

3 機械フローショップ・スケジューリング問題 への強化学習の適用

5 J-6

田中 雄介, 吉田 武稔

北陸先端科学技術大学院大学

はじめに

フローショップ・スケジューリングに関して、ジョブの総処理時間最小化問題を考える。2機械の場合にはジョンソン則による最適解獲得が証明されているが、3機械以上の場合、総処理時間を最小化するスケジューリング方策は、小規模な問題を除いては、計算量の観点から求めるのが困難である。以上を背景に、この問題に対する実行解を求めるために、強化学習を適用する。学習が効果を上げる鍵は、問題の定式化にある。よって、本研究における目的は、

- いまだ理論解のない本問題に対し、強化学習問題としての定式化を考案する。
- 実装評価により、その有効性を吟味する。

強化学習問題としての定式化

2機械フローショップ・スケジューリングに関して、総処理時間を最小化する解法であるジョンソン則は、2機械 M_1, M_2 の順に処理される n 個のジョブ $j_i \in J (i = 1, 2, \dots, n)$ の、機械 M_1, M_2 各々の処理時間を p_{1i}, p_{2i} とするとき、2ジョブ $j_i, j_k (i, k = 1, 2, \dots, n; i \neq k)$ について、 $\min[p_{1i}, p_{2k}] \leq \min[p_{1k}, p_{2i}]$ を満たすようにジョブ j_i を j_k に先行して処理するルールである (等号成立時の仕掛順は任意) [1]。

この解法を基に、3機械問題への適用を前提にして、強化学習問題としての定式化を考案する。強化学習は、環境、エージェント、政策、報酬、価値関数の5構成要素に分けられる。環境は、スケジューリン

グの対象になる機械やジョブ集合である。エージェントは、ライン投入をするディスパッチャーとした。フローショップの総処理時間最小化問題でかつ、ジョブが予め与えられている2機械の静的スケジューリング問題では、第2機械前の待ち行列に対し、スケジューリングを行う必然性はない [1]。

次に、スケジューリングでのジョブ仕掛順の意思決定が、強化学習における逐次意思決定へどのように適用されるかが焦点になる。文献 [6] では、ジョブの順序制約のみを考慮し、資源制約が違反した状態のスケジュールから、ジョブの移動や資源割当の変更を行いつつ、実行可能なスケジュールへ到達する過程と位置付けている。また、文献 [3] では、日々製品別生産量を決定し、完成品在庫をいかにしてなくならないようにかつ最小化するかを永遠に繰り返す過程と位置付けている。本研究では、文献 [3] と同様の逐次意思決定問題としての定式化を提案する。なぜなら、時刻経過とともにジョブが新規に到着する、動的過程を想定した定式化であり、実際のスケジュール実行時のトラブル等を状態遷移の不確実性にそのまま解釈できるメリットが、今後の研究で享受できらうと思われるからである。

政策は、エージェントが知覚する状態に対して、その行為を決定する。エージェントの学習とは、最適政策を獲得することだが、状態と行為の定式化次第で、問題の解性能が大きく左右される。前段のとおり、スケジュール実行時間軸上での逐次意思決定とする結果、エージェントの行為は、ライン投入前の待ち行列から、仕掛けるジョブ1つを選択することになる。待ち行列内全てのジョブを選択の対象にすることは汎用性を持たせる点では利点がある。他方、問題を解くヒューリスティクスが予め分かっている場合、選択対象を絞り込むことがルール獲得に役立つ。そこで、3種類の選択候補となるジョブ集合を作った。

- a1) 最大2ジョブが候補になる. $p_{1i}+p_{2i} \leq p_{2i}+p_{3i}$ のジョブの中で $p_{1i}+p_{2i}$ 最小, $p_{1i}+p_{2i} > p_{2i}+p_{3i}$ のジョブの中で $p_{2i}+p_{3i}$ 最大, のもの. これは, ジョンソン則から得られる仕掛順の規則性をほとんど利用したものとなっている.
- a2) 最大8ジョブが候補になる. $p_{1i}+p_{2i} \leq p_{2i}+p_{3i}$ のジョブの中で $p_{1i}+p_{2i}, p_{2i}+p_{3i}$ それぞれについての最小と最大のもの, $p_{1i}+p_{2i} > p_{2i}+p_{3i}$ のジョブの中でも同様に選ぶ. つまり, 2^3 個のジョブが候補になる. 最初のジョブ集合の条件を緩和したものとなっている.
- a3) 待ち行列内全てのジョブが候補になる.

次に, 本問題の状態記述として, 問題毎による差異を正規化するため, 0-1 区間で表される指標を, 7つ考案した. なお, 前工程とは第1, 第2工程を, 後工程とは第2, 第3工程をさす.

- s1) 全ジョブ中, 待ち行列内の残りジョブ数
- s2) 待ち行列内のジョブのうち, 前工程時間が後工程時間より長いものの数
- s3) 前工程で, 全ジョブに対する, 待ち行列内ジョブの残り負荷
- s4) 後工程で, 全ジョブに対する, 待ち行列内ジョブの残り負荷
- s5) 待ち行列内のジョブで, 前工程処理時間が全ジョブ中最小のジョブが残っているかどうか
- s6) 待ち行列内のジョブで, 後工程処理時間が全ジョブ中最小のジョブが残っているかどうか
- s7) 後工程の機械稼働率

総処理時間最小化問題であるから, 報酬は, 最終状態において総処理時間を与えればよい. 価値関数の学習には $TD(\lambda)$ 法を用い, 人工ニューラル・ネットワークによる関数近似を行った [4]. また, 強化学習単独では最適政策獲得までに多くの訓練回数を必要とするため [5], 文献 [2] を参考に, 遺伝的アルゴリズムを組合せ, その優良解を価値関数に学習させることで, 訓練回数の短縮を図るモデルを考案した.

実装評価

まず, 予備的に2機械問題で, つまり最適解獲得のアルゴリズムが存在する問題で解いた場合, a1) とs2) の行為と状態の定式化の組み合わせでは, 最適アルゴリズムと同一解を得られることが確認で

きた. また, それ以外の組合せで最適解を得られないまでも, 最適解の7%増程度の最大処理時間を実現する学習結果も観測された. このように定式化が不完全でも学習できることは, 最適解が獲得できない問題領域でも学習が有効となる可能性を示唆している.

さらに, 3機械問題で全探索以外最適解獲得のアルゴリズムが存在しない場合でも, 一定の学習効果をあげることができた. ただ, ジョンソン則による解の近傍に最適解が存在することが多いことから [1], ジョンソン則を参考にした本研究の定式化による学習において, 2機械問題と同程度の成績をあげることが, 相当程度予想されていた. なお, ジョンソン則を参考にした定式化である以上, ジョンソン則による解を上回る性能を実現することはできなかった. いくつかの特定の問題で, ジョンソン則を上回る解を出す状態行為の組合せがあったものの, 平均においては明らかに劣るものであるから, 学習した結果をメタな知識として検討する価値はないものと推測される. また, 遺伝的アルゴリズムによる優良解を学習させたモデルでは, より少ない訓練回数で同程度の安定した解を得ることができた.

参考文献

- [1] Conway, R.W., Maxwell, W.L., and Miller, L.W.: *Theory of scheduling*, Chap.5, Addison-Wesley(1967).
- [2] Kim, G.H., Lee, C.S.G.: Genetic reinforcement learning approach to the heterogeneous machine scheduling problem. *IEEE Transactions on robotics and automation* 14(6), pp.879-893(1998).
- [3] Schneider, J.G., Boyan, J.A., and Moore, A.W.: Value function based production scheduling. *Machine learning: proc. of the fifteenth international conference (ICML '98)*, pp.522-530(1998).
- [4] Sutton, R.S., and Barto, A.G.: *Reinforcement learning*, The MIT Press(1998).
- [5] 田中 雄介, 吉田 武稔. フローショップ・スケジューリング問題への強化学習の適用. 人工知能学会知識ベースシステム研究会資料, SIG-KBS-9901, pp.13-18, 1999.
- [6] Zhang, W., and Dietterich, T.G.: A reinforcement learning approach to job-shop scheduling. *Proc. of the fourteenth international joint conference on artificial intelligence*, pp.1114-1120(1995).