行動蓄積過程に基づく行動知識の獲得と 行動ネットワークによる知識表現* (2次元積み合せ問題への適用)

諏訪 晴彦^{*1}, 荒木 奈美^{*2}
舩岡 征浩^{*3}, 藤井 進^{*4}

Heuristics Acquisition from Problem Solving Process and Knowledge Representation by State-Action Network (Application to Two-Dimensional Packing Problems)

Haruhiko SUWA^{*5}, Nami Araki,
Masahiro MASUOKA and Susumu FUJII

^{*5} Department of Industrial and Systems Engineering, Setsunan University,
17-8 Ikedanaka-machi, Neyagawa-shi, Osaka, 572-8508 Japan

This paper proposes a new model and method for knowledge acquisition for planning problems in manufacturing, such as production planning, packing problems, and so forth, based on a problem solving process of a planner. In our proposed model, knowledge used for solving the objective problem is obtained through construction of *state-action network*. A state-action network consists of states describing some phases of the problem solving process and planner's actions in each phase. Moreover, reinforcement learning is used to refine the obtained state-action network and to evaluate the problem solving process. We apply the proposed method to two-dimensional packing problems and demonstrate its applicability and effectiveness through some computational experiments.

Key Words: Knowledge Acquisition, Reinforcement Learning, Two-Dimensional Packing Problem, Heuristics

1. はじめに

近年の生産の場では、工業技術やIT (Information Technology) の急速な進歩により、機械による作業の無人化・自動化が実現可能な場面が増えつつある。作業の無人化・自動化は、作業時間の短縮や加工精度の向上、事故率の減少などに大きく貢献している。しかし一方では、依然として熟練技能者の経験的知識に依存する局面が数多く残されているのも現状である⁽¹⁾。この局面は、人間による製造が絶対不可欠の場合を除き、人間の意思決定でのみ仕様を満足できる場合と、無人化・自動化は可能であるがコスト面に難がある場合に大別できる⁽¹⁾⁽²⁾。通常の製造プロセスにおいて

は、前者に関する知識を技能(暗黙的な知識)、後者に関する知識を技術(明示的に表現可能な知識)と捉えられることができる。これらの知識は必ずしも互いに独立するものではなく、重複する部分が多々あると考えられる⁽²⁾⁽³⁾。この重複した部分は、技能から技術への置換が可能であると考えられる領域であり⁽²⁾、人工知能分野における知識ベースシステムがその代表的なものといえる。知識ベースシステムは、生産管理の実用レベルにおいて数多く開発されており⁽⁴⁾⁽⁵⁾、特に長期/中期の生産計画や、日単位の生産スケジューリングを対象とした知識ベースシステムが実用化されている^{(6)~(8)}。生産場面における知識ベースシステムは、一般に生産や管理の効率性向上、計画者の負担を軽減するなどの役割を担う。しかしながら、知識獲得のボトルネックや環境変化にともなう知識ベースの更新など、知識抽出の不確実性やメンテナンスの負担が課題となっているのが現状である。

本研究では、知識獲得の新たなアプローチとして、計画者の試行錯誤の知識(行動知識)蓄積プロセスに内在する行動の特性を定量的に表現することにより、

* 原稿受付 2000年12月27日。

^{*1} 正員、摂南大学工学部(☎572-8508 寝屋川市池田中町17-8)。

^{*2} (株)富士総合研究所(☎134-8609 東京都江戸川区西葛西3-22-36 西葛西センター富士総研)。

^{*3} CTC伊藤忠テクノサイエンス(株)(☎532-0011 大阪府淀川区西中島6-1-1)。

^{*4} 正員、神戸大学工学部(☎657-8501 神戸市灘区六甲台町1-1)。

E-mail: suwa@isc.setsunan.ac.jp

計画者の知識を自動的に獲得することを試みる。ここで述べる行動とは、計画者が対象とする問題の状態とそれに対する計画者の行為の組であるものとする。計画者は行動の試行錯誤の段階で、同じ状況において様々な行動を取り、有用な知識を得た後でも常に同じ行動をとるとは限らない。また、表現可能な状態の各々が独立しているのではなく、過去や将来の状態を考慮して現在の状態における適切な行為を選択すると考えられる。本研究ではこのような観点から、行為と状態が有機的に結合したネットワークにより、計画者の知識を表現することを考える。以下では、このような計画者の行動知識を表現するネットワークを行動ネットワーク (state-action network) と呼ぶこととする。

ここでは、幾何的情報に基づく視覚的な判断の容易性に着目して、2次元積み合せ問題⁽¹²⁾⁽¹³⁾を取り上げる。計画者の計画立案の過程から、それに内在すると考えられる知識を獲得し、行動ネットワークとして表現する。本研究では、積み合せ計画に対する行動の蓄積が可能な実験システムをコンピュータ上に構築し、行動ネットワークの獲得を行うものとする。また、数値シミュレーションにより、獲得した行動ネットワークを問題に適用し、獲得知識の有用性を検討する。

2. 行動知識の獲得

2.1 ネットワークによる知識表現 本研究では、対象計画問題が計画者に与えられたとき、計画者の意思決定を「行為」、行為の対象となる計画状況を「状態」と定義する。したがって計画者の行動とは、ある計画状況における計画者の意思決定という事例の一つに相当する。ここで、対象問題に起こり得る状態 s すべてを含む状態の表現空間を S 、可能な行為 a すべてを含む行為の表現空間を A とする。また、ある状態 $s \in S$ での可能な行為の集合を $A(s) (\subseteq A)$ とする。このとき、対象問題に対する計画者の行動を $(s, a) (s \in S, a \in A(s))$ で表す。

状態および行為の記述は、対象問題からの特徴抽出の作業（以下、抽象化）であり、人間によって事前に与えておかねばならない。対象問題の全ての特徴を列挙するのは極めて困難な作業であり、抽象化は必然的に情報の劣化をとともなう。一方、抽出すべき特徴が少なすぎると、状態と行動の記述を過度に一般化する恐れがある。そこで、本研究では状態および行動の記述の汎用性を重視し、対象問題のモデル化（数理モデル、シミュレーションモデルなど）する際に必要となる基本的な情報のみで抽象化を行うものとする。

状態 s および行為 a は、各々を特徴づける m_s 個

の状態属性 $X_1^s, X_2^s, \dots, X_{m_s}^s$ 、および m_a 個の行為属性 $X_1^a, X_2^a, \dots, X_{m_a}^a$ により表現されるものとする。状態属性 X_i^s ($i = 1, \dots, m_s$) および行為属性 X_j^a ($j = 1, \dots, m_a$) のとりうる値をそれぞれ、状態属性値 v_i^s 、行為属性値 v_j^a と呼ぶこととする。このとき、計画者のある行動 (s, a) は、

$$(v_1^s v_2^s \cdots v_{m_s}^s, v_1^a v_2^a \cdots v_{m_a}^a)$$

と表現される。所与の計画問題について、計画者が m 回の行動で実行可能な計画を立案した際、初期状態を s_1 として、 m 個の行動事例 $(s_1, a_1), (s_2, a_2), \dots, (s_m, a_m)$ ($s_1, \dots, s_m \in S, a_1, \dots, a_m \in A$)、および最終の状態 s_{m+1} が得られるものとする。ただし、失敗による行為のやり直し（バックトラック）は考慮しないものとする。行動事例を獲得するプロセスは、計画者からの行動蓄積のプロセスに相当するものであり、各行動事例には、「状態 s_k ($k = 1, 2, \dots, m$) での行為 a_k の結果、状態 s_{k+1} となる」という因果関係が存在する。また、各行動 (s_k, a_k) について、計画者により何らかの満足度 r_k ($\in R$) が与えられ、満足度に基づいた評価関数により行動の評価が与えられるものとする。行動の評価については次節で述べる。

本研究では、状態を表すノード（状態ノード）の集合 $S_N (\subseteq S)$ 、および行為を表す有向アーク（行為アーク）の集合 $A_N (\subseteq A)$ で表される状態遷移のネットワーク $N(S_N, A_N)$ を行動ネットワークと呼ぶこととする。行動ネットワークは、上述の計画者の行動事例に基づいて獲得される。

2.2 強化学習による行動の評価 行動の評価は、計画者の行動特性をネットワークに反映させる役割を持つ。本研究では、強化学習⁽¹⁰⁾に基づいて行動に定量的評価を与えることを考える。強化学習は、未知なる環境における適切な行為を報酬と経験から獲得する機械学習の一つである⁽¹⁰⁾。強化学習を実現する様々な方法が存在するが、ここでは、状態-行為の組そのものを評価することが可能であると考えられる Q-learning⁽¹⁰⁾⁽¹¹⁾を利用することとする。

ある状態 $s \in S$ での行為 $a \in A(s)$ に対する評価関数 $f(s, a)$ を次の (1) 式で定義する。

$$f(s, a) = Q(s, a) \times R(s, a) \quad (1)$$

$Q(s, a)$ は状態 s で計画者の行為が a であったときの Q 値⁽¹¹⁾を、 $R(s, a)$ は行動 (s, a) の重みを表す。本研究では、 $R(s, a)$ は計画者が状態 s で a を選択した回数（行為の経験回数）を表すものとする。なお、計画

Fig. 1 Framework of knowledge acquisition

者にとって行動事例 (s, a) が初めての試み（経験）である場合は、 $Q(s, a) = \alpha \times r$, $R(s, a) = 1$ とする。 r は行動 (s, a) に対する計画者の満足度を表す。

行動 (s, a) によって次の状態 $s' (\in S)$ に遷移したとき、 $Q(s, a)$ および $R(s, a)$ は次の (2), (3) 式にしたがって更新される。

$$Q(s, a) \leftarrow (1 - \alpha)Q(s, a) + \alpha(r + \gamma \max_{b \in A(s')} Q(s', b)) \quad (2)$$

$$R(s, a) \leftarrow R(s, a) + 1 \quad (3)$$

ただし、 α ($0 \leq \alpha \leq 1$) は学習率、 γ は次の状態 s' での Q 値（の見積もり）の最大値に対する割引率を表す。(1) 式では、 Q 値のみ（計画者の満足度のみ）の行動評価が考えられるが、行為の経験数を Q 値の重みとして位置付けることによって、評価に柔軟性を持たせることが可能となる。

2.3 知識ネットワーク獲得の枠組み 本研究で提案する計画者の知識獲得プロセスの枠組みを図1に示す。まず、事例生成機構（Case Base Unit）では、対象問題の属性情報（problem information）および計画者の意思決定から、行動事例 $(s_1, a_1), (s_2, a_2), \dots$ を生成する。これと同時に、各行動に対する計画者の満足度を記憶する。次に、ネットワーク構築機構（Network Construction Unit）に各行動事例を逐次的に入力し、対象計画者の行動ネットワークを獲得する。

行動ネットワークの基本的な構造は、所与の行動事例に対して、図2(a)に示すようなフロー型をとる。全ての状態の記述が異なり ($s_j \neq s_{j+1}, j = 1, \dots, m$)、かつ各状態での計画者の行為が一意に定まる場合のみ、図2の行動ネットワークが獲得される。しかし、対象問題の特徴を完全に記述することは一般に困難であ

り、計画者の行為は常に一つではないと考えられるため、一つの状態に複数の行為が属することとなる（図2(b)参照）。図2(b)では、行動事例 (s_k, a_k) を現在の行動ネットワークに入力する際、 s_k と等価な状態 $s_{k'}$ ($k' \in \{1, 2, \dots, k-1\}$) が既に存在するため、状態 $s_{k'}$ に行為 a_k を追加した例を表している。この場合、状態 $s_{k'}$ において2つの行為 $a_{k'}, a_k$ が存在することとなる。

行動ネットワーク獲得の手続きを以下に示す。

Step 1 $S_N = \phi, A_N = \phi, j \leftarrow 1$.

Step 2 行動事例 (s_j, a_j) について、 s_j と一致する状態ノード $s \in S_N$ を検索する。存在する場合は Step 3 へ、そうでない場合は、

$$S_N \leftarrow S_N \cup s_j, A_N \leftarrow A_N \cup a_j$$

とし、 $f(s_j, a_j)$ を更新する。

Step 3 状態ノード s において、 a_j と等価な行為アーク $a \in A_N(s)$ を検索する。行為アーク a が存在する場合は $f(s, a)$ を更新した後 Step 4 へ、存在しない場合は、

$$A_N \leftarrow A_N \cup a_j$$

とし、 $f(s, a_j)$ を更新する。

Step 4 $j = m$ ならば、

$$S_N \leftarrow S_N \cup s_{m+1}$$

として、終了する。そうでなければ、 $j \leftarrow j+1$ とし Step 2 へ。

学習機構（Learning Unit）は、上述の Q-learning アルゴリズムを利用して、各行動を評価し行動ネットワークを強化する。最終的に獲得した行動ネットワークは、知識ベース（Knowledge Base）に蓄えられ実際の問題解決に利用することとなる。

最後に、行動ネットワークの特徴を示す。

- 計画者から得られる行動事例を逐次的に行動ネットワークに組み込んでいくことにより、計画者の各意思決定場面での判断結果を把握することが可能となる。
- 一つの状態で常に同じ行為を行うとは限らない、という人間の行動を表現することが可能である。
- 状態ノードを時系列に結合していくため、時間的な状態-行為の因果関係を把握することが可能である。

Fig. 2 Construction of action-state network

3. 2次元積み合せ問題への適用

本研究では、幾何情報により視覚的な判断で行動がとりやすいと考えられる2次元形積み合せ問題を取り上げる。計画者による積み合せの計画立案の行動を蓄積し、立案過程に存在すると考えられる行動知識の獲得を試みる。

3.1 2次元積み合せ問題 2次元積み合せ問題は、ある空間的広がりの中に与えられた有限個数の矩形をもっとも効率よく積み合わせることを目的としている。このような問題は、所定のパレットやコンテナに貨物を積み込む場合や、所定の大きさの板材から種々の形状の部材を切り出す場合に生じる。ここでは、2次元積み合せ問題を以下のように定義する⁽¹³⁾。 n 個の矩形 $p_1, p_2, \dots, p_n$ (以下、矩形リスト $P = (p_1, p_2, \dots, p_n)$) を、それぞれが重なることなく幅 K 、高さが無限大のパレットに積み合わせるものとする。矩形 p_i ($i = 1, 2, \dots, n$) の短辺および長辺をそれぞれ a_i, b_i ($0 < a_i \leq b_i \leq K$) とする。パレットは横方向 $[0, K]$ 、縦方向 $[0, \infty)$ の2次元座標空間で表わすものとする。パレットの左下頂点を原点とし、各矩形の位置はパレットの原点方向に位置する頂点の座標 (x_i, y_i) で表わすものとする。また、所与の矩形リスト P の積み合せ計画のある段階で、パレットに配置した矩形の集合を $P_r (\in P)$ 、未配置の矩形の集合を $\bar{P}_r (= P \setminus P_r)$ とする。

このとき、2次元積み合せ問題を「積み合せの高さ L を可能な限り低くするように、各矩形の位置座標と向き（横あるいは縦）を決定する問題である」と定義する。図3に矩形リストの可能な積み合せ例を示す。

矩形リストの最適な積み合せを見つけることはNP困難であることが知られている⁽¹²⁾。したがって、分枝限定法などの厳密解法により実時間で最適解を導出することは困難であり、様々な近似解法が報告されている。2次元積み合せ問題に対する有効なヒューリスティック・アルゴリズムとして、NFDH (Next-Fit Decreasing Height)⁽¹²⁾、FFDH (First-Fit Decreasing Height)⁽¹²⁾、あるいは、UD (Up-Down)

Fig. 3 Example of a possible packing of a list $(p_1, p_2, p_3, p_4, p_5, p_6)$

アルゴリズム⁽¹³⁾などが存在する。

3.2 実験システムの概要 図4に、本研究で構築した実験システムのインタフェースを示す。実験システムは、コンピュータ上に積み合せ問題に対して視覚的な判断で行動をとることが可能である。本実験システムを用いて積み合せ問題を行う被験者を擬似的な「計画者」と呼ぶこととする。計画者は、矩形をパレットに可能な限り低く積み合わせることを目的とする。本システムでは、計画者にとってその目的を視覚的・直感的に理解しやすいように、パレットに目標レベル（図4参照）を表示させている。

本実験システムにおける計画者の可能な操作は、矩形の移動 (MOVE)、パレットの縦方向への積み合せ (MOVE Y)、パレットの横方向への積み合せ (MOVE X)、および矩形の向きの変更 (HENKAN) の4種類である。システムの出力データは、計画者の積み合せ計画過程における矩形配置状態、矩形の移動履歴と最終の配置座標である。矩形の積み合せは、パレットの左下方向から詰めていくものとし、積み合せ途中でのやり直し (バックトラッキング) は許可するものとする。

3.3 矩形の配置状況の抽象化 2.3で述べたように、状態と行動はそれぞれ、有限個数の状態属性値、行為属性値の列で表現する。ここでは、計画者の感覚入力として最も重要であると考えられる視覚情報に着目する。矩形配置状況の抽象化を行い、この抽象化した幾何的情報を行動ネットワークの状態として定量的に表現する。

所与の2次元積み合せ問題において、パレット幅 K の最小分割単位を K/M とする。ただし、 M は P 内の矩形短辺の最大公約数とする。パレットの分割幅 K/M に基づいて、パレットの横座標を $(jK/M, 0)$ ($j = 0, 1, \dots, M$) で表すものとし、積み合せ計画の

Fig. 4 Experiment system for two-dimensional packing

過程において各分割区間 $[(j-1)K/M, jK/M)$ ($j = 1, 2, \dots, M$) の高さを y'_j ($y'_j \geq 0$) とする. このとき, $j = 1, 2, \dots, M$ に対して $y_j = y'_j - \min_k \{y'_k\}$ としたとき, $(y_1, y_2, \dots, y_M)$ をパレットの抽象化形状と定義する. また, 抽象化形状における段数 n_y を次の (4) 式で表すこととする.

$$n_y = \sum_{j=1}^{M-1} u_j \quad (4)$$

$$u_j = \begin{cases} 0 & \text{if } (y_j = y_{j+1}) \\ 1 & \text{if } (y_j \neq y_{j+1}) \end{cases} \quad (5)$$

図 5(a) に, 3つの矩形 p_1, p_2, p_3 を幅 K のパレットに積み合せた場合の抽象化形状の抽出の一例 ($M = 4$) を示す. 図 5(a) の例では, $y'_1 = y'_2 = y'_3 = 12, y'_4 = 5$ であり, 抽象化形状は $(7, 7, 7, 0)$ で表され, 抽象化形状の段数 $n_y = 1$ となる.

抽出した抽象化形状における矩形の配置領域を A_p ($= (K/M) \sum y_j$), 未配置領域を A_o ($= \max\{y_j\}K - A_p$) とする. また, $(A_p + A_o)$ の領域を $(\bar{j}K/M, 0)$ ($\bar{j} = [M/2]$) で縦に分割したとき, その領域の左半面および右半面での配置領域をそれぞれ, A_p^l, A_p^r で表すこととする. ただし, $[*]$ は $*$ を越えない最大の整数とする. 抽象化形状において, 未配置矩形を配置することが可能な場所の基準となる座標を, パレット横方向について原点に近いものから, $q(1), q(2), \dots, q(v)$ で表し, これらを配置可能点と呼ぶこととする. 図 5(a) での矩形の配置状況に対する各領域 A_p, A_o, A_p^l , および A_p^r を図 5(b) に示す. このとき配置可能な箇所は $q(1)$ と $q(2)$ であり, 抽象化形状では, 矩形の頂点を $q(1)$ ないしは $q(2)$ に一致させて配置させることとなる.

3.4 状態と行為の表現 本研究では, パレットの抽象化形状および矩形属性から得られる情報に基づい

(a) Extraction of Shape in Packing

(b) Extracted Shape

Fig. 5 Example of "shape" extraction

て, 図 6(a) に示す 4つの状態属性 $X_1^s \sim X_4^s$ を考える. X_1^s は段数に基づく形状の複雑さに関する特徴, X_2^s および X_3^s は, 未配置領域の大きさとその位置に関する特徴, X_4^s は未配置矩形に関する特徴を表現している. 一方, 行為属性は「どの矩形をどの向きでどの場所に配置したか」を表現することが最小限必要である. そこで, 矩形の形状, 配置箇所および未配置領域などの基本情報により, 未配置の矩形 $p_w (\in \bar{P}_r)$ を選択したときに, 図 6(b) に示す 4つの行為属性 $X_1^a \sim X_4^a$ を与えるものとする. これらの行為属性では, $X_3^a = X_4^a = 0$ の場合を除き, 計画者の選択した矩形の形状および向きを表現することが可能である. そこで, $X_3^a = X_4^a = 0$ となる場合には, 図 6(b) に示すように $X_5^a \sim X_9^a$ に示す補助的な行為属性を与えるものとする.

4. 数 値 実 験

4.1 行動ネットワークの獲得 以下に示す問題例 P1 について, 二人の計画者 A, B に積み合せ計画を行わせた.

問題例 P1:

矩形数 n : 20

長辺 a_i / 短辺 b_i : $10 \times U(1, 10)$

パレット幅 K : 150 (分割数 $M = 15$)

なお, $U(\alpha, \beta)$ は $[\alpha, \beta]$ ($\alpha \leq \beta$) の一様乱数であり, 各矩形の配置箇所を決定した際の計画者の満足度 r は, 最も満足した場合を 1 として, $[-1, 1]$ の実数値で与

X_1^s	$0 (n_v < k) \text{ or } 1$
X_2^s	$0 (A_p < A_o) \text{ or } 1$
X_3^s	$0 (A_p' < A_p'') \text{ or } 1$
X_4^s	$0 (A_o < a) \text{ or } 1 \quad a = \sum_{p_i \in P_i} a_i b_i / \bar{P}_i $

(a) State Attributes

$X_1^a = \{00, 01, 10\}$ (placed on $q(1), q(n_v)$ or otherwise)

$X_2^a = \{00, 01, 10\}$ (placed on $y_{\min}, y_{\max}$ or otherwise)

$X_3^a = \{0, 1\}$ (fit on the width of A_o , or not)

$X_4^a = \{0, 1\}$ (fit on the height of A_o , or not)

$X_5^a = \{0, 1\}$ (sideways, or not)

$X_6^a = \{0, 1\}$ ($b_w < \text{width of } A_o$, or not)

$X_7^a = \{0, 1\}$ ($b_w < \sum b_i/n$, or not)

$X_8^a = \{0, 1\}$ ($a_w < \sum a_i/n$, or not)

$X_9^a = \{0, 1\}$ ($a_w b_w < \sum a_i b_i/n$, or not)

(b) Action Attributes

Fig. 6 Attributes and their values

えることとした。各計画者に20回の積み合せ計画を試行させ、それぞれについて行動事例数 $m = 20$ を記録し、計400個の行動事例を実験システムにより生成した。

実験システムで記録した矩形移動の履歴から、計画者には以下の傾向があることがわかった。

- 計画者A,Bとも大きい矩形を優先して積み合わせる。
- 計画者Aは適当数の矩形を組み合わせ、パレット幅に合うような一つの方形ブロックを考え、逐次的に方形ブロックを積み合わせる。
- 計画者Bはパレットの左側に矩形を横向きに積み合わせることを最優先する。
- 計画者Bは矩形の移動回数(バックトラッキングも含む)が少なく、計画者Aに比べて意思決定での「迷い」が少ない。

各計画者から得られた行動事例を入力データとして、2.3のアルゴリズムにより各計画者に対する行動ネットワークを獲得した。なお、学習率 $\alpha = 0.8$ 、割引率 $\gamma = 0.8$ とした。以下、計画者AおよびBから得られた行動ネットワークをそれぞれ N_A, N_B とする。表1に、行動ネットワーク N_A, N_B における状態ノード数 (STATE)、行為ノード数 (ACTION)、各状態での

Table 1 Obtained state-action networks

	STATE	ACTION	AVG
Network N_A	15	64	4.3
Network N_B	12	52	4.3

STATE :Total number of states

ACTION:Total number of actions

AVG :Average number of actions for each state

行為ノード数の平均 (AVG) を示す。表1から、 N_B の状態ノード数および行為ノード数は N_A のそれに比べて、少ないことがわかる。このことは、計画者Bが比較的簡単な規則で矩形を配置していると考えられ、行動ネットワークにより、傾向(d)の特徴を抽出しているといえる。また、計画者A,Bとも各状態での行為のパターンが単一ではなく、行動ネットワークを用いることにより、ある配置状況での計画者の行為が常に同じではないことが表現可能であるといえる。

次に、行動ネットワークを獲得する際に顕著に出現した状態に注目し、獲得した行動ネットワークから各計画者の行為の特徴を考察する。表2は、2つの状態0000, 0110での各計画者の行為、および行動の評価値を表している。状態0000では、計画者Aは計画者Bに比べて多様な行為を有しており、特に計画者Aは行為 a_1 を選好するが、この行為は計画者Bには存在しない。この状態での行為 a_1, a_2 はそれぞれ、「配置可能点 $q(1)$ に、比較的大きい矩形を縦向きに配置する」、「配置可能点 $q(1)$ に、比較的大きい矩形を横向きに配置する」という意味を持つ。すなわち、計画者Bは状態0000で横向きに矩形を配置することを重視するが、計画者Aは矩形の向きは重視していないと判断できる。計画者Bについては、状態0110で横向きに配置することを表す行為 a_{11} を多用していることから、傾向(c)が行動ネットワーク N_B に反映されているといえる。

一方、行動(0000, a_2), (0100, a_{10}), (0110, a_{11}) および (0110, a_{12}) は計画者A,Bともに行動評価が高く、積み合せ計画に関する有効なヒューリスティックスである可能性が高いといえる。次節では、数値シミュレーションにより、獲得した行動ネットワークのヒューリスティックスとしての有用性を検討する。

4.2 行動ネットワークによる積み合せ実験 ここでは、4.1で獲得した行動ネットワーク N_A および N_B の2次元積み合せ問題のヒューリスティック解法としての有用性を検討する。行動ネットワークを利用した積み合せ計画の手続きを以下に示す。

Step 1 初期状態を s_0 とする。 $s \leftarrow s_0$ 。

Step 2 状態 s での行為の一つ a を選択する。

Table 2 An example of obtained behavior patterns(f , R and Q represent respectively evaluation of behavior, number of experienced actions and Q -values.)

State	Planner A				Planner B			
	Action	f	R	Q	Action	f	R	Q
0000	a_2 :000010*****	174.9	35	5.0	a_2 :000010*****	369.4	74	5.0
	a_{35} :00000000111	18.3	7	2.6	a_{35} :00000000111	43.8	9	4.9
	a_1 :000001*****	199.9	40	5.0				
	a_{45} :00000010010	3.9	1	3.9				
	a_{50} :00000010111	25.0	5	5.0
0110	a_{12} :010011*****	104.9	21	5.0	a_{12} :010011*****	209.6	42	5.00
	a_{11} :010010*****	204.9	41	5.0	a_{11} :010010*****	125.5	26	4.83
	a_{10} :010001*****	16.9	4	4.2	a_{10} :010001*****	24.7	5	4.95
	a_{146} :01000010111	2.0	1	2.0	a_{150} :01000011011	4.3	2	2.15
					a_{126} :01000000011	3.6	1	3.60

Step 3 行動 (s, a) にしたがって、次の状態 s' に遷移する。 $s \leftarrow s'$ 。

Step 4 全ての矩形を積み合せた時点で終了。そうでなければ、Step 2 へ。

Step 2 における行為の選択は、行動の評価値に基づいて確率的に行うものとする。状態 s での行為 a について、選択可能な行為を $a^i (i = 1, 2, \dots, a^q, a^i \in A(s))$ としたとき、行為 a^i を選択する確率 $Pr(a = a^i)$ を次式 (6) で定義する。

$$Pr(a = a^i) = f(s, a^i) / \sum_{j=1}^q f(s, a^j) \quad (6)$$

4.1 と同一の2次元積み合せ問題に対し、行動ネットワーク N_A, N_B を100回適用した。表3に数値シミュレーションの結果を示す。表中の“Best”は100回の試行での最良解を、“Average”は解の平均を表し、カッコ内は各計画者の最良解と解の平均を表している。表3では、獲得した行動ネットワークにより良好な積み合せ結果が得られたことがわかる。特に、行動ネットワーク N_A は、計画者Aが20試行中の最良解と等しい解を導出しており、行動ネットワークによる積み合せ解法は、有用なヒューリスティクスであるといえる。一方、行動ネットワークによる解の平均は計画者のそれに比べて大きくなっている。この理由の一つとして、計画者は20回の試行の各々においてほぼ同様の積み合せパターンを生成したことに対し、行動ネットワークによる積み合せは確率的に行われるため、評価の低い行動が選択される場合があることが考えられる。

図7に行動ネットワーク N_A, N_B による積み合せ結果を示す。図7(a)の積み合せ前半部分では、いくつかの矩形をまとめて高さを揃えるよう積み合わせる傾向が見られ、傾向(b)を反映する結果となった。図7(b)では、パレット左側に矩形を横向きに積み合わせる傾向

Table 3 Results of packing

	Best		Average	
N_A	430	(430)	466.3	(437.5)
N_B	440	(430)	462.9	(447.5)

が見られ、行動ネットワーク N_B は、計画者Bの特徴(傾向(c))を反映しているといえる。また、行動ネットワークによる積み合せは、矩形が隙間なく配置されており、効率的な計画を提供している。以上のことから、獲得した行動ネットワークは計画者の特徴をとらえた有用な知識であるといえる。なお、(6)式以外の効果的な行為の選択方法を含めた行動ネットワークの利用方法や解のロバスト性について、今後詳細に検討する必要がある。

4.3 他の積み合せ問題への適用 獲得した行動ネットワークを他の積み合せ問題に適用することにより、行動ネットワークの汎用性を検討する。ここでは、計画者Aから抽出した行動ネットワーク N_A を取り上げ、 N_A を以下に示す問題例 P2, P3 に適用した。

問題例 P2: $n = 30, K = 150$

問題例 P3: $n = 20, K = 300$

各矩形の長辺および短辺は、問題例 P1 と同様、一様乱数に基づいて与えた。また、各問題例について、計画者Aに各々20回の積み合せ計画を試行させた。表4に、計画者Aの各問題例に対する20回の試行の最良解と、行動ネットワーク N_A を100回適用した際の最良解を示す。表4から、計画者Aによる解とほぼ同等の解が得られたことがわかる。また、このときの積み合せ結果は矩形をほぼ隙間なく積み合わせており、効率性の高い積み合せ計画を提供していることがわかった。

Fig. 7 Packing results by obtained network

Table 4 Results of packing for P_2 and P_3

	P_2	P_3
Network N_A	620	440
Planner A	610	440

5. お わ り に

本研究では、計画者の試行錯誤的な行動蓄積プロセスに内在する行動の特徴を定量的に表現することにより、計画者の知識を自動的に獲得することを試みた。計画者のある行動を状態と行為の組と捉え、行為の選択による状態の遷移を表現するネットワーク－行動ネットワーク－により計画者の知識を表現することを提案した。2次元積み合せ問題を対象として、知識獲得の実験システムを構築し、計画者の行動事例に基づいて行動ネットワークの獲得を試みた。状態と行為は、矩形の属性データなど対象問題から得られる基本情報により表現することを考えた。実験の結果、計画者の際立った行動の特徴をネットワーク表現することが可能であることがわかった。また、獲得した行動ネットワークを用いた積み合せ計画のシミュレーションを行った結果、行動ネットワークにより良好な解を生成することが可能であることがわかった。以上、提案手法は計画者の知識獲得の手法として有用であるといえる。なお、生産計画など他の計画問題への適用、ネッ

トワークの洗練化などは今後の検討課題とする。

文 献

- (1) 山本 孝, “組立ラインにおける技能伝承型の生産システム研究”, 経営システム, 8 - 2, 53 - 60 (1998)
- (2) 川野 常夫, “技能・技術の伝承の現状と問題点”, 機械の研究, 51 - 1, 7 - 15 (1999)
- (3) 三好 隆志, “熟練技能の技術化・コンピュータ化－金型自動磨き作業を例に－”, 日本機械学会講演論文集, No.994 - 1, 27 - 28 (1999)
- (4) A. Kusiak and M. Chen, “Expert Systems for planning and scheduling manufacturing systems”, *European Journal of Operational Research*, 34, 113 - 130 (1998)
- (5) J. Hynynen, “Using artificial intelligence technologies in production management”, *Computers in industry*, 19, 21 - 35 (1992)
- (6) A. Artiba and C. Tahon, “Production planning knowledge-based system for pharmaceutical manufacturing lines”, *European Journal of Operational Research* 61, 18 - 29 (1992)
- (7) S. Minton, *Machine learning methods for planning*, Morgan Kaufman, CA (1993)
- (8) Z. Doulgeri, G. D'alessandro and N. Magaletti, “A hierarchical knowledge-based scheduling and control for FMSs”, *Int. J. Computer Integrated Manufacturing*, 6-3, 191 - 200 (1993)
- (9) 佐藤知正, 森武俊, “人と機械の知的協調システム－行動蓄積機能に基づいた人間行動適合型機械システム－”, 計測と制御, 38 - 6, 357-362 (1999)
- (10) 畝見達夫, “強化学習法”, 人工知能学会誌, 9-6, 830-836 (1994)
- (11) C.J.C.H. Watkins and P. Dayan, “Technical note : Q-learning”, *Machine Learning*, 8 - 3, 279-292 (1992)
- (12) E.G. Coffman, M.R. Garey, D.S. Johnson and R.E. Tarjan, “Performance bounds for level-oriented two-dimensional packing algorithm”, *SIAM journal on computing*, 9 - 4, 808 - 826 (1980)
- (13) B.S. Baker and H.P. Katseff, “A 5/4 algorithm for two-dimensional packing”, *Journal of Algorithms*, 2, 348 - 368 (1981)