Selective Laser Thermoregulation 法における走査パスを提案する強化学習 AI の学習方法の検討

Study on the learning method of reinforcement learning AI that proposes scanning paths in the Selective Laser Thermoregulation Method

宇井 翔太[1)]
指導教員　大久保 友雅[1)]，研究協力者　河原崎 祐作[1)]

1）東京工科大学大学院 工学研究科 サステイナブル工学専攻 光・エネルギー(大久保)研究室

Selective Laser Thermoregulation(SLT)法におけるレーザの走査パスを提案する AI として Monte Carlo (MC)法による学習をする AI と Temporal Difference(TD)法による学習をする AI を開発し比較したところ学習結果に大きな差はなく，学習に要した時間が短かった TD 法が適していることがわかった.

キーワード：レーザ加熱試験，SLT 法，強化学習，MC 法，TD 法

１． はじめに

代表的な非酸化物系セラミックス複合材料（CMC：Ceramic Matrix Composite）である SiC/SiC は，耐熱性に優れた高比強度材であるため，次世代航空機用ガスタービンエンジン用の材料として期待されている．しかし，高温の物理的特性に関しては不明な点が多い．高温下での物理的特性を得るためには SiC/SiC を 1400 ℃以上の高温まで高速かつ均一に加熱する手段の確立が必要である．その実現のために著者らは，加熱対象にレーザを照射しその照射点をガルバノスキャナを用いて高速で走査することにより任意の形状の加熱を行う SLT 法を開発している[1]．しかし，この手法はレーザの走査パスによって試験片表面の温度分布が変化するため，所望の温度分布を実現するための最適なパスを実験や数値計算で模索するには膨大な試行が必要である．そこで本研究では，温度分布を均一に近づけるようなレーザの移動方向を強化学習により求めるために，MC 法[2]による学習を行う AI と TD 法[2]による学習を行う AI を開発し，それぞれの提案する走査パスと学習に要した時間を比較しどちらの学習法が適しているか評価した．

２．開発する AI の概要

強化学習ではエージェントと環境が互いに影響を与え合う．エージェントの行動により環境に変化を与え，環境は与えられた変化を評価して報酬を与える．そして，エージェントはこの報酬がより多く獲得できるように期待値の高い行動パターンに改善して再度行動する．このとき，期待値の高い行動パターンを学習したい一方で，報酬が最大となる 1 つの行動だけを学習すると，全ての行動のサンプルデータを得ることができない．そのため，ϵ-greedy 法[3]と呼ばれる，一定確率ϵでランダムに行動を選択する探索手法を用いた．

本研究ではこのエージェントをレーザの照射点，環境を照射物の温度分布，報酬は温度分布がどの程度均一かつ高温であるかとし評価した．具体的には，レーザの照射点が移動し，温度分布が変化する．こうして変化した温度分布を評価し，報酬を与えることでレーザの照射点が移動方向を学習し，均一かつ高温な温度分布を実現するまで移動を続ける．この一連の処理を繰り返すことで AI はレーザの走査パスを学習することができる．また， MC 法と TD 法の学習方法によって期待値の計算方法が異なる．MC 法は均一かつ高温な状態に到達した際の知識すなわち期待値を更新する手法である．一方，TD 法は照射点が移動するたびに期待値を更新する手法である．MC 法と TD 法の期待値の計算式をそれぞれ式(1)と式(2)に示す．時刻tにおける照射点の座標をS_t，移動方向をA_tとしたときの期待

値を$Q(S_t, A_t)$，更新後の期待値をQ'，学習率をα，報酬をR，その合計をG，報酬の優先度をγとした．

$$Q'(S,A) = Q(S,A) + \alpha\{G - Q(S,A)\} \quad (1)$$

$$Q'(S,A) = Q(S_t,A_t) + \alpha\{R_t + \gamma Q(S_{t+1},A_{t+1}) - Q_\pi(S_t,A_t)\} \quad (2)$$

加熱と冷却の温度分布の計算については図 1 に示す形状の SUS304 の試験片の評価領域にレーザ照射を行い，その最大温度の時間変化を計測[1]しその温度変化の様子を指数関数でフィッティングして得られた数式を使用した．ビーム半径 8 mm，出力 400 Wのレーザを走査速度 10 m/sで 115 秒間往復させた際の最高温度の推移を図 2 に示す．また，フィッティングで得られた時刻tにおける加熱時の温度T_tと冷却時の温度T_tの数式を式(3)と式(4)に示す．本研究では式(3)と式(4)，さらに熱伝導による温度変化を考慮して計算した．計算負荷軽減のために 2 mm四方の空間を 4×4 のセルに区切り，その小さい空間を照射点が移動した際の温度分布の変化を計算した．

$$T_t = -606.0e^{-0.1137t} + 1418 \quad (3)$$

$$T_t = 644.9e^{-0.0819t} + 573.8 \quad (4)$$

報酬値Rはどの程度均一であるかを温度の空間的な標準偏差を用いて，どの程度高温であるかを空間的な平均温度を用いて評価するために式(5)に示すような数式で報酬値Rを計算した．

$$R = 2 - \frac{\text{現在のセル温度}}{\text{目標温度}} - \frac{\text{現在の標準偏差}}{\text{目標標準偏差}} \quad (5)$$

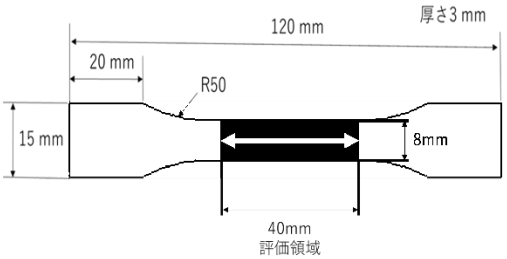

図 **1** 実験時の試験片の寸法と評価領域

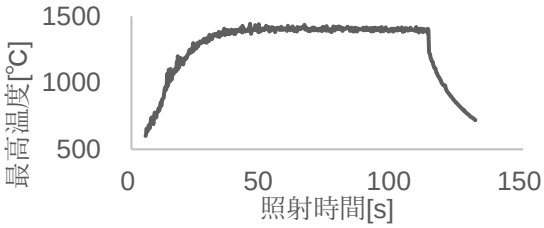

図 **2** 実験で計測した最高温度の推移

3．学習結果の比較

MC 法と TD 法を用いて空間的な温度分布の平均温度が 400 ℃，空間的な温度の標準偏差が 40 ℃以下になるようにそれぞれ学習させた．MC 法の学習結果と TD 法の学習結果を図 3 に示す．図 3 は X 軸と Y 軸を加熱領域の座標とし，縦軸を移動回数とすることで目標の温度分布に到達するまでの軌跡を表現している．図 3 よりそれぞれの学習方法による違いは螺旋の向き程度であり大きな差がない．また，学習前と MC 法と TD 法で目標の温度分布に到達するまでに要した移動回数については学習前は約8000回であったのに対してどちらの学習法も約 200 回であり，学習前より約 40 倍少ない移動回数で達成することができた．

学習に要した時間は MC 法では 1044 秒に対して TD 法は 97 秒と約 10 倍短い時間で走査パス移動方向を提案することができた．以上のことから，TD 法を用いることでより短い移動回数で目標の温度分布に到達し，より短い時間で学習を終えることができる AI の開発に成功した．

4．参考文献

[1] H. Koshiji *et al.*: J. Laser Micro Nanoeng.15 (2020) pp.174-177.

[2] Richard S. Sutton, Andrew G. Barto.: Reinforcement learning: An introduction. MIT press(2018)

[3] P. Auer *et al.*: P. Finite-time Analysis of the Multiarmed Bandit Problem. Machine Learning 47, (2002) pp.235-256.

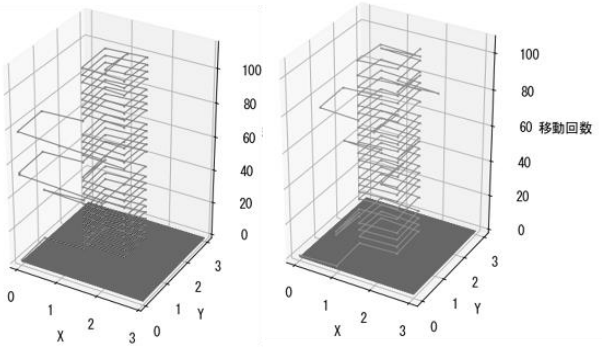

図 **3** (左)**MC** 法の学習結果と(右)**TD** 法の学習結果