

遺伝的アルゴリズムと深層強化学習に基づく ジョブスケジューリング最適化問題

—PCB 加工プロセスをモデルとして—

汪 永豪[†]
関西学院大学[†]

山田 孝子[‡]
関西学院大学[‡]

1 はじめに

本研究では、プリント基板 (PCB) 製造工場を対象に、同一加工機械を複数回使用する生産プロセスをモデル化した上で、スケジューリング最適化に関する検討を行った。静的な環境においては、遺伝的アルゴリズム (GA) および深層強化学習に基づく Deep Q-Network (DQN) を用いて最適化を行い、両手法の性能比較を実施した。さらに、機械故障等による環境変動が発生する動的な環境で高速にスケジュールを生成可能な DQN により、適応性能の向上を図った。

2 生産管理問題のモデル

本研究では受注生産を行う工場の製造工程、受注型ジョブショップスケジューリング問題 (Job Shop Scheduler problem: JSSP) を最適化問題として解く。JSSP を制約付き最適化問題として定式化し、目標関数としては:

$$\text{opt. } JSSP = \min \left\{ \max_s TJ_s \right\} \quad (1)$$

ここで、 TJ_s は第 s 種別製品の最終加工工程の完了時刻を表す。

したがって、本最適化問題は、ジョブセットに記載された全製品種別の最終加工完了時刻の最大値を求める。様々なスケジューリング案の最終加工完了時刻の中で最小のものを本論文では *makespan* と呼ぶ、すなわち、最適化とは、この *makespan* の最小化である。

ジョブショップスケジューリングの目的は、各加工機械をできるだけ連続的に稼働させるとともに、各製品の加工工程を待ち時間なく加工することにより、総加工時間を最小化することである。したがって、JSSP は待機している加工機械に対して加工対象となるジョブを割り当てる逐次意思決定問題とみることができる。この意思決定過程は次の図 1 に示します。

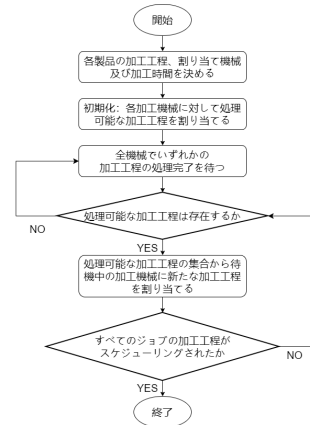


図 1: ジョブスケジューリング問題フローチャート

3 実験用データ自動生成

モデルとしてクライアント側から提供される注文書が最適化の対象となり、このようなデータは不可欠である。PCB 製造に関する実データは、企業の営業機密や顧客情報に直結するデータのためこうしたデータは入手困難で、共同研究であっても外部に具体的な数値の利用は難しい。そのため本研究では、先行研究に記載されている PCB 製造工程を参考にして、各加工機械に対して加工役割を割り当てた人工的なデータをまず生成する。そこで文献 [5] に記載された多層 PCB の製造工程に基づく各製品種別ごとの加工工程と使用機械をもとに、注文書のばらつきやロット数を確率的にデータとして生成する

し、ジョブセットを得る。

4 遺伝的アルゴリズムによる解の改善

本研究では、JSSP に対して、GA を用いたスケジューリングの最適解探索手法を採用する。

各染色体は、全ジョブに属する加工工程を一度ずつ含む遺伝子列として表現され、各遺伝子は「ジョブ J_s の第 q 工程」を表す。遺伝子列の順序に従い、制約を満たすようスケジュールを構成する。各ジョブは処理可能なジョブからランダムに工程を選択し配置することで初期染色体を生成する。各染色体をデコードし、得られた総加工時間 *makespan* を適応度として評価する。トーナメント選択により最良個体を選択し、親染色体を決定する。制約を保持可能な交差手法を用いて子個体を生成する。子個体を生成後に修復処理を行い、解の実行可能性を保証する。進化計算を繰り返し、得られた最良染色体を最終解として採用する。最終解に対応するカントチャートと世代ごとの *makespan* の折れ線グラフで視覚的に評価する。

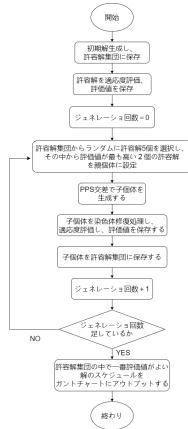


図 2: 遺伝的アルゴリズムフローチャート

加工順序が厳密に定まる問題なので、PPS(Precedence Preservative Crossover) という交差手法を採用する。

5 深層強化学習による解の改善

本研究では、価値関数に基づく強化学習手法である Deep Q-Network(DQN) を用いる。DQN では、各状態 s_t における行動 a_t の価値を表す行動価値関数 $Q(s_t, a_t)$ をニューラルネットワークにより近似し、各状態における行動選択を Q 値に基づいて行う。

扱う JSSP 環境は、加工工程の開始および完了と

いったイベントによって状態が更新されるイベント駆動型の環境として構成されている。

DQN における 1 step とは、時刻 t においてエージェントが状態 s_t を観測し、行動 a_t を選択した後、環境が次状態 s_{t+1} と報酬 r_t を返すまでの連続的な意思決定単位を指す。

6 まとめ

参考文献

- [1] K.T. Chung, C.K.M.Lee and Y.P. Tsang, "Neural combinatorial optimization with reinforcement learning in industrial engineering: a survey" *Artif Intell Rev.* vol. 58, 130(2025).
- [2] S. Lee, Y. Cho and Y.H. Lee, "Injection Mold Production Sustainable Scheduling Using Deep Reinforcement Learning" *Sustainability.* vol. 12, 8718(2020).
- [3] L.B. Wang, X. Hu, Y. Wang, S. Xu, S. Ma, K. Yang, Z. Liu and W. Wang, "Dynamic job-shop scheduling in smart manufacturing using deep reinforcement learning" *Computer Networks.* vol. 190, 107937(2021).
- [4] 村山 昇, 川田 誠一, "遺伝的アルゴリズムを用いた加工機械と複積載 AGV の同時スケジューリング" 日本機械学会論文集 (C 編). vol. 12, 712, pp. 3638-3643(2005-12).
- [5] Andrzej Kiernich, Jerzy Kalenik, Wojciech Steplewski, Marek Koscielski and Aneta Chołaj, "Impact of Particular Stages of the Manufacturing Process on the Reliability of Flexible Printed Circuits" *Sensors.*25, 140, (2025).