日本船舶海洋工学会講演会論文集　（この行は学会にて記入します）

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

学生員　白　石　耕一郎*　　正　員　木　村　　元**

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

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

Key Words: *Delivering of Steel Plates, Rehandling of Steel Plates, Multiple Decision Problem, Markov Decision Process, Hierarchical Reinforcement Learning*

1. 序　　論

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り，これが鋼板搬出作業を停滞させる一因となっている．板繰り量は，取り出す鋼板の順序によって大きく変動するため，板繰り作業を考慮した適切な鋼板の搬出順序を見つけ出すことによって作業時間の短縮が可能である．しかしながら，板繰り量を最小にする鋼板の搬出順序を見つけ出すことが可能なプランニング手法は，著者が知る限り，確立されていない．なぜなら，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 [1,2]やタブーサーチ [3]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だからである．そこで，本研究では，鋼板搬出作業の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ることによって有用な解を求める新しいアプローチを提案する．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．

2. 鋼板ストックヤードにおける鋼板搬出作業

本研究で対象とする鋼板ストックヤードでは，各鋼板は使用される工程や集配日に応じてグループ化され，その「グループ(Group)」ごとに適切な順序で次の工程へ出庫される．しかし，各鋼板は工程計画で定められた順序ではなく，ばらばらに入庫されるため，出庫時に仕分け作業が必要となる．また，その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，板繰りによって必要な鋼板を取り出す作業が生じる．この板繰り作業が，鋼板搬出作業に要する時間に大きく影響している．したがって，鋼板搬出作業の時間短縮を行うには，板繰りの作業量を最小化する鋼板の取り出し順序を見つけ出すことが可能な鋼板搬出作業のプランニングが必要である．本研究で取り扱う鋼板搬出作業における入庫と出庫の流れを Fig.1 に示す．また，鋼板ストックヤードにおいて，鋼板を 1 ヶ所に積み重ねて保管する場所をヤードと呼ぶ.

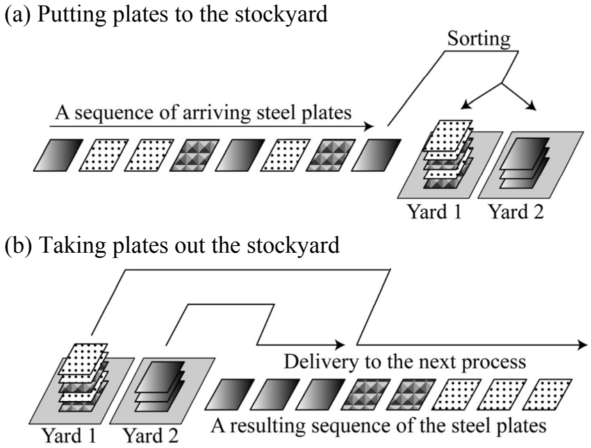

Fig. 1: An outline of the stockyard.

2. 1 入庫作業

入庫した鋼板を保管するヤードが 1 ヶ所であれば，そのヤードに鋼板を積み重ねていけばよいが, 2 ヶ所以上ある場合は，鋼板を分けて保管することも可能となる．本研究では，そのような場合のヤードの仕分け方法として，木村らの方法 [4]を用いる．この方法は，LPT 則(Longest processing time rule)と呼ばれる経験則的方法に従って，グループごとに鋼板を仕分けるというものである．この方法によって, 到着した順に鋼板を積み上げる場合に比べ 4 割程度の板繰り量の低減が可能である．また, 1 つのグループの鋼板が複数のヤードに分かれて保管されるということは無いので, 1 つのヤードだけに着目して鋼板搬出作業を行うことができる.

2. 2 出庫作業

鋼板を出庫する場合，どのグループの鋼板から出庫するのか，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 3 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，最上層が出庫する鋼板のグループを定める出庫グループ(Delivery group)の決定，中間層が出庫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

---

*　九州大学大学院 工学府 海洋システム工学専攻
**　九州大学大学院 工学研究院 海洋システム工学部門

原稿受付　(学会にて記入します)
秋季講演会において講演　(学会にて記入します)

くヤードを定める一時ヤード(Temporary yard)の決定となる．そして，全ての鋼板が入庫された状態(出発点)から，全ての鋼板を次の工程へ出庫した状態(目的点)へ至るまでの階層的な意思決定を，作業時間が最小となるように定めることが鋼板搬出作業のプランニング問題となる．

2.3 鋼板搬出作業のプランニング問題の定式化

プランニング問題の定式化のために，各階層における意思決定を次のように定める．出庫グループの順序を $G=\{g_1,g_2,\cdots,g_l\}$ 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．出庫グループが 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，出庫グループが 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出庫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．

プランニングで考慮するコスト関数 $Cost(g_i,y_j,y_k)$ を出庫グループが 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時間を 1 ステップとして考え，この作業時間をコストとする．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，エネルギー消費を考慮した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

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

Subject to

$$g_i \in G,\ y_j \in Y_d(g_i),\ y_k \in Y_t(g_i,y_j) \tag{2}$$

鋼板搬出作業のプランニング問題は，コスト関数の総和(1)を最小にするような，出庫グループの順序 G，取出しヤードの順序 Y_p，一時ヤード Y_t の順序を決定する最適化問題である．この最適化問題を解くことによって，作業時間が最小となるプランニングを求めることができる．しかし，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

本研究における鋼板の取り出し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．

<u>鋼板搬出規則</u>

(1) 出庫グループの全ての鋼板の出庫が完了するまで，出庫グループ以外の鋼板は出庫しない．
(2) 取出しヤードに存在する出庫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出庫しない．
(3) 取出しヤードの最上段の鋼板が出庫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作業時間を最小にするものを見つけ出す．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

3.1 マルコフ決定過程としてのモデル化

マルコフ決定過程[5]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おいて，状態を観測しながら最適な行動を決定するための数学的手法である．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．

3.2 プランニング問題のモデル化

鋼板搬出作業の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，本問題で取りう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，全ての実現可能な状態遷移は確率 1 で行われる．報酬は，前章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，行動価値関数 $Q(s,a)$ 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．本問題では，現在の状態を観測し，その状態で最も良いと考えられるヤードへ鋼板を移動させる．したがって，マルコフ性が成立するので，本問題はマルコフ決定過程としてモデル化できる．

3.2 動的計画法

マルコフ決定過程では，最適な行動価値関数 $Q^*(s,a)$ において次式の Bellman 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 方程式は，動的計画法[5,6]を用いることによって効率良く解くことが可能である．

3.3 強化学習

動的計画法は，本問題のような高次元問題に用いる場合，状態数が指数関数的に増大し，計算が非常に困難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，モデルフリーの動的計画法である強化学習[5,6]を適用する．モデルフリーの場合，逐次的に状態と報酬を観測することで状態遷移確率を推定するため，全状態における状態遷移確率や報酬の保持が不要である．そのため，計算に必要な記憶量を大幅に減少させることが可能である．そして，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．また，強化学習における行動選択手法として，ボルツマン行動選択手法を用いる．この手法では，次式で表されるボル

ツマン分布に従って，確率的に行動を選択する．

$$p(a \mid s) = e^{Q(s,a)/T} \Big/ \sum_{a_i \in A} e^{Q(s,a_i)/T} \qquad (4)$$

$p(a \mid s)$ は状態 s で行動 a を選択する確率，T は温度パラメータである．T によって行動選択を価値関数にしたがってグリーディに行うか，ランダムに行うかを調整する．

3．4 関数近似

強化学習を用いる場合でも，Bellman 方程式を解くためには，全状態の価値関数を保持する必要がある．状態空間が大きく，全状態における価値関数の保持が困難な場合は，何らかの関数近似を用いて価値関数を保持する必要がある．そこで，関数近似の方法として線形アーキテクチャ(Linear architecture)[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，ヤードに積み重なった鋼板の枚数と同じ数の要素をもち，鋼板の順序と要素の順序を対応させた特徴ベクトルを鋼板のグループごとに生成する．そして，各要素の対応する位置に，対象とする特徴ベクトルのグループの鋼板が存在する場合には，その位置の要素に 1 を入力し，そうでない場合は 0 を入力する．この操作によって，各グループの鋼板配置の特徴量を抽出した特徴ベクトルを生成する．

3．5 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問題となるため，学習するのに非常に時間がかかってしまう．そこで，本研究では，複雑なタスクを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．具体的には，各階層における価値関数の表現方法として Dietterich が提案する MAXQ[8]という手法を用いる．そして，強化学習アルゴリズムとして MAXQ と Q-Learning と組み合わせた MAXQ Q-Learning[8]を用いる．MAXQ Q-Learning は最適解への収束が証明されている．また，アルゴリズムも単純であるので実装が比較的容易であるという特徴をもつ．

4．数値実験

4．1 プランニングの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出庫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 $C_r(g_i, y_j)$ 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．

$$C_r(g_i, y_j) = H(y_j) - I(g_i, y_j) - N(g_i, y_j) + 1 \qquad (5)$$

$H(y_j)$ 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．ヒューリスティクスでは，この板繰りコスト関数を用いて，次式で表される規則に従って，出庫グループ g_d，取出しヤード y_p，一時ヤード y_t を決定する．

<u>出庫グループの決定</u>

$$g_d = \min_i \sum_j C_r(g_i, y_j) \qquad (6)$$

<u>取出しヤードの決定</u>

$$y_p = \max_j N(g_d, y_j) \qquad (7)$$

<u>一時ヤードの決定</u>

$$y_t = \min_k H(y_k) \qquad (8)$$

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出庫する．そして，出庫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．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，出庫グループ，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 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(Proposed Method)と，簡易なヒューリスティクスを用いた方法(Heuristics Method)を比較することによって提案手法の有用性を検証する．プランニング問題はグループ数や鋼板の枚数によって，問題の難しさが変わってくる．本研究では，グループ数を固定し，各グループの鋼板の枚数を変えることによって，3 種類の数値実験を行った．全ての数値実験において，入庫される鋼板は全て 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，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 2，3 となっても，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 と設定した．

また，提案手法には確率的要素が含まれるため，探索ごと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，同じ初期配置での解の探索を 10 回行い，その平均を結果としている．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解の変動を考慮するために，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 種類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している．

各数値実験で用いた鋼板のグループとその枚数の関係を Table1 に示す．また，強化学習に関するパラメータは，学習率を $\alpha = 0.1$，割引率を $\gamma = 1.0$，学習における最大のエピソード数を Max Episode = 30000 とした．計算機環境は，CPU: Pentium 4 3.0GHz, Memory: 2038MB, OS: Windows XP, プログラム言語: JAVA となっている．

Table1: The number of plates per group.

Experiment	A	B	C	D	E	F	G	H	I	J
1	5	5	5	5	5	5	5	5	5	5
2	1	2	3	4	5	6	7	8	9	10
3	5	6	7	8	9	10	11	12	13	14

4.3 実験結果

(1) 数値実験 1: 鋼板枚数 50 枚

この実験は，各グループの鋼板枚数が均等な場合における提案手法の性能を調べるために行った．Table2 に数値実験 1 の結果を示す．Table2 より，提案手法は，ヒューリスティクスに比べ，作業時間を約 10% 短縮している．鋼板の初期配置による解のばらつきを考慮しても，提案手法の方が十分に優れた解を得ていることが分かる．またヒューリスティクスは，得られる解の標準偏差が大きいことから，得られる解の質が鋼板の初期配置によって大きく変化していることが確認できる．

Table2: Results of Experiment 1.

Method	Cost [step]	Calculation Time
Heuristic	151.1±5.11	0.075±0.021 [sec]
Proposed	138.39±1.6	17.23±0.73 [min]

(2) 数値実験 2: 鋼板枚数 55 枚

この実験は，各グループの鋼板枚数にばらつきがある場合における提案手法の性能を調べるために行った．Table3 に数値実験 2 の結果を示す．Table3 より，提案手法は，ヒューリスティクスに比べ，作業時間を約 20%短縮している．また，提案手法とヒューリスティクスの解を数値実験 1 と数値実験 2 で比較すると，数値実験 2 の作業時間の方が長くなっている．これは，グループ数の鋼板枚数にばらつきがある方が，板繰り作業が増える傾向があるからだと考えられる．

Table3: Results of Experiment 2.

Method	Cost [step]	Calculation Time
Heuristic	173.3±15.3	0.071±0.014 [sec]
Proposed	143.6±1.9	17.66±0.14 [min]

(3) 数値実験 3: 鋼板枚数 95 枚

この実験は，種類ごとの鋼板枚数にばらつきがあり，さらに取り扱う鋼板枚数が 95 枚とすることで，大規模問題における提案手法の性能を調べるために行った．Table4 に数値実験 3 の結果を示す．Table4 より，提案手法はヒューリスティクスに比べ，作業時間を約 20% 短縮している．この結果より，提案手法は，処理する鋼板が 100 枚近くなる大規模問題おいても，十分な性能を発揮できることが分かる．ただし，処理する鋼板の枚数が増えるにつれて，学習に要する時間も長くなるため，実際の現場で用いるためには，計算時間を考慮した適切な利用方法の考案が必要である．

Table4: Results of Experiment 3.

Method	Cost [step]	Calculation Time
Heuristic	383.30±15.86	0.089±0.054 [sec]
Proposed	313.07±4.67	62.25±6.5 [min]

4.4 考察

数値実験を通して，提案手法によって鋼板搬出に要する時間を最低でも約 10% 短縮することが可能であることを確認した．さらに，提案手法によって得られた解の標準偏差が小さいことから，鋼板の初期配置に依存せずに，安定した性能を発揮することが分かる．また，提案手法は鋼板の枚数が多く，グループごとの鋼板枚数にばらつきがあり，解くのが困難な問題であるほど，ヒューリスティクスより優れた解を探索できるという特徴をもつことを確認した．

提案手法は，計算時間が長いのが欠点であるが，鋼板搬出作業では一日の作業量を基準に考えるので，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. 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時間を最小化する鋼板搬出作業の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．また，本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．

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通して提案手法の有用性を検証することを考えている．最終的には，現場で利用できるアプリケーションの開発に取り組むことを考えている．

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) 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

3) 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

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, MIT Press, 1996

8) 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