

Q学習を用いた MTS/MTOハイブリッド生産システム のモード切り替え方策

広島大学大学院 先進理工系科学研究科
先進理工系科学専攻 電気システム制御プログラム
生産システム工学研究室
稲田景太

目次

1. MTS/MTOハイブリッド生産システムについて
2. 研究背景
3. 前提条件
4. 従来モデル
5. 提案モデル
6. 数値実験
7. まとめ・今後の課題

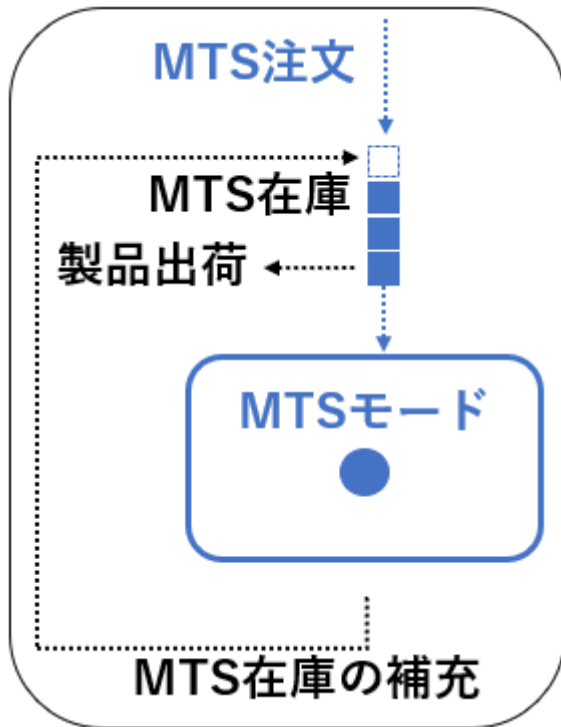
目次

- 1. MTS/MTOハイブリッド生産システムについて**
2. 研究背景
3. 前提条件
4. 従来モデル
5. 提案モデル
6. 数値実験
7. まとめ・今後の課題

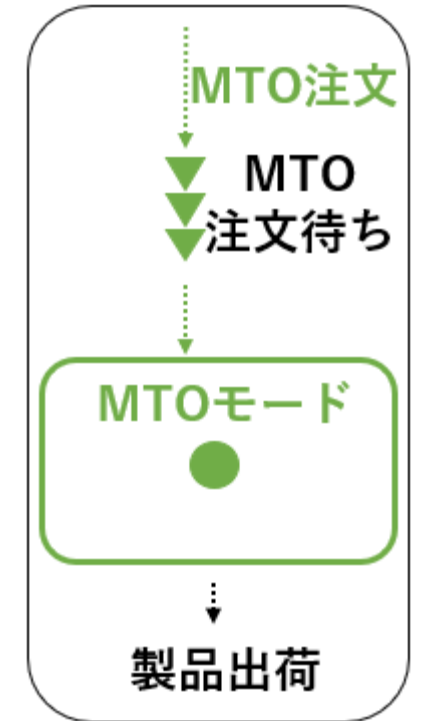
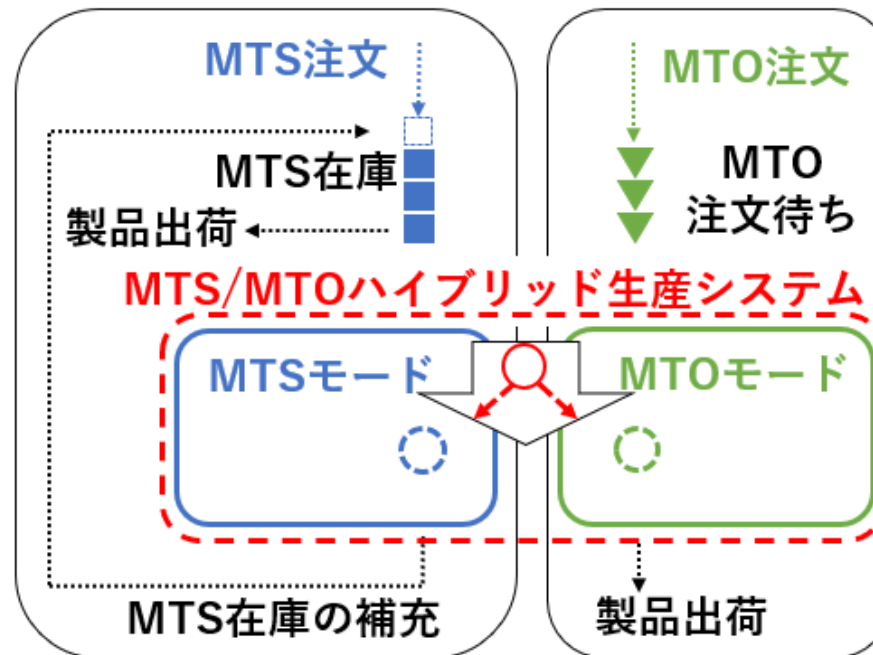
MTS/MTOハイブリッド生産システム

MTS生産 (make-to-stock)
注文到着前に生産し在庫を保持
一般品の生産に適している

MTO生産 (make-to-order)
注文到着後に注文に応じて生産
オーダーメイド品等の生産に適している



MTS/MTOハイブリッド生産システム
メリット：MTS/MTOモードの適切な切り替えによりコスト削減可能
デメリット：管理・運用の複雑化



目次

1. MTS/MTOハイブリッド生産システムについて
- 2. 研究背景**
3. 前提条件
4. 従来モデル
5. 提案モデル
6. 数値実験
7. まとめ・今後の課題

研究背景

Beemsterboerら(2017)

MTS/MTOハイブリッド生産マシンの切り替え方式

- ✓ マルコフ決定過程で単一マシンの生産システムをモデル化
- ✓ 大規模な状態空間等を扱う際にはモデル化が複雑になり、計算的に非常に困難

Huangら(2020)

PM問題における強化学習を用いたマシンのオン/オフ切り替え

- ✓ 生産システムの複雑さは非常に大きな状態空間に繋がるため、適した強化学習を採用する必要がある
- ✓ シミュレーションを使用

本研究の方針

強化学習を適応させたMTS/MTOハイブリッド生産システムを提案し
より近似最適なモデルをシミュレーションを用いて比較検討

研究背景

表1：生産分野における強化学習を用いた従来研究

著者	概要	強化学習を用いる際の提案
Mohsenら(2017)	人的資源計画と生産在庫管理の複合的な問題への提案	保管コストと機械損失コストのトレードオフを確率論的方策でモデル化
Zhengら(2020)	製造派遣の遅刻・遅刻報酬機能を設計	遅刻等の生産時間指標に関連するペナルティを含む報酬を追加
Chienら(2020)	需要予測モデルに基づく行動選択のためのフレームワーク構築	過去の状態を観測して行動を決定
Chengら(2022)	製造セルのスケジューリング	Q学習に基づく遺伝的アルゴリズムを設計

MTS/MTOハイブリッド生産システムでは過去のデータが蓄積されており、最大限活用することが重要である



本研究では過去の状態を観測し行動を決定する強化学習を適応させた
MTS/MTOハイブリッド生産システムモデルを提案する

研究背景

方策オン/オフ型：状態価値関数を更新する際に、方策 π を使い/使わないで更新する手法

表2：代表的な強化学習とMTS/MTOハイブリッド生産システムにおける環境

種類	方策オン/オフ型	適した状態と行動	リアルタイム性	高次元な状態空間
マルコフ決定過程	方策オン型	離散	×	△
モンテカルロ法	方策オフ型	離散	×	○
TD法	方策オン型	離散	○	○
Q学習	方策オフ型	離散	○	○
SARSA	方策オン型	離散	△	○
MTS/MTOハイブリッド生産システム の環境	方策オン型は制約等 が多く存在するため、 方策オフ型が適切	特定の行動を選択 する 離散的な環境	ランダムな需要到着等、 リアルタイムで対応 することが不可欠	多様な 高次元

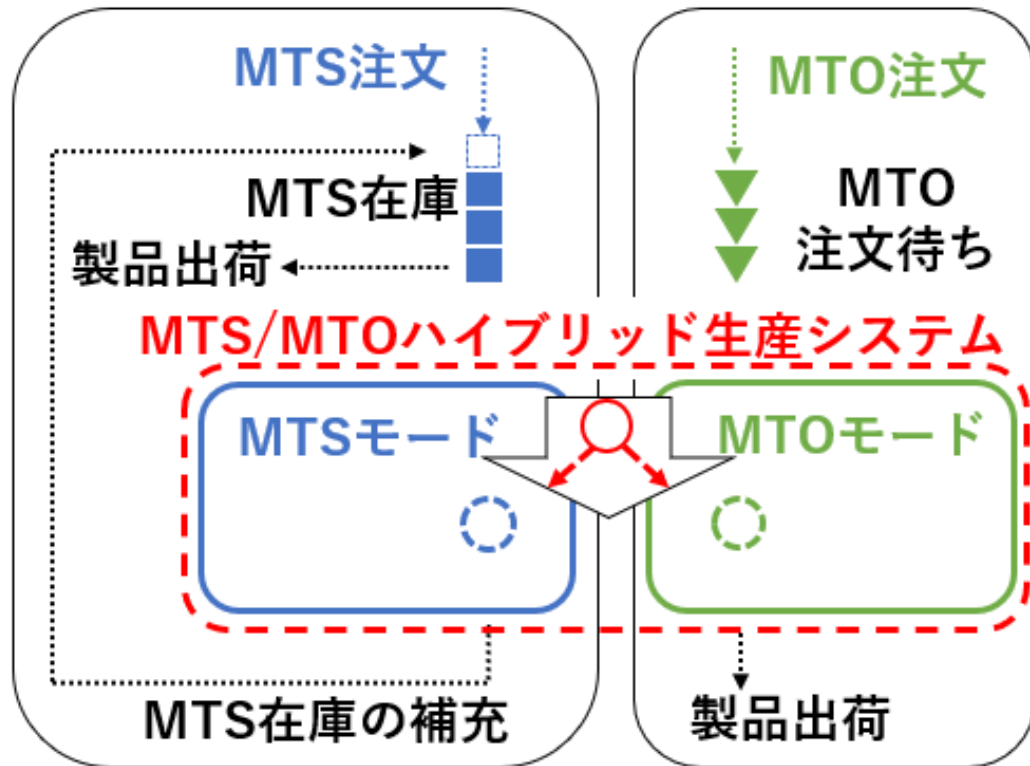


本研究では強化学習の中でもQ学習を適応させる

目次

1. MTS/MTOハイブリッド生産システムについて
2. 研究背景
- 3. 前提条件**
4. 従来モデル
5. 提案モデル
6. 数値実験
7. まとめ・今後の課題

前提条件



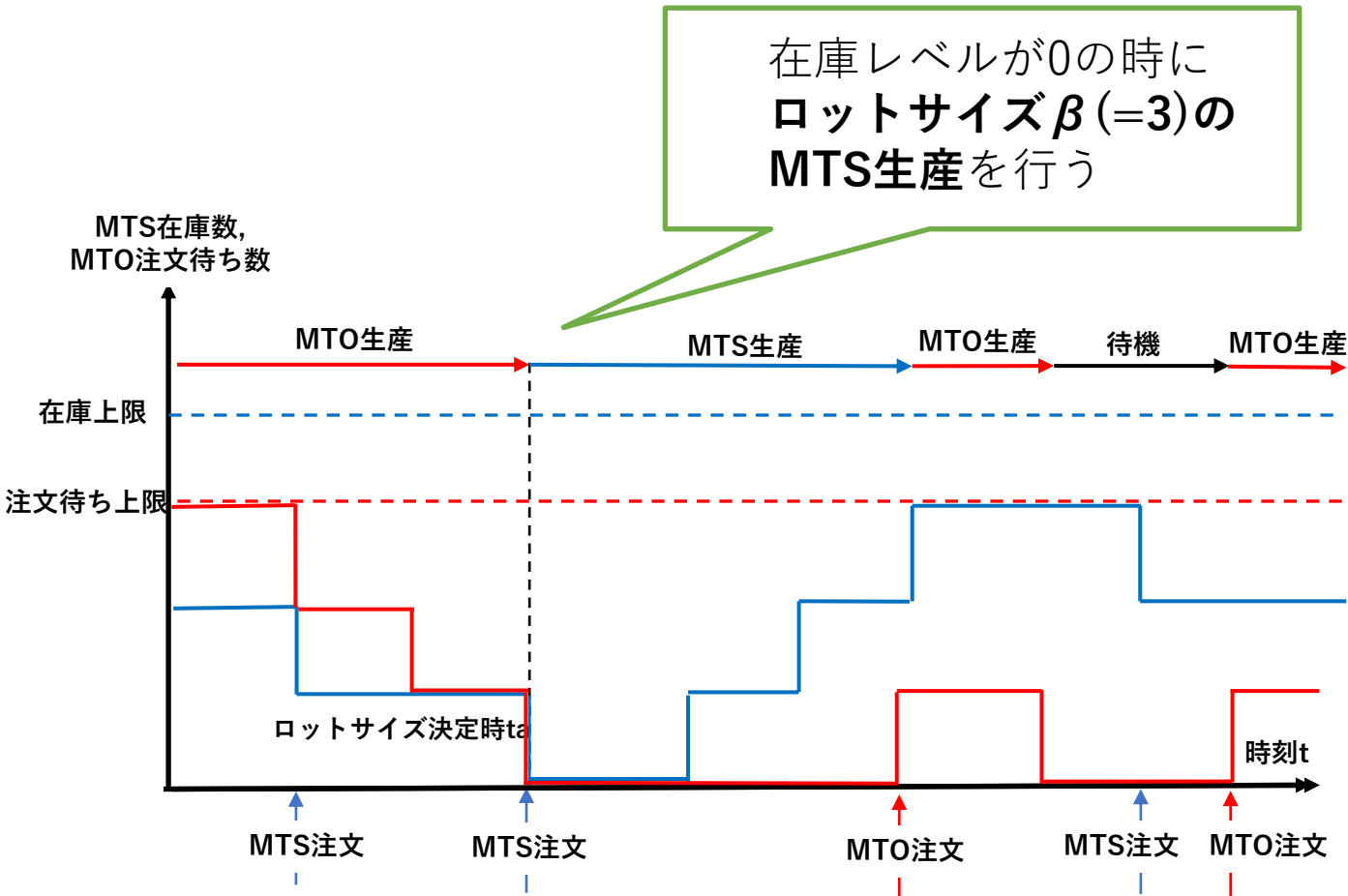
MTS/MTOハイブリッド生産システムの概要図

- システムは**単一の生産マシン**から成る
- MTS/MTO注文の**到着間隔は $1/\lambda_s, 1/\lambda_o$ の指数分布**に従う
- MTS生産は**在庫上限H**が存在し、**在庫保管コスト**がかかる
- MTS在庫が0の際にMTS注文が到着した場合は**MTS機会損失**となる
- MTS製品のロット生産は最初に**セットアップ時間 τ_s** がかかり、その後1製品ごとに**一定の生産時間 l_s** がかかる
- MTO注文は**待ち上限V**が存在し、**注文待ちコスト**がかかる
- 待ち注文が上限Vに達している際にMTO注文が到着した場合は**MTO機会損失**となる
- MTO製品は**セットアップ時間 τ_o と生産時間 l_o** を合わせて形状母数k, 平均 $1/\mu$ のアーラン分布に従う
- セットアップコストは**連続で同じモードを選択した際にはかからない**(ex.MTSモード→MTSモード)
- 出荷時間は考慮しない

目次

1. MTS/MTOハイブリッド生産システムについて
2. 研究背景
3. 前提条件
- 4. 従来モデル**
5. 提案モデル
6. 数値実験
7. まとめ・今後の課題

固定ロットサイズ方策



固定ロットサイズ方策

MTS/MTO生産が終わった時点で次の
いずれかを判断

➤ MTO生産

- ✓ 生産終了後、到着済みのMTO注文がある場合はMTO生産を行う

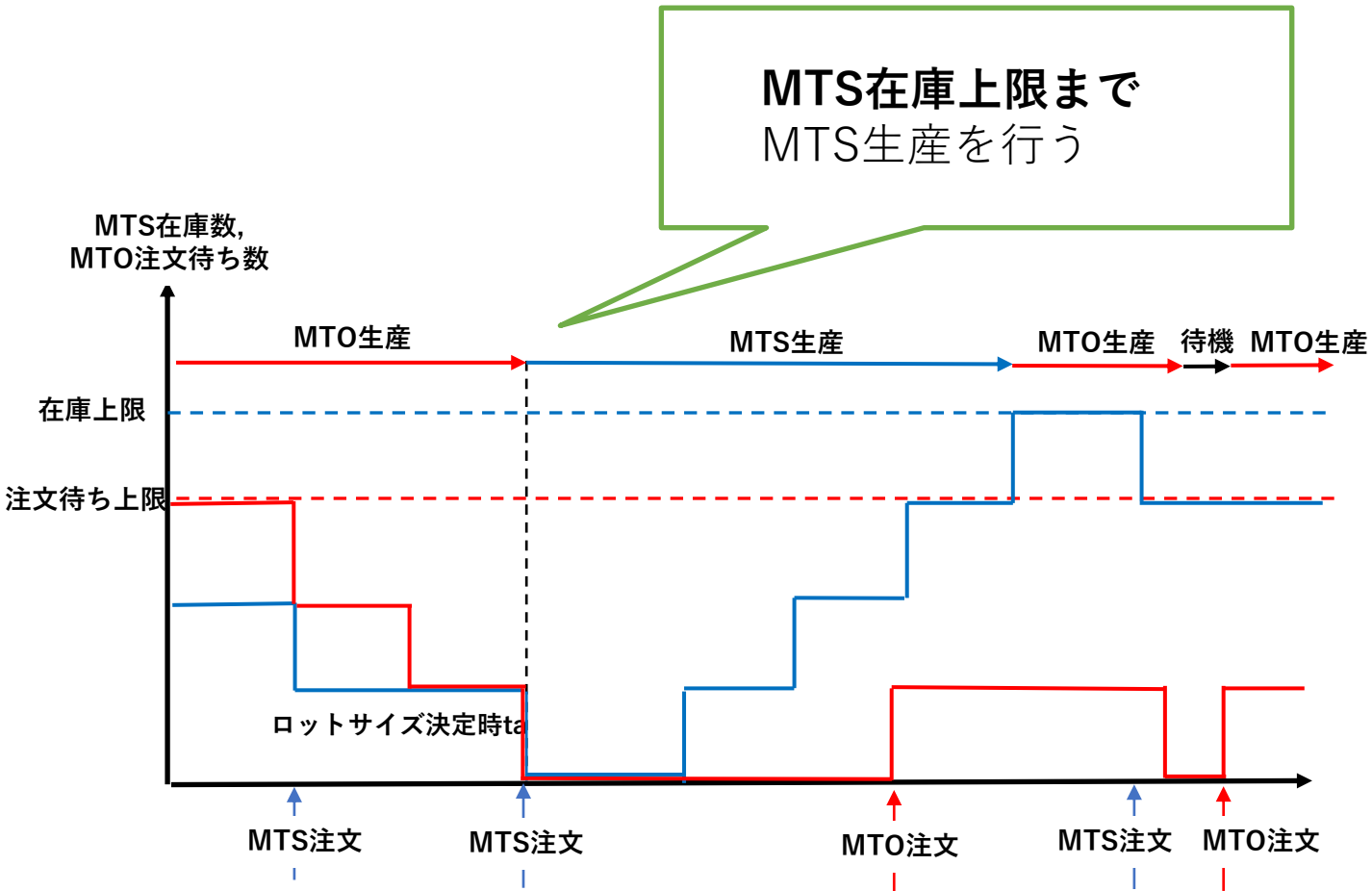
➤ MTS生産

- ✓ MTO注文が到着しておらず、在庫レベルが0である場合にロットサイズ β のMTS生産を行う

➤ 待機

- ✓ 在庫レベルが1以上でありMTO注文が到着していない場合

MTS優先方策



MTS優先方策

MTS/MTO生産が終わった時点で次のいずれかを判断

➤ MTS生産

- ✓ MTS在庫が0である場合にロットサイズH(MTS在庫上限)のMTS生産を行う

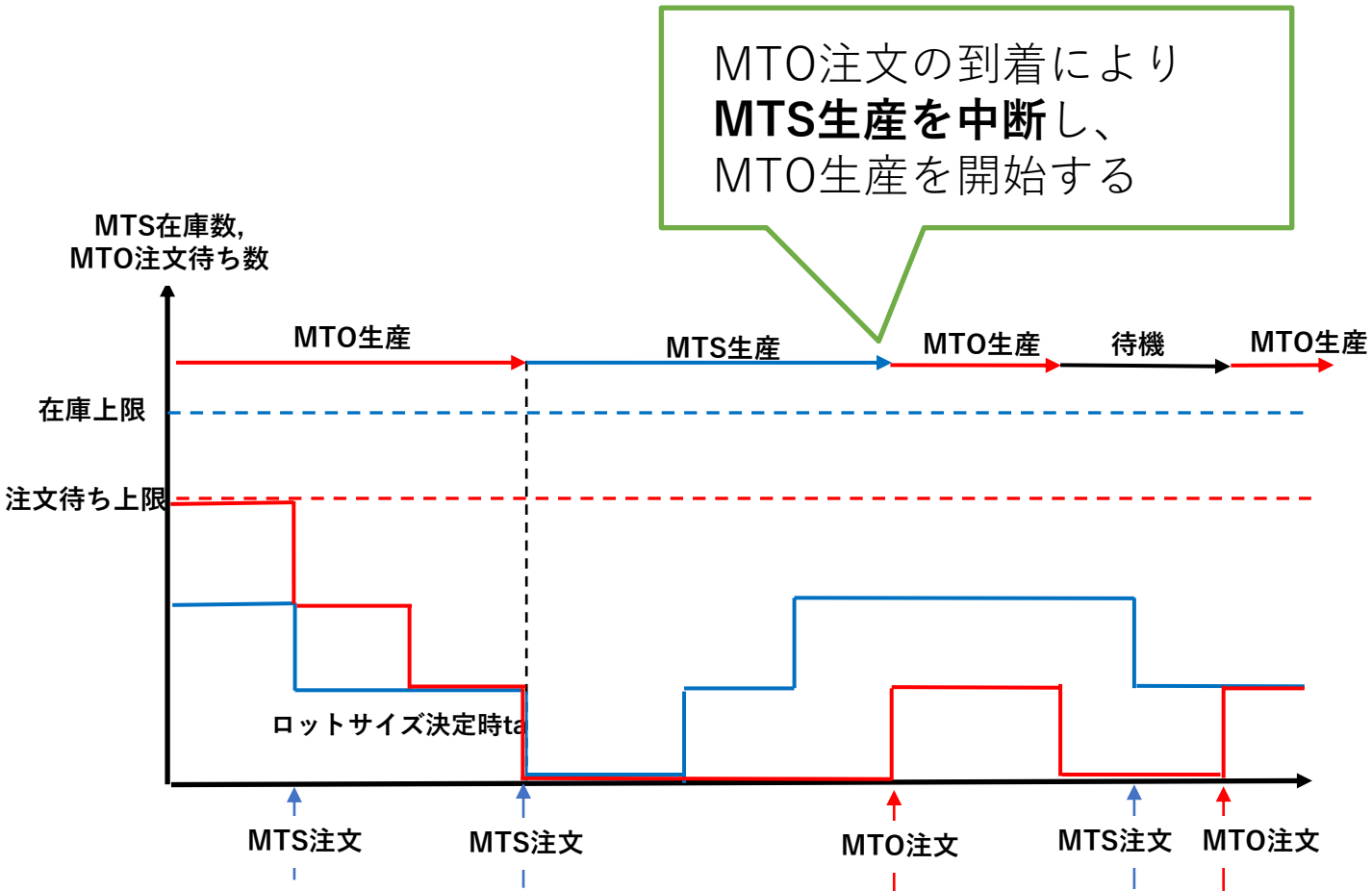
➤ MTO生産

- ✓ MTS在庫が1以上でありMTO注文が到着している場合、MTO生産を行う

➤ 待機

- ✓ 在庫レベルが1以上であり、MTO注文が到着していない場合に待機する

MTO優先方策



MTO優先方策

MTS/MTO生産が終わった時点で次のいずれかを判断

➤ MTO生産

- ✓ MTO注文が到着している場合、MTO生産を行う

➤ MTS生産

- ✓ MTO注文が到着しておらずMTS在庫が0である場合はロットサイズ δ のMTS生産を行う
- ✓ MTS生産中にMTO注文が到着した場合、MTS生産を中断する

➤ 待機

- ✓ MTO注文が到着しておらず、MTS在庫が0以上である場合は待機

目次

1. MTS/MTOハイブリッド生産システムについて
2. 研究背景
3. 前提条件
4. 従来モデル
- 5. 提案モデル**
6. 数値実験
7. まとめ・今後の課題

提案モデル(行動価値関数、状態、行動)

行動価値関数(Q値)

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha(r_{t+1} + \gamma \max_a Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)).$$

t : 現在の時間
 a_t : 行動
 r_t : 即時報酬
 s_t : 状態
 γ : 割引率
 α : 学習率
($0 \leq \gamma, \alpha \leq 1$)

状態s

過去j期間のMTS在庫数n・MTO注文待ち数m、
j期前に決定されたセットアップ状態zにより定義

$$s_t = [(n_{t-j}, \dots, n_t), (m_{t-j}, \dots, m_t), (z_{t-j}, \dots, z_t)]$$

z	セットアップ状態
0	MTSモードにセットアップ中
1	MTOモードにセットアップ中

行動a

期間始めの状態に応じて決定される

$$a_t = \begin{cases} 0, & \text{setup machine to MTS mode} \\ 1, & \text{setup machine to MTO mode} \end{cases}$$

a	行動
0	MTSモードにセットアップ
1	MTOモードにセットアップ

※連続で同じモードを選択した場合はセットアップを行わない(ex.a{0→0})

提案モデル(報酬関数)

報酬関数 r

- ※セットアップ：MTS→MTSモード等の際はかからない
- ※ペナルティ：MTS在庫が上限時にMTSモードを選択
MTO注文待ちが0の時にMTOモードを選択

状態 s で行動 a を決定した時にかかるコスト

$$r_t = - \left(\begin{array}{l} \boxed{C_1 l_{st}} \\ \text{MTS機会損失コスト} \end{array} + \begin{array}{l} \boxed{C_2 i_t} \\ \text{MTS在庫保管コスト} \end{array} + \begin{array}{l} \boxed{C_3 u_t} \\ \text{セットアップコスト} \end{array} \right. \\ \left. + \begin{array}{l} \boxed{C_4 l_{ot}} \\ \text{MTO機会損失コスト} \end{array} + \begin{array}{l} \boxed{C_5 q_t} \\ \text{MTO注文待ちコスト} \end{array} + \begin{array}{l} \boxed{C_6 p_t} \\ \text{ペナルティ} \end{array} \right)$$

r_t : 即時報酬

t : 行動選択を行う時刻

C_1 : MTS製品の機会損失コスト

l_{st} : MTS製品の機会損失数

C_2 : MTS製品の在庫保持コスト

i_t : MTS製品の在庫数

C_3 : セットアップコスト

u_t : セットアップ回数

C_4 : MTO製品の機会損失コスト

l_{os} : MTO製品の機会損失数

C_5 : MTO製品の注文待ちコスト

q_t : MTO製品の注文待ち数

C_6 : ペナルティコスト

p_t : ペナルティ回数

提案モデル(アルゴリズム)

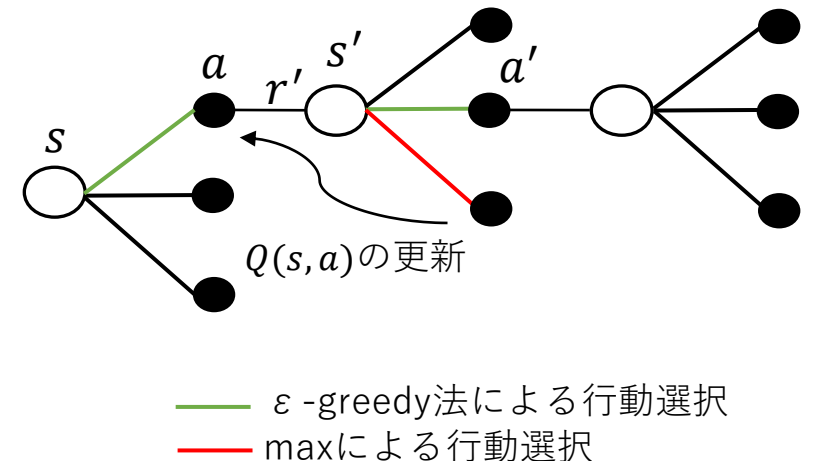
行動価値関数(Q値)

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha(r_{t+1} + \gamma \max_a Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)).$$

t : 現在の時間
 a_t : 行動
 r_t : 即時報酬
 s_t : 状態
 γ : 割引率
 α : 学習率
 ε : ε -greedy法における行動選択率
($0 \leq \gamma, \alpha \leq 1$)

Q学習のアルゴリズム

1. 全ての $s \in S, a \in A(s)$ に対して, Q値を初期化する
2. エピソードの各ステップを繰り返す
 - I. Q値から導出された方策を使用して状態 S から行動 A を選択
 - II. 行動 A を取り, 報酬 R を受け取り, 状態 S を観察
 - III. 行動価値関数 Q を更新
 - IV. 状態 s を更新 ($S \leftarrow S'$)



目次

1. MTS/MTOハイブリッド生産システムについて
2. 研究背景
3. 前提条件
4. 従来モデル
5. 提案モデル
- 6. 数値実験**
7. まとめ・今後の課題

評価・実験方法

コスト関数F

$$F = C_1 L_s + C_2 I + C_3 U + C_4 L_o + C_5 Q$$

MTS機会
損失コスト

MTS在庫
保管コスト

セットアップ
コスト

MTO機会
損失コスト

MTO注文
待ちコスト

実験方法

- ✓ 固定ロットサイズ方策・MTS優先方策・MTO優先方策を比較方策とする
- ✓ 10000のシミュレーション期間、MTS在庫上限HとMTO注文待ち上限Vと、各方策におけるパラメータを変化させた結果のうち、最も総コストが小さい解について、総コストと各コストの内訳を比較検討
- ✓ 学習率 α ・MTS/MTO注文の到着率を変化させた場合についての提案方策の有用性を比較検討

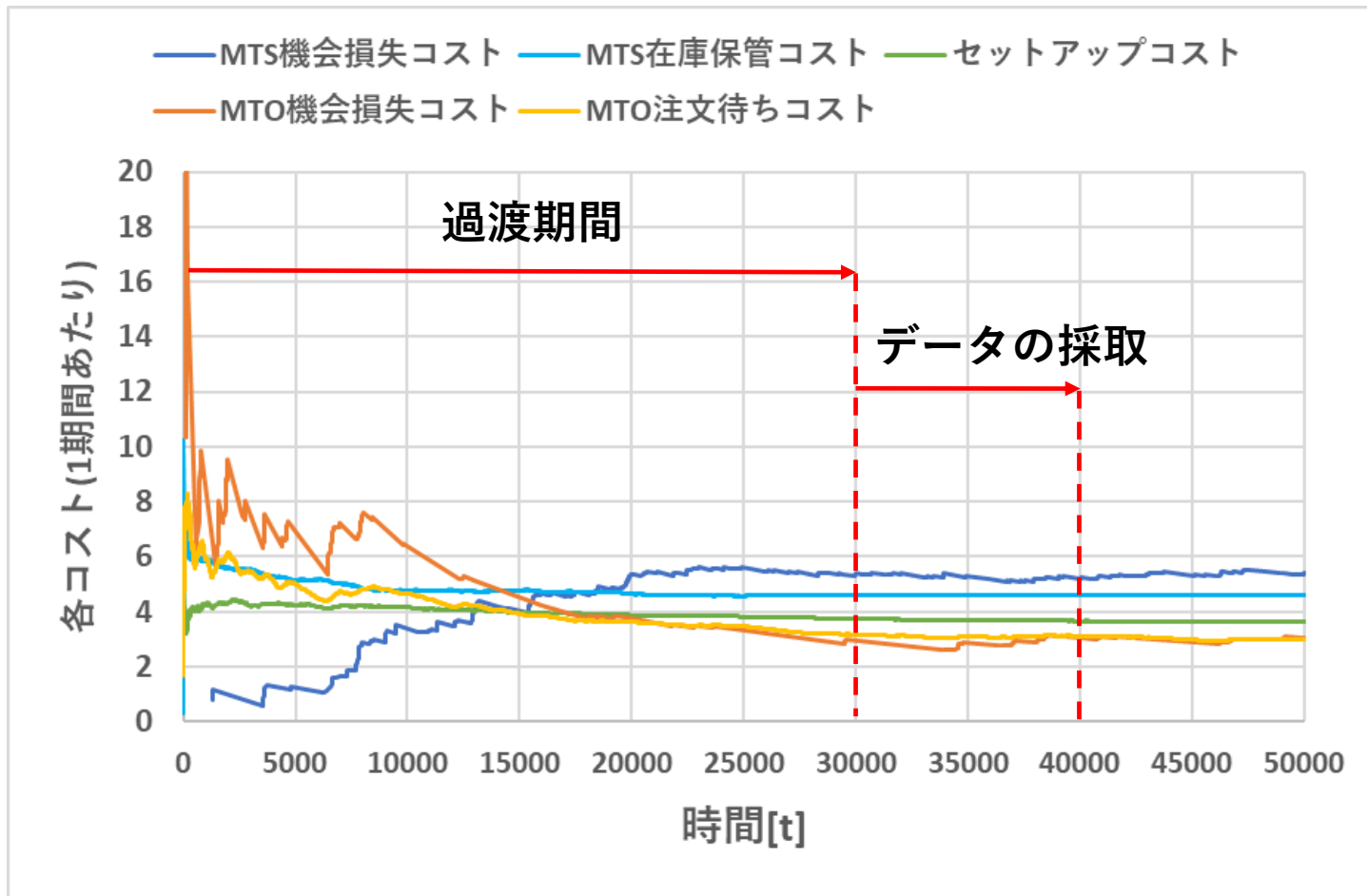
C_1 : MTS注文の機会損失単位コスト
 L_s : MTS注文の平均機会損失数
 C_2 : MTS注文の在庫保持単位コスト
 I : MTS注文の平均在庫数
 C_3 : セットアップ単位コスト
 U : セットアップの平均回数
 C_4 : MTO注文の機会損失単位コスト
 L_o : MTO注文の平均機会損失数
 C_5 : MTO注文の待ち単位コスト
 Q : MTO注文の平均待ち数

数値実験のパラメータ設定

表2：各パラメータ値

パラメータ	実験条件	パラメータ	実験条件
MTS需要到着率 λ_s	0.15, 0.20, 0.25	即時報酬の割引率 γ	0.7
MTO需要到着率 λ_o	0.15, 0.20, 0.25	ϵ -greedy法における ϵ	0.01
観測する過去の状態数 j	0, 1, 2	学習率 α	0.01, 0.03, 0.06, 0.1, 0.2, 0.3, 0.45, 0.6, 0.8, 0.99
MTS製品機会損失コスト C_1	500	MTO生産時間 l_o	$k=2, 1/\mu=1.5$ のアーラン分布
MTS製品の在庫保持コスト C_2	1	MTS生産時間 l_s	1
セットアップコスト C_3	50	固定方策ロットサイズ β	1, 2, 3, 4, 5, 6, 7, 8, 9, 10
MTO製品機会損失コスト C_4	500	MTO優先方策 ロットサイズ δ	1, 2, 3, 4, 5, 6, 7, 8, 9, 10
MTO注文待ちコスト C_5	1	MTS在庫上限 H	1, 2, 3, 4, 5, 6, 7, 8, 9, 10
セットアップ時間 τ_s	2	MTO注文待ち上限 V	1, 2, 3, 4, 5, 6, 7, 8, 9, 10

数値実験1(各コストの収束)



- ✓ 提案方策(j=2)は(j=0, 1)と比べ、状態空間が大きく収束に時間かかる
- ✓ 図1より各コストの収束に約30,000期間を要するため、そこから10,000のシミュレーション期間のデータを採取する

図1：提案方策(j=2)における各コストの収束($\lambda_s = 0.20, \lambda_o = 0.20$)

数値実験2(学習率 α の変更)

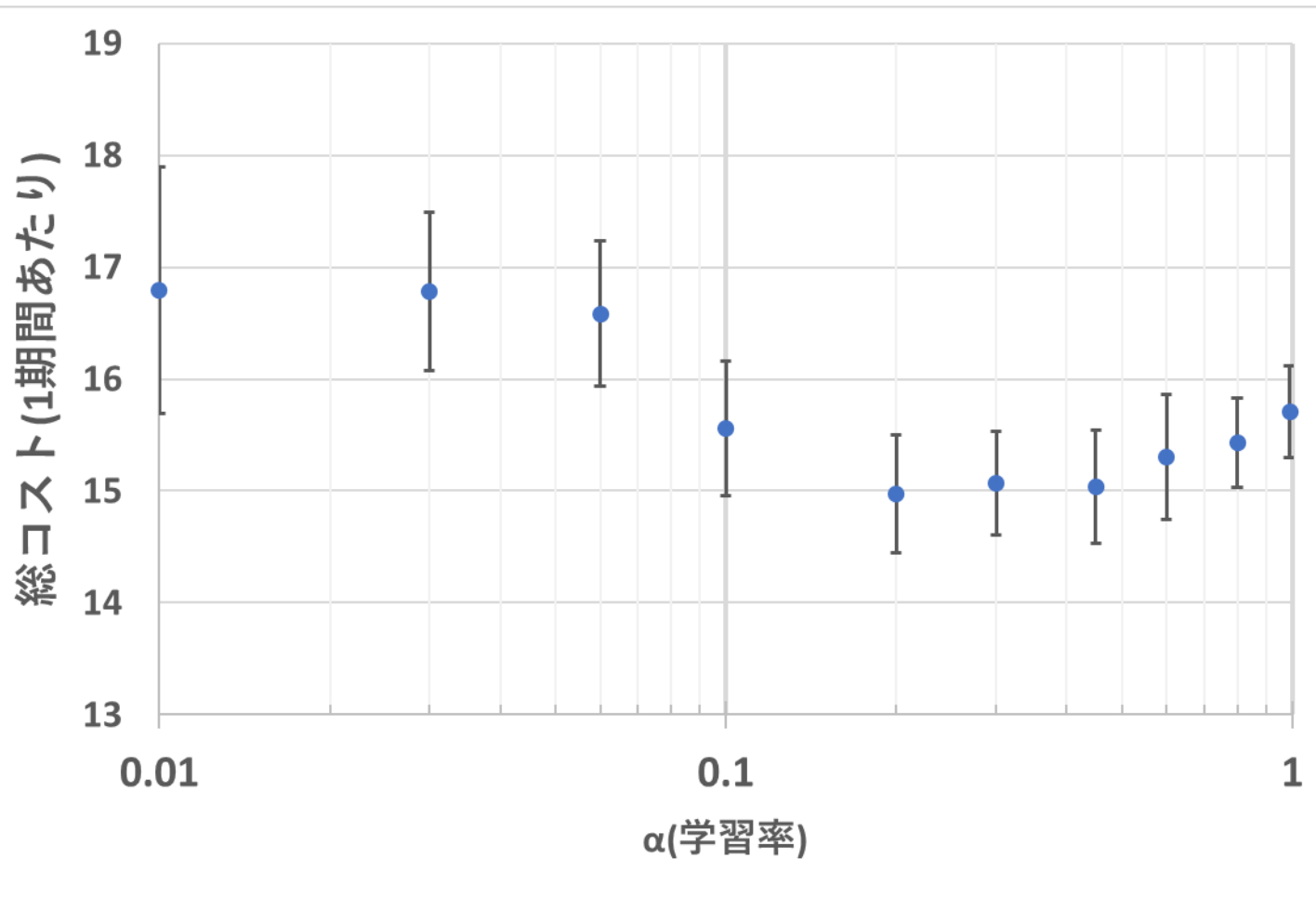


図2：学習率 α ($0 \leq \alpha \leq 1$) を変更した場合の提案方策の比較
(総コストの期待値と95%信頼区間)

学習率 α が $0.2 \leq \alpha \leq 0.5$ 付近の値をとるとき、総コストは小さくなった

➤ 学習率が小さいとき

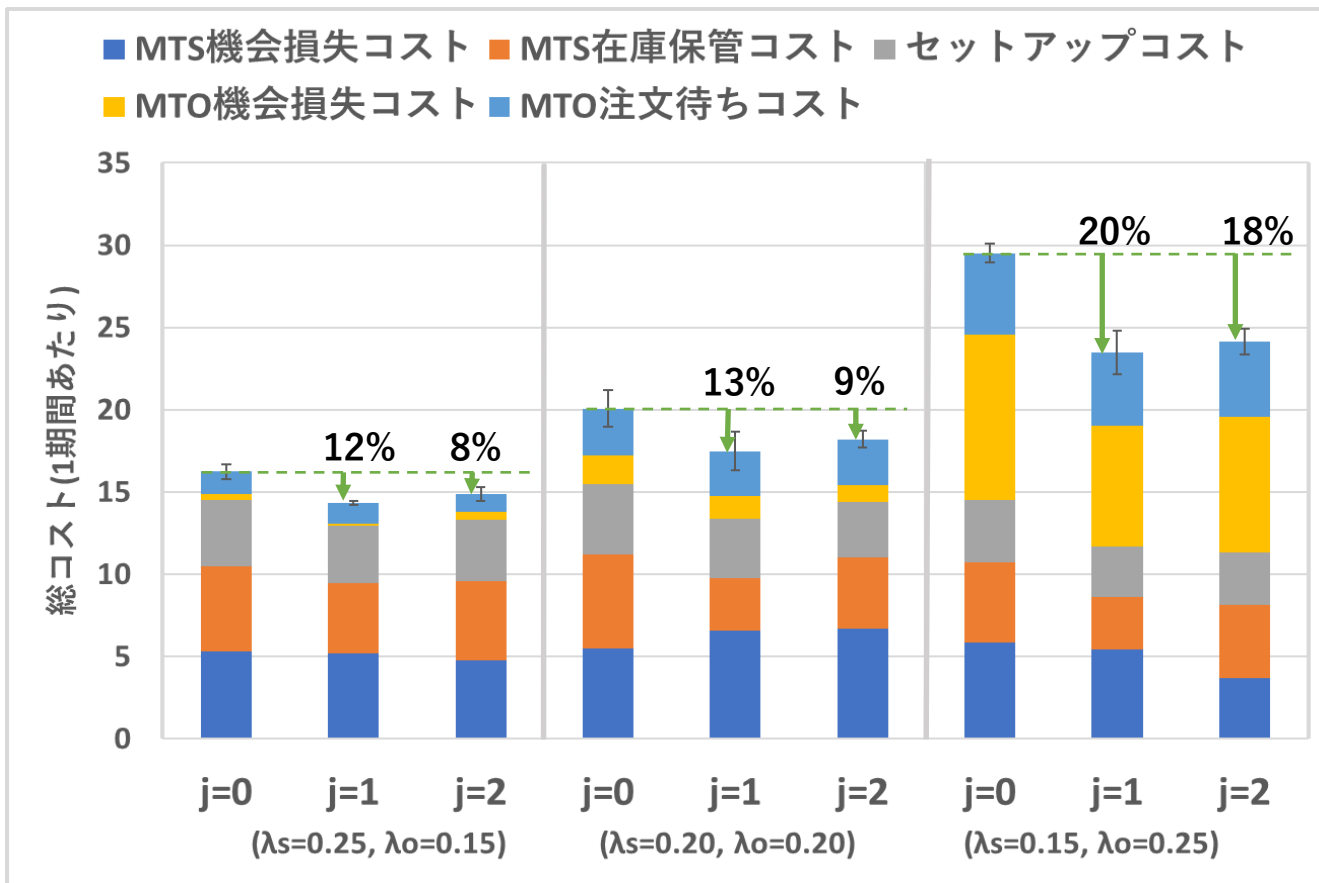
- ✓ Q値の収束値の精度 → 向上
- ✓ Q値の収束時間 → 長い

➤ 学習率が大きいとき

- ✓ Q値の収束値の精度 → 低下
- ✓ Q値の収束時間 → 短い

早く適切な行動価値関数を得ることと、精度の高い行動価値関数を得ることの間にトレードオフの関係がある

数値実験3(提案方策の比較)



✓ 現在の状態のみを見る方策(j=0)と比べ、過去を見る方策(j=1, j=2)は**総コストが小さくなった**

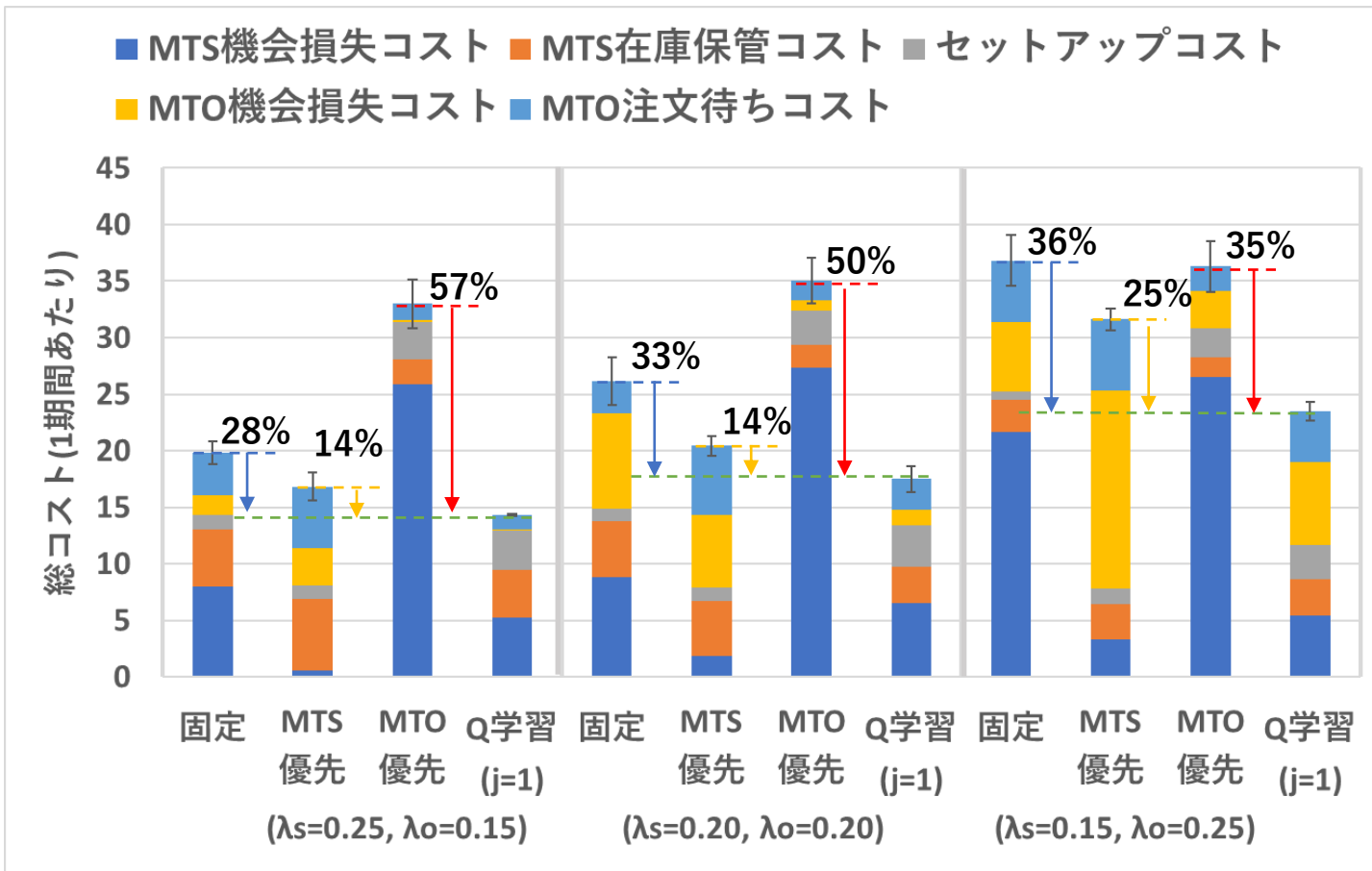
✓ j=1とj=2比較すると、状態として観測する期間を増加させてもコストが減少するとは限らない

➤ 到着する需要がランダムなため

✓ 総コストの値が異なる場合でも95%信頼区間が重なっておらず、提案手法の優位性が認められる

図3：需要到着率を変更した場合の提案モデルの比較
(総コストの期待値と95%信頼区間)

数値実験4(従来方策との比較)



✓ 提案方策(強化学習)は従来方策よりコストが**最大57%減少した**

✓ 従来方策と比べ、**状況に応じてセットアップを行う**ことにより他コストを抑える

✓ 総コストの値が異なる場合でも95%信頼区間が重なっておらず、提案手法の優位性が認められる

図4：需要到着率を変更した場合の提案方策と従来方策との比較 (総コストの期待値と95%信頼区間)

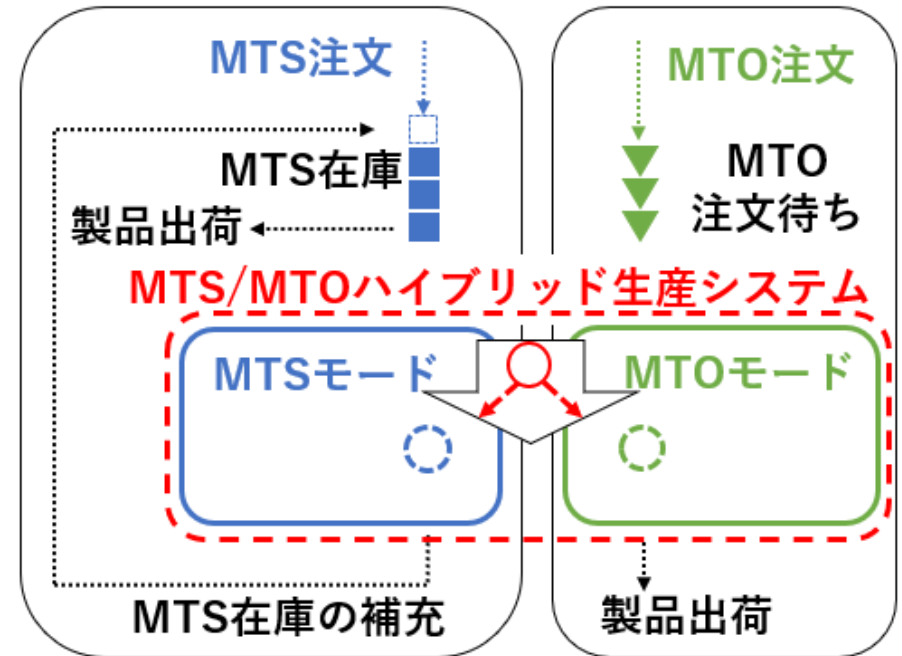
目次

1. MTS/MTOハイブリッド生産システムについて
2. 研究背景
3. 前提条件
4. 従来モデル
5. 提案モデル
6. 数値実験
- 7. まとめ・今後の課題**

まとめ・今後の課題

まとめ

- ✓ 過去の状態までを観測し行動を決定するQ学習を適応させたMTS/MTOハイブリッド生産システムのモード切り替え方策を提案
- ✓ 観測する過去の状態数を変更した結果、現在の状態のみを観測する方策と比べ、過去の状態までを観測する方策は総コストが減少した
- ✓ 注文到着率を変更した場合でも、従来方策と比べ総コストが減少することを示した



今後の課題

- ✓ MTO注文における納期を考慮した環境に適応できる方策の提案
- ✓ 複数台数の生産機械を考慮した環境に適応できる方策の提案

参考文献

- B.Beemsterboer, M. Land, R. Teunter, "Flexible lot sizing in hybrid make-to-order/make-to-stock production planning," European Journal of Operational Research, Vol.260, pp.1014-1023, 2017.
- 白須充啓, "複数台のハイブリッド生産マシンにおける切り替え方式のマルコフ決定過程解析," 広島大学大学院先進理工系科学研究科, 令和2年度修士論文, 2020.
- Huang, Jing, Q.Chang, and J.Arinez. "Deep reinforcement learning based preventive maintenance policy for serial production lines." Expert Systems with Applications Vol.160, 113701, 2020.
- Y.Sato, H.Maeda, R.Toshima, K,Nagasawa, K.Katsumi, K.Takahashi, "Switching decisions in a hybrid MTS/MTO production system comprising multiple machines considering setup." International Journal of Production Economics Vol.263, 108877, 2023.
- Karimi-Majd, Amir-Mohsen, Masoud Mahootchi, and Amir Zakery. "A reinforcement learning methodology for a human resource planning problem considering knowledge-based promotion." Simulation Modelling Practice and Theory 79 (2017): 87-99.

参考文献

- Zheng, Shuai, Chetan Gupta, and Susumu Serita. "Manufacturing dispatching using reinforcement and transfer learning." Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2019, Würzburg, Germany, September 16–20, 2019, Proceedings, Part III. Springer International Publishing, 2020.
- C-F.Chein, Y-S.Lin, and S-K.Lin. "Deep reinforcement learning for selecting demand forecast models to empower Industry 3.5 and an empirical study for a semiconductor component distributor." International Journal of Production Research Vol.58, pp.2784-2804, 2020.
- Cheng, Lixin, et al. "Scheduling flexible manufacturing cell with no-idle flow-lines and job-shop via Q-learning-based genetic algorithm." Computers & Industrial Engineering 169 (2022): 108293.
- 戸島隆晟, “MTS/MTOハイブリッド生産システムに対する動的ロットサイジングによる生産方策,” 広島大学工学部第二類電気システム制御過程令和2年度卒業論文, 2021.
- 数理学とコンピュータサイエンスのメディア MSIISM by NTTデータ数理システム, 「S4で始める強化学習」, <https://www.msiism.jp/article/s4-rl-3.html>, (参照 2023-11-30)

謝辞

本研究の一部は、JSPS科研費 JP20H02390 の助成を受け、実施したものであります。

また本研究の実施にあたり、S⁴ Simulation Systemを貸与して頂いた上、ご担当頂いた皆様には度々の質問に対し真摯にご対応頂き、これまで研究を進めることができました。
この場を借りて、NTTデータ数理システム株式会社、またご担当頂いた皆様に深く感謝申し上げます。