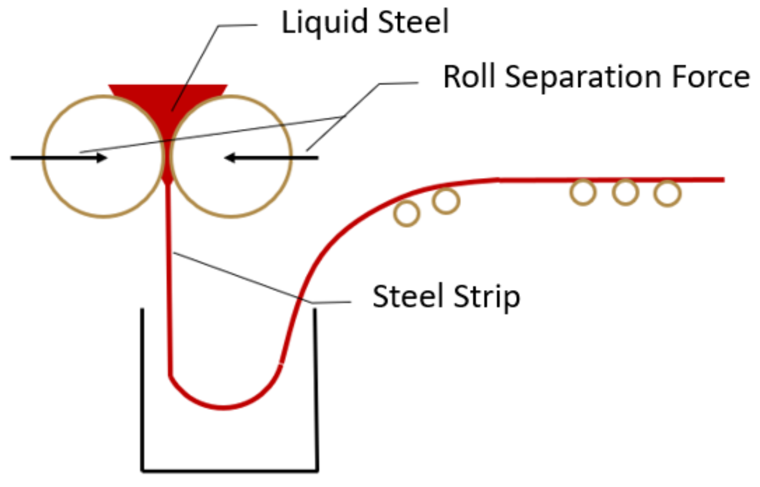
Determining Optimal Decision-Making Sequence for Castrip® Startup Process

Research Assistants: Jianqi Ruan
Principal Investigator(s): Dr. Neera Jain and Dr. George Chiu
Contact email: *ruan27@purdue.edu*

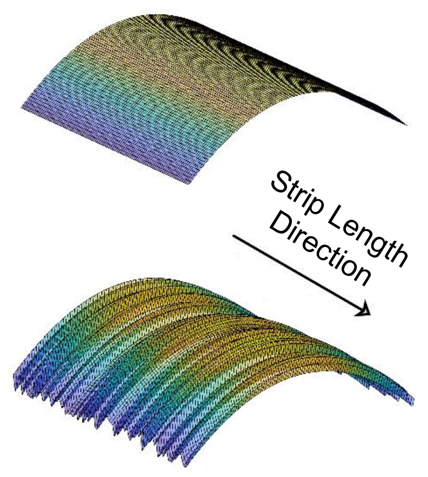
Project Objective

Problem Description:

- During startup, human operators adjust process setpoints to drive the system to steady-state operation that satisfies quality metrics.
- The roll separation force setpoint is the most frequently adjusted setpoint during startup.
- However, each operator uses their own policy for adjusting the force setpoint, introducing variations in product quality.



The roll separation force acts on the rolls and affects different strip characteristics, foremost the strip chatter.



Strip chatter: strip thickness variation along the strip length direction

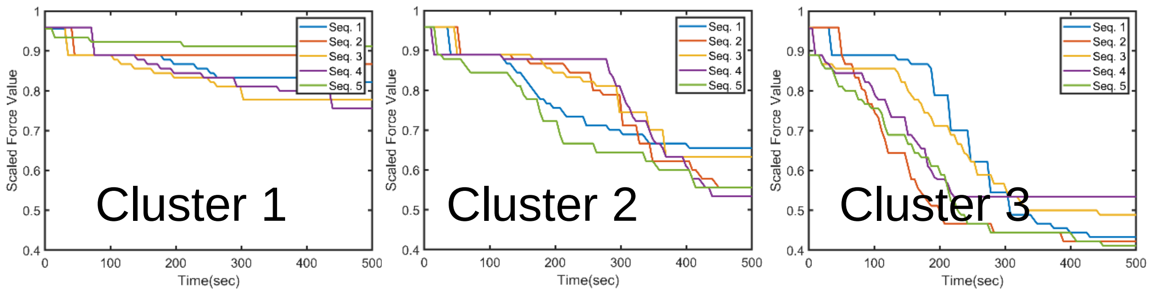
Project Objectives:

- Determine the optimal control policy according to both explicit control objectives (e.g., short startup time, low chatter level) and implicit objectives revealed by human operator behavior.

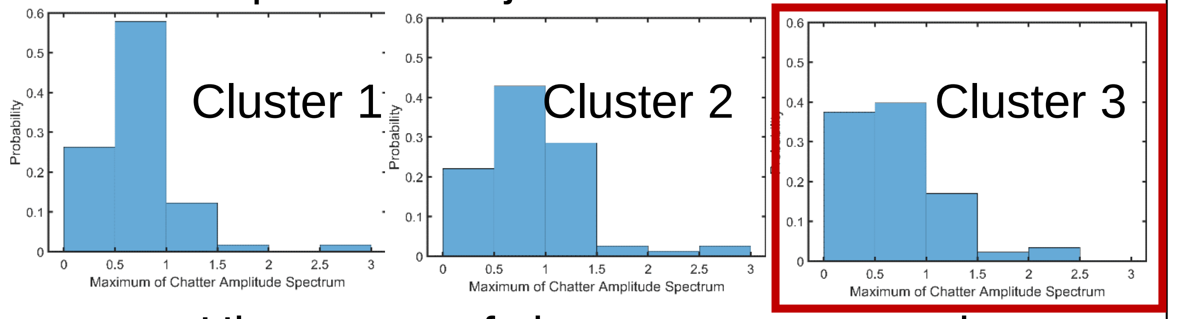
Approach

Operator Behavior Analysis

- Employ unsupervised clustering method to classify sequences based on force setpoint trajectory.
- Select a preferred behavior based on user-defined objective metrics.



Example force trajectories in different clusters



Histograms of chatter spectrum peak

Normalized time performance

Cluster	1	2	3
Startup mean	0.99	1.00	0.94
Startup st.dev.	0.78	1.00	1.21

🟥 Cluster 3 is preferred because of its time and chatter performance

Approach (cont.)

Policy Searching

- Employ a modified deep Q network (DQN) to estimate the value of each state-action pair.

Set state:
$S_k = F_k, C_k, \delta(F_k), \delta(C_k)$

Set action:
$A_k = \delta(F_{k+1})$

Q network:
$Q(S_k, A_k) = q_k$
$q \in \mathbb{R}$

$\delta(x_k) = x_k - x_{k-1}$
F force, C chatter

- Define reward function:

$$R_k = R(C_k, T_{st}, P)$$

Reward is a function of current step chatter value, startup time, and if the behavior is marked as preferred.

- Update target value q_k where $q_k = R_k + \gamma \max_A Q(S_{k+1}, A)$
- At the end of training, the Q network should approximate the long-term total reward:

$$Q(S_k, A_k) \approx \mathbb{E}\left[\sum_{i=k}^{\infty} \gamma^{i-k} R_i\right]$$

PURDUE UNIVERSITY®

The Ray W. Herrick Laboratories

Results

Policy Implementation

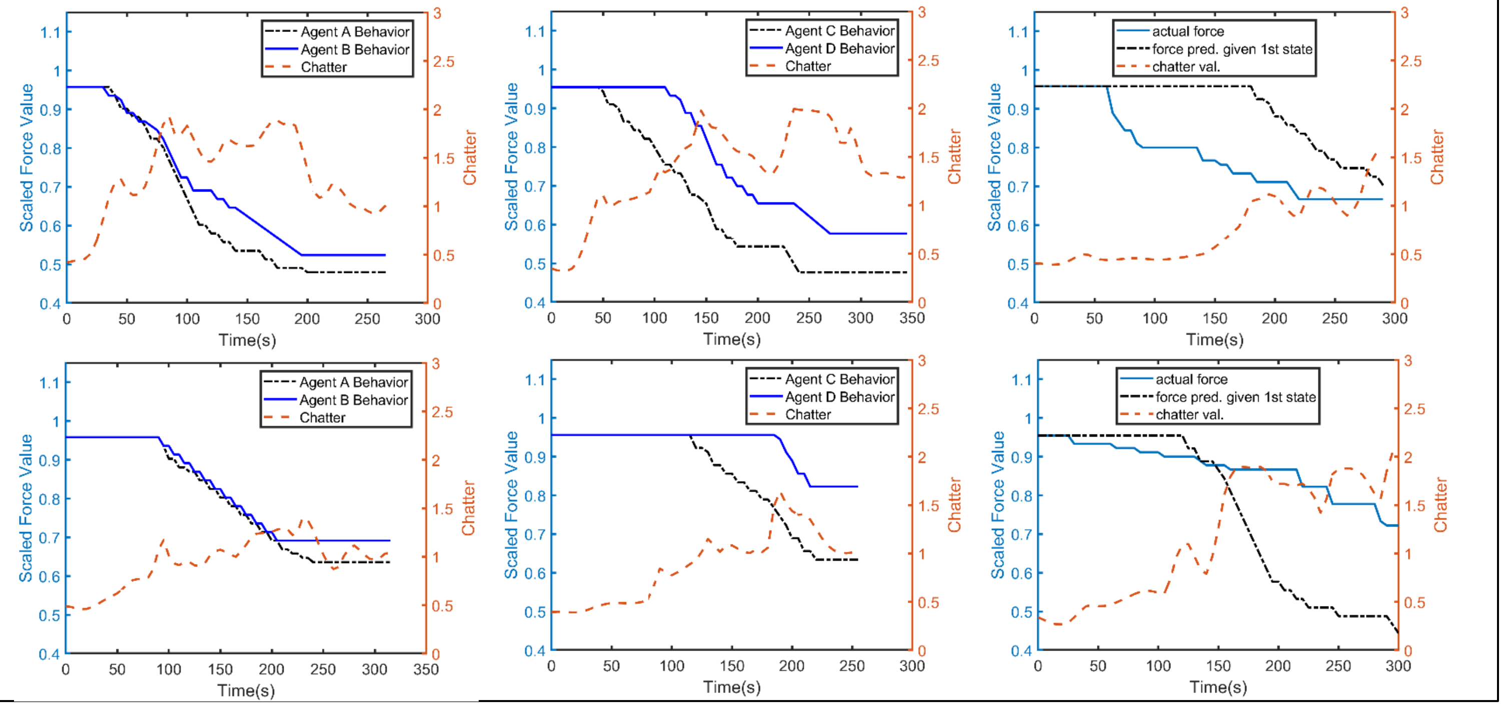
- The agent is tested by using cast sequence data.
- It is required to independently adjust the force setpoint according to the trained Q network: $A_k = \text{argmax}_A(Q(S_k, A))$
- The agent's decision on force adjustment affects the force components at the next state, but the chatter components are imported from the recorded cast sequence.

Sensitivity to Variations in the Reward Function

Agent A is trained with reward only preferring Cluster 3, and Agent B is trained with reward preferring a mixed group of Clusters 2 and 3.

Agent C is trained with a reward function whose chatter tolerance is lower (to assign negative chatter reward as chatter value is lower)

In some cases, an agent behavior can be more preferrable than the operators in the sense of reacting to chatter value change.



* Learned from Castrip engineers, lowering the roll separation force may reduce the chatter level. Hence, we consider that the proper behavior of an agent is to reduce the force setpoint when the chatter exceeds the tolerance and/or has a strong increasing trend.

Future Work

- Consider other casting parameters potentially affected by the roll separation force and therefore affecting the force setpoint decision-making.
- Extend the optimal policy determination to a steady-state setpoint adjustment scenario.

Acknowledgements

We want to thank Castrip LLC for supporting this research, and Wal Blejde and his colleagues for lending their expertise on the twin-roll casting process.