DATA DRIVEN DYNAMIC MODELING AND REINFORCEMENT LEARNING BASED CONTROL OF BATCH PROCESSES

TANUJA JOSHI

DEPARTMENT OF CHEMICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY DELHI
DECEMBER 2022

DATA DRIVEN DYNAMIC MODELING AND REINFORCEMENT LEARNING BASED CONTROL OF BATCH PROCESSES

by

TANUJA JOSHI

Department of Chemical Engineering

Submitted

in fulfilment of the requirements of the degree of Doctor of Philosophy

to the

INDIAN INSTITUTE OF TECHNOLOGY DELHI

DECEMBER 2022

THESIS CERTIFICATE

This is to certify that the thesis titled **Data Driven Dynamic Modeling and Reinforcement Learning based Control of Batch Processes**, submitted by **Tanuja Joshi (2018CHZ8100)**, to the Indian Institute of Technology Delhi, for the award of the degree of **Doctor of Philosophy**, is a bonafide record of the research work done by her under my supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Prof. Hariprasad Kodamana
Research Supervisor
Associate Professor
Dept. of Chemical Engineering &
Yardi School of Artificial Intelligence
IIT-Delhi, 110016

Date: 7 July 2022

ACKNOWLEDGEMENTS

Foremost, I would like to express my sincere gratitude to my advisor Prof. Hariprasad Kodamana for the continuous support during my Ph.D study, for his patience, motivation, enthusiasm, and immense knowledge. His guidance, timely advise and scientific approach is a source of inspiration and helped me in all the time of research and writing of this thesis. His dedication and keen interest and above all his overwhelming attitude to always help his students had been mainly responsible for completing this thesis.

Besides my advisor, I would like to thank my research committee members Prof. Manojkumar C. Ramteke, Prof. Anupam Shukla, and Prof. Subashish Datta for their encouragement and insightful comments. I especially want to thank Prof. Harikumar Kandath, IIIT Hyderabad, for his valuable inputs, vast knowledge and motivating guidance in this project. Many thanks to Prof. Niket Kaisare, IIT Madras to participate in my research by providing invaluable feedback and make this work possible.

In addition, I owe my thanks to my current fellow labmates, Avan, Reena, Nikita, Jyoti, Umang, Ahtesham, Nasre, Deepak and all of my past labmates, Devansh, Abhyansh, Shikhar, Vishesh, Vikas, Mayuna in the Control & Analytics of Process System (CAPS) Lab, IITD for the stimulating discussions and the positive environment they have created in the lab. Furthermore, I would like to express my sincere thanks to all the office staff members of the the Chemical Engineering Department, IITD.

Last but not the least I would like to thank the almighty god, my parents, family members, and my friends for their unconditional love, prayers and constant support throughout my life.

ABSTRACT

Development of control relevant models with excellent predictive capability is a challenging task for a batch process due to the non-linear dynamics and varying operating conditions within the batch and batch-to-batch. First principles-based models do not provide a remedy to this problem in most cases as they are tedious to develop and strenuous to solve for complex processes. Incidentally, industrial processes are warehouses of a large amount of multivariate data. Hence, data-driven modelling approaches can be conveniently used in this juncture. Traditional approaches of the data-based modeling methods focus on global approaches, however, developing a single global model for the entire batch process may not be a reasonable option because operating conditions change dynamically in industrial processes and may warrant updating the model online. The first part of the thesis focuses on the development of a novel data-driven dynamic model based on the local modeling approach i.e., Just-in-Time Learning (JITL) framework for batch process modeling. The accuracy of the underlying model under the JITL framework is based on the 'similarity measure' used to extract the relevant data from the massive historical database similar to the query point, and the 'weighting strategies' adopted. The proposed formulation incorporates a 'query profile' instead of a 'query point', with a view to bring in the dynamic modeling capabilities. A new 'searching strategy' based on 'profile similarity' is proposed for taking into account the time-varying dynamics of the batch processes. Further, a new 'weighting strategy' is introduced to ensure that the complete dynamics of the batch processes are captured and also to accommodate the outliers in the historical database.

The second part of this thesis focuses on the data-driven control of batch processes.

Control of batch processes is difficult due to their complex nonlinear dynamics and unsteady-state operating conditions within batch and batch-to-batch. Advanced control strategies like Model Predictive Control (MPC) used in the process industries for batch process control employs mathematical programming to solve a constrained, possibly non-convex, optimisation problem. The key issue here is that the online computational burden at each time step to obtain the optimal control input profile is very high. This limits the implementation of these approaches for complex nonlinear, high-dimensional dynamical systems, despite the advancement in computational hardware and numerical methods. Further, the performance of a model-based controller is highly dependent on the availability of an accurate process model. For a complex and nonlinear process, the availability of such a model is a limitation as it requires significant prior knowledge and expertise. Plant-model mismatch occurs even if there is a slight variation in the real process and its approximate model, leading to inaccurate predictions of the performance variable, such as product yield. Even if the process model is available, the controller performance deteriorates in the presence of uncertainties and process drifts. Batch-to-batch variations that occur due to raw material fluctuations, cleaning, etc., are a source of uncertainty that further deteriorate the closed-loop performance. In this juncture, the development of a control strategy that does not entirely rely on the knowledge of the accurate dynamics of the system and can handle the stochastic dynamics and plant-model mismatches is extremely useful. Model-free Reinforcement Learning (RL), where the agent (analogous to the controller) learns the optimal control action (analogous to control input) by directly interacting with the operational environment (analogous to the process), offers a potential alternative to traditional model-based approaches for process control. RL frameworks with actor-critic architecture have recently become popular for the control process systems where both the state and action spaces are continuous. Subsequent works focus on the development of Actor-Critic RL based controller by developing two novel Actor-Critic RL algorithms, namely, (i) Twin Actor Twin Delayed Deterministic Policy Gradient (TATD3), a deterministic RL algorithm, and (ii) Twin Actor Soft Actor-Critic (TASAC), a stochastic RL algorithm, for batch process control by incorporating an 'en-

semble of actors' in the Actor-Critic algorithms for training the policies to achieve an overall optimal policy for batch process control.

The efficacies of the developed approaches in this work are evaluated by simulation studies on batch process case studies. In a nutshell, the overall objective is to develop a data-driven dynamic model and a model-free RL-based controller for complex, nonlinear batch processes to ensure optimal operation of the batch processes.

KEYWORDS: data-driven model; just-in-time learning; reinforcement learning; deep-Q-learning; deep deterministic policy gradient; actor-critic algorithms; batch process control

सार

उत्कृष्ट भविष्यवाणी क्षमता वाले नियंत्रण प्रासंगिक मॉडलों का विकास, गैर-रैखिक गतिशीलता और बैच और बैच-टू-बैच के भीतर बदलती परिचालन स्थितियों के कारण बैच प्रक्रिया के लिए एक चुनौतीपूर्ण कार्य है। पहले सिद्धांत-आधारित मॉडल ज्यादातर मामलों में इस समस्या का समाधान नहीं प्रदान करते हैं क्योंकि वे जटिल प्रक्रियाओं के लिए विकसित करने और हल करने के लिए कठिन हैं। संयोग से, औद्योगिक प्रक्रियाएं बड़ी मात्रा में बहुभिन्नरूपी डेटा के गोदाम हैं। इसलिए, इस मोड़ पर डेटा-संचालित मॉडलिंग दृष्टिकोणों का आसानी से उपयोग किया जा सकता है। डेटा-आधारित मॉडलिंग विधियों के पारंपरिक दृष्टिकोण वैश्विक दृष्टिकोण पर ध्यान केंद्रित करते हैं, हालांकि, संपूर्ण बैच प्रक्रिया के लिए एकल वैश्विक मॉडल विकसित करना एक उचित विकल्प नहीं हो सकता है क्योंकि परिचालन की स्थिति औद्योगिक प्रक्रियाओं में गतिशील रूप से बदलती है और मॉडल को ऑनलाइन अपडेट करने की आवश्यकता हो सकती है। थीसिस का पहला भाग बैच प्रोसेस मॉडलिंग के लिए स्थानीय मॉडलिंग दृष्टिकोण यानी जस्ट-इन-टाइम लर्निंग (JITL) फ्रेमवर्क के आधार पर एक उपन्यास डेटा-संचालित गतिशील मॉडल के विकास पर केंद्रित है। जेआईटीएल ढांचे के तहत अंतर्निहित मॉडल की सटीकता 'समानता माप' पर आधारित है जिसका उपयोग क्वेरी बिंदु के समान विशाल ऐतिहासिक डेटाबेस से प्रासंगिक डेटा निकालने के लिए किया जाता है, और अपनाई गई 'भार रणनीति'। गतिशील मॉडलिंग क्षमताओं को लाने की दृष्टि से प्रस्तावित सूत्रीकरण में 'क्वेरी पॉइंट' के बजाय 'क्वेरी प्रोफाइल' शामिल है। बैच प्रक्रियाओं की समय-भिन्न गतिशीलता को ध्यान में रखते हुए 'प्रोफ़ाइल समानता' पर आधारित एक नई 'खोज रणनीति' प्रस्तावित है। इसके अलावा, यह सुनिश्चित करने के लिए एक नई 'वेटिंग स्ट्रैटेजी' पेश की गई है कि बैच प्रक्रियाओं की पूरी गतिशीलता पर कब्जा कर लिया गया है और ऐतिहासिक डेटाबेस में आउटलेयर को भी समायोजित किया गया है।

इस थीसिस का दूसरा भाग बैच प्रक्रियाओं के डेटा-संचालित नियंत्रण पर केंद्रित है। बैच और बैच-टू-बैच के भीतर उनकी जटिल अरैखिक गतिशीलता और अस्थिर-स्थिति परिचालन

स्थितियों के कारण बैच प्रक्रियाओं का नियंत्रण मुश्किल है। बैच प्रक्रिया नियंत्रण के लिए प्रक्रिया उद्योगों में उपयोग किए जाने वाले मॉडल प्रिडिक्टिव कंट्रोल (एमपीसी) जैसी उन्नत नियंत्रण रणनीतियाँ गणितीय प्रोग्रामिंग को एक विवश, संभवतः गैर-उत्तल, अनुकूलन समस्या को हल करने के लिए नियोजित करती हैं। यहां मुख्य मुद्दा यह है कि इष्टतम नियंत्रण इनपुट प्रोफ़ाइल प्राप्त करने के लिए प्रत्येक चरण में ऑनलाइन कम्प्यूटेशनल बोझ बहुत अधिक है। कम्प्यूटेशनल हार्डवेयर और संख्यात्मक तरीकों में प्रगति के बावजूद, यह जटिल गैर-रैखिक, उच्च-आयामी गतिशील प्रणालियों के लिए इन दृष्टिकोणों के कार्यान्वयन को सीमित करता है। इसके अलावा, मॉडल-आधारित नियंत्रक का प्रदर्शन सटीक प्रक्रिया मॉडल की उपलब्धता पर अत्यधिक निर्भर है। एक जटिल और अरेखीय प्रक्रिया के लिए, ऐसे मॉडल की उपलब्धता एक सीमा है क्योंकि इसके लिए महत्वपूर्ण पूर्व ज्ञान और विशेषज्ञता की आवश्यकता होती है। प्लांट-मॉडल बेमेल तब भी होता है जब वास्तविक प्रक्रिया और उसके अनुमानित मॉडल में थोड़ी भिन्नता होती है, जिससे उत्पाद की उपज जैसे प्रदर्शन चर की गलत भविष्यवाणी होती है। भले ही प्रक्रिया मॉडल उपलब्ध हो, अनिश्चितताओं और प्रक्रिया के बहाव की उपस्थिति में नियंत्रक का प्रदर्शन बिगड़ जाता है। कच्चे माल में उतार-चढ़ाव, सफाई आदि के कारण होने वाली बैच-टू-बैच विविधताएं अनिश्चितता का एक स्रोत हैं जो बंद-लूप प्रदर्शन को और खराब करती हैं। इस समय, एक नियंत्रण रणनीति का विकास जो पूरी तरह से सिस्टम की सटीक गतिशीलता के ज्ञान पर निर्भर नहीं करता है और स्टोकेस्टिक गतिशीलता और प्लांट-मॉडल बेमेल को संभाल सकता है, अत्यंत उपयोगी है। मॉडल-मुक्त सुदृढीकरण सीखना (आरएल), जहां एजेंट (नियंत्रक के अनुरूप) परिचालन वातावरण (प्रक्रिया के अनुरूप) के साथ सीधे बातचीत करके इष्टतम नियंत्रण क्रिया (इनपुट को नियंत्रित करने के अनुरूप) सीखता है, पारंपरिक मॉडल के लिए एक संभावित विकल्प प्रदान करता है। प्रक्रिया नियंत्रण के लिए आधारित दृष्टिकोण। अभिनेता-आलोचक वास्तुकला के साथ आरएल ढांचे हाल ही में नियंत्रण प्रक्रिया प्रणालियों के लिए लोकप्रिय हो गए हैं जहां राज्य और क्रिया स्थान दोनों निरंतर हैं। इसके बाद के कार्यों में दो उपन्यास अभिनेता-आलोचक आरएल एल्गोरिदम विकसित करके अभिनेता-आलोचक आरएल आधारित नियंत्रक के विकास पर ध्यान केंद्रित किया गया है, अर्थात्, (i) ट्विन एक्टर ट्विन डिलेड डिटर्मिनिस्टिक पॉलिसी ग्रेडिएंट (टीएटीडी3), एक नियतात्मक आरएल एल्गोरिथ्म और (ii) ट्विन एक्टर सॉफ्ट ट्विन एक्टर सॉफ्ट एक्टर-क्रिटिक (TASAC), एक स्टोकेस्टिक आरएल एल्गोरिथम, बैच प्रक्रिया नियंत्रण के लिए एक समग्र इष्टतम नीति प्राप्त करने के लिए नीतियों को प्रशिक्षित

करने के लिए अभिनेता-आलोचक एल्गोरिदम में अभिनेताओं के समूह को शामिल करके बैच प्रक्रिया नियंत्रण के लिए।

इस कार्य में विकसित दृष्टिकोणों की प्रभावशीलता का मूल्यांकन बैच प्रक्रिया मामले के अध्ययन पर सिमुलेशन अध्ययन द्वारा किया जाता है। संक्षेप में, समग्र उद्देश्य बैच प्रक्रियाओं के इष्टतम संचालन को सुनिश्चित करने के लिए जटिल, गैर-रैखिक बैच प्रक्रियाओं के लिए एक डेटा-संचालित गतिशील मॉडल और एक मॉडल-मुक्त आरएल-आधारित नियंत्रक विकसित करना है।

कीवर्ड: डेटा-संचालित मॉडल; समय-समय पर सीखना; सुदृढीकरण सीखना; डीप-क्यू-लर्निंग; गहरी नियतात्मक नीति प्रवणता; अभिनेता-आलोचक एल्गोरिदम; बैच प्रक्रिया नियंत्रण

Contents

List of Tables

List of Figures

ABBREVIATIONS

JITL	Just-in-Time Learning
RL	Reinforcement Learning
MPC	Model Predictive Control
TD	Temporal Difference
MDP	Markov Decision Process
MC	Monte Carlo
DQN	Deep Q Network
DNN	Deep Neural Network
PG	Policy Gardient
DDPG	Deep Deterministic Policy Gradient
TD3	Twin Delayed Deep Deterministic Policy Gradient
TATD3	Twin Actor Twin Delayed Deep Deterministic Policy Gradient
SAC	Soft Actor Critic
TASAC	Twin Actor Soft Actor Critic
FAME	Fatty Acid Methy Ester
MSE	Mean Squared Error
TV	Target Value

NOTATION

s, a, r, s'	State, action, reward, next state
γ	Discount factor
G_t	Discounted return
$\phi_{\mathbf{C}_j}$	Q-network weight
$\phi_{\mathbf{C}_j,\mathbf{T}}$	Target Q-network weight
$\phi_{\mathbf{A}_i}$	i^{th} Actor network weight
$\phi_{\mathbf{A}_i,\mathbf{T}}$	Target Actor Network weight
$\mu_{\phi_{\mathbf{A}_i}}$	Deterministic parameterised policy
π	Policy
π_{ϕ_A}	Policy corresponding to parameter ϕ_A
π^*	Optimal policy
$\pi_{\phi_A}(\cdot\|s_t)$	Stochastic policy parameterised with parameter ϕ_A
$\mu_{\phi_A}(s_t)$	Deterministic policy parameterised with parameter ϕ_A
$\tilde{a_i}$	Target action
τ	Target update rate
β_C	Critic learning rate
β_A	Actor learning rate
k_i	Rate constant
T_r	Reactor temperature
T_j	Jacket temperature
T_{jin}	Jacket inlet temperature
α	Temperature parameter

$\mathbf{H}(\pi(\cdot|s_t))$ Entropy of policy π