法の両方が自明な解を探索できているからである．

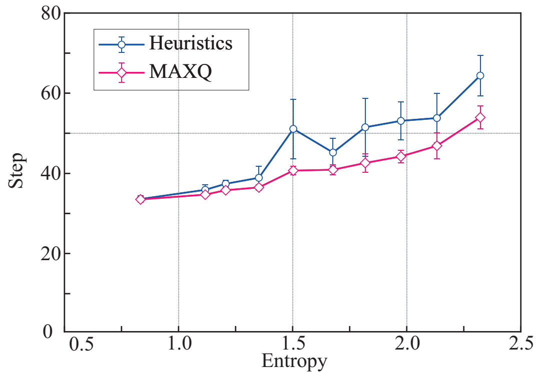

Fig. 8 The results of the costs for Simulation (2-1): The number of groups is 5 and the number of plates is 30.

(2-2) グループ数 10，鋼板の総枚数 50 の場合

この設定は，実際に鋼板ストックヤードで 1 日当たりに処理する鋼板の枚数の平均より 20% 程度少なくしている．数値実験 (2-2) の結果を Fig.9に示す．Fig.9より，全ての鋼板の組み合わせにおいて提案手法が優れた解が探索できていることが分かる．この場合も数値実験 (2-1) と同様にエントロピーが小さい場合は，解に大きな差が生まれていない．ただし，この場合はエントロピーが最大となる場合も解に大きな差が出ていない．この原因は，エントロピーが最大になる場合は，各グループの鋼板の配置が均一になり，どのグループから取り出しても作業時間が大きく変わらなくなくなるからである．しかし，実問題において，エントロピーが最小，または最大となるような鋼板の組み合わせが発生する確率はとても低いと考えられる．したがって，エントロピーが最大・最小以外の場合は，提案手法の方が優れた解を探索できるので，提案手法の優位性が確認できる．

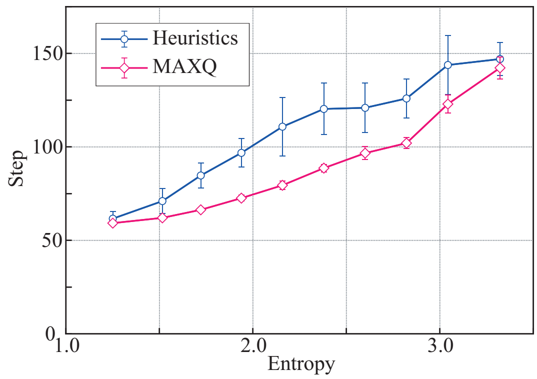

Fig. 9 The results of the costs for Simulation (2-2): The number of groups is 10 and the number of plates is 50.

(2-3) グループ数 15，鋼板の総枚数 100 の場合

この設定は，実際に鋼板ストックヤードで 1 日当たりに処理する鋼板の平均グループ数と平均枚数よりも大きくなるようにしている．したがって，この設定においても提案手法の優位性が確認できれば，実問題においても提案手法が有用であることが確認できる．数値実験 (2-2) の結果を Fig.10に示す．Fig.10より，エントロピーが最小・最大以外の場合は，全ての鋼板の組み合わせにおいて提案手法が優れた解が探索できていることが分かる．この結果より，提案手法が実問題においても有用であることが確認できる．

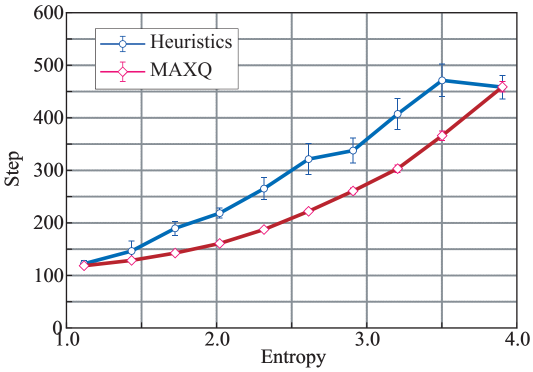

Fig. 10 The results of the costs for Simulation (2-3): The number of groups is 15 and the number of plates is 100.

(3) 実際の鋼板データへの適用

鋼板ストックヤードの実問題においても提案手法が有用であるか検証するために，実際の鋼板データを用いた数値実験を行う．造船所・鋼板ストックヤードにおける実際に 1 日で処理する鋼板データとして，木村らが用いた鋼板データ [4] を本研究では用いる．実際の鋼板データの詳細を Table 5に示す．これら 8 日分の実データにおいて数値実験を行い，実問題における提案手法の有用性を検証する．数値実験結果を Fig.11の棒グラフに示す．

Fig.11より，8 日分の全ての鋼板データにおいて提案手法が優れた解を探索できていることが分かる．この結果より，提案手法は実際の鋼板ストックヤードにおいても有用であることが確認できる．

4.5 考察

数値実験 (1) から (3) の結果より，提案手法によって鋼板搬出作業に要する時間を約 10% 以上も短縮することが可能であることを確認した．その一方で，提案手法はヒューリスティクスに比べ計算時間が長いという結果が得られた．しかし，提案手法の計算時間は処理する鋼板の枚数が 50 枚程度なら約 10 分，100 枚程度なら約 30 分であるので，実際に用いる場合には問題ならない程度である．以上の結果より，提案手法は鋼板搬出作業の作業時間を大幅に短縮でき，現場で利用できる有用なプランニング手法であることが確認できる．

また，本研究では提案手法によって得られるプランニングにどのような特徴があるのか調べるために，数値実験 (1) で得られたプランニングの解析を行った．その結果，提案手法のプランニングの特徴として，大きく分けて次の 3 つの特徴を確認した．

1 つ目は，鋼板の枚数が多いグループから搬出していくというものである．これは，枚数が多いグループから搬出していくことで，積み重なっている鋼板の枚数を減らすことができ，他のグループの鋼板を取り出す場合の板繰り回数を減らすことができるからだと考えられる．

2 つ目は，ばらばらに並んでいる鋼板をできる限り同じグループの鋼板がまとまるように並び替える傾向である．これによって，同じグループの鋼板を取り出すときに，邪魔な鋼板が少なくなり板繰り回数を減らすことできるからだと考えられる．

3 つ目は，次に搬出するグループの鋼板ができるだけ積み上がった鋼板の上の方に配置されるように，搬出中のグループの鋼板を取り出したり，邪魔な鋼板を一時ヤードに移動させるというものである．これによって，次に搬出するグループの板繰り回数が減らし，全てのグループにこの操作を行うことで鋼板搬出作業全体の板繰り回数を減らすことが可能である．

これら 3 つ特徴は，ヒューリスティクスには無く，階層型教科学習によって得られた新しい鋼板搬出規則のひとつだと考えられる．今後，鋼板のグループ数や枚数を変えてシミュレーションを行い，得られたプランニングの解析を行うことで，作業時間を大幅に短縮する鋼板搬出規則を見つけ出し，新しいヒューリスティクスを開発することも可能であると考えられる．

Table 5 The number of groups and the number of plates per group and the entropy of the real data.

	Groups	Plates	Entropy
Data1	18	204	3.55
Data2	17	96	3.81
Data3	17	84	3.84
Data4	14	83	3.19
Data5	13	80	3.32
Data6	12	77	3.13
Data7	12	73	3.11
Data8	12	49	3.31
Average	14.38	93.25	3.41

Table 6 The results of the costs for Simulation (3): Using the actual data of the steel plates in the shipbuilding yard.

	Heuristics	MAXQ	Difference
Data1	1283.9 ± 58.8	1071.8 ± 22.9	16.52 %
Data2	460.1 ± 29.0	416.5 ± 13.9	9.48 %
Data3	362.2 ± 20.6	355.6 ± 9.0	1.83 %
Data4	302.3 ± 26.0	247.3 ± 4.3	18.19 %
Data5	338.4 ± 32.0	264.3 ± 6.2	21.89 %
Data6	279.9 ± 32.2	217.4 ± 8.1	22.33 %
Data7	244.5 ± 11.1	208.0 ± 5.6	14.92 %
Data8	150.1 ± 21.6	130.5 ± 6.5	13.04 %
Average	427.7	363.9	14.78 %

5. 結論

5.1 本研究のまとめ

本研究では，鋼板搬出作業に要するコストを最小化するプランニング問題をマルコフ決定過程として新しいモデル化を行った．

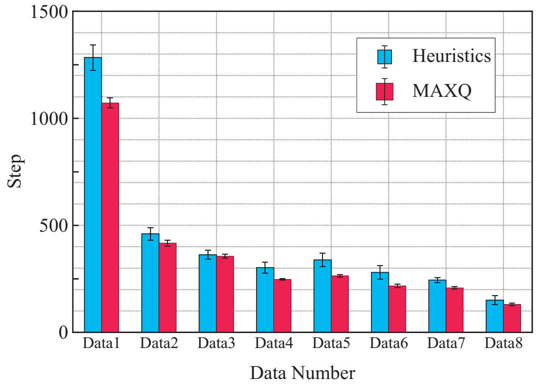

Fig. 11 The results of the costs for Simulation (3): Using the actual data of the steel plates in the shipbuilding yard.

そして，プランニング問題が有する意思決定の階層性を取り扱うために，階層型強化学習を用いたプランニング手法を開発し，数値実験を通して提案手法の有用性の検証を行った．その結果，ヒューリスティクスに比べ，提案手法を用いることによって，作業時間を約 10% 以上短縮することを可能にした．また，搬入される鋼板の枚数のばらつきを表す指標としてエントロピーという概念を導入し，あらゆる鋼板の枚数の組み合わせにおいても提案手法が優れていることを確認した．そして，実際の造船所・鋼板ストックヤードにおける鋼板データを用いた数値実験を行うことによって，実問題においても提案手法が有用であることを確認した．ただし，提案手法はヒューリスティクスに比べ計算時間が長くなるという欠点が存在する．しかし，提案手法の計算時間は，処理する鋼板の枚数が 50 枚程度なら約 10 分，100 枚程度なら約 30 分なので，実際に用いる場合には問題ならない程度である．また，計算時間の問題は，プランニング手法の利用方法を工夫することによって解決することが可能である．具体的には，勤務時間後の夜間を利用してプランニングの計算を行い，次の日の朝から鋼板搬出作業に取り掛かるといった手順を組めば，鋼板ストックヤードの現場でも十分に利用できる手法であると考えられる．

5.2 今後の課題

本研究で取り扱った問題は，提案手法の検証実験のために著者らが作成した問題であるため，実問題から多少異なる部分が存在する．今後，造船所や鋼板配送センターの鋼板搬出作業における問題調査を行い，得られた情報を参考に実問題を反映したシミュレーションを開発し，そのシミュレーションを用いて提案手法の有用性を検証することを考えている．そして，鋼板ストックヤードにおける鋼板のデータベースとリンクさせたシミュレータを開発することによって，現場で利用できるアプリケーションを開発することを考えている．また，本研究で提案したプランニング手法は鋼板ストックヤードだけではなく，コンテナヤードや工場の倉庫などにも応用できる可能性が存在する．今後は他分野への応用も目指した新しい問題設定を行い，提案手法の有用性を検証することも考えている．

参 考 文 献

1) Guangmin Wang, Chun Jin, Xiaoyi Deng: Modeling and optimization on steel plate pick-up operation scheduling on stackyard of shipyard, Proceedings of the IEEE International Conference on Automation and Logistics (ICAL 2008), pp.548-553, 2008

2) 荒井誠，西原弘之：造船業におけるクレーン問題への遺伝的最適化アルゴリズムの応用，日本造船学会論文集，No.196，pp.1-7，2004

3) 阿瀬始，勘定義弘：鋼材ヤードにおける鋼板搬出スケジューリング，オペレーションズ・リサーチ：経営の科学，48(7)，pp.522-524，2003

4) 木村元，中尾洋一，梶原宏之：造船所鋼板ストックヤードにおける仕分け作業の効率化，西部造船会会報 No.109，pp.185-193，2005

5) R. S. Sutton, A. G. Barto: Reinforcement Learning: An Introduction, Bradford Books, 1998

6) D. P. Bertsekas, J. N. Tsitsiklis: Neuro-Dynamic Programming, Athena Scientific, 1996

7) R. S. Sutton: Generalization in Reinforcement Learning: Successful Examples Using Sparse Coarse Coding, Advances in Neural Information Processing Systems 8, pp.1038-1044, 1996

8) A. G. Barto and Sridhar Mahadevan Recent Advances in Hierarchical Reinforcement Learning, Discrete Event Dynamic Systems archive, Volume 13, Issue 4, pp.341-379, 2003

9) T. G. Dietterich: Hierarchical Reinforcement Learning with the MAXQ Value Function Decomposition, The Journal of artificial intelligence research, volume 13, pp.227-303, 2000

10) T. G. Dietterich: An Overview of MAXQ Hierarchical Reinforcement Learning, Proceedings of the 4th International Symposium on Abstraction, Reformulation, and Approximation, pp.26-44, 2000

11) 今井秀樹，情報・符号・暗号の理論，コロナ社, 2004