Development of a Reinforcement Learning-Based Control Strategy for Load Following in Supercritical Pulverized Coal (SCPC) Power Plants

Elijah Hedrick[1], Katherine Reynolds[1], Parikshit Sarda[1], Debangsu Bhattacharyya[1], Stephen E. Zitney[2], and Benjamin Omell[3]

[1] *Department of Chemical and Biomedical Engineering, West Virginia University, 395 Evansdale Drive, Morgantown, WV 26506-6070, USA*

[2] *National Energy Technology Laboratory, 3610 Collins Ferry Road, Morgantown, WV 26507, USA*

[3] *National Energy Technology Laboratory, 626 Cochrans Mill Road, Pittsburgh, PA 15236, USA*

Clearwater Clean Energy Conference – Clearwater, Fl

16-21 June 2019

West Virginia University
BENJAMIN M. STATLER COLLEGE OF ENGINEERING AND MINERAL RESOURCES

NATIONAL ENERGY TECHNOLOGY LABORATORY

U.S. DEPARTMENT OF ENERGY

Outline

- Introduction
- Theory
- Control Application
- Algorithm
- Results
- Conclusions
- Bibliography

Outline

- **Introduction**
- Theory
- Control Application
- Algorithm
- Results
- Conclusions
- Bibliography

Introduction

- Recent years have seen ever increasing availability of renewables in the energy market
  - While these sources are taking up a larger and larger market share, they suffer from intermittency with naturally fluctuating weather conditions
- This increase in overall production with short term dry spells has posed a problem for traditional production facilities, where the real-time demand can now fluctuate significantly – and quickly – necessitating operational changes

Load-Following

- In order to account for renewables, traditional plant must now cycle their loads in order to meet the real-time demand
  - This is both costly and operationally difficult in plants that were designed for static operational at the nominal load
- Fast changes in plant conditions are also causing lasting damage to key processes equipment
  - Notably in the boiler, where fluctuations in the main (final) steam temperature can cause creep and fatigue
- Effective load-following necessitates the development of better control strategies to quickly move the plant from one load to another, while controlling key process variables tightly

Load-Following

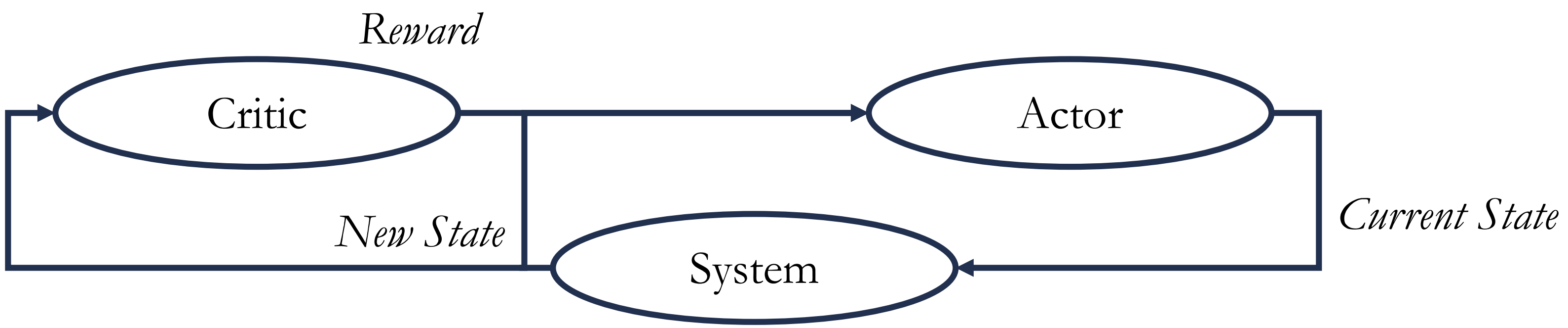
- While each individual load change will clearly be different, we seek to approach the problem by asking how they are similar
  - If each load change shares certain characteristics, how can we learn from them (and incorporate them dynamically into our control system) in order to obtain improved control over a large range of conditions?
- As the equipment performance changes over time, the control system must adapt for maximizing its performance
- In the existing control system designs, 'lessons learnt' from current or past control strategies and actions are not utilized in the future even when same or similar control challenges appear
  - Repetitive mistakes are made
- Can the control system learn from its performance for a given load-following task and then adapt when 'similar' task appear in the future?

Outline

- Introduction
- **Theory**
- Control Application
- Algorithm
- Results
- Conclusions
- Bibliography

Reinforcement Learning

- RL is a model-free process whereby an actor takes action on a Markov decision process, from which the transitioned state is evaluated by a critic[1]
- As this process continues, the knowledge base of the learner is converged, yielding a mapping of states to optimal policies
- This mapping is often recorded as a lookup table that has been learned offline, though online implementation is possible



Q-Learning

- Q-learning is an iterative tabular method for the evaluation of a *quality* value for a given state-action pair[1,7]
  - Convergence of this table yields the optimal policy mapping
- For each of the state-action pairs, the quality can be updated as:

$$\boldsymbol{Q}(s_t, a_t) \leftarrow \boldsymbol{Q}(s_t, a_t) + \alpha \left[r_t + \gamma \max_a \boldsymbol{Q}(s_{t+1}, a) - \boldsymbol{Q}(s_t, a_t) \right]$$

New value of Q | **Current value of Q** | **Learning Rate** | **Reward** | **Discount Rate** | **Maximum future value of Q**

The Curse of Dimensionality

- Q-learning algorithm involves no optimization and can be easily solved faster than real-time thus making it very practical for real-life application in small to medium state-space systems
- For large state-space systems such as the power plants, the number of computations required to fill out the Q-Table can increase combinatorically (the Curse of Dimensionality[9]) making the algorithm computationally intractable
- Here we propose to use a clustering approach to collapse the state-space into a number of tractable number of clusters from the perspective of their control space[8]
    - The control space here consists of a 3xN dimensional space of discrete control parameter selections (i.e. K_c, τ_I, τ_D)

K-means Clustering

- K-means clustering is a data management algorithm used to minimize mean square error of a data set relative to a given number of cluster centers[8]
  - These cluster centers can then be taken to represent the data, with each center representing its member states
  - Each new state is evaluated with reference to the existing cluster centers
  - The resulting cluster centers are used directly in place of the state space
  - Given N cluster centers, C and a new state, S:

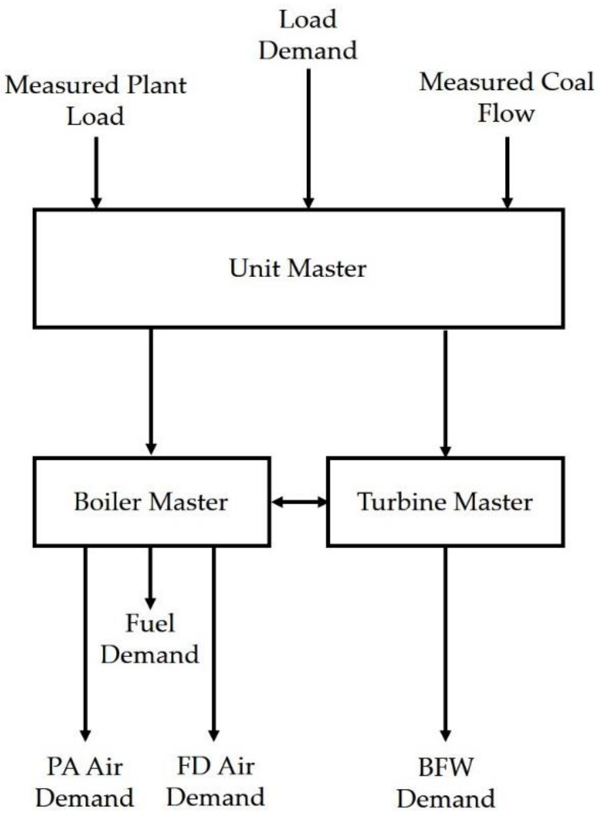
1. $d_{NS} = \|C_i - S_i\|$
2. If $d_{NS} \leq \rho$, the state is already represented → Use the nearest C_i and keep N centers
3. If $d_{NS} > \rho$, the state needs representation → Use as kernel for new C_i and take N+1 centers
4. Perform new clustering considering S and N or N+1 cluster centers

Outline

- Introduction
- Theory
- **Control Application**
- Algorithm
- Results
- Conclusions
- Bibliography

Coordinated Control System

- The coordinated control system is developed in SCPC plants to control the entire plant as a function of the load[6]
- Years of industrial experience have yielded a robust structure
- However, static tuning and structure hinder the system in transient and off-design operation
- The system proposed here could be implemented alongside the CCS with minimal intrusion to the existing structure

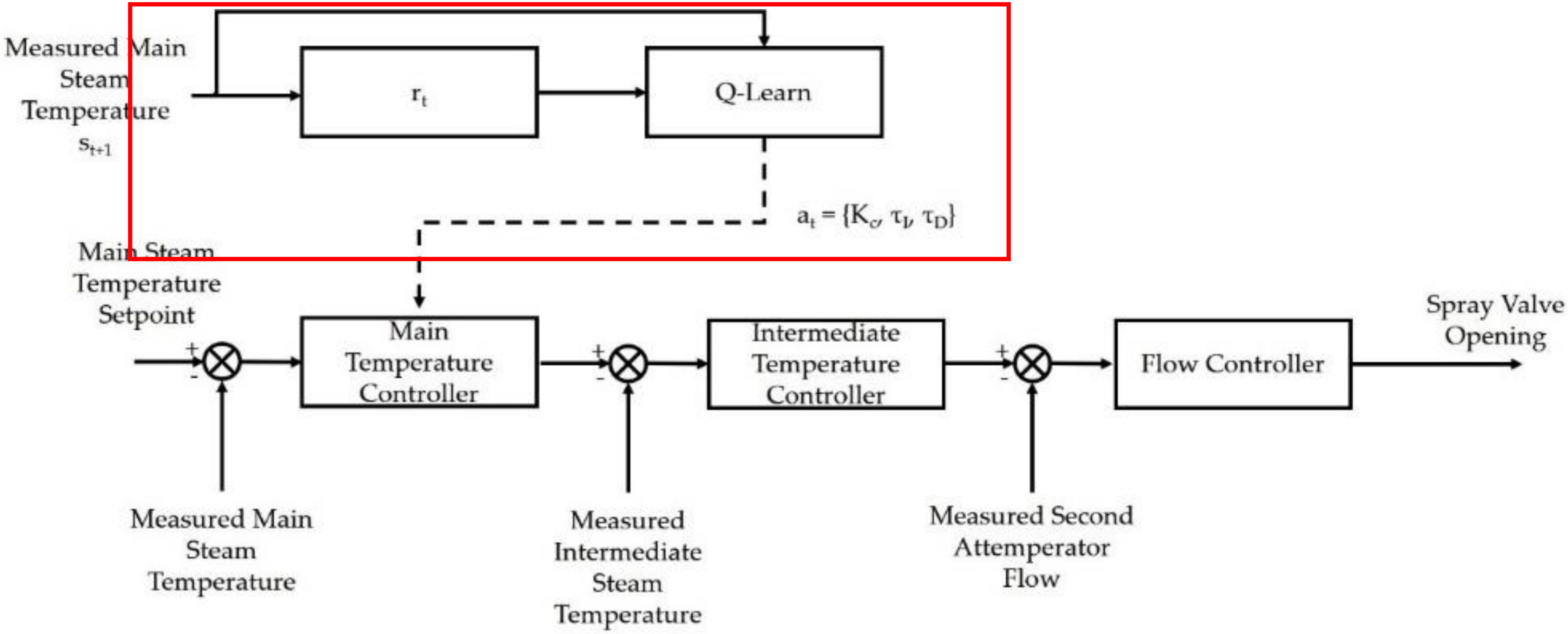


Used with permissions from [6]

Main Steam Temperature Control

- Control of the main steam temperature is process critical
  - Deviation causes efficiency losses along with long-term damage to the steam boiler and turbine
- Control problem also exhibits significant time-delay[6]
- Little is known about off design or transient conditions and operating parameters
  - This point yields a system where a RL superstructure could be effective, given that no model is needed to gain information about these operating states
- It is desired to develop strategies that work within the existing CCS

Control Diagram

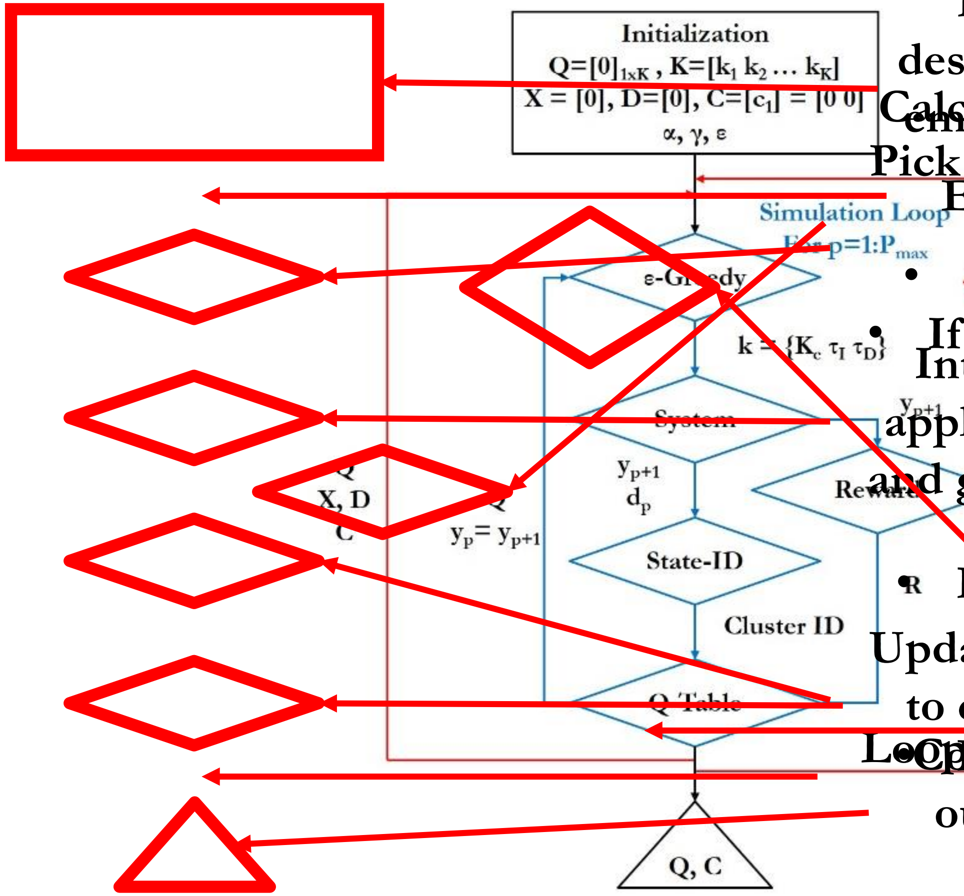


Citations: [6]

Outline

- Introduction
- Theory
- Control Application
- **Algorithm**
- Results
- Conclusions
- Bibliography

Algorithm Flow Chart



Initialize with a fully described action space, an empty state space, and an empty Q-Table

Enter Simulation Loop

Pick an action under random probability

- If P< ε, pick randomly
- If P> ε, pick a_j at max$\{Q_j\}$

Interact with the system, applying tuning parameters and getting the updated state

Calculate gaussian reward based on deviation from set point

- $(y - y_{sp}) \rightarrow 0$, $r \rightarrow 0$
- $(y - y_{sp}) \rightarrow 0$, $r \rightarrow 1$

Identify new state

- If distance from cluster centers is small, add to nearest center
- If distance is large, identified outlier kernel for new cluster

Update the Q-Table according to cluster membership and reward

Loop for the simulation time allowed

Output Q-Table and identified cluster centers

Use Episodic loop under random changes to y_{sp} to converge the Q-Table under reducing ε, picking best control move more often

Resulting System Dimensionality

$s_i \in \boldsymbol{S}$ for the state space
$d_i \in \boldsymbol{D}$ for the disturbance space
$c_i \in \boldsymbol{C}, C \leq S$ for the cluster centers
$a'_i \in \boldsymbol{A'}$ for the condensed action space

$$a'_i = [K_{c,i} \quad \tau_{I,i} \quad \tau_{D,i}]^T$$

$Q_{i,j} \in \boldsymbol{Q}$ for the Q-table

$$\begin{bmatrix} K_{c,1} & \cdots & K_{c,A} \\ \tau_{I,1} & \cdots & \tau_{I,A} \\ \tau_{D,1} & \cdots & \tau_{D,A} \end{bmatrix} \quad \boldsymbol{A}: 3xA$$

$$[a'_1 \quad \cdots \quad a'_A] \quad \boldsymbol{A'}: 1xA$$

$$\begin{bmatrix} s_1 \\ \vdots \\ \vdots \\ s_S \end{bmatrix} \quad \boldsymbol{S}: Sx1$$

$$\begin{bmatrix} d_1 \\ \vdots \\ \vdots \\ d_D \end{bmatrix}$$

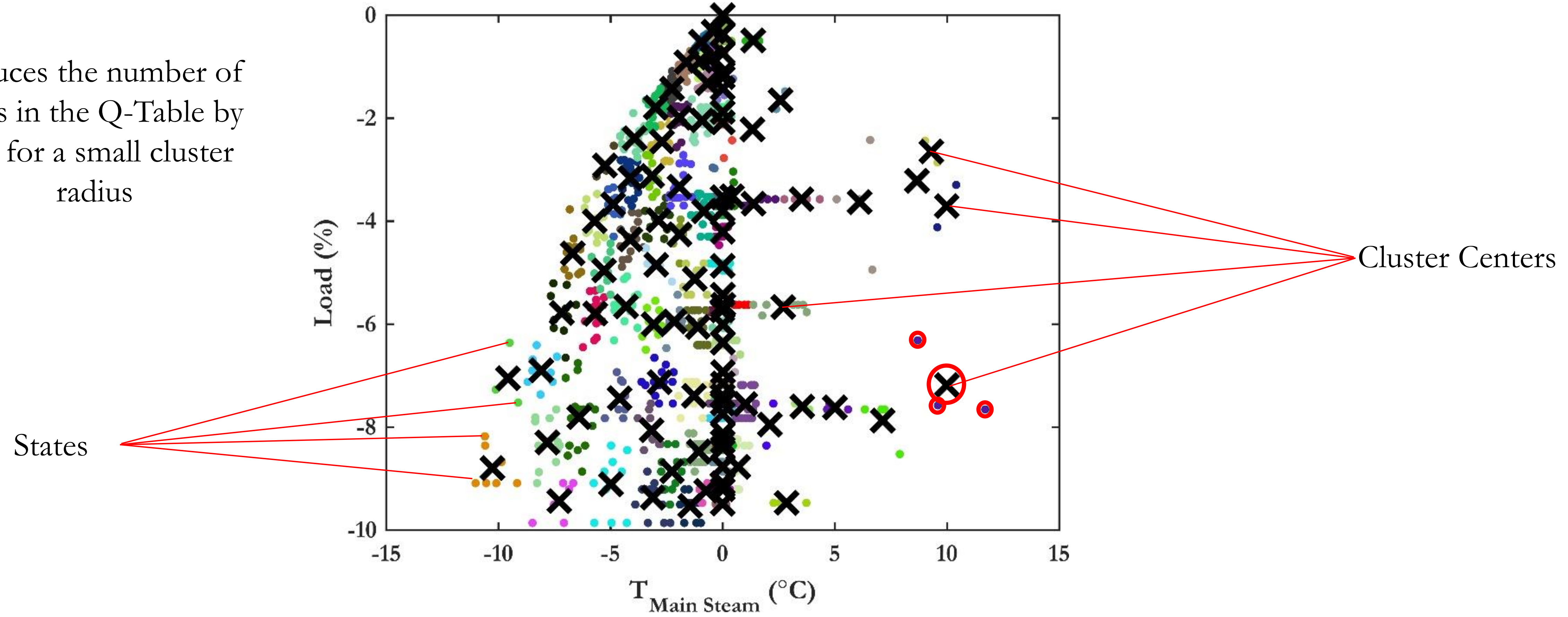
$$\longrightarrow \underset{Sx2}{\begin{bmatrix} s_1 & d_1 \\ \vdots & \vdots \\ s_S & d_S \end{bmatrix}} \longrightarrow \text{K-Means} \longrightarrow \underset{\boldsymbol{C}: Cx1}{\begin{bmatrix} c_1 \\ \vdots \\ c_C \end{bmatrix}} \quad \underset{\boldsymbol{Q}: CxA}{\begin{bmatrix} Q_{11} & \cdots & Q_{1A} \\ \vdots & \ddots & \vdots \\ Q_{C1} & \cdots & Q_{CA} \end{bmatrix}}$$

Outline

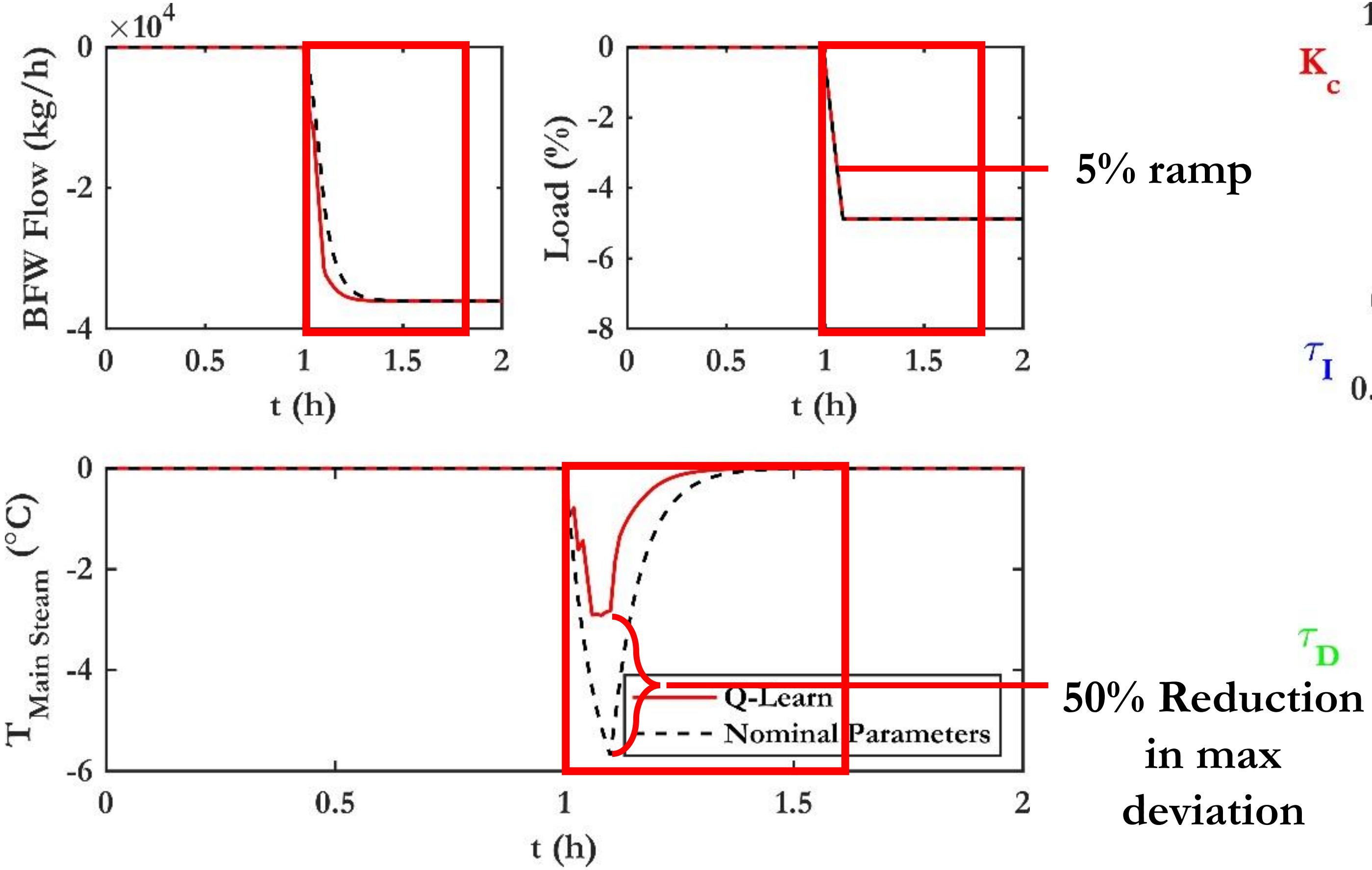
- Introduction
- Theory
- Control Application
- Algorithm
- **Results**
- Conclusions
- Bibliography

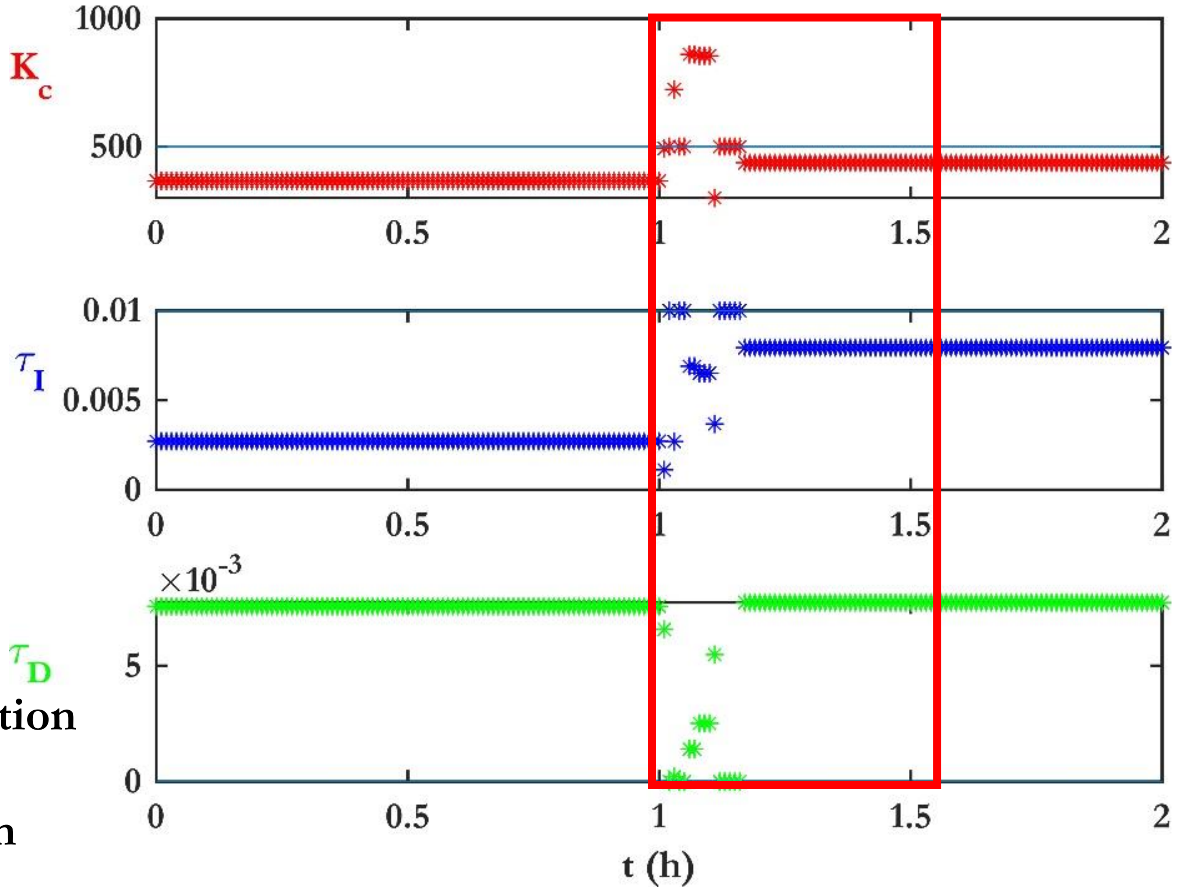
Temperature-Load Clustering

Reduces the number of rows in the Q-Table by 10^2 for a small cluster radius



Control Response and Selected Parameters





Outline

- Introduction
- Theory
- Control Application
- Algorithm
- Results
- **Conclusions**
- **Bibliography**

Conclusions

- Developed an algorithm for a Q-Learning augmented PID control strategy for online tuning
  - This algorithm included the use of k-means clustering for reduction of the size of the state space
- Algorithm was tested on the control of the main steam temperature in an SCPC plant
  - 50% reduction in maximum deviation from the setpoint was achieved

Bibliography

1. R. S. Sutton and A. G. Barto, *Reinforcement Learning - An Introduction*, 2nd ed. United States of America: Westchester Publishing Services, 2018.
2. T. A. Badgwell, J. H. Lee, and K.-H. Liu, "Reinforcement Learning – Overview of Recent Progress and Implications for Process Control," in *Computer Aided Chemical Engineering*, vol. 44, M. R. Eden, M. G. Ierapetritou, and G. P. Towler, Eds. Elsevier, 2018, pp. 71–85.
3. L. A. Brujeni, J. M. Lee, and S. L. Shah, "Dynamic tuning of PI-controllers based on model-free Reinforcement Learning methods," in *ICCAS 2010*, 2010, pp. 453–458.
4. I. Carlucho, M. De Paula, S. A. Villar, and G. G. Acosta, "Incremental Q-learning strategy for adaptive PID control of mobile robots," *Expert Syst. Appl.*, vol. 80, pp. 183–199, Sep. 2017.
5. J. E. Morinelly and B. E. Ydstie, "Dual MPC with Reinforcement Learning," *11th IFAC Symp. Dyn. Control Process Syst. Biosyst. DYCOPS-CAB 2016*, vol. 49, no. 7, pp. 266–271, Jan. 2016.
6. P. Sarda, E. Hedrick, K. Reynolds, D. Bhattacharyya, E. S. Zitney, and B. Omell, "Development of a Dynamic Model and Control System for Load-Following Studies of Supercritical Pulverized Coal Power Plants," *Processes*, vol. 6, no. 11, 2018.
7. C. Watkins, "Learning from Delayed Rewards," Dissertation, Cambridge, London, 1989.
8. S. P. Lloyd, "Least Squares Quantization in PCM," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 129–137, Mar. 1982.
9. R. Bellman, *Dynamic Programming*. Courier Corporation, 2013.

Acknowledgements

- This project was supported by National Energy Technology Laboratory (NETL) through KeyLogic Systems, Inc. Grant#PO-5000-074-001.
- This project was also supported by the Benjamin M. Statler Fellowship

Thank you

Questions?