공학박사 학위논문

Data-driven Soft Sensor Development for Multi-grade and Time-varying Processes

다등급 및 시변 공정을 위한

데이터 기반 소프트 센서 개발

2023년 2월

서울대학교 대학원

화학생물공학부

송 민 준

Data-driven Soft Sensor Development for Multi-grade and Time-varying Processes

지도교수 이 종 민

이 논문을 공학박사 학위논문으로 제출함

2023년 2월

서울대학교 대학원

화학생물공학부

송 민 준

송 민 준의 공학박사 학위논문을 인준함

2023년 2월

위 원 장 이 원 보 (인)

부위원장 이 종 민 (인)

위 원 남 재 욱 (인)

위 원 김 연 수 (인)

위 원 손 상 환 (인)

Abstract

Data-driven Soft Sensor Development for Multi-grade and Time-varying Processes

Min Jun Song

School of Chemical and Biological Engineering

The Graduate School

Seoul National University

The prediction of the qualities and properties of chemical products and reactor conditions during process operation is essential for chemical processes where real-time online quality measurement is not available. In order to decrease off-specification products and increase process efficiency, the prediction is required to be more accurate and rapid. Therefore, there has been a great effort to build an accurate and fast soft sensor for industrial chemical processes using first principles or machine learning regression methods. However, there exist several challenges which make it much more difficult to accurately predict qualities in industrial chemical processes.

First, a prediction model needs to be able to deal with the high nonlinearity of process dynamics resulting from complex interactions between chemical species, chemical reactions, energy and mass transportations, and phase transitions. Additionally, the high nonlinearity makes the construction and interpretation of a soft sensor model significantly more difficult. Second, industrial chemical processes exhibit time-varying process dynamics. Many chemical processes are operated

alternately in multiple modes to satisfy fluctuating market demands. However, unsteady state operations including grade transition and startup result large settling time and overshoot. In most batch processes, furthermore, the dominant reaction taking place in the reactor is constantly changing as the reaction progresses and the concentrations of chemical species change. In addition, industrial chemical processes are subject to process drifts such as catalyst deactivation and fouling. The process drifts may lead to the model-plant mismatch and large discrepancy between the predictions from the trained soft sensor model and actual quality measurement. Therefore, in order to obtain more accurate and robust predictions, the aforementioned problems must be dealt with during modeling procedure.

In this regard, this thesis proposes novel soft sensor development methods with improved accuracy and robustness by addressing modeling problems in industrial chemical processes. First, a hybrid soft sensor modeling framework for online quality prediction of polymer products from industrial polymerization process is proposed. When compared to conventional first principles models and data-only machine learning models, the proposed hybrid model provides more accurate prediction by combining prior process knowledge including reaction mechanisms and data-driven regression methods such as neural networks. Additionally, the generalizability of a machine learning prediction model is improved by applying the proposed hybrid modeling approach. The hybrid model has more robust prediction performance when the process undergoes grade changeover and unsteady state operation, compared to data-only machine learning models.

Second, a time series-based just-in-time learning soft sensor modeling method combined with dynamic time warping is proposed. The process data obtained from a multi-grade or time-varying process shows multimodal distribution, which is difficult for a single global model to learn. Rather than training a global model with the full training dataset, a just-in-time learning model makes prediction based on a local model trained only with selected training samples which are similar to the query

sample, allowing a soft sensor to adapt to process changes. Additionally, the proposed model takes a time series as an input rather than a data point at a certain time in order to account for dynamic behavior of a process. The prediction performance of a just-in-time learning model is strongly affected by how to select similar samples among large database. The proposed modeling method utilizes dynamic time warping method in calculation of similarities between the query and stored data samples. Dynamic time warping is one of the most widely used synchronization methods for data sequences. Additionally, the similarity between two distinct data sequences can be calculated with dynamic time warping by stretching the sequences.

The proposed modeling approach achieves two main contributions by combining just-in-time learning and dynamic time warping in soft sensor development for multi-grade and time-varying chemical processes. First, complex process dynamics is accounted in the proposed model by taking time series process data as an input. Second, more accurate selection of similar data sequences from the database is achieved by applying dynamic time warping in similarity calculation. The improvements in the prediction performance achieved by the proposed modeling approach are demonstrated using case studies of multi-grade processes.

In addition, the effects of the model hyperparameters on the prediction performance are thoroughly investigated via simulation studies. As a result of the sensitivity analysis, the optimal input length is determined by the cross-correlation between the input and output variables. Furthermore, prediction accuracy and computational speed can be improved by applying the width constraint on the dynamic warping path. Therefore, this thesis provides hyperparameter-optimized just-in-time learning modeling approach combined with dynamic time warping by introducing width constraint on the warping path and calculating cross-correlations between the input and output.

In summary, this thesis proposes novel soft sensor development methods with

improved prediction accuracy and generalizability for industrial chemical processes which are highly nonlinear and exhibit multi-grade or time-varying process dynamics. The proposed modeling approaches demonstrate improved modeling performance compared to conventional modeling methods when applied to simulation data and actual polymerization process data.

Keywords: Soft sensor, Machine learning, Hybrid model, Just-in-time learning, Dynamic time warping

Student Number: 2017-28864

*This thesis is based in part on the previously published articles listed below.

1. M. J. Song, S. H. Ju, S. Kim, S. H. Oh, and J. M. Lee, “Hybrid modeling approach for polymer melt index prediction”, *Journal of Applied Polymer Science*, **2022**, vol. 139, no. 41, e52987.

2. M. J. Song, S. H. Ju, and J. M. Lee, “Soft sensor development based on just-in-time learning and dynamic time warping for multi-grade processes”, *Korean Journal of Chemical Engineering*, accepted.

3. M. J. Song, S. Kim, S. H. Oh, P. S. Jo, and J. M. Lee, "Soft Sensor for Melt Index Prediction Based on Long Short-Term Memory Network", *IFAC-PapersOnLine*, **2022**, vol. 55, no. 7, pp. 857-862.

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

1.1. Research motivation and previous works

The quality variables of the chemical products provide essential information to determine not only the properties and grade of the product but also whether the product is off-specification or not. For example, the molecular characteristics of a thermoplastic polymer determines the mechanical properties and processability of the polymer product [1]. However, it is difficult to measure the quality variables of chemical and biological products online during the operation in most industrial processes because of the lack of real-time measurement technology or its high cost. Instead, offline laboratory analysis is conducted infrequently to measure the quality variables of a sample collected from the final product. However, infrequent product analysis imposes several challenges in process monitoring and control. First, there exists time lag between the analyzed product sample and the currently produced product, resulting in quality deviation. Second, many chemical processes are operated in multiple mode to produce multiple grades of products. Additionally, industrial processes exhibit time-varying process dynamics due to the process drifts such as catalyst deactivation and fouling. However, offline quality measurement has difficulty in capturing fast process changes due to its long time delay between each measurement.

Therefore, there is a growing interest in the developments of soft sensors for various types of industrial processes. In earlier studies, most soft sensor models have been developed using first principles obtained from chemical and thermodynamic knowledge. Therefore, the first principles models provide robust predictions when process dynamics has already been thoroughly investigated and available. Additionally, the first principles models are easier to interpret compared to black-box or data-driven models because the model prediction is the result of known

process dynamics. However, the first principles modeling approach is not suitable for some types of industrial processes. For instance, it is impossible to build a first principles model if precise chemical reaction mechanisms or interaction between chemical species are not available because of the lack of prior investigation. Furthermore, the model-plant mismatch where there exists a difference between model assumption and actual plant may result in large deviations in quality prediction. For instance, most of the first principles are derived from the assumptions that the interactions between chemical species are performed in ideal condition, such as perfect mixing. Lastly, the development of a first principles model for a large-scale industrial process requires a lot of time and computations.

Therefore, with the rapid growth of data storage and computer technologies, it has become more popular to build a data-driven prediction model using machine learning regression methods rather than to develop a soft sensor model using process knowledge. Data-driven approaches only require process and quality measurement data obtained from the sensors and laboratory analyses. Thus, a prediction model can be developed without prior process knowledge such as reaction kinetics and thermodynamic properties of chemical species. Soft sensor models based on latent variables methods such as principal component analysis (PCA) and partial least squares (PLS) have been developed for chemical processes [2-6]. Additionally, machine learning methods such as vector machines (SVM) [7-10] and Gaussian process regression (GPR) [11-15] have been applied to soft sensor developments. In recent years, soft sensors based on various types of neural networks, including radial basis function neural networks (RBFNNs) [16-21], fuzzy neural networks (FNNs) [22, 23], wavelet neural networks (WNNs) [24], fuzzy wavelet neural networks (FWNNs) [25], deep belief networks (DBNs) [26], and dilated convolution neural networks (DCNNs) [27], have demonstrated good prediction performance in MI prediction problems for polymerization processes.

However, when it comes to industrial large-scale processes, there are several

challenges which make it significantly more difficult to build an accurate data-driven soft sensor model. First, process data variables obtained from the operation of industrial processes show highly nonlinear and complex relationships derived from numerous chemical reactions, which imposes difficulties in training and interpreting machine learning models. In order to build a soft sensor which can account for nonlinear dynamics, nonlinear regression methods including SVM, GPR, and neural networks have been widely applied to soft sensor development for chemical and biological processes [7-27]. However, these nonlinear machine learning models are susceptible to overfitting and require hyperparameter optimization to obtain optimized prediction performance.

Second, most industrial processes have time-varying process dynamics derived from multiple operation modes and process drifts. The modeling performance of a global machine learning model trained using fixed training dataset can be severely deteriorated when the process dynamics change since data-driven regression models have inherently low extrapolation and generalization abilities. Therefore, a machine learning model that are able to continuously adapt to process changes have been developed for time-varying systems. For example, soft sensor models, including recursive PCA [28-30] and recursive PLS [31-33], that are updated iteratively with new measurement data have been developed to account for time-varying process dynamics in quality prediction. However, recursive models have difficulty in modeling chemical processes with rapidly changing dynamics because their adaptability are only guaranteed for slow time-varying dynamics.

Therefore, aforementioned challenges in industrial-scale soft sensor developments need to be investigated and addressed for more accurate and robust predictions. In other words, there is an academic and industrial demands for soft sensor models with improved modeling and generalization performance that can account for time-varying dynamics as well as nonlinear relationships between data variables.

1.2. Statement of contributions

The main objective of the thesis is to propose novel soft sensor modeling approaches with improved prediction accuracy and generalizability for industrial processes by solving modeling problems stated in subsection 1.1. In order to accomplish the research objectives, two modeling approaches are proposed: hybrid soft sensor modeling and just-in-time learning soft sensor modeling based on dynamic time warping.

The first work of this thesis, the hybrid soft sensor modeling approach for an industrial polymerization process, achieves quality prediction for styrene-acrylonitrile polymer product with improved prediction accuracy and generalizability by combining mechanistic modeling and data-driven modeling approaches. The target process is a commercial polymerization process which produces multiple grades of styrene-acrylonitrile polymer products. Therefore, the process experiences frequent grade changeovers and operation condition changes, resulting in large deviations in product qualities. Furthermore, infrequent quality measurements make it more difficult to monitor and control the qualities of the products produced during unsteady states, which affects process efficiency and profitability negatively.

In earlier studies, most soft sensors have been developed based solely on the first principles modeling approach or data-driven modeling approach. However, each modeling approach has its pros and cons. For instance, the first principles modeling approach provides more fundamental insight into the system and how the properties of the final products are determined. However, the model-plant mismatch and thus large prediction error would occur when the model assumptions are not applicable to the actual process because of the changes in the process dynamics. On the other hand, data-driven soft sensor models have been applied to various types of industrial processes and have demonstrated good prediction performance without prior process

knowledge, given sufficient measurement data. However, data-driven approaches are inherently susceptible to new data obtained outside training region and therefore have weak generalization ability.

Therefore, a novel hybrid modeling approach which integrates advantages of the first principles approach and the data-driven approach is proposed to obtain a soft sensor model with improvements in both prediction accuracy and generalizability. The proposed hybrid model is constructed using a serial model structure where mechanistic modeling of the target process and polymerization reaction is performed first, followed by training MI prediction model based on black-box machine learning regression modeling methods. The proposed model is verified using multi-grade polymerization process data and its prediction performance is compared to the models based solely on first principles or data-driven approaches. As a result, the proposed model has achieved better prediction accuracy and robustness even when the target process suffers rapid and abrupt changes due to grade changeovers and process failures.

The second work of this thesis, soft sensor modeling based on just-in-time learning and dynamic time warping, proposes a novel modeling framework which is able to address multi-grade and time-varying characteristics of industrial processes. It is very challenging to build an accurate global soft sensor for multi-grade processes because the process data exhibit multimodal distribution. Furthermore, already trained data-driven model's prediction may deviate largely from actual measurement data when process dynamics changes along with new operation mode. For a soft sensor to address aforementioned challenges in quality prediction of multi-grade and time-varying processes, it requires adaptation ability to newly updated data.

In the proposed modeling approach, three main contributions are achieved. First, the proposed model takes time series as an input by augmenting raw process data variables into time series segments in order to account for process dynamics. Second, dynamic time warping is implemented in calculating similarities between different

time series data, resulting in more appropriate training dataset for local model training in the just-in-time learning modeling. Third, the effects of the model parameters on the modeling performance are discussed and a soft sensor modeling approach with the hyperparameter optimization is proposed. As a result of aforementioned contributions, the proposed model has demonstrated improved modeling performance for frequently changing processes when verified using several case studies of chemical processes.

In summary, this thesis contributes to industrial-scale soft sensor modeling by proposing novel modeling methods with improved prediction accuracies and generalization abilities.

1.3. Outline of the thesis

The remainder of this thesis is organized as follows. In Chapter 2, the research background and preliminaries necessary for the later developments of the proposed modeling approaches are provided, including just-in-time learning and dynamic time warping.

Chapter 3 proposes a novel hybrid modeling method that combines first principles and machine learning modeling approaches for MI prediction in an industrial styrene-acrylonitrile polymerization process. The improved prediction accuracy of the proposed hybrid model is demonstrated using about 17 months of process measurement data, compared to those of classical first principles models and data-only machine learning models.

In Chapter 4, just-in-time soft sensor modeling approach combined with dynamic time warping is proposed. It is very challenging to build an accurate soft sensor model for a process with multi-grade characteristics and time-varying process dynamics. In the proposed model, a local soft sensor model is trained using only data samples that are similar to the query sample, which makes the prediction model adapt

to process changes. Additionally, data sequences are used as a model input and the similarity between two different sequences are calculated using dynamic time warping. The proposed modeling approach's modeling performance is evaluated using simulation case studies. As a result, the proposed model has demonstrated improved prediction accuracy and robustness compared to classical Euclidean distance-based just-in-time learning soft sensor models. Additionally, the effects of the modeling parameters on prediction performance are analyzed and the modeling method with correlation-based parameter optimization is proposed.

Finally, Chapter 5 provides concluding remarks derived from the research results introduced in the previous chapters, and suggests future works for further improvement of modeling performance.

Chapter 2. Background and Preliminaries

This chapter introduces fundamental background and preliminaries required for the novel soft sensor modeling approaches proposed in later chapters, including just-in-time learning modeling method, dynamic time warping, and machine learning regression models.

2.1. Just-in-time learning modeling

Just-in-time learning modeling approach is a modeling method developed to accurately model systems having time-varying process dynamics and multimodal process data. Most classical modeling approaches aim to build a global model which can explain general process dynamics of a system using entire training dataset. However, when it comes to processes which have multiple operation modes or time-varying process dynamics, it is very challenging for a single global model to accurately simulate entire process dynamics. Therefore, instead of modeling a whole process with a single model, just-in-time learning modeling approach trains a local model for every query data in an online manner using data samples similar to the query sample.

By training a local model rather than a global model, just-in-time learning models have two main advantages in process modeling. First, a new model is trained every time new query data is obtained. Therefore, just-in-time learning models have adaptability to process changes. On the other hand, the prediction of a data-driven global model deviates significantly from actual measurement data when the operating condition is outside training region because of the low extrapolation ability.

Second, a local model is trained using only similar samples from the database. Therefore, the required time for training a machine learning model is much shorter when compared to using an entire training dataset. Additionally, the training samples

are selected from the database consisting of past process data, which indicates that just-in-time learning soft sensors have ready-to-use memories of previous operating conditions. On the other hand, other adaptive modeling approaches, including recursive and moving window approaches, either have slow adaptability or past process data are rarely considered.

Figure 2.1 and **Figure 2.2** illustrate the differences between basic algorithms of classical global learning approaches and just-in-time learning approaches. As shown in **Figure 2.1**, the classical global learning approaches train an offline model using fixed historical database. Then, the predictions are made using the trained offline global model. Therefore, the prediction performance of a classical global model is only guaranteed for training region. The prediction model is not updated even if the query sample is obtained outside the training region.

On the contrary, as shown in **Figure 2.2**, just-in-time learning modeling for soft sensors consists of three main steps. First, when a new query sample is obtained from the online sensors, similarities between the query sample and samples in the historical database are calculated. Next, the fixed number of most similar historical samples are selected and used to train a local prediction model. Lastly, the prediction for the query sample is made using the local model trained during the second step. After above three steps are performed, the trained local model is discarded and the historical database is updated using actual measurement data. Since a new model is trained every time a new query data is obtained, just-in-time learning models are able to adapt to process changes. Therefore, just-in-time learning soft sensors have been applied to various chemical processes and demonstrated good prediction performance and adaptabilities [12, 13, 34-46].

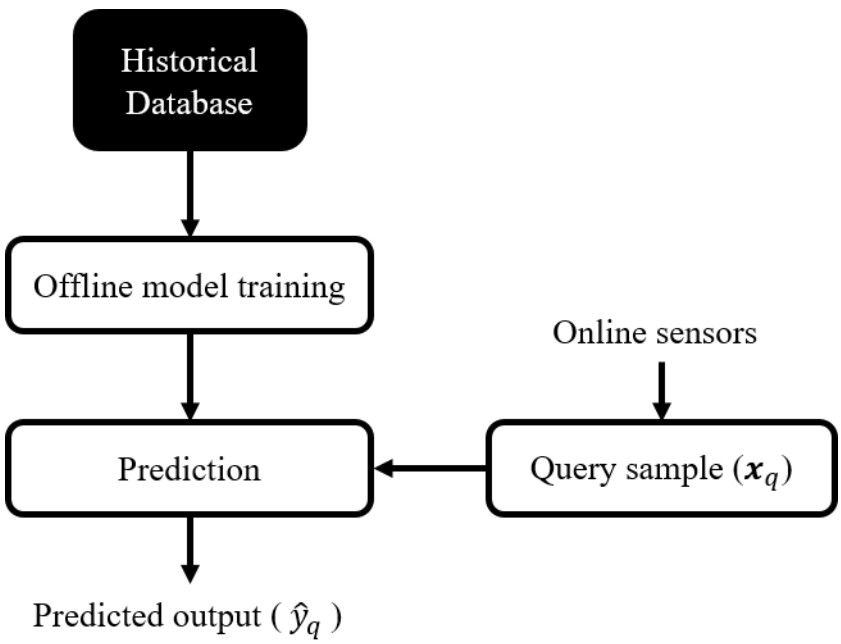


Figure 2.1. Basic algorithm of classic global learning approach.

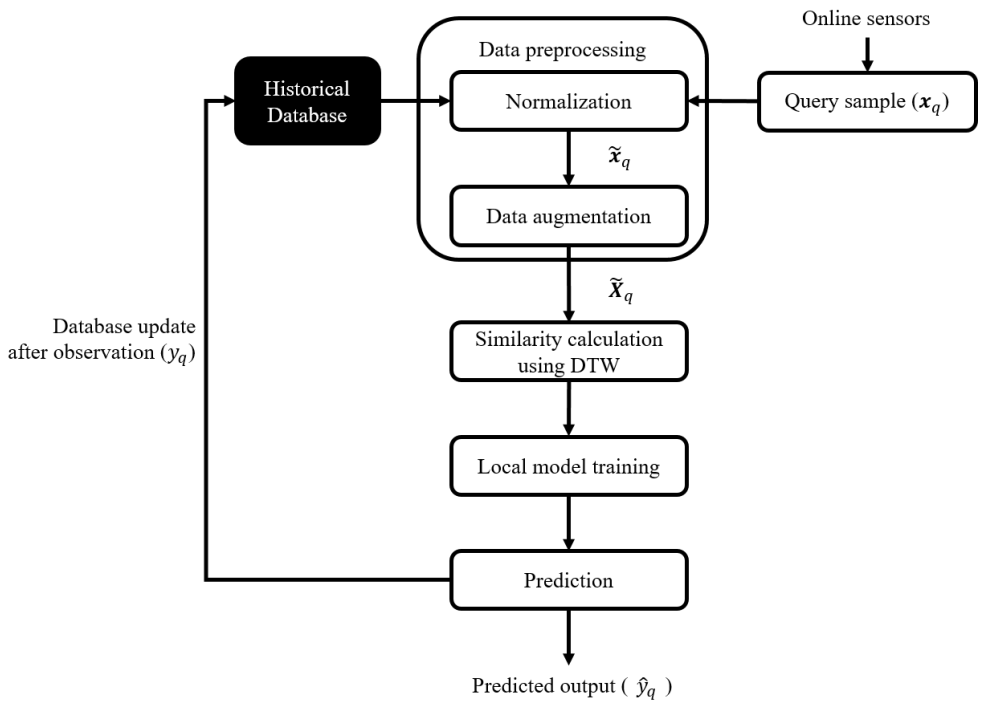


Figure 2.2. Basic algorithm of just-in-time learning approach.

One of the distinct features of the just-in-time learning models is selection of similar samples from historical database. Since a local model is trained using the dataset of similar samples only, the modeling performance of just-in-time learning models is drastically affected by the way of how to calculate similarities different data samples.

One of the most widely used similarity metrics in the just-in-time learning models is the Euclidean distance because of its simplicity and short calculation time. The Euclidean distance calculates the distance between two different data points without transforming data variables. The Euclidean distance, $\boldsymbol{ED}$, between a query sample $\boldsymbol{x}_q$ and a sample $\boldsymbol{x}_n$ from the database is calculated as in (2.1).

$$\boldsymbol{ED}(\boldsymbol{x}_q, \boldsymbol{x}_n) = \sqrt{\sum_{i=1}^{n_{var}} (x_{q,i} - x_{n,i})^2} \tag{2.1}$$

where n_{var} is the number of variables of the data vector $\boldsymbol{x}$, and $x_{q,i}$ is the i-th variable of the vector $\boldsymbol{x}_q$.

Additionally, another similarity metric where accounting for not only distance between the points but also the angle between them is applied to similarity calculation in just-in-time learning models. The cosine value of the angle rather than the angle itself is used in similarity calculation, which is calculated as in (2.2).

$$\cos(\theta_{qn}) = \frac{\langle \boldsymbol{x}_q, \boldsymbol{x}_n \rangle}{\|\boldsymbol{x}_q\|_2 \|\boldsymbol{x}_n\|_2} \tag{2.2}$$

where θ_{qn} is the angle between the query data sample $\boldsymbol{x}_q$ and data sample $\boldsymbol{x}_n$ of the historical database. In just-in-time learning modeling, only data samples having positive cosine values to the query sample are considered in similar sample selection

because the negative cosine values indicate that two samples are significantly different.

In contrast to above metrics where the information from two data points is considered only, similarity calculation methods based on statistical information of multiple data points have been also applied in just-in-time learning modeling. Among statistical similarity measures, the Mahalanobis distance and Kullback-Leibler divergence are the most widely studied and applied in just-in-time learning models.

The Mahalanobis distance is a measure of the distance between a data point and a distribution. The Mahalanobis distance, $\boldsymbol{MD}$, between a point $\boldsymbol{x}$ and a distribution $\boldsymbol{P}$ is calculated as in (2.3).

$$\boldsymbol{MD}(\boldsymbol{x}, \boldsymbol{P}) = \sqrt{(\boldsymbol{x} - \mu)^{\mathrm{T}} S^{-1} (\boldsymbol{x} - \mu)} \tag{2.3}$$

where μ and S are the sample mean and covariance matrix of the distribution $\boldsymbol{P}$, respectively. Additionally, the Mahalanobis distance between two data points, $\boldsymbol{x}$ and $\boldsymbol{y}$ can be calculated with respect to a distribution $\boldsymbol{P}$ as in (2.4).

$$\boldsymbol{MD}(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{P}) = \sqrt{(\boldsymbol{x} - \boldsymbol{y})^{\mathrm{T}} S^{-1} (\boldsymbol{x} - \boldsymbol{y})} \tag{2.4}$$

Another statistical similarity metric is Kullback-Leibler divergence which is also known as relative entropy. The Kullback-Leibler divergence measures dissimilarity between two different distributions. The Kullback-Leibler divergence, $\boldsymbol{KLD}$, between two continuous probability distributions, $\boldsymbol{P}$ and $\boldsymbol{Q}$, is calculated as in (2.5).

$$\boldsymbol{KLD}(\boldsymbol{P}||\boldsymbol{Q}) = \int_{-\infty}^{\infty} p(x) \frac{\log\big(p(x)\big)}{\log\big(q(x)\big)} \tag{2.5}$$

where p and q are the probability densities of $\boldsymbol{P}$ and $\boldsymbol{Q}$, respectively. In order to calculate the Kullback-Leibler divergence, the probability distributions of variables need to be known in advance. Therefore, the transformation of the process variables into the latent variables whose probability distributions are expressed by Gaussian distribution is first performed in just-in-time learning soft sensors based on the Kullback-Leibler divergence [45-47].

2.2. Dynamic time warping

Dynamic time warping is one of the most widely used synchronization methods for different data sequences. The distance between two sequences with different lengths is unable to measure using classic distance measures including the Euclidean distance. Sequences need to be synchronized first before measuring the distance. Dynamic time warping synchronizes data sequences to identical length by stretching them. Additionally, the distance between synchronized sequences can be measured. With its ability to synchronize data sequences with different lengths, dynamic time warping has been applied to various types of dynamic time series problems, including handwriting recognition [48], gesture recognition [49], traffic speed prediction [50], state of health estimation [51], fault detection [52], and uneven batch trajectory synchronization [53].

Dynamic time warping finds the optimal warping path $W = \{w_1, w_2, \dots, w_k\}$ of length k, which minimize cumulative distance between warped data sequences. For two time series of different lengths, $X = \{x_1, x_2, \dots, x_m\}$ of length m and $Y = \{y_1, y_2, \dots, y_n\}$ of length n, the warping path W must satisfy the following constraints. First, the boundary conditions at the initial and final points must be satisfied.

$$w_1 = (1,1) \tag{2.6}$$

$$w_k = (m,n) \tag{2.7}$$

where $w_i = (a,b)$ indicates that X_a corresponds to Y_b in the path W. Second, the warping path must be continuous. If $w_i = (a,b)$ and $w_{i+1} = (a',b')$, then $a' - a \leq 1$ and $b' - b \leq 1$. Lastly, the warping path must be monotonic in order to prevent the warping path from moving backward in time. Therefore, if $w_i = (a,b)$ and $w_{i+1} = (a',b')$, then $a' - a \geq 0$ and $b' - b \geq 0$. The above two constraints can be combined and expressed as follows.

$$0 \leq a' - a \leq 1 \tag{2.8}$$

$$0 \leq b' - b \leq 1 \tag{2.9}$$

which indicates that w_{i+1} can have only three values, including $(a+1,b)$, $(a,b+1)$ or $(a+1,b+1)$. However, there exist numerous warping paths satisfying the above constraints. Therefore, dynamic time warping finds the warping path with minimum cumulative distance using dynamic programming. Finally, the distance, $\boldsymbol{DTW}$, between two data sequences X and Y calculated using dynamic time warping is defined as follows.

$$\boldsymbol{DTW}(X,Y) = \min_{W} \sum_{i=1}^{k} d_i \tag{2.10}$$

where d_i is the distance between $X(a)$ and $Y(b)$ when a warping path passes through $w_i = (a,b)$. The distance d_i is calculated using the Euclidean distance as in (2.11).

$$d_i = \|X(a) - Y(b)\|_2 \tag{2.11}$$

When data sequences become longer, the performance of dynamic time warping algorithm suffers from drastically growing computational requirement and outliers in the sequences. For instance, a single outlier in a sequence can severely affect warping path, resulting in excessive stretching of the sequences. Under the circumstances of excessive warping, the dynamics and characteristics of the original data sequences are too distorted to measure similarity precisely. Therefore, another constraint on the warping path has been introduced in order to prevent excessive calculation and stretching in dynamic time warping. Limiting the maximum warping window width to δ reduces the number of possible warping paths and makes the warping paths closer to diagonal connecting $w_1 = (1,1)$ and $w_k = (m,n)$. As a result, the computational requirement for finding the optimal warping path is reduced and excessive stretching of the sequences is prevented. **Figure 2.3** and **Figure 2.4** illustrate the differences between the alignments of two different time series using dynamic time warping with and without the warping window width constraint. The example time series used in **Figure 2.3** and **Figure 2.4** are based on sine functions of different frequencies and delays. Without warping window width constraint, DTW stretches both data sequences excessively in order to synchronize them as much as possible as shown in **Figure 2.3**. As a result, the warped time series perfectly matches each other. On the other hand, only the nearby peaks in original time series are synchronized when the maximum warping window width is constrained to a fixed value, as shown in **Figure 2.4**. In order to prevent excessive warping, peaks that are apart from each other are not synchronized.

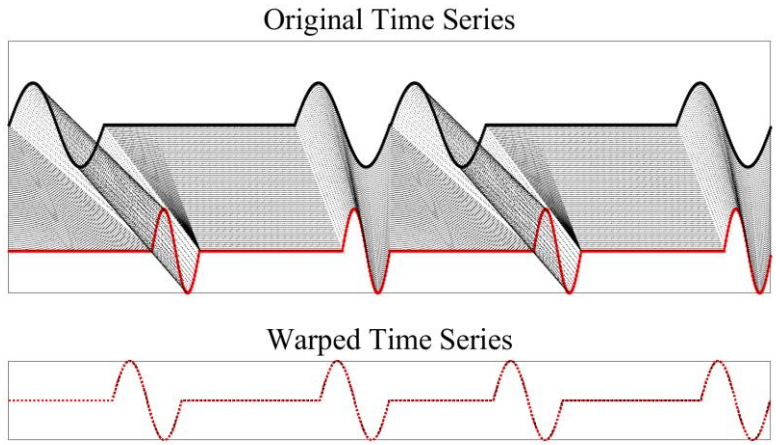


Figure 2.3. Example of warping two data sequences using DTW without warping window width constraint.

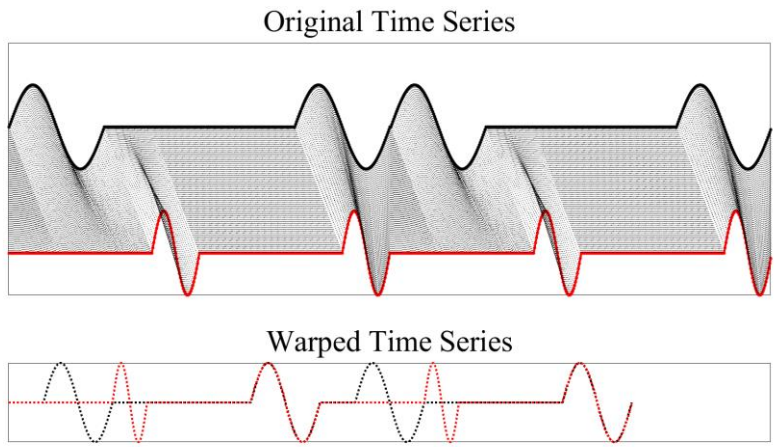


Figure 2.4. Example of warping two data sequences using DTW under warping window width constraint.

2.3. Data-driven regression models

In this section of Chapter 2, the preliminaries and background of the regression models applied in the proposed modeling approaches in later chapters for the reader's understanding. In the proposed modeling approaches of this thesis, a prediction model is trained based on widely studied data-driven regression models. A total of four regression models are introduced including partial least squares (PLS), support vector machine (SVM), and Gaussian process regression (GPR).

2.3.1. Partial Least Squares

Partial least squares (PLS) model is one of the latent variable models where raw data variables are projected onto the lower-dimensional latent variable space. By transforming original variables into the lower-dimensional latent variables that are independent to each other, two main advantages are achieved. First, the dimensions of the process data obtained from industrial processes are very high in general, which requires more calculation and time in data preprocessing and training a model. The computational requirement can be reduced by using the latent variables which are fewer than raw process variables. Second, process variables are correlated with each other. The multicollinearity inherent in the process data makes it more difficult to interpret the modeling result. Additionally, the modeling performance of a linear model can be drastically reduced when data variables are highly correlated. PLS method transforms highly correlated variables into the linearly independent variables, addressing the multicollinearity-related problems.

While principal component regression (PCR), another most widely used regression model based on a latent variable method, only transforms the input variables into the latent variables, PLS transforms both the input and output variables. For a matrix of process variables $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n] \in \mathbb{R}^{n \times p}$ and a matrix of the

quality variables $\boldsymbol{Y} = [\boldsymbol{y}_1, \boldsymbol{y}_2, \dots, \boldsymbol{y}_n] \in \mathbb{R}^{n \times q}$, PLS decomposes $\boldsymbol{X}$ and $\boldsymbol{Y}$ so that the covariance between the projections of them are maximized. The decompositions of the $\boldsymbol{X}$ and $\boldsymbol{Y}$ matrices are expressed as follows.

$$\boldsymbol{X} = \boldsymbol{T}\boldsymbol{P}^{\mathrm{T}} + \boldsymbol{E} \tag{2.12}$$

$$\boldsymbol{Y} = \boldsymbol{U}\boldsymbol{Q}^{\mathrm{T}} + \boldsymbol{F} \tag{2.13}$$

where $\boldsymbol{T}$ and $\boldsymbol{U}$ are scores that are the projections of the $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively. $\boldsymbol{P}$ and $\boldsymbol{Q}$ are orthogonal loading matrices and $\boldsymbol{E}$ and $\boldsymbol{F}$ are residual matrices. There are several algorithms, including nonlinear iterative PLS algorithm (NIPALS) [54] and SIMPLS [55], to find score and loading matrices. Since detailed descriptions of the statistical algorithms are beyond the scope of this thesis, the details on algorithms are not provided and can be found in the literature [54-56]. With its simplicity and ability to address multicollinearity, PLS has been widely applied in process monitoring and soft sensing of a variety of chemical processes [3-6, 31-33].

2.3.2. Support vector machine

Support vector machine (SVM) is a machine learning algorithm developed based on Vapnik-Chervonenkis (VC) theory and has been demonstrated good modeling performance in both classification and regression problems [57-59]. When SVM algorithm is applied to a regression problem, it is also referred as support vector regression (SVR). While classical modeling methods employ the empirical risk minimization in general, SVM employs structural risk minimization. In contrast to the empirical risk minimization where a learning algorithm minimizes a loss function of the empirical risk, the structural risk minimization considers both minimization of the training error and overall generalization performance [60]. Therefore, an SVM

model is obtained by balancing data fitting and the model's complexity, which addresses the overfitting problem common in machine learning algorithms. A general empirical risk minimization problem can be expressed as in (2.14).

$$\hat{h} = \underset{h}{argmin}\left(\frac{1}{n}\sum_{i=1}^{n} L(h(x_i), y_i)\right) \tag{2.14}$$

where h and L are hypothesis and loss function, respectively. On the other hand, a general structural risk minimization problem optimizes both the training error and weight regularization simultaneously as in (2.15).

$$\hat{\theta} = \underset{\theta}{argmin}\left(\frac{1}{2n}\sum_{i=1}^{n} L(h(x_i), y_i) + \frac{\lambda}{2}\sum_{j=1}^{m} \theta_j^2\right) \tag{2.15}$$

where θ and λ are model parameter and regularization parameter, respectively. Smaller λ relaxes regularization, and if λ is zero then the model is obtained by only minimizing the training error. On the other hand, larger λ regularize model parameters more tightly while reducing the model's fitting to the training dataset.

A linear ϵ-insensitive SVR problem, which is to find the optimal linear function $f(x) = x^T\theta + b$ whose error is less than or equal to a margin ϵ while regularizing the model parameter θ, can be formulated using the structural risk minimization problem as follows.

$$\min_{\theta,\xi,\xi^*} J(\theta,\xi,\xi^*) = \frac{1}{2}\theta^{\mathrm{T}}\theta + C\sum_{i=1}^{n}(\xi_i + \xi_i^*)$$

subject to

$$\begin{aligned} y_i - \left(x_i^{\mathrm{T}}\theta + b\right) &\le \epsilon + \xi_i \\ \left(x_i^{\mathrm{T}}\theta + b\right) - y_i &\le \epsilon + \xi_i^* \\ 0 &\le \xi_i \\ 0 &\le \xi_i^* \end{aligned} \tag{2.16}$$

where ξ and ξ^* are slack variables and $C \ge 0$ is the box constraint. The Lagrange dual formulation of the above linear ϵ-insensitive SVR problem is expressed as in (2.17).

$$\begin{aligned} \min_{\alpha,\alpha^*} J(\alpha,\alpha^*) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\alpha_i - \alpha_i^*)\left(\alpha_j - \alpha_j^*\right)x_i^{\mathrm{T}}x_j \\ + \epsilon\sum_{i=1}^{n}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{n} y_i(\alpha_i^* - \alpha_i) \end{aligned}$$

subject to

$$\begin{aligned} \sum_{i=1}^{n}(\alpha_i - \alpha_i^*) &= 0 \\ 0 \le \alpha_i &\le C \\ 0 \le \alpha_i^* &\le C \end{aligned} \tag{2.17}$$

where α and α^* are the Lagrangian multipliers. The above dual formula for a linear ϵ-insensitive SVR problem can be extended to a nonlinear problem by introducing a nonlinear kernel function, $K(x_i, x_j)$. The Lagrangian dual formulation of a nonlinear ϵ-insensitive SVR problem is formulated by replacing the inner product with a nonlinear kernel function as follows.

$$\min_{\alpha,\alpha^*} J(\alpha,\alpha^*) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\alpha_i-\alpha_i^*)\left(\alpha_j-\alpha_j^*\right)K(x_i,x_j)$$
$$+\epsilon\sum_{i=1}^{n}(\alpha_i+\alpha_i^*)+\sum_{i=1}^{n}y_i(\alpha_i^*-\alpha_i)$$
subject to (2.18)
$$\sum_{i=1}^{n}(\alpha_i-\alpha_i^*)=0$$
$$0\le\alpha_i\le C$$
$$0\le\alpha_i^*\le C$$

The above optimization problems can be solved using quadratic programming (QP) techniques, such as decomposition methods and sequential minimal optimization (SMO) [5, 61-63].

2.3.3. Gaussian process regression

Gaussian process regression (GPR), also known as Kriging in geostatistics field, is data-driven regression method based on an assumption that the output variable is expressed as a regression function with a Gaussian prior distribution [64, 65]. In contrast to other black-box data-driven regression models, a GPR model provides the probabilistic information of its prediction.

A Gaussian process (GP) is defined as a collection of the finite number of random variables which have a joint Gaussian distribution [65]. Consider the training dataset of the input $\boldsymbol{X}=\{\boldsymbol{x_i}\}$ and output $\boldsymbol{y}=\{y_i\}$. Since the variables of a GP have the joint Gaussian distribution, the mean and covariance functions of a GP, $\{f(\boldsymbol{x})|\boldsymbol{x}\in\boldsymbol{X}\}$, can be expressed as follows.

$$m(\boldsymbol{x}) = \mathbb{E}[f(\boldsymbol{x})] \tag{2.19}$$

$$k(\boldsymbol{x}, \boldsymbol{x}') = \mathbb{E}\big[\big(f(\boldsymbol{x}) - m(\boldsymbol{x})\big)\big(f(\boldsymbol{x}') - m(\boldsymbol{x}')\big)\big] \tag{2.20}$$

where $m(\boldsymbol{x})$ and $k(\boldsymbol{x}, \boldsymbol{x}')$ are the mean function and covariance function, respectively. A linear regression model, $g(\boldsymbol{x})$ can be formulated using explicit basis functions, $h(\boldsymbol{x})$ and a GP, $f(\boldsymbol{x})$.

$$g(\boldsymbol{x}) = f(\boldsymbol{x}) + h(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{\beta} \tag{2.21}$$

where $\boldsymbol{\beta}$ is the basis function coefficient vector and $f(\boldsymbol{x}) \sim GP(0, k(\boldsymbol{x}, \boldsymbol{x}'))$, which indicates that the mean function of $f(\boldsymbol{x})$ is zero. The probability of y_i given $f(\boldsymbol{x_i})$ and $\boldsymbol{x_i}$ can be expressed as in (2.22).

$$P(y_i | f(\boldsymbol{x_i}), \boldsymbol{x_i}) \sim N\big(\boldsymbol{y_i} \big| h(\boldsymbol{x_i})^{\mathrm{T}}\boldsymbol{\beta} + f(\boldsymbol{x_i}), \sigma^2\big) \tag{2.22}$$

where σ^2 is the noise variance assumed that the noise of the regression function, $y_i = g(x_i) + \epsilon$, follows Gaussian distribution. For a new test data $\boldsymbol{x}_{new}$, the predictive distribution is a Gaussian distribution with mean and variance expressed as follows.

$$\widehat{\boldsymbol{y}}(\boldsymbol{x}_{new}) = \boldsymbol{k}^{\mathrm{T}}(\boldsymbol{x}_{new})\boldsymbol{K}^{-1}\boldsymbol{y} \tag{2.23}$$

$$\sigma^2_{\widehat{\boldsymbol{y}}}(\boldsymbol{x}_{new}) = c(\boldsymbol{x}_{new}, \boldsymbol{x}_{new}) - \boldsymbol{k}^{\mathrm{T}}(\boldsymbol{x}_{new})\boldsymbol{K}^{-1}\boldsymbol{k}(\boldsymbol{x}_{new}) \tag{2.24}$$

where $\boldsymbol{k} = [c(\boldsymbol{x}_{new}, \boldsymbol{x}_1), \dots, c(\boldsymbol{x}_{new}, \boldsymbol{x}_n)]$ is the vector of covariance between the newly obtained data and the training data. $\boldsymbol{K}$ is the covariance matrix where $\boldsymbol{K}_{ij} = c(\boldsymbol{x}_i, \boldsymbol{x}_j)$. The covariance function, $c(\boldsymbol{x}_i, \boldsymbol{x}_j)$, is approximated using a kernel

function in general. A detailed description of GPR and the derivations of probability distributions can be found in the literature [64, 65].

Chapter 3. Hybrid Modeling Approach for Industrial Polymerization Process[1]

3.1. Introduction

The melt index (MI), also known as the melt flow index (MFI), of a thermoplastic polymer is a critical quality variable that affects the flow and mechanical properties of the polymer product. MI is defined as the weight of melted polymer samples flowing through a die over a specified period of time, typically 10 min [66-69]. Thus, MI is inversely proportional to the viscosity of melted polymer and is used as an indirect measure of molecular weight. MI is used to determine the grade of a polymer product and whether it is off-spec or not for the majority of commercial polymerization processes. Therefore, the monitoring and control of MI is essential for the operation of a polymerization process. However, due to the lack of online measurement technology, MI cannot be measured in real-time during polymerization. Instead, MI is typically measured offline in a laboratory only once every 2 or 4 hours, which is much less frequent than the measurement of process variables such as reactor temperature, pressure, and flow rate. MI measurements are delayed and infrequent, posing challenges for real-time quality control of products. Additionally, polymerization processes produce a continuous stream of different grades of polymer products. The process suffers from significant settling time and overshoot during grade changeover, complicating the monitoring and control of polymer properties. As a result, it is critical to estimate MI as accurately and rapidly as

[1] This chapter is an adapted version of M. J. Song, S. H. Ju, S. Kim, S. H. Oh, and J. M. Lee, "Hybrid modeling approach for polymer melt index prediction", *Journal of Applied Polymer Science*, **2022**, vol. 139, no. 41, e52987.

possible from the easy-to-measure process variables and to control based on the estimation in order to minimize off-specification products and maximize process efficiency.

Numerous studies have been conducted over the last several decades to infer the MI of a polymer product using readily available process variables. Previously, mechanistic models derived from first principles such as material and energy balances and reaction kinetics, as well as empirical models derived from experimental data, were developed primarily on the basis of the relationship between MI and molecular weight or viscosity. As a result, models for instantaneous and cumulative MI inferences have been constructed using the logarithm of the linear combination of the process variables [66, 67, 70, 71]. In addition, MI prediction models have been proposed using a mathematical modeling [72] and a continuum mechanics method [73]. While carefully developed mechanistic models make accurate MI predictions, the complex correlations and high nonlinearities inherent in industrial polymerization processes make developing accurate mechanistic models more difficult and time-consuming.

Recent advances in computer and data technologies have enabled the availability of high-dimensional process measurement data, allowing for more extensive research on data-driven soft sensor models for MI prediction. Only historical operation data obtained from the process and laboratory quality measurements is used to build an inference model in data-driven models. Thus, data-driven models are significantly easier to develop than mechanistic models, as they do not require prior knowledge of the process, such as reaction mechanisms and relationships between variables. In order to address the multicollinearity and computational load associated with high-dimensional process data, PLS method which projects both process and quality data onto the lower-dimensional latent space has been widely applied for MI prediction [4-6, 33]. However, the PLS model's performance in predicting MI is limited because it is based on linear relationships between variables,

whereas polymerization processes are highly nonlinear.

Rather than that, nonlinear machine learning modeling methods such as SVM [5, 6], least squares SVM [8-10], relevance vector machines [74], and GPR [11-15] have been used. In addition, MI inference models based on various types of artificial neural networks (ANNs) [75, 76] have been developed including radial basis function neural networks (RBFNNs) [16-21], fuzzy neural networks (FNNs) [22, 23], wavelet neural networks (WNNs) [24], fuzzy wavelet neural networks (FWNNs) [25], deep belief networks (DBNs) [26], and dilated convolution neural networks (DCNNs) [27]. While the aforementioned nonlinear machine learning soft sensor models demonstrated good prediction performance for industrial polymerization processes, there are still academic and industrial demands for a prediction model with improved performance and generalization ability.

Rather than developing a model solely through mechanistic or data-driven modeling, there has been a growing interest in developing a hybrid model through the combination of prior process knowledge and data-driven modeling methods, particularly in chemical and biochemical engineering [77]. Hybrid modeling offers several advantages over traditional mechanistic and machine learning modeling techniques, including increased prediction accuracy, improved calibration properties, and enhanced extrapolation properties. In the field of chemical engineering, hybrid models based on neural networks and first-principles calculations have been developed and demonstrated improved modeling performance for a polymerization process [78], fed-batch bioreactor [79, 80], continuous stirred tank reactor (CSTR) [81], hydrocracking unit [82], and fluidized catalytic bed reactor (FBR) [83].

In this chapter, a hybrid modeling approach for predicting MI in an industrial polymerization process is proposed. The proposed modeling approach is based on the idea of enhancing prediction and generalization performance through the integration of prior process knowledge and machine learning modeling methods. The proposed hybrid model is constructed in a serial structure, beginning with a white-

box submodel ending with a black-box submodel. First, the white-box submodel, which is based on the mechanistic modeling method, calculates information about the current state of the process and the polymer product. White-box submodel's information is then passed along with the process measurement data to the black-box submodel, which is a machine learning MI inference model used to predict the MI of the polymer product. The inclusion of polymerization information from the white-box submodel is critical for more accurate prediction of polymer quality. The proposed approach is evaluated in terms of prediction performance using a commercial styrene-acrylonitrile (SAN) polymerization process in South Korea and compared to the performances of data-only soft sensor models and mechanistic soft sensor models. In addition, to demonstrate that the proposed hybrid modeling approach's improvement in prediction performance is not limited to a single black-box modeling method, a total of five machine learning methods are used as the black-box submodel. The models considered in this study are the most widely used regression models in machine learning, including PLS, decision tree, SVM, ANN, and GPR. This chapter demonstrates that the proposed hybrid modeling approach outperforms conventional machine learning modeling approaches in terms of prediction accuracy.

The remainder of this chapter is organized as follows. Subsection 3.2 briefly describes the target process and the data and variables obtained from the process. In Subsection 3.3, the overall structure of the proposed hybrid model is first introduced and the modeling procedures for the white-box and black-box submodels are presented. Finally, Subsection 3.4. contains the modeling results and a comparison of the proposed hybrid modeling approach to the traditional data-only modeling methods.

3.2. Process and data description

Styrene-acrylonitrile resin is a thermoplastic polymer that is widely used in a variety of industrial products due to its high resistance to heat and chemicals. The purpose of this paper is to predict the MI of a SAN polymer product that is commercially produced in South Korea. In contrast to commonly produced SAN resin, α-methylstyrene is added as a third monomer during the target process in order to increase the polymer's heat resistance. Thus, the polymerization reaction involves a total of three monomers: styrene, acrylonitrile, and α-methylstyrene. **Figure 3.1** shows the simplified schematic of the SAN polymerization process which consists of two reactors, two devolatilizers, and a pelletizer. First, three different types of monomers are mixed with the initiator and the monomers and solvent recovered from the devolatilizers. The reaction mixture is then subjected to sequential free radical bulk polymerization in the liquid phase in the first and second reactors. Due to the high reactor temperature and exothermic polymerization reaction in the second reactor, a portion of the reaction mixture is vaporized. The vaporized molecules are transferred to a tank connected to the reactor, where they are condensed and reintroduced. Additionally, a small amount of monomer-initiator mixture is fed to the tank to keep the monomer concentrations in the second reactor high. Consequently, the second reactor produces more polymers. After achieving the monomer conversion of about 65% in the reactors, the reaction mixture is conveyed to the devolatilizers for drying. The reaction mixture, excluding polymer particles, is vaporized and recycled in the devolatilizers. The final step is to pelletize the polymer products and transport them to the storage silos. The SAN polymerization process produces two distinct grades of SAM polymer products with varying MI values, mechanical and chemical properties.

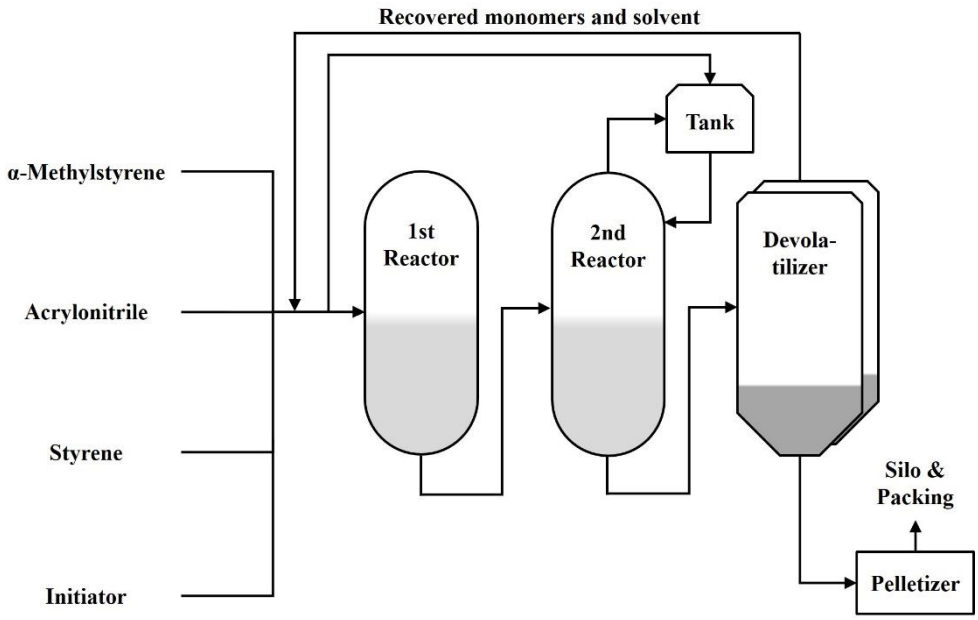


Figure 3.1. Schematic of SAN polymerization process.

The measurement data obtained from the process contains two types of variables: process variables and quality variables. Process variables are readily measurable online via sensors and provide information about the current operating condition of the process. Every hour, a total of 27 process variables are measured in the SAN polymerization process. The process measurement includes temperature, pressure, and level of the reactors and tank. Additionally, the flowrate measurements of monomers, initiator, and recycled stream are available. Finally, the required current for pump operation in the process is measured. On the other hand, quality variables are typically measured offline every 4 hours via laboratory analysis, which is much less frequent than process variables. MI and color of polymer samples collected from the pelletizer are considered to be quality variables. If one of the quality variables in a polymer product does not meet the specification standard, it is considered off-specification. MI of a SAN polymer sample is measured in a laboratory analysis where molten polymer samples are flowing through a die over 10 min. Process and quality measurement data were collected for approximately 17 months from January 2, 2020 to May 23, 2021 for the research. After removing mismeasured and shutdown data, a total of 2285 MI measurement data are available for modeling.

3.3. Proposed hybrid modeling approach

In this subsection, a hybrid modeling procedure for MI prediction in the commercial SAN polymerization process described in Subsection 3.2 is proposed. First, the structure of the proposed model is introduced. Then, the procedures for modeling white-box and black-box submodels are discussed.

3.3.1. Hybrid model structure

A hybrid MI inference model is proposed in this study. It consists of a white-box submodel that is a mechanistic model and a black-box submodel that is a machine learning model. A hybrid model composed of a white-box and a black-box submodel can be structured in one of two ways: parallel or serial [77]. As illustrated in **Figure 3.2**, the proposed hybrid model is constructed on a serial structure, with the output of the white-box submodel serving as an input to the black-box submodel. In this study, the disadvantages of two modeling approaches are compensated by adopting the serial structure. While the polymerization mechanisms of radical bulk polymerization have been extensively studied in the past, the relationships between MI and process or kinetic variables remain unknown or inaccurate because of the deviations between reality and assumptions, imposing difficulties in the accurate prediction of MI using first principles. The data-driven black-box regression model compensates for the low accuracy of white-box MI prediction model derived from the model-plant mismatch. On the other hand, the black-box models have low extrapolation abilities, which is compensated by considering process knowledge in first principles models.

In the proposed hybrid model, the white-box submodel calculates polymerization-related variables such as molecular weights, polymerization rates, and degree of polymerization using process knowledge such as reaction kinetics and mass and energy balances. The white-box submodel predictions and process data are then used as inputs to the black-box submodel that predicts the value of MI. The output of the white-box submodel provides information about the process and its current state to the black-box submodel, which improves prediction accuracy and generalizability when compared to a black-box model based solely on measurement data.

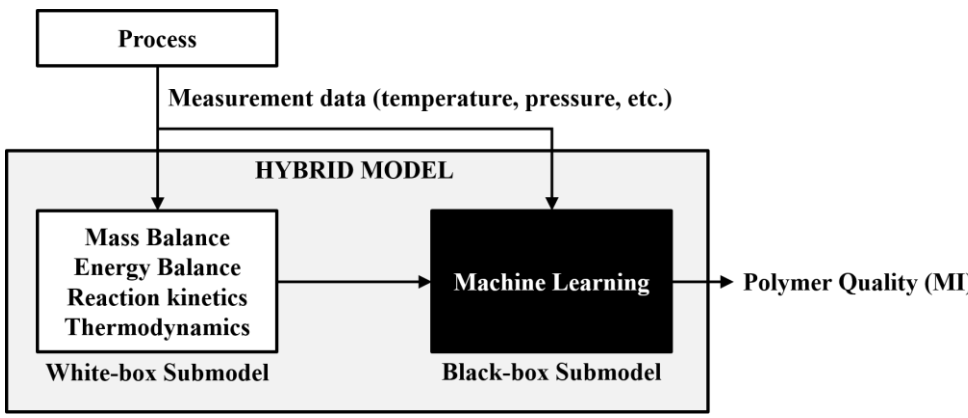


Figure 3.2. Proposed hybrid model structure.

3.3.2. White-box modeling

The mechanistic modeling procedure of the white-box submodel is based on the dynamic modeling of a modified SAN polymerization process [84]. For dynamic simulation of the polymerization process, Aspen Dynamics is used to apply the segment-based polymer non-random two-liquid (polymer-NRTL) activity coefficient model based on the local composition model developed by Chen [85], the group contribution method and the method of moments. The nomenclature for the modeling is listed in **Table 3.1**. The white-box submodel employs a segment-based approach, in which each polymer chain in the system is defined as a sequence of segments or repeating units. A polymer's segments are made of several functional groups. For substances whose thermophysical properties have not been instigated are thus unavailable, the properties of the segments such as heat capacity and density are calculated by the Van Krevelen group contribution method and the functional group properties. Additionally, the properties of a polymer can be calculated by examining the polymer segments. UNIFAC, another group contribution method, is used to estimate the binary interaction parameters of polymer-NRTL. The UNIFAC group contribution method is a method for predicting liquid-phase nonelectrolyte activity coefficients in nonideal liquid mixtures [86]. Using group contribution methods, the phase equilibria of the SAN polymerization system are calculated using the poylmer-NRTL activity coefficient model and the Soaves-Redlich-Kwong (SRK) cubic equation of state.

Table 3.1. Nomenclature for white-box modeling.

Symbol	Description
I	Initiator
S	Solvent
$R\cdot$	Primary radicals
M_i	Monomer of type i
P_n^i	Live polymer chain of length n having an active segment of type i
D_n	Dead polymer chain of length n
$D_n^{i=}$	Dead polymer chain of length n having a terminal double bond of type i
k	Reaction rate constant
k_0	Pre-exponential factor
E_a	Activation energy
ΔV	Activation volume
T_{ref}	Reference temperature for reaction rate constant
x_i	Segment-based liquid phase mole fraction
r_i	Degree of polymerization
n_I	Number of moles
m_I	Ratio of the free volume of polymer
ϕ_I	Segment mole fraction
τ_{ij}	Binary interaction parameter
α_{ij}	Non-randomness factor
γ_I	Activity coefficient of the species I

In the polymer-NRTL model, the Gibbs energy of mixing for a polymer solution is defined as the sum of entropy of mixing and the enthalpy of mixing.

$$\frac{\Delta G_{mixing}}{RT} = \frac{\Delta H_{mixing}^{NRTL}}{RT} - \frac{\Delta S_{mixing}^{FH}}{R} \tag{3.1}$$

The preceding equation for the Gibbs energy of mixing may be modified by substituting the entropy of mixing and the enthalpy of mixing from the Flory-Huggins and NRTL theories.

$$\frac{\Delta G_{mixing}}{RT} = \sum_s n_s \frac{\sum_j x_j G_{js} \tau_{js}}{\sum_j x_j G_{js}} + \sum_p n_p \sum_i r_{i,p} \frac{\sum_j x_j G_{ji} \tau_{ji}}{\sum_j x_j G_{ji}} + \sum_I n_I \ln(\phi_I) \tag{3.2}$$

Likewise, the activity coefficient of each species can be expressed as the sum of two contributions from the Flory-Huggins and NRTL theories.

$$\ln(\gamma_I) = \ln(\gamma_I^{NRTL}) + \ln(\gamma_I^{FH}) \tag{3.3}$$

where,

$$\ln(\gamma_I^{FH}) = \ln\left(\frac{\phi_I}{x_I}\right) + 1 - m_I \sum_J \left(\frac{\phi_J}{m_J}\right) \tag{3.4}$$

For the solvent and polymer activity coefficients, $\ln(\gamma_I^{NRTL})$ is expressed as follows.

$$\ln(\gamma_{I=s}^{NRTL}) = \frac{\sum_j x_j G_{js}\tau_{js}}{\sum_j x_j G_{js}} + \sum_j \frac{x_j G_{js}}{\sum_k x_k G_{kj}}\left(\tau_{sj} - \frac{\sum_k x_k G_{ks}\tau_{ks}}{\sum_k x_k G_{kj}}\right) \tag{3.5}$$

$$\ln(\gamma_{I=p}^{NRTL}) = \sum_i r_{i,p}\left(\frac{\sum_j x_j G_{ji}\tau_{ji}}{\sum_k x_k G_{ki}} + \sum_j \frac{x_j G_{ji}}{\sum_k x_k G_{kj}}\left(\tau_{ij} - \frac{\sum_k x_k G_{kj}\tau_{kj}}{\sum_k x_k G_{kj}}\right)\right) \tag{3.6}$$

The binary parameter, τ_{ij}, and non-randomness factor, α_{ij}, are considered to be dependent on the temperature.

$$\tau_{ij} = a_{ij} + \frac{b_{ij}}{T} + e_{ij}\ln(T) + f_{ij}T \tag{3.7}$$

$$\alpha_{ij} = c_{ij} + d_{ij}(T - 273.15) \tag{3.8}$$

where a_{ij}, b_{ij}, c_{ij}, d_{ij}, e_{ij}, and f_{ij} are the NRTL parameters.

The terpolymerization reaction of the SAN process is free radical bulk polymerization. Thus, chain initiation, propagation, chain transfer, terminal double bond polymerization, and termination reaction mechanisms are considered in modeling the white-box submodel. In chain initiation, initiator decomposition and thermal initiation of styrene monomers resulting from high reaction temperature are considered to produce primary radicals which lead to primary chain initiation reaction. Then, live polymer chains propagate by reacting with monomers.

Initiator decomposition	$I \rightarrow 2R \cdot$	(3.9)
Styrene thermal initiation	$3M_1 \rightarrow R \cdot$	(3.10)
Chain initiation	$R \cdot + M_i \rightarrow P_1^i$	(3.11)
Propagation	$P_n^i + M_j \rightarrow P_{n+1}^j$	(3.12)

Chain transfer reactions to monomer, solvent, and polymer have the following mechanisms. While chain transfer to small molecules such as monomers and solvent decreases the molecular weight of the polymer, chain transfer to polymer produces a long chain branch in the polymer molecule.

Chain transfer to monomer	$P_n^i + M_j \rightarrow D_n + P_1^j$	(3.13)
Chain transfer to solvent	$P_n^i + S \rightarrow D_n + R \cdot$	(3.14)
Chain transfer to polymer	$P_n^i + D_m \rightarrow D_n + P_m^j$	(3.15)

In terminal double bond reaction, growing polymer chains react with dead polymer chains that contain terminal double bonds, resulting in the formation of long chain branches.

Terminal double bond polymerization	$P_n^i + D_m^{j=} \rightarrow P_{n+m}^j$	(3.16)

Finally, polymer growth stops when one of the following termination reactions occur.

Termination by combination	$P_n^i + P_m^j \rightarrow D_{n+m}$	(3.17)
Termination by disproportionation	$P_n^i + P_m^j \rightarrow D_n + D_m^=$	(3.18)

It is assumed that the reaction rate constants of all chemical reactions follow the Arrhenius equation.

$$k = k_0 e^{\left[-\frac{E_a+\Delta VP}{R}\left(\frac{1}{T}-\frac{1}{T_{ref}}\right)\right]} \tag{3.19}$$

The NRTL parameters and kinetic parameters including k_0, E_a and ΔV are fitted in an offline manner using molecular weight distribution data obtained during the process, except for the parameters which are already available from previous studies. For instance, the kinetic parameters for the SAN polymerization reaction, whose values are not available, are initially guessed from the results of previous copolymerization studies and fitted using process data.

The chain length distribution has always been most demanded data, since it has direct effect on the mechanical and fluid properties of polymer products. The method of moments is one of the most widely used techniques for estimating the average properties of a polymer. While calculating the whole chain length distribution throughout the polymerization process is time-consuming, the method of moments is significantly faster. The moments of the chain length distribution are defined as in (3.20).

$$\lambda_n = \sum_{k=1}^{\infty} k^n [D_k] \tag{3.20}$$

Due to the fact that the number of live polymer chains is significantly fewer than the number of dead polymer chains due to their high reactivity, the moments of the dead chains are assumed to be equal to the moments of the bulk polymer chains. The moments can be used to compute the average polymer properties such as the number-average and weight-average degrees of polymerization, as well as polydispersity index (PDI). For example, the degree of polymerization on a number-average and

weight-average basis is computed as $DP_n = \lambda_1/\lambda_0$ and $DP_w = \lambda_2/\lambda_1$. The kinetics and moments are used to calculate a total of 17 polymerization-related output variables for the white-box submodel. The descriptions of the output variables of the white-box submodel are listed in **Table 3.2**.

Table 3.2. Output variables of the white-box submodel.

Symbol	Description
λ_n	n^{th} moment of chain length distribution, $n = 0,1,2$
λ_1^i	Mole flow of segment of type i
F^i	Mole fraction of segment of type i
DP_n	Number-average degree of polymerization
DP_w	Weight-average degree of polymerization
MW_n	Number-average molecular weight
MW_w	Weight-average molecular weight
PDI	Polydispersity index
R_p	Polymerization rate
LCB	Number of long chain branches
$\Delta FLCB$	Long chain branching frequency

The majority of the white-box output variables represent the molecular structure of the polymer product. These molecular characteristics of the polymer are known to affect the polymer's properties and processability. For instance, mechanical and chemical properties such as transparency, tensile strength, cold resistance, and chemical resistance are all dependent on the molecular weight and its distribution [1]. In addition, long chain branching of the polymer affects its tensile and impact strengths, as well as processability characteristics such as bubble stability and extrusion torque [1]. Furthermore, it has been extensively examined and established that the MI of a thermoplastic polymer is inversely proportional to its molecular weight [67, 69]. For instance, a relationship between MI and molecular weight has been developed for high density polyethylene [87].

$$\mathrm{MI} = a(bMW_w + cMW_n)^d \tag{3.21}$$

where a, b, c, and d are constant model coefficients. Additionally, a simpler relationship between MI and molecular weight has been developed based on the Pouiselle Equation [67].

$$\mathrm{MI}^{-1} = GMW_w^x \tag{3.22}$$

where G and x are constant model coefficients. Another modeling study established the relationship between MI and process variables using the linear combination of logarithms [72, 88]. For instance, a theoretical model for predicting MI in a high density polyethylene reactor has been developed based on the reactor temperature and ratio of monomer concentrations [1].

$$\ln(\mathrm{MI}) = \beta + \alpha_1 \, ln\left(\frac{[H_2]}{[M_1]}\right) + \alpha_2 \, ln\left(\frac{[M_2]}{[M_1]}\right) + \alpha_3 \, ln\left(\frac{[M_3]}{[M_1]}\right) + \alpha_4 \, ln([R]) + \alpha_5 \, ln(T) \tag{3.23}$$

where α_i and β are model parameters and $[M_i]$, $[H_2]$, and $[R]$ represent the concentrations of the i^{th} monomer species, hydrogen, and cocatalyst, respectively. In conclusion, MI has relationships with the white-box output variables acquired from the mechanistic modeling. Therefore, the white-box output variables aid in more precise prediction of MI in the proposed hybrid modeling approach.

3.3.3. Black-box modeling

Both the measurement data for the process variables and the output from the white-box submodel were utilized as predictors in the black-box submodel to estimate MI using machine learning modeling methods. The modeling procedure for the black-box submodel consists of two steps: data preprocessing and model training. All variables, including process and quality variables, as well as the white-box output, were normalized first to have zero means and unit variances during the data preprocessing step. Then, because the process and MI measurements use different sampling times, the input variables to the black-box submodel were averaged over a 4-hour period. Finally, the averaged input variables were subjected to principal component analysis (PCA) in order to obtain the newly constructed variables that are linearly independent of one another. The process variables that comprise measurement data are highly correlated. For instance, there are four temperature measurements for the first reactor, including the average reactor temperature. In addition, the output variables of the white-box submodel, such as the weight-average

and number-average molecular weights, are highly correlated. Thus, by applying PCA in the data preprocessing step, multicollinearity between the input variables is eliminated. Additionally, the computational cost of training a machine learning model was reduced by using only 10 principal components (PCs) out of a total for 44 PCs as inputs, which makes sense given that the first 10 PCs account for approximately 93.1% of the variance in the data, as illustrated in **Figure 3.3**.

After preprocessing the data, a MI prediction model was trained using machine learning. Approximately 80% of the sample data obtained from the process was used for training, while remaining 20% was utilized to evaluate the model's performance. The black-box submodel in this work was constructed using five popular machine learning modeling methods: PLS, decision tree, SVM, ANN, and GPR. **Table 3.3** summarizes the features and hyperparameters of the black-box submodels. To avoid further linear transformation of the variables, the PLS modeling technique omits the PCA preprocessing step. Instead of PCs, the PLS model predicts MI using 10 linearly independent PLS components. To determine the optimal approach to split nodes, the binary decision tree model was trained using the standard classification and regression tree (CART) algorithm. As a result, the tree structure with the lowest mean squared error (MSE) was obtained for prediction. For the kernel functions of the SVM and GPR models, Gaussian and exponential kernels were used, respectively. The ANN model structure is based on a multi-layer perceptron with an activation function of hyperbolic tangent. The weights of the ANN network were initialized by Glorot initialization, and each layer's initial bias was set to 0. Finally, the ANN model was trained for 1000 iterations using the LBFGS algorithm. Except for the PLS model, the hyperparameters of the black-box models presented in **Table 3.3** were tuned using Bayesian optimization. As a result, for each model, the set of hyperparameters with the minimum 5-fold cross validation error was obtained.

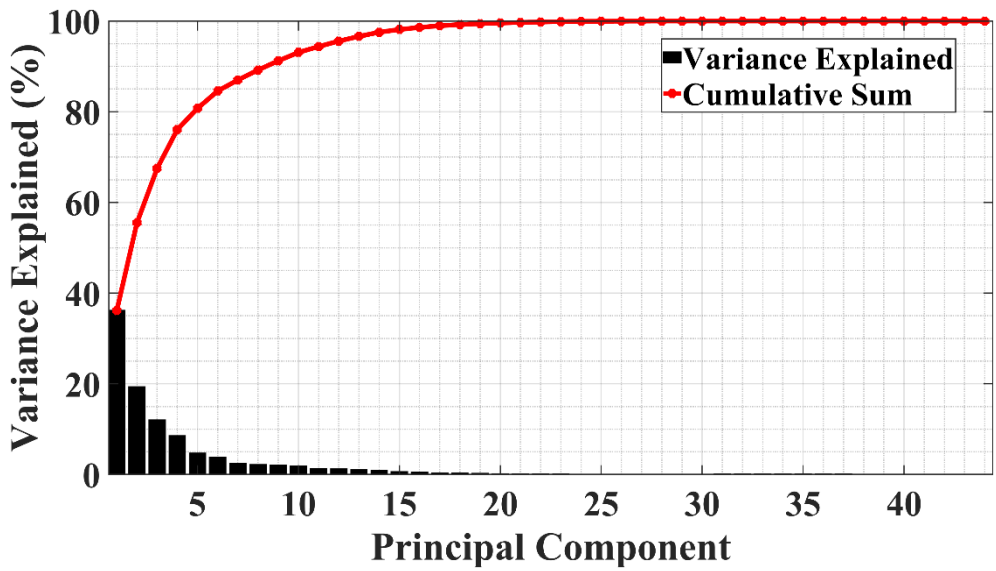


Figure 3.3. Scree plot of variances explained by PCs.

Table 3.3. Machine learning models used for training of the black-box submodel.

Model	Property	Hyperparameter
PLS	Projection of data onto linear latent space	Number of PLS components
Decision tree	Binary regression	Minimum leaf size
SVM	Nonlinear regression using Gaussian kernel	Box constraint Kernel scale Width of the ϵ-insensitive band Regularization strength
ANN	Multi-layer perceptron	Number of hidden layers Number of nodes
GPR	Exponential covariance kernel function	Initial value of the noise standard deviation

3.4. Results and discussions

To compare the prediction performance of the proposed modeling approach to that of a conventional data-driven approach, five hybrid models based on the proposed approach and five data-only models were trained using PLS, decision tree, SVM, ANN, and GPR. The modeling procedure for a data-only model is identical to the procedure described in Subsection 3.3.3, except that the model inputs are limited to measurement data for process variables. Over a 4-hour period, moving averages of the normalized process variables were taken and then linearly transformed into the 10 PCs. In addition, the data-only models optimized the same set of hyperparameters using Bayesian optimization.

A total of 2285 data samples from the industrial plant were available for this study. A total of 457 data samples, or 20% of the whole dataset, were randomly chosen as a testing dataset for performance comparison. On the testing dataset, the models' prediction abilities were evaluated using four statistical indices: root mean squared error (RMSE), mean absolute percentage error (MAPE), Theil's inequality coefficient (TIC), and standard deviation of absolute error (STD).

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(y_k - \hat{y}_k)^2} \tag{3.24}$$

$$\mathrm{MAPE} = 100 \times \frac{1}{N}\sum_{k=1}^{N}\left|\frac{y_k - \hat{y}_k}{y_k}\right| \tag{3.25}$$

$$\mathrm{TIC} = \frac{\sqrt{\sum_{k=1}^{N}(y_k - \hat{y}_k)^2}}{\sqrt{\sum_{k=1}^{N} y_k^2} + \sqrt{\sum_{k=1}^{N} \hat{y}_k^2}} \tag{3.26}$$

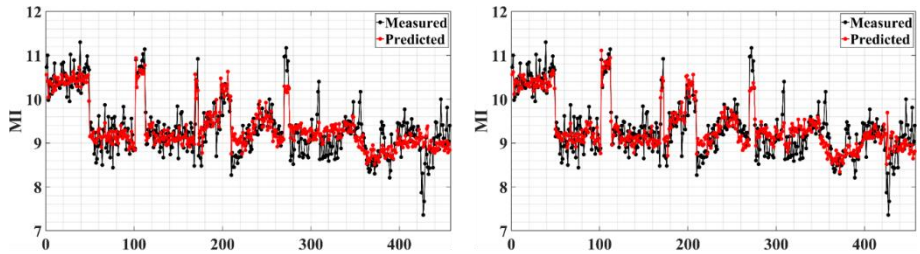
$$\mathrm{STD} = \sqrt{\frac{1}{N-1}\sum_{k=1}^{N}(e_k - \bar{e})^2} \tag{3.27}$$

where N is the number of test samples, $e_k = y_k - \hat{y}_k$, $\bar{e} = \frac{(\sum_{k=1}^{N} e_k)}{N}$, and y_k, $\hat{y}_k$ represent the measured and predicted value of MI, respectively. The RMSE and MAPE measure the prediction accuracy of a model. The TIC evaluates the agreement between the trained model and the process. The TIC value becomes zero when the predicted and measured sequences are identical. Smaller STD values indicate that a model is more stable.

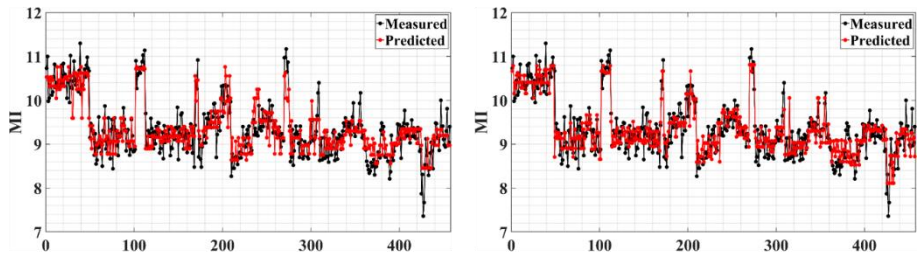
Table 3.4 summarizes the prediction performance of the hybrid and data-driven models on the testing dataset, as well as their differences. It is clear that hybrid models outperform data-only models in all modeling methodologies. PLS models perform the lowest in terms of prediction accuracy, whereas GPR models perform the best in terms of prediction accuracy for both hybrid and data-only modeling methods. Due to the highly nonlinear and dynamic nature of the polymerization process, PLS models have difficulty in capturing the process behavior, resulting in the worst prediction performance. However, the hybrid PLS model outperforms the data-only PLS model by 5.442%, 3.922%, 5.427%, and 5.434% in terms of RMSE, MAPE, TIC, and STD. The improved prediction performance of the hybrid PLS model suggests that mechanistic knowledge derived from the white-box submodel can be advantageous even when using linear modeling approaches. The binary decision tree models predicted MI more accurately than the PLS models, and the hybrid modeling approach also improved the prediction accuracy of the tree model, with percentage decreases in the RMSE, MAPE, TIC and STD of 8.787%, 6.325%, 8.784%, and 8.782%, respectively.

Table 3.4. MI prediction performances of all trained models on the testing dataset.

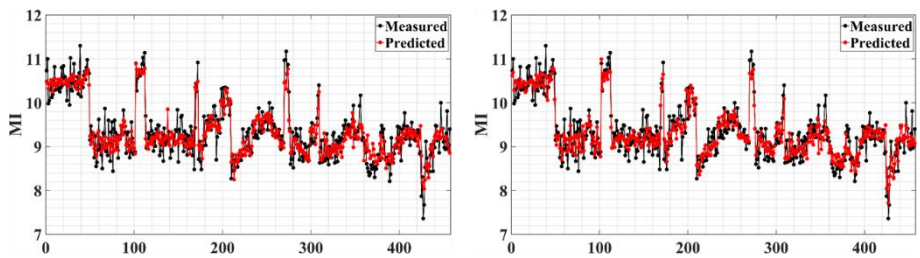
Modeling method		RMSE	MAPE	TIC	STD
PLS	Data-only	0.3600	2.948	0.01918	0.3604
	Hybrid	0.3404	2.832	0.01814	0.3408
	(Difference)	**-5.444%**	**-3.922%**	**-5.427%**	**-5.434%**
Decision tree	Data-only	0.3509	2.871	0.01869	0.3512
	Hybrid	0.3200	2.689	0.01705	0.3204
	(Difference)	**-8.787%**	**-6.325%**	**-8.784%**	**-8.782%**
SVM	Data-only	0.2845	2.413	0.01517	0.2846
	Hybrid	0.2544	2.157	0.01357	0.2545
	(Difference)	**-10.56%**	**-10.59%**	**-10.56%**	**-10.59%**
ANN	Data-only	0.2854	2.415	0.01521	0.2857
	Hybrid	0.2553	2.179	0.01361	0.2556
	(Difference)	**-10.56%**	**-9.745%**	**-10.57%**	**-10.55%**
GPR	Data-only	0.2637	2.212	0.01406	0.2640
	Hybrid	0.2366	1.994	0.01261	0.2368
	(Difference)	**-10.30%**	**-9.873%**	**-10.30%**	**-10.30%**



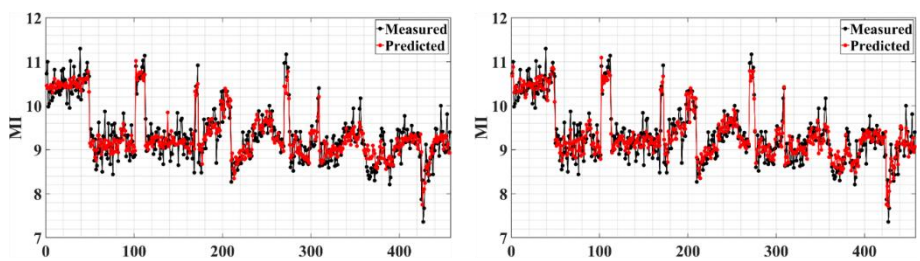
(a) data-only (left) and hybrid (right) PLS models



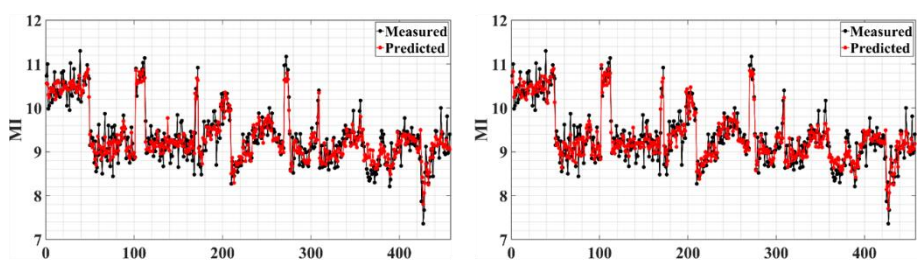
(b) data-only (left) and hybrid (right) decision tree models



(c) data-only (left) and hybrid (right) SVM models



(d) data-only (left) and hybrid (right) ANN models



(e) data-only (left) and hybrid (right) GPR models

Figure 3.4. Prediction performances of data-only and hybrid models.

Nonlinear machine learning methods such as SVM, ANN, and GPR perform significantly better than PLS and decision tree models in terms of prediction performance. Additionally, the improvements in the prediction performance are greater for the nonlinear machine learning approaches. For instance, the RMSE, MAPE, TIC, and STD of the hybrid GPR model are 0.2366, 1.994, 0.01261, and 0.2368, respectively, resulting in a reduction of 10.30%, 9.873%, 10.30%, and 10.30%, respectively, when compared to the data-only GPR model.

Figure 3.4 shows visual comparisons of the measured and predicted MI values of the testing dataset for each modeling approach. The red curves with circles represent predictions from trained hybrid and data-only models, whereas the black curves with circles represent the measured MI values. Apparently, the predictions of the hybrid models correspond more closely with the measurements than the data-only models' forecasts do. The testing data samples with a sample number greater than 424 were collected when the process experienced rapid changes in operating conditions and process failures, resulting in the creation of off-spec polymer products. For instance, during polymerization, the process's temperature control system failed for almost 19 hours. In such unsteady and unstable circumstances, precise MI prediction is critical for polymer quality control in order to minimize economic losses caused by off-spec products.

Table 3.5 summarizes the prediction performance of the hybrid and data-driven models under unsteady states using testing data samples 424-457. The prediction performance on the unsteady testing dataset is lower than the prediction performance on the overall testing dataset. Since the number of data samples obtained under unsteady states is not large enough to learn from, the data-driven models exhibit lower prediction accuracy. Additionally, the operating condition during the unsteady states is changing rapidly, imposing difficulties in training the data-driven models. However, with the exception of the PLS model, the disparities in prediction performance between the data-only and hybrid models are much larger in the

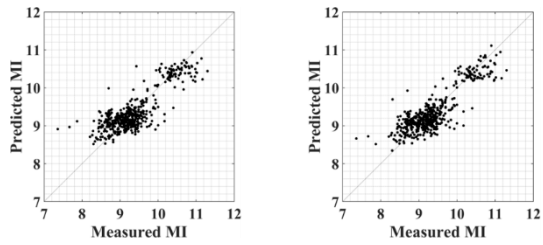
unsteady condition. For example, the RMSE, MAPE, TIC, and STD of the GPR models differ by -20.30%, -21.45%, -20.16%, and -19.58%, respectively. This is also true for the remaining three machine learning models: decision tree, SVM, and ANN. The results from the unsteady state testing dataset reveal that the benefits of the proposed hybrid modeling technique not only increase prediction accuracy but also aid in achieving higher degree of generality in predicting MI.

Additionally, **Figure 3.5** illustrates the measured and predicted MI values in a two-dimensional space. When the points are dispersed about the $y = x$ line, the model and process are said to be in good agreement. For example, the plot of the hybrid GPR model that is most consistent with the process demonstrates that the points on the plot are close to $y = x$ line than the points on the other models. On the other hand, the points on the plots of data-only models are more widely scattered around the plane.

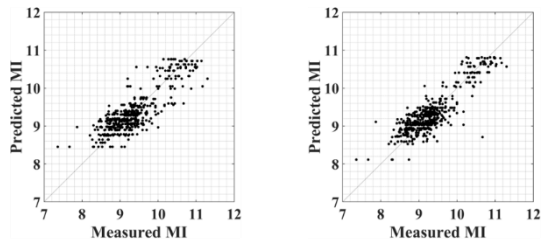
Figure 3.6 illustrates the prediction errors for the data-only model and hybrid model. The blue solid lines with circles indicate the prediction errors of the data-only models, whereas the red dashed lines with circles represent the prediction errors of the hybrid models. As illustrated in **Figure 3.6**, the prediction errors of the hybrid models are closer to zero than those of the data-only models. In summary, the hybrid models outperform the data-only models in terms of prediction accuracy. Furthermore, the GPR models demonstrated the best prediction performance with the smaller number of hyperparameters compared to other data-driven models.

Table 3.5. MI prediction performances of all trained models on the testing dataset under unsteady state.

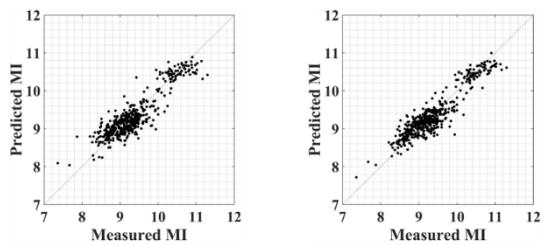
Modeling method		RMSE	MAPE	TIC	STD
PLS	Data-only	0.5981	5.329	0.03343	0.1621
	Hybrid	0.5624	5.269	0.03150	0.1532
	(Difference)	**-5.962%**	**-1.130%**	**-5.788%**	**-5.461%**
Decision tree	Data-only	0.4513	3.955	0.02525	0.1225
	Hybrid	0.4000	3.552	0.02242	0.1091
	(Difference)	**-11.36%**	**-10.20%**	**-11.24%**	**-10.96%**
SVM	Data-only	0.3542	2.964	0.01984	0.0963
	Hybrid	0.2663	2.175	0.01491	0.0721
	(Difference)	**-24.83%**	**-26.63%**	**-24.84%**	**-25.13%**
ANN	Data-only	0.3857	3.451	0.02159	0.1048
	Hybrid	0.2706	2.320	0.01517	0.0739
	(Difference)	**-29.86%**	**-32.76%**	**-29.73%**	**-29.49%**
GPR	Data-only	0.3263	2.644	0.01826	0.0882
	Hybrid	0.2601	2.077	0.01458	0.0709
	(Difference)	**-20.30%**	**-21.45%**	**-20.16%**	**-19.58%**



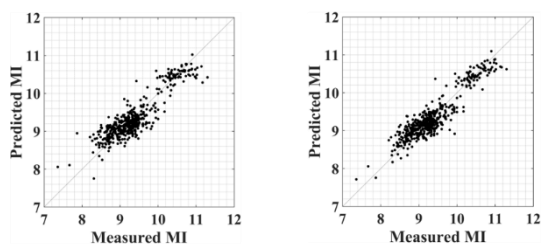
(a) data-only (left) and hybrid (right) PLS models



(b) data-only (left) and hybrid (right) decision tree models



(c) data-only (left) and hybrid (right) SVM models



(d) data-only (left) and hybrid (right) ANN models

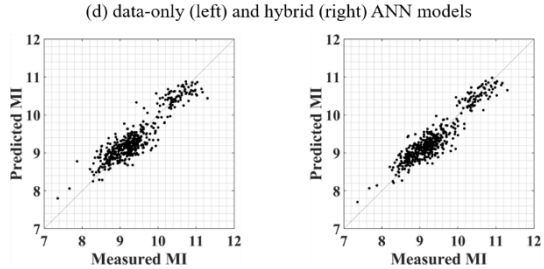


Figure 3.5. Parity plot of measured and predicted MI values of data-only and hybrid models.

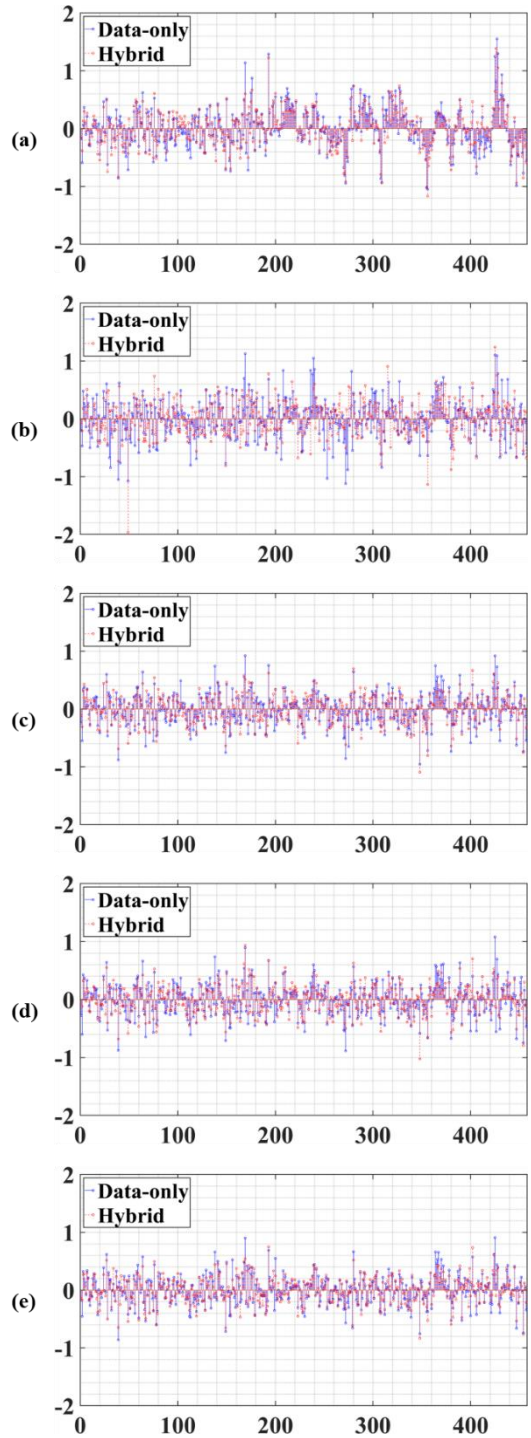


Figure 3.6. Prediction errors of all trained models on the testing dataset (a) PLS (b) decision tree (c) SVM (d) ANN (e) GPR.

Additionally, two mechanistic MI prediction models were developed in order to compare the accuracy of their predictions. First, mechanistic model is developed based on the relationship between MI and weight-average molecular weight as in (3.22). Due to the lack of molecular weight measurement data for the SAN polymerization process, the weight-average molecular weight derived from the white-box submodel was used for modeling. The model parameters were determined using 1828 training data samples. G and x have respective parameter values of 4.0315×10^{10} and -1.9143. Another mechanistic model was developed on the basis of the linear combination of the logarithms of the monomer concentration ratio and the temperature, as in (3.23). Since the concentrations of monomers in the reactors are unavailable, the inputs to the model were adjusted. Instead of monomer concentrations, the ratio of flowrates of monomer feed streams was used as a model input. In addition, temperatures of both the first and second reactors served as input variables. The modified mechanistic model for the SAN polymer is finally defined as follows.

$$\ln(\mathrm{MI}) = \beta + \alpha_1 \, ln\left(\frac{[M_2]}{[M_1]}\right) + \alpha_2 \, ln\left(\frac{[M_3]}{[M_1]}\right) + \alpha_3 \, ln\left(T_{avg}\right) \qquad (3.28)$$

where T_{avg} is the average temperature of the two reactors. The trained parameter values are -23.226, -0.608427, 0.31785, and 5.4202 for β, α_1, α_2, and α_3, respectively.

Table 3.6 provides a summary of the predictions made by the two mechanistic models. In comparison to data-driven models, the predictions of mechanistic models are significantly less reliable.

Table 3.6. MI prediction performances of mechanistic models on the testing dataset.

Input variables	RMSE	MAPE	TIC	STD
MW_w	0.6296	5.097	0.03362	0.6299
Ratio of monomer flowrates and reactor temperature	0.4180	3.570	0.02229	0.4184

In addition, **Figure 3.7** depicts the outcomes of two mechanistic models' predictions on the testing dataset, where there are significant differences between the predictions and measurements. One of the main reasons for the low prediction accuracy of mechanistic models is that they were designed based on assumptions that are unsuitable for actual industrial plants. For instance, polymer samples are assumed to have constant density in the mechanistic model whose input is weight-average molecular weight [67]. The other mechanistic model assumes that the structures of the polymer samples are comparable and determined by the instantaneous operating conditions [88]. However, a variety of polymers with different structures and properties are produced in real polymerization processes. As a result, the predictions of mechanistic models are less accurate than those of the data-driven and hybrid models. A cumulative MI prediction model has been developed from the instantaneous model given in (3.23) for mixed polymer samples with different MI [88]. However, the cumulative model requires residence time of a polymer at each reactor, which is impossible to measure in industrial plants. In summary, the proposed hybrid models outperform mechanistic models and conventional data-only machine learning models in terms of prediction performance and generalizability.

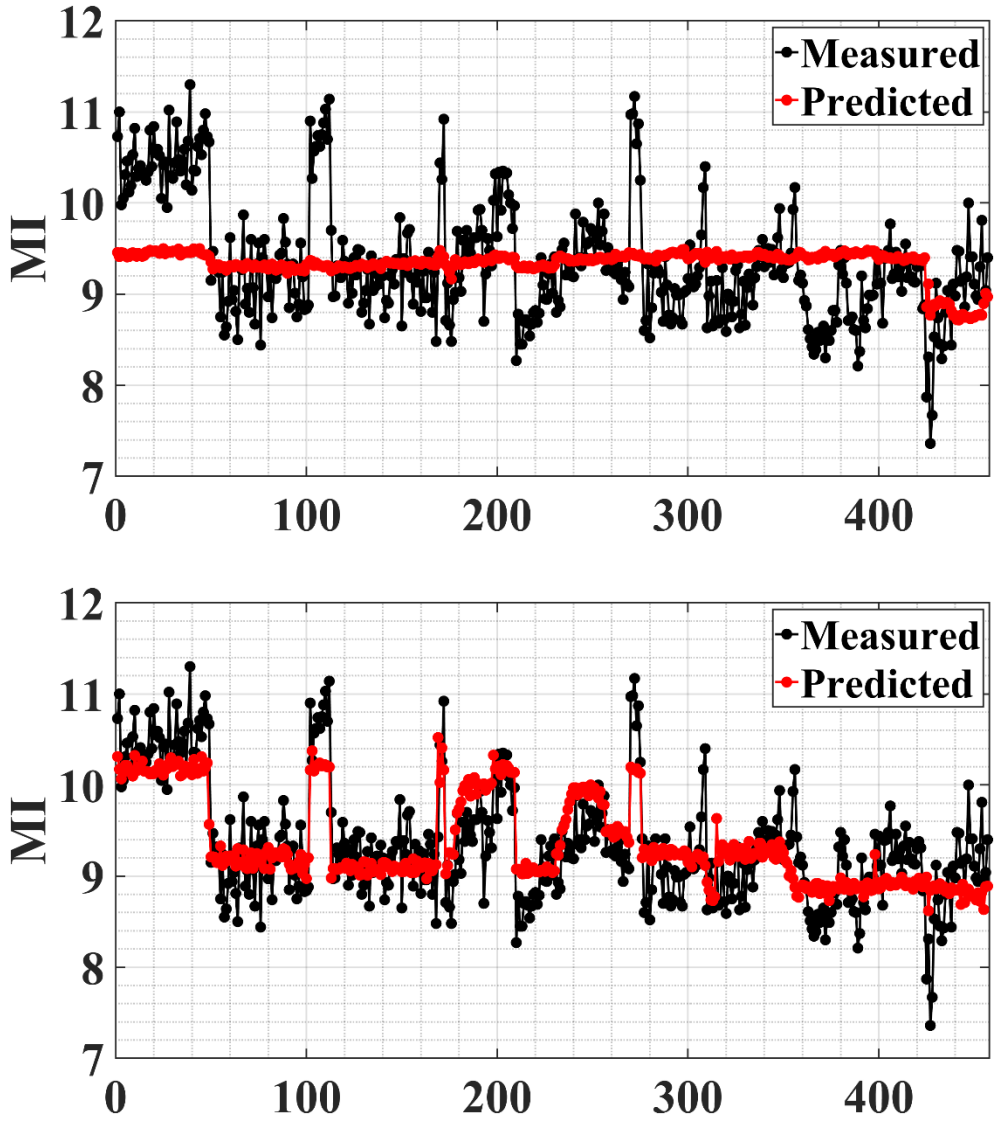


Figure 3.7. Prediction performances of mechanistic models based on molecular weight (above) and monomer flowrate ratios (below).

Chapter 4. Just-in-time Learning Modeling Approach Combined with Dynamic Time Warping[2]

4.1. Introduction

There is an increasing market need in the chemical sector for the manufacturing of a variety of chemical products with specific applications. To satisfy fluctuating demands of the market, many chemical processes produce multiple grades of products with distinct qualities within the same facility. For example, thermoplastic polymer products are manufactured in a variety of grades with varying chemical and mechanical properties, such as melt index, tensile strength, and transparency [1, 89, 90]. Due to the frequent grade transitions in multi-grade operations, fast changes in operating conditions are unavoidable. As a result, large settling times and overshoots in product quality occur during process transitions, resulting in off-specification products. Additionally, online measurements of the quality of chemical products are not available in most industrial processes. Therefore, a soft sensor model with high prediction accuracy is necessary for quality monitoring and control of multi-grade processes.

There are two main types of soft sensor modeling approaches: mechanistic modeling and data-driven modeling. The mechanistic models require process knowledge, including reaction mechanisms and thermodynamic properties. However, development of accurate mechanistic models for multi-grade processes is difficult due to the process's strong nonlinearity and time-varying dynamics.

---

[2] This chapter is an adapted version of M. J. Song, S. H. Ju, and J. M. Lee, "Soft sensor development based on just-in-time learning and dynamic time warping for multi-grade processes", *Korean Journal of Chemical Engineering*, accepted.

Data-driven approaches, as opposed to mechanistic modeling approaches, utilizes simply process data and quality measurements to construct a soft sensor. Due to the rapid growth of data technology in recent decades, it has become possible to collect massive quantities of high-dimensional process data from industrial processes. In order to construct soft sensor models for the quality prediction of chemical processes, multiple data-driven modeling approaches have been implemented. Latent variable methods such as principal component analysis (PCA) [91-93] and partial least squares (PLS) [5, 6, 94-97], which transform process variables into linearly independent latent variables, are among the most common data-driven modeling approaches. Additionally, machine learning modeling approaches based on support vector machines (SVM) [5, 6, 98], Gaussian process regression (GPR) [99-101], artificial neural network (ANN) [5, 102, 103], and long short-term memory (LSTM) network [104, 105] have been used to predict the quality of chemical and biochemical processes.

However, multi-grade process modeling presents various issues that are difficult for a single global model to address. First, one grade's operating conditions are distinct from those of other grades. As the number of product grades increases, it becomes more challenging for a single soft sensor to precisely mimic all process variables for each grade. When estimating the quality of a new product grade that was not included in the training dataset, for instance, a model's prediction performance could be drastically reduced. Furthermore, industrial processes exhibit time-varying dynamics due to the drift of process characteristics such as catalyst deactivation, instrument degradation, and fouling. These process drifts result in the gradual degradation of the prediction performance of a soft sensor model. The wide disparity between the amount of samples of each grade poses an additional difficulty for training a global soft sensor model. To build an accurate model of a soft sensor, sufficient samples of each grade are necessary. However, the measurement data from a specific grade may be very limited because the operating mode changes based on

market demands. Due to the short duration of a grade changeover, only a small number of measurement samples are available for simulating the transient dynamics.

Therefore, numerous investigations have been conducted to build soft sensors that can account for multi-grade and time-varying characteristics of chemical processes. One approach is to iteratively update a soft sensor model with new measurement data. Examples include recursive PCA [28-30] and recursive PLS [31-33], which are based on latent variables method. While recursive models have demonstrated enhanced prediction performance for systems with slow time-varying dynamics, they are not suitable for multi-grade chemical processes where abrupt and rapid changes in operating conditions occur during grade changeovers.

In recent years, just-in-time learning (JITL) soft sensors have been used to a variety of chemical processes and have exhibited good prediction performance with various data-driven modeling methods [12, 13, 34-47, 89, 90]. In the JITL framework, a local model is trained online only using the most similar historical data samples to the query sample. Thus, the similarity measurement employed in JITL modeling has a substantial effect on the performance of a model. One of the most widely used metrics for calculating the similarity between two samples is the Euclidean distance. Additionally, the information regarding the angle between two samples was combined with the Euclidean distance to determine similarity [40]. JITL soft sensors have embraced additional similarity metrics, including the Mahalanobis distance [43, 44] and the Kullback-Leibler divergence [45-47]. However, temporal correlations inherent in process data are not considered in the computations of similarity described above. The process data from industrial chemical processes exhibit highly nonlinear, complicated, and temporally correlated dynamics. In order to build an accurate JITL soft sensor for multi-grade chemical processes, a similarity measurement that takes temporal correlations of process variables into account is required.

In this chapter, a JITL soft sensor modeling framework where the similarity

between two multivariate time series data is calculated using dynamic time warping (DTW) is proposed. DTW is a similarity measurement method for two different data sequences [106, 107]. The similarity between two time series of different lengths can be calculated using DTW by stretching or compressing a times series to match another. In recent years, DTW has been utilized for a variety of dynamic time series problems, including handwriting recognition [48], gesture recognition [49], traffic speed prediction [50], state of health estimation [51], fault detection [52], and batch trajectory synchronization [53].

Three main contributions are accomplished in the proposed DTW-based JITL modeling approach. First, the process dynamics and temporal correlations in the process data are considered by training a machine learning model with time series data as opposed to data points. Previous studies have shown that the prediction performance of latent variable models and machine learning models can be improved by augmenting input data with time-lagged data [108, 109]. Second, the ability of selecting the most relevant time series from the historical database is improved by utilizing DTW as a similarity measurement. Third, the implications of the DTW path constraint and input-output cross-correlation on the modeling performance are discussed and a DTW-based JITL soft sensor modeling approach with the hyperparameter optimization is proposed. The prediction performance of the proposed DTW-based JITL soft sensor is evaluated with three multi-grade dynamic simulation studies and compared to that of the conventional JITL method based on the Euclidean distance.

The remainder of this chapter is structured as follows. In Subsection 4.2, a DTW-based JITL soft sensor modeling method for multi-grade processes is proposed. Subsection 4.3 provides the results of sensitivity analysis of the hyperparameters of the proposed modeling approach. In Subsection 4.4, a DTW-based JITL modeling approach with the optimized model hyperparameters is proposed and the modeling results is discussed.

4.2. DTW-based JITL modeling approach

In this subsection, a JITL soft sensor modeling method for multi-grade processes is developed by applying DTW to calculating similarities between data sequences. There are two main advantages of the proposed modeling approach. First, the data sequence is used as a model input in order to consider process dynamics and temporal correlations in the process data. By augmenting the input variables with time-lagged data, the modeling performance of a data-driven model for a dynamic system may be increased. Second, the similarity between two data sequences is measured using DTW. The similarity measurements introduced in Section 2, including DTW, are summarized in **Table 4.1**. The Euclidean distance and angle between two data samples are easier to calculate, but temporal correlations in the data are ignored. The statistical methods, the Mahalanobis distance and Kullback-Leibler divergence, require the probability information of the variables. DTW measures distance between two data sequences without any probabilistic information or transformation of variables. Additionally, DTW can calculate similarity even when a data sequence is distorted by frequency change or process drift.

The core algorithm of the proposed modeling approach is illustrated in **Figure 4.1**. First, the query sample, $\boldsymbol{x}_q$, is collected from online sensors. Then the historical and query data are normalized so that the variables have zero means and unit variances, as described below.

$$\tilde{x}_{q,i} = \frac{x_{q,i} - \mu_i}{\sigma_i} \tag{4.1}$$

where $x_{q,i}$ is the i-th variable of x_q. μ_i and σ_i are the mean and standard deviation of the i-th variable, respectively. The next step is to augment a single data sample $\tilde{\boldsymbol{x}}_q$ into a data sequence $\tilde{\boldsymbol{X}}_q$ with time-lagged data as in (4.2).

$$\widetilde{\boldsymbol{X}}_q = \left[\widetilde{\boldsymbol{x}}_{q-n+1}, \widetilde{\boldsymbol{x}}_{q-n+2}, \ldots, \widetilde{\boldsymbol{x}}_q\right] = \begin{bmatrix} \widetilde{\boldsymbol{x}}_{q-n+1,1} & \widetilde{\boldsymbol{x}}_{q-n+2,1} & \cdots & \widetilde{\boldsymbol{x}}_{q,1} \\ \widetilde{\boldsymbol{x}}_{q-n+1,2} & \widetilde{\boldsymbol{x}}_{q-n+2,2} & \cdots & \widetilde{\boldsymbol{x}}_{q,2} \\ \vdots & \vdots & \ddots & \vdots \\ \widetilde{\boldsymbol{x}}_{q-n+1,n_{var}} & \widetilde{\boldsymbol{x}}_{q-n+2,n_{var}} & \cdots & \widetilde{\boldsymbol{x}}_{q,n_{var}} \end{bmatrix} \tag{4.2}$$

where n is the fixed window length for data augmentation. As a result of data preprocessing, data samples are transformed into data sequences of length n. In the second step of JITL modeling, the similarities between the normalized data sequence $\widetilde{\mathbf{X}}_q$ and the sequences of the database are calculated using DTW. Then a local machine learning soft sensor model is trained with the historical data sequences which are most similar to $\widetilde{\mathbf{X}}_q$. Two data-driven modeling methods, SVM and GPR, are used to construct a local model in this study. Finally, the prediction $\hat{y}_q$ for the quality of the query sample x_q is made with the trained local model. The local model is discarded and the above modeling steps are repeated when a new query sample is acquired.

Table 4.1. Comparison of distance measures used in JITL modeling.

Method	**Data type**	**Definition**
Euclidean distance	Data point	$ED(\boldsymbol{x}_q, \boldsymbol{x}_n) = \sqrt{\sum_{i=1}^{n_{var}} (x_{q,i} - x_{n,i})^2}$
Euclidean distance with angle	Data point	$\cos(\theta_{qn}) = \frac{\langle \boldsymbol{x}_q, \boldsymbol{x}_n \rangle}{\| \boldsymbol{x}_q \|_2 \ \| \boldsymbol{x}_n \|_2}$
Mahalanobis distance	Data point & distribution	$MD(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{P}) = \sqrt{(\boldsymbol{x} - \boldsymbol{y})^T S^{-1} (\boldsymbol{x} - \boldsymbol{y})}$
Kullback-Leibler divergence	Distribution	$\mathrm{KLD}(P \| \| Q) = \int_{-\infty}^{\infty} p(x) \frac{\log(p(x))}{\log(q(x))}$
Dynamic time warping	Data sequence	$DTW(X, Y) = \min_{W} \sum_{i=1}^{k} d_i$

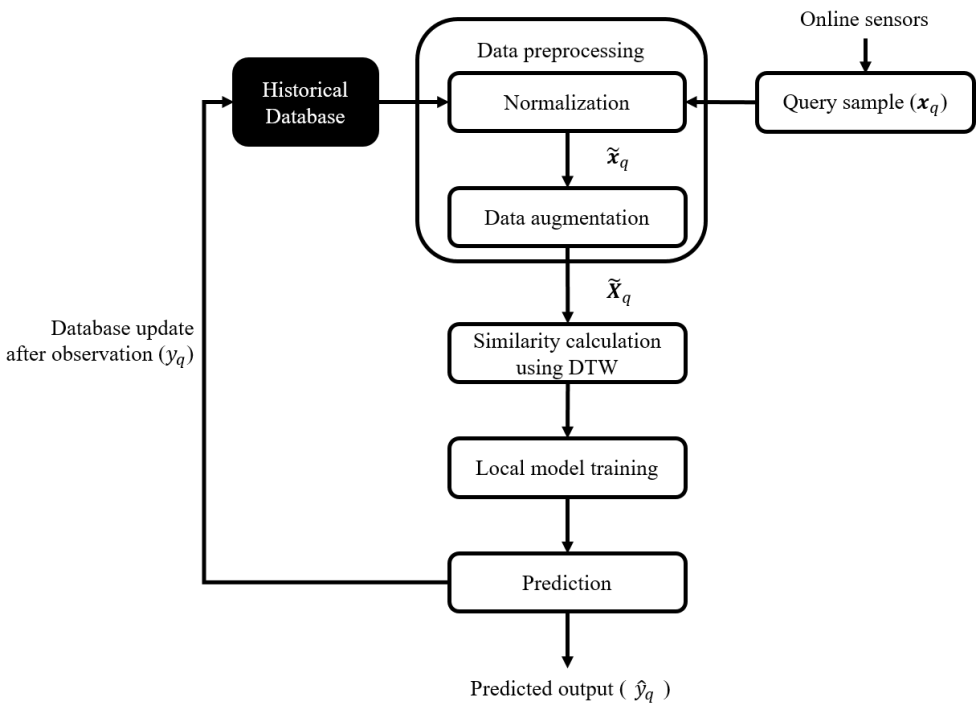


Figure 4.1. Algorithm of the proposed DTW-based JITL model.

4.3. Sensitivity analysis of model hyperparameters

This section evaluates the effects of model hyperparameters on the performance of the proposed DTW-based JITL model by simulating two multi-grade systems. The prediction accuracy and computational requirement for the proposed DTW-based JITL modeling approach are highly dependent on three hyperparameters: the input length, the width of the warping window, and the number of similar samples chosen from the database. Two multi-grade simulation case studies are considered in order to undertake sensitivity analysis on these hyperparameters.

The first simulation case study is a simple nonlinear dynamic system where the output is determined by the past 10 samples of the inputs. Additionally, the inputs change periodically depending on the grade to simulate multi-grade characteristics. The system is defined as follows.

$$\mathrm{y}_1 = A_1 \cdot X_1 + B_1 \cdot X_2 + C_1 \cdot X_1 \cdot X_2 \tag{4.3}$$

$$\mathrm{y}_2 = A_2 \cdot X_1 + B_2 \cdot X_2 + C_2 \cdot X_1 \cdot X_2 \tag{4.4}$$

$$y = \log(y_1 \cdot y_2) \tag{4.5}$$

where $\mathrm{X}_1 = [x_{1,t-9}, x_{1,t-8}, \dots, x_{1,t}]$ and $x_{i,k}$ is the i-th input variable at time k. A, B, and C are the coefficient vectors of length 10. Gaussian random noise is added to the inputs and output. The values of coefficients, including A, B, and C, are present in **Table 4.2**. A total of 520 data samples were obtained as a simulation result, which is shown in **Figure 4.2**. The first 50 percent of the samples were used as the original training dataset, and the remaining 50 percent of the samples were used as the testing dataset.

Table 4.2. Values of coefficient vectors of numerical example.

Symbol	Value
A_1	$[0.0403, 0.0543, 0.1291, 0.1995, 0.3456, 0.4454, 0.6008, 0.8935, 0.9363, 0.9462]$
A_2	$[0.1578, 0.1715, 0.3566, 0.4147, 0.4849, 0.5166, 0.6569, 0.6865, 0.9870, 0.9916]$
B_1	$[0.0579, 0.1318, 0.1556, 0.2551, 0.6964, 0.7609, 0.8176, 0.8575, 0.8957, 0.9820]$
B_2	$[0.0184, 0.1433, 0.2459, 0.3627, 0.3914, 0.4347, 0.7435, 0.7720, 0.8422, 0.8483]$
C_1	$[0.0112, 0.0244, 0.0272, 0.0450, 0.0581, 0.0684, 0.0711, 0.0718, 0.0736, 0.0933]$

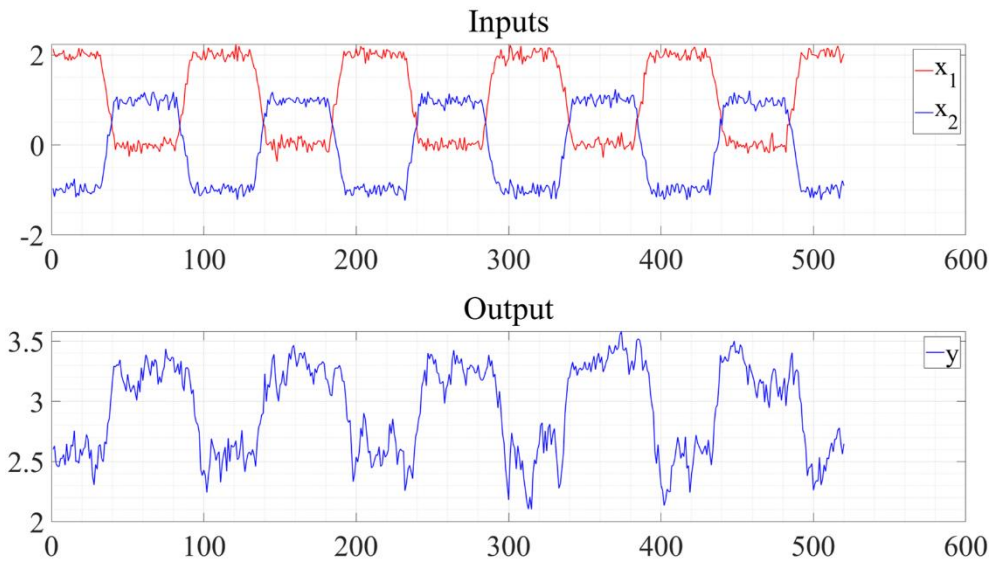


Figure 4.2. Simulation data obtained from the nonlinear numerical example.

In the second case study, a sequence of three continuous stirred tank reactors (CSTRs) was used to simulate a multi-grade chemical system with time delay and multi-grade products. The simple schematic of the CSTR system is illustrated in **Figure 4.3** and the nomenclature and nominal operating conditions are summarized in **Table 4.3**.

The process model of each reactor consists of two nonlinear ordinary differential equations [110]. Two chemical species, A and B, exist in the reactors and undergo irreversible and exothermic chemical reaction, A $\rightarrow$ B.

$$\dot{C}_{Ai} = \frac{q_i}{V_i}\left(C_{A(i-1)} - C_{Ai}\right) - k_0 C_{Ai}\, exp\left(-\frac{E}{RT}\right) \tag{4.6}$$

$$\begin{aligned}\dot{T}_i = {} & \frac{q_i}{V_i}(T_{i-1} - T_i) + \frac{(-\Delta H)k_0 C_{Ai}}{\rho C_p}\exp\left(-\frac{E}{RT}\right) \\ & + \frac{\rho_c C_{pc}}{\rho C_p V} q_c[1 - \exp(-\frac{hA}{q_c \rho_c C_{pc}})](T_{cf} - T_i)\end{aligned} \tag{4.7}$$

Since the flowrates of all streams are set equal, the reactor volumes are equal and remain unchanged during simulation. Additionally, transport delays of 3 minutes are added between the reactors. The manipulated variable of the system is the flowrate of the coolant for the first reactor, q_{in}. It is assumed that only the manipulated variable, q_{in}, and the temperature of the first reactor, T_1, are measured. The output of the process is the effluent concentration of A from the third reactor, C_{A3}. **Figure 4.4** illustrates the simulation result of the CSTR system. A total of four grades and their grade changeovers were simulated by manipulating the coolant flowrate or the first reactor. Additionally, the Gaussian random noise was added to the input and output variables. The simulation time was 500 minutes and a total of 1001 data samples were obtained as a result. The first 501 samples were used as the original training dataset and the remaining 500 samples were used as the testing dataset.

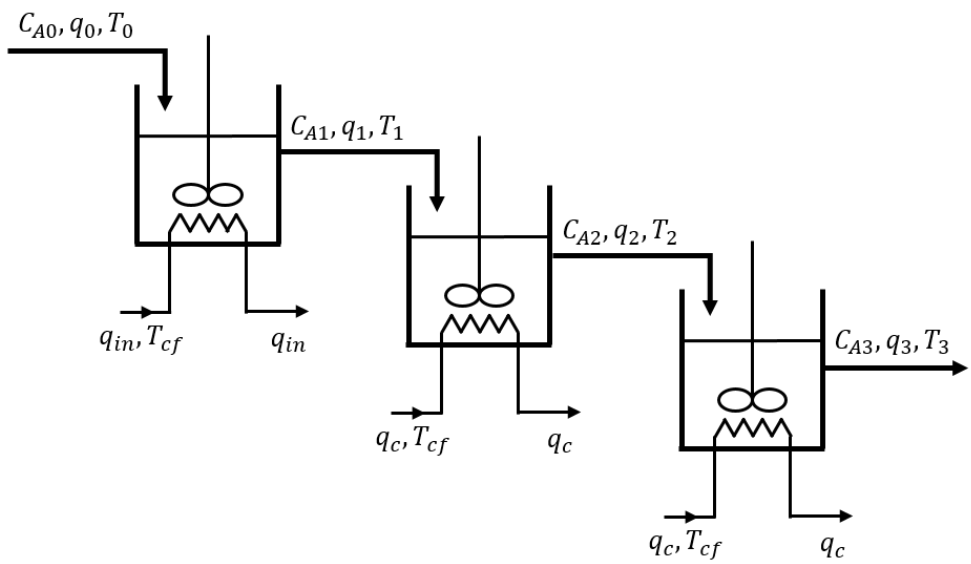


Figure 4.3. Schematic of the chemical system consisting of three CSTRs in series.

Table 4.3. Nomenclature and nominal operating conditions of the CSTR system.

Symbol	Description	Operating condition	Unit
C_{A0}	Concentration of A in the feed stream into the first reactor	1.00	mol l^{-1}
C_{Ai}	Concentration of A in the i-th reactor	0.0836	mol l^{-1}
C_p	Heat capacity of the reaction mixture	1.00	cal g^{-1} K^{-1}
C_{pc}	Heat capacity of the coolant	1.00	cal g^{-1} K^{-1}
E/R	Fraction of the activation energy divided by the gas constant	9.95×10^3	K
hA	Product of the heat transfer coefficient and heat transfer area	7.00×10^5	cal min^{-1} K^{-1}
k_0	Pre-exponential factor	7.20×10^5	min^{-1}
q_c	Flowrate of coolant for the second and third reactor	103.41	l min^{-1}
q_i	Flowrate of feed stream into the i-th reactor	100	l min^{-1}
q_{in}	Flowrate of coolant for the first reactor	103.41	l min^{-1}
T_0	Temperature of the feed stream into the first reactor	350	K
T_c	Temperature of coolant	350	K
T_i	Temperature of the i-th reactor	440.2	K
V_i	Volume of the i-th reactor	100	l
ΔH	Heat of reaction	-2.00×10^5	cal mol^{-1}
ρ	Density of the reaction mixture	1000	g l^{-1}
ρ_c	Density of coolant	1000	g l^{-1}

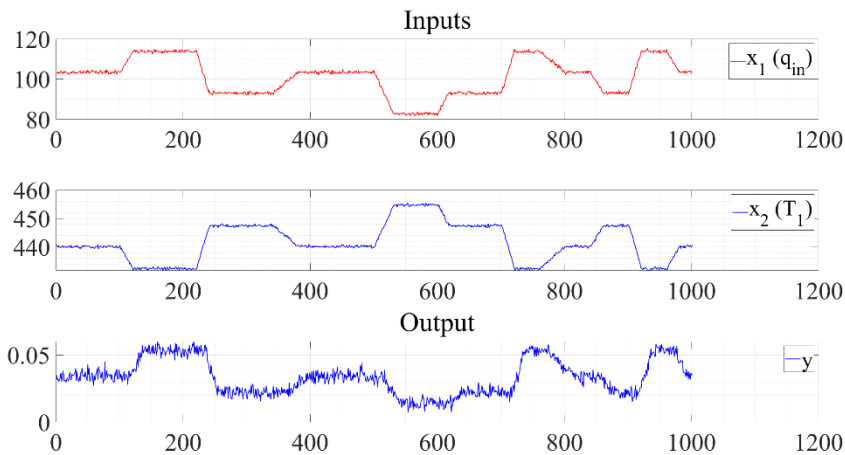


Figure 4.4. Simulation data obtained from the CSTR system.

SVM and GPR models were trained for the sensitivity analysis of the hyperparameters for both case studies. The prediction accuracy of the soft sensors was evaluated using the root mean squared error (RMSE) defined as in (18).

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{4.8}$$

where N is the number of samples, y_i and $\hat{y}_i$ are the measured and predicted outputs, respectively. Additionally, the CPU time was measured on an Intel Core i7-8700 CPU @3.20 GHz in order to evaluate the computational requirement.

4.3.1. Input length and number of similar samples

The length of the input sequence and the number of similar samples used in local modeling affect not only the required calculation time but also the prediction accuracy of the proposed model. Therefore, the effects of the hyperparameters on the prediction performance of the DTW-based JITL models is analyzed for both case studies in this subsection. **Figure 4.5** and **4.6** depict the prediction accuracy of the DTW-based JITL models with varying input lengths and the number of similar samples on the testing datasets of the numerical and CSTR examples, respectively. Both the SVM and GPR models for the numerical example reach the lowest RMSE values at the input length of 6, regardless of the number of samples used for local modeling. On the other hand, the DTW-based JITL models for the CSTR system achieve the best prediction accuracy at the input length close to 18.

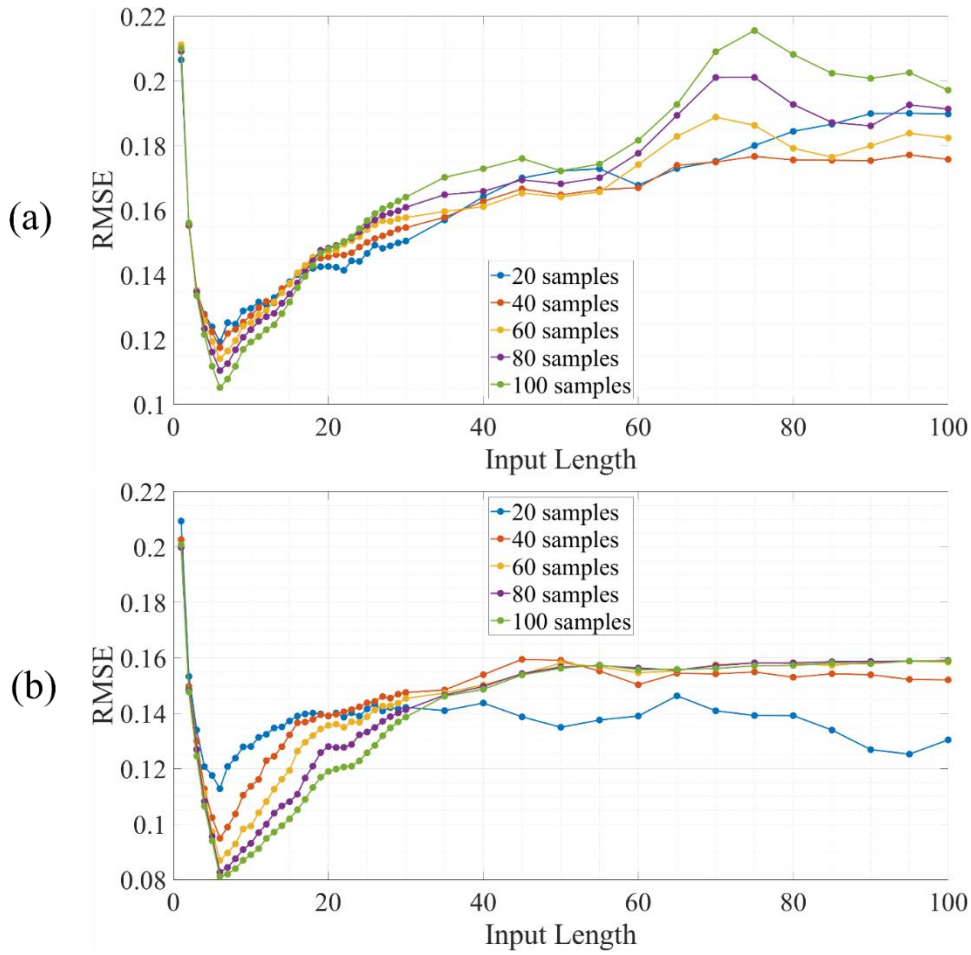


Figure 4.5. Prediction accuracy of models with different input lengths and number of similar samples for the numerical example (a) SVM models (b) GPR models.

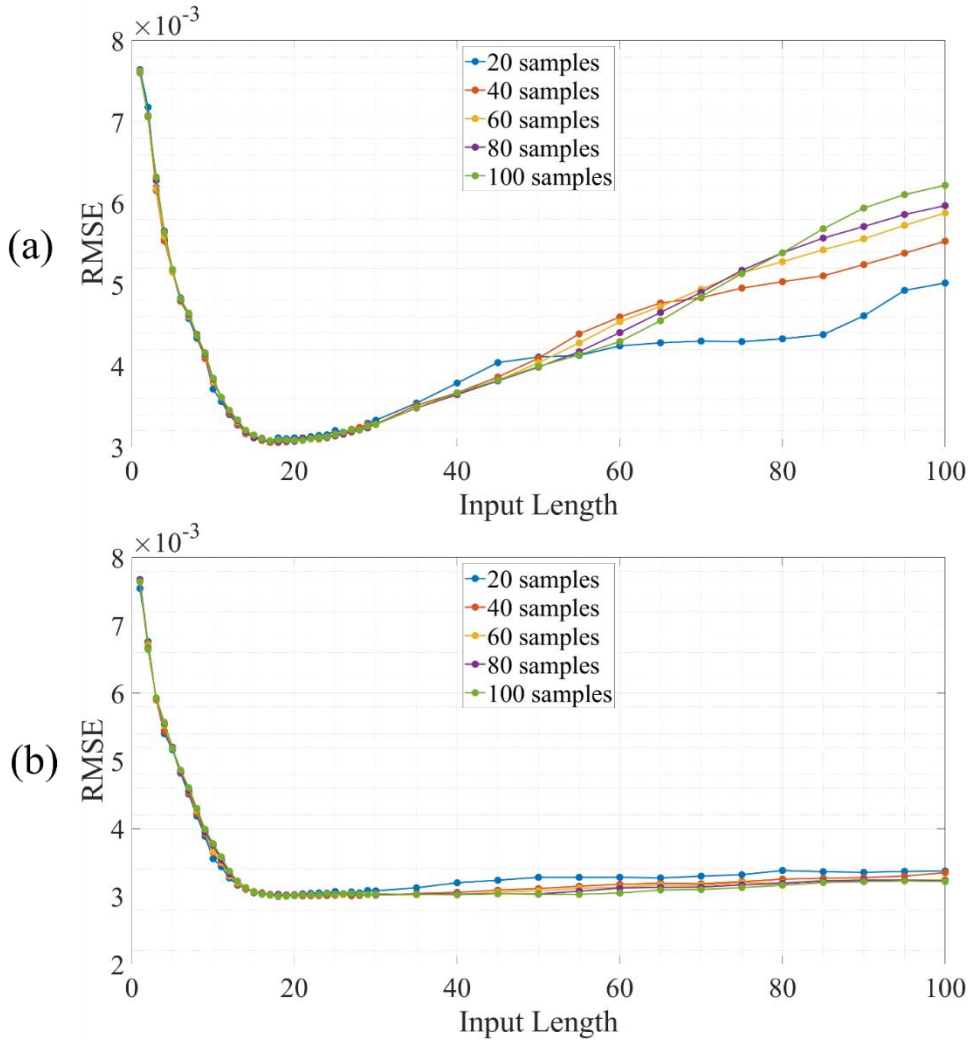


Figure 4.6. Prediction accuracy of models with different input lengths and number of similar samples for the CSTR example (a) SVM models (b) GPR models.

The cross-correlations between the input and output variables of the numerical example and CSTR system are shown in **Figure 4.7** and **4.8**, respectively. The absolute values of cross-correlations reach their maxima at time lags of 4 and 54 for the numerical example. Furthermore, the cross-correlations exhibit periodical changes because the grade and input variables change periodically with a period of 50 time steps. On the other hand, the output variable of the CSTR system is most correlated with the first and second input variables at time lags of 16 and 15, respectively.

The results from varying input lengths and cross-correlations suggest that the best modeling performance is achieved when a model takes time series of length slightly longer than the time lags of maximum cross-correlation between input and output. Additionally, the modeling performance of the JITL models decreases as the input length become longer or shorter than the optimal time lags as shown in Figure 7 and Figure 8, which represents the importance of finding the optimal length for input data sequences. If the input data sequence is longer than the cross-correlation between the input and output, the sequences are more likely to lose their original characteristics and dynamics through warping. On the other hand, if the input time series is too short, the temporal correlations of the variables are not sufficiently considered in modeling, resulting in poor prediction performance. Therefore, it is necessary to find the optimal length for input augmentation by calculating the cross-correlation between the input and output variables.

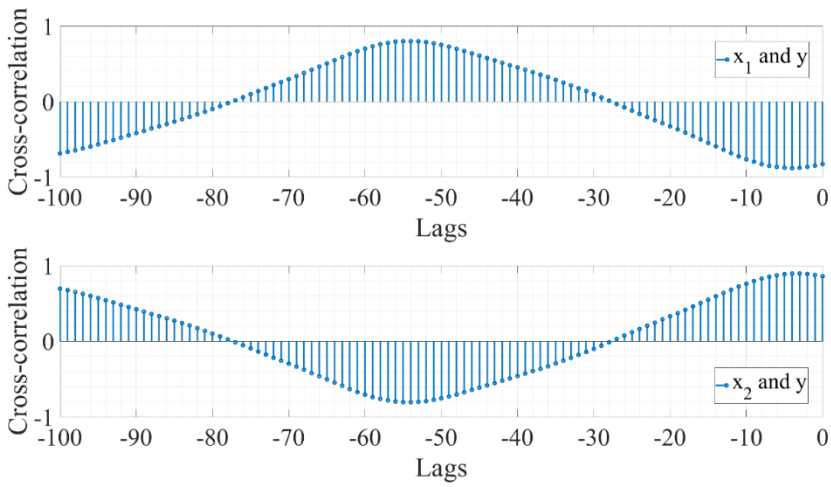


Figure 4.7. Cross-correlation between input and output variables of the numerical example.

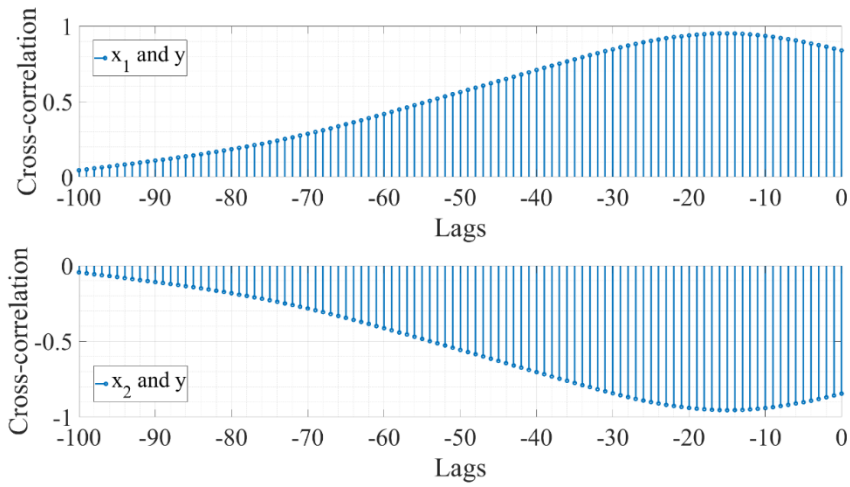


Figure 4.8. Cross-correlation between input and output variables of the CSTR example.

In addition, Figure 4.5 and 4.6 illustrate the effect of the number of similar samples selected from the database on the prediction accuracy of the proposed model. The prediction accuracy of the proposed model increases as the more samples are used in the numerical example for the optimal input length, 6. Since the nonlinear system changes periodically between two distinct operation mode, there exist many samples for each mode. Therefore, the local models trained with more samples predict the output more accurately. Additionally, the numerical example's system is much simpler than the reactor systems, making it easier for regression models to learn when provided more data samples.

On the other hand, the DTW-based JITL models for the CSTR example achieved the best prediction performance when 40 similar samples with an input length of 18 were utilized. When too many samples are used in model training, it is more likely that the training dataset will contain samples that are not similar or relevant to query data, particularly for grade changeovers whose duration is short compared to steady-state operations, thereby decreasing the performance of the model. In addition, **Figure 4.9** illustrates the CPU required to compute a whole testing procedure for the proposed GPR-based model. As the number of training samples increases, the CPU time required by the proposed model increases. In order to achieve the optimal prediction accuracy in an acceptable amount of time, it is crucial to examine the tradeoff between prediction accuracy and computation time by selecting an adequate number of similar samples for each target system.

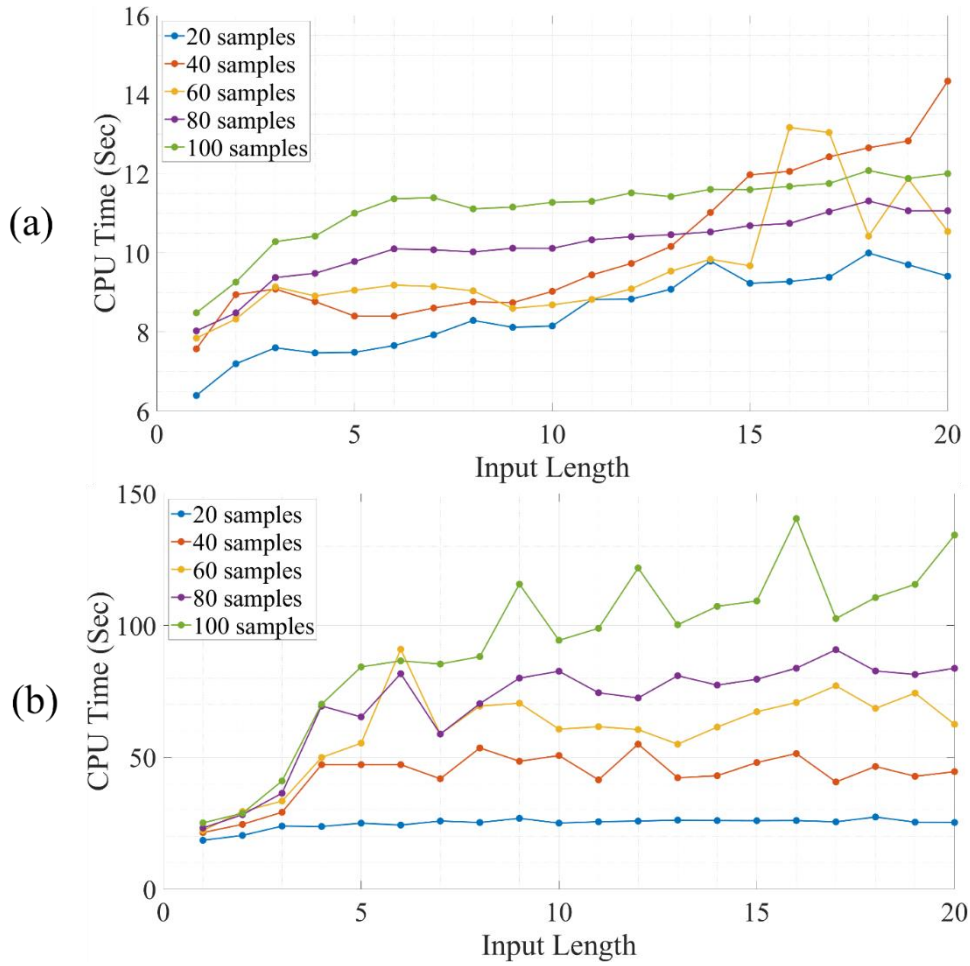


Figure 4.9. CPU time spent calculating the proposed modeling approach based on GPR (a) numerical example (b) CSTR example.

4.3.2. Warping path constraint

In this subsection, the effect of the temporal constraint on the warping window size of DTW on the modeling performance in the DTW-based JITL modeling approach is analyzed. Without the warping path constraint, DTW stretches and compresses data sequences as much as possible to make the warped sequences as similar as possible. However, the unconstrained warping is susceptible to outliers in the time series. Additionally, the unconstrained warping may distort data sequences excessively that the temporal correlations and dynamics of the raw sequences are rarely remained in the warped sequences. Therefore, the accuracy of distance measure and computational efficiency can be improved by constraining the warping window width, δ, in a fixed range [107].

Figure 4.10 and **4.11** show the average CPU time required to calculate the similarities between the query and stored data sequences using the Euclidean distance, unconstrained DTW, and constrained DTW for the numerical and CSTR examples, respectively. For both case studies, the computation times for the Euclidean distance were less than 0.001 seconds, which is significantly less than those of DTW. On the other hand, unconstrained DTW calculations for data sequences of length 10 required 0.011571 and 0.022339 seconds in the numerical and CSTR examples, respectively. As illustrated in Figure 4.10 and 4.11, the constrained DTW with a narrower warping window width takes less computation time than the unconstrained DTW because only a small number of warping pathways must be determined. Moreover, longer data sequences necessitate additional calculation time, which is especially evident for the unconstrained DTW. Nevertheless, unconstrained DTW computations with an input length of 100 required 0.022965 and 0.047334 seconds for the numerical and CSTR examples, respectively, demonstrating that the proposed modeling approach can be utilized for online applications with short sampling periods.

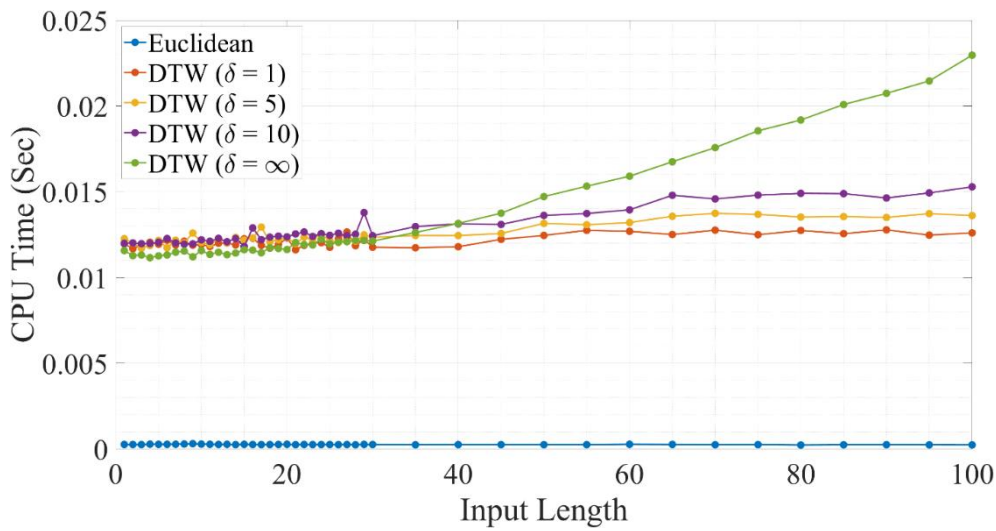


Figure 4.10. CPU time spent calculating similarities between samples of the numerical example.

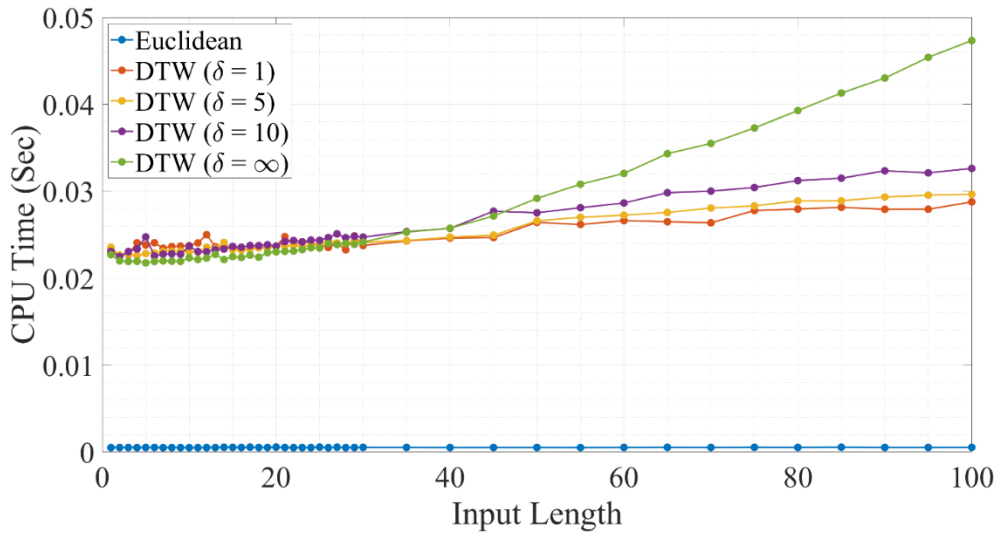


Figure 4.11. CPU time spent calculating similarities between samples of the CSTR example.

Figure 4.12 and **4.13** illustrate, for numerical and CSTR examples, respectively, a comparison of the prediction accuracy of the proposed models with different similarity measures and constraints as measured by root mean square error (RMSE). The models were trained using 40 similar samples and the input data sequences with a length of 6 and 18 for the numerical and CSTR examples, respectively. For both case studies, the prediction accuracy of the models improves as the width of the warping window increases. On the other hand, a constraint on the warping window width that is overly stringent may reduce prediction performance because only warping paths close to the raw time series are evaluated. As shown in Figure 4.12 and 4.13, RMSE decreases and reaches plateaus when the warping window constraint is approximately half the input length. Therefore, by limiting the warping window width to half the length of the input sequence, it is possible to attain optimal prediction accuracy with a reduced computing demand.

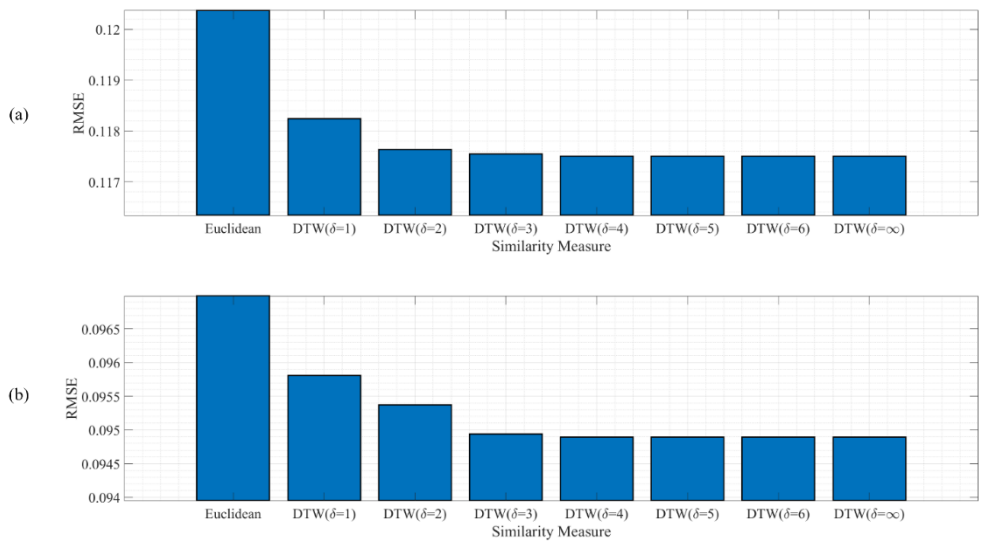


Figure 4.12. Prediction accuracy of models using the Euclidean distance and DTW with different warping window width constraint for the numerical example (a) SVM models (b) GPR models.

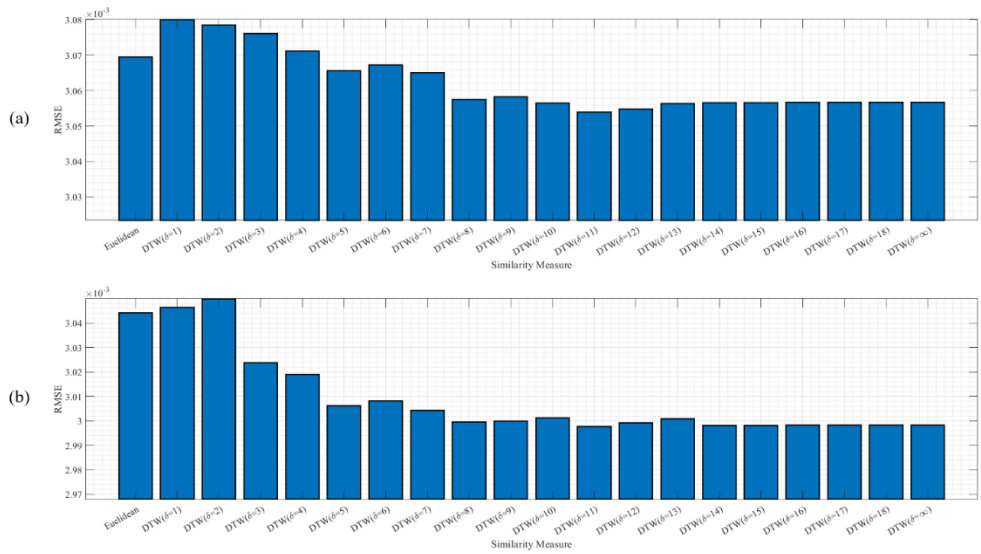


Figure 4.13. Prediction accuracy of models using the Euclidean distance and DTW with different warping window width constraint for the CSTR example (a) SVM models (b) GPR models.

4.4. Optimized DTW-based JITL modeling approach

4.4.1. Proposed modeling approach

In this section, we propose a DTW-based JITL soft sensor modeling approach with the optimized model hyperparameters. The results from Section 4 reveal that the prediction accuracy and computational requirement of the DTW-based JITL model depend on three hyperparameters: input length, number of similar samples, and width constraint of the warping window. **Figure 4.14** illustrates the algorithm of the proposed modeling approach. The proposed modeling algorithm differs from the previous DTW-based JITL modeling algorithm introduced in Subsection 4.3 in that the input length and DTW constraint are optimized by analyzing cross-correlation between the input and output variables. First, the cross-correlation coefficients between the input and output variables are calculated using the stored data samples in the historical database. Then, the data samples are augmented to data sequences of length slightly longer than the time lags of maximum cross-correlation. In this study, the length of the augmented input is determined by adding two to the time lag of maximum cross-correlation, which is the optimal value from the results of Subsection 4.3.1. Then, the similarities between the augmented data sequences are calculated using DTW under the constraint of the warping window width. The maximum warping width, δ, is determined to be half the length of the input sequence. The subsequent steps are identical to the original DTW-based JITL modeling procedure.

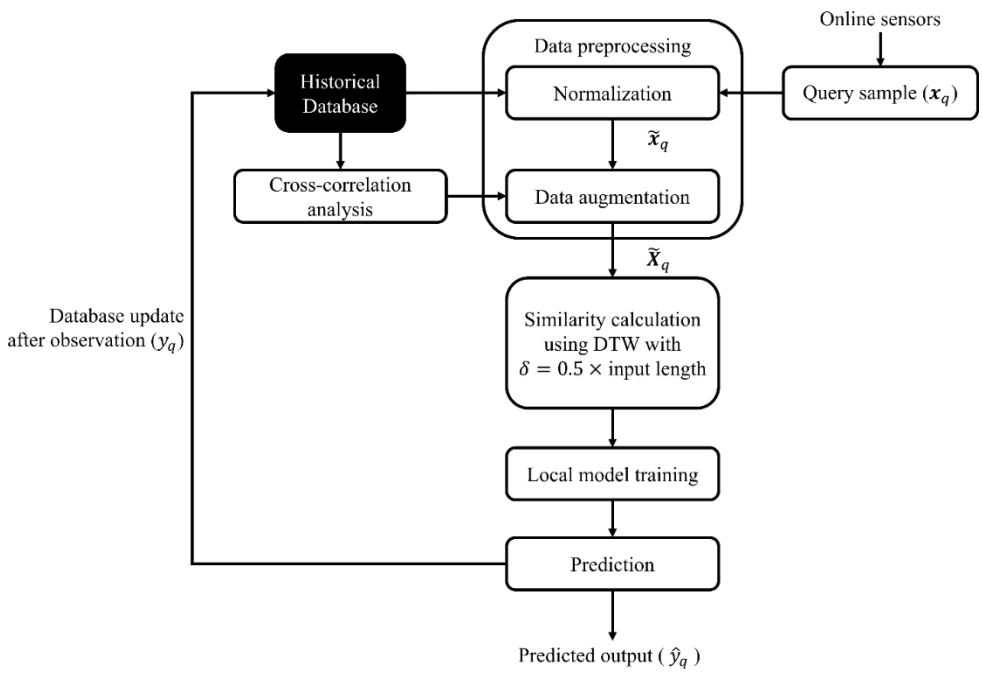


Figure 4.14. Algorithm of the proposed DTW-based JITL model with hyperparameter optimization.

4.4.2. Results and discussions

In this section, multi-grade case studies were simulated and used to evaluate the prediction performance of the proposed DTW-based JITL model to verify the effectiveness of the proposed modeling approach. Additionally, conventional JITL soft sensors based on the Euclidean distance were developed for performance comparison. The prediction accuracy of the soft sensors was evaluated using four statistical indices: root mean squared error (RMSE), mean absolute percentage error (MAPE), Theil's inequality coefficient (TIC), and coefficient of determination (R^2).

$$\text{MAPE} = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{4.9}$$

$$\text{TIC} = \frac{\sqrt{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}}{\sqrt{\sum_{i=1}^{N} y_i^2}\sqrt{\sum_{i=1}^{N} \hat{y}_i^2}} \tag{4.10}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \tag{4.11}$$

where N is the number of samples, $\bar{y}$ is the mean value of y, y_i and $\hat{y}_i$ are the measured and predicted output variable, respectively. For both case studies, two machine learning regression models, SVM and GPR, are utilized as local predictive models and the number of similar samples drawn from the database for local model training is 40. The augmented input lengths are 6 and 18 for the numerical and CSTR examples, respectively. Thus, the respective maximum DTW window width are 3 and 9 for the numerical and CSTR examples.

Table 4.4 and **4.5** summarize modeling results of the proposed optimized DTW-

based JITL models and Euclidean distance-based JITL models for the numerical and CSTR examples, respectively. Since the simulated output was determined by both current and past inputs, the models taking time series as input outperform the models taking data points as input. Additionally, the DTW-based JITL models demonstrate better prediction accuracy than the Euclidean distance-based JITL models. For instance, the RMSE, MAPE, TIC, and R^2 of the constrained DTW-based GPR model for the CSTR system are 3.000, 10.41, 0.04424, and 0.9527, respectively, which represents the better prediction accuracy than the Euclidean distance-based GPR model whose RMSE, MAPE, TIC, and R^2 are 3.044, 10.65, 0.04490, and 0.9513, respectively. Additionally, the prediction performance of a constrained DTW-based JITL model is comparable to that of an unconstrained DTW-based JITL model.

In order to further demonstrate the efficiency of the proposed model, a distillation process where methylcyclohexane (MCH) is recovered from a mixture was simulated. **Figure 4.15** illustrates the schematic of the process in which the purity of MCH, which is the output variable, is controlled by manipulating the flowrate of the phenol inlet stream, which is the input variable. The sampling time and total simulation time were 0.01 h and 30 h, respectively. Thus, a total of 3000 samples were obtained, of which 50% were used as the testing dataset. **Table 4.6** summarizes prediction performance of the proposed models and conventional Euclidean distance-based JITL models on the testing dataset of the distillation example. The window length calculated from the cross-correlation analysis and warping window width for the proposed model were 30 and 15, respectively. For both data-driven regression models, the proposed constrained DTW-based JITL models demonstrate better prediction accuracy compared to the Euclidean distance-based models. In summary, the modeling results indicate that improved prediction accuracy with reduced computation can be achieved by the proposed DTW-based JITL modeling approach with hyperparameter optimization.

Table 4.4. Prediction results of all trained models on the testing dataset of the numerical example.

Data-driven model	**Input length**	**Similarity measure**	**RMSE**	**MAPE (%)**	**TIC**	R^2
SVM	1	Euclidean distance	0.2067	5.425	0.03497	0.7312
SVM	6	Euclidean distance	0.1204	3.421	0.02024	0.9089
SVM	6	DTW ($\delta = 3$)	0.1175	3.320	0.01977	0.9131
SVM	6	DTW ($\delta = \infty$)	0.1175	3.318	0.01976	0.9131
GPR	1	Euclidean distance	0.2056	5.617	0.03473	0.7341
GPR	6	Euclidean distance	0.09699	2.737	0.01631	0.9408
GPR	6	DTW ($\delta = 3$)	0.09494	2.676	0.01598	0.9433
GPR	6	DTW ($\delta = \infty$)	0.09489	2.675	0.01597	0.9433

Table 4.5. Prediction results of all trained models on the testing dataset of the CSTR example.

Data-driven model	**Input length**	**Similarity measure**	**RMSE** $(\times \mathbf{10^3})$	**MAPE (%)**	**TIC**	$\boldsymbol{R^2}$
SVM	1	Euclidean distance	7.634	21.93	0.1127	0.6934
SVM	18	Euclidean distance	3.069	10.81	0.04530	0.9504
SVM	18	DTW $(\delta = 9)$	3.058	10.71	0.04512	0.9508
SVM	18	DTW $(\delta = \infty)$	3.057	10.71	0.04510	0.9509
GPR	1	Euclidean distance	7.529	22.13	0.1118	0.7018
GPR	18	Euclidean distance	3.044	10.65	0.04490	0.9513
GPR	18	DTW $(\delta = 9)$	3.000	10.41	0.04424	0.9527
GPR	18	DTW $(\delta = \infty)$	2.998	10.41	0.04421	0.9527

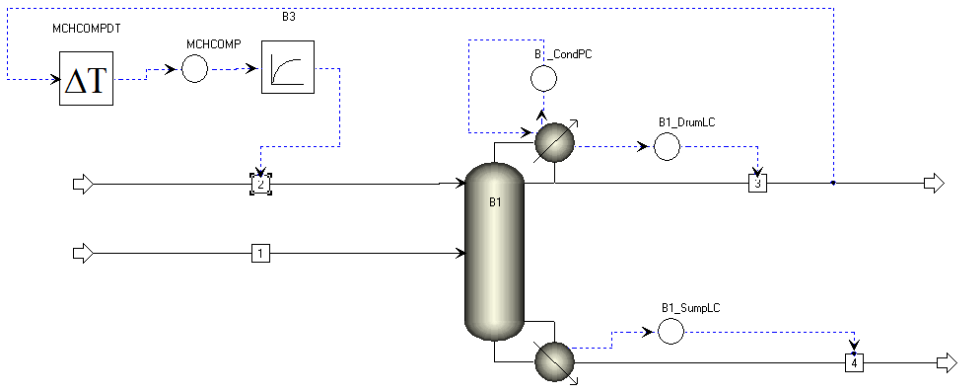


Figure 4.15. Schematic of the MCH distillation process.

Table 4.6. Prediction results of all trained models on the testing dataset of the distillation example.

Data-driven model	**Input length**	**Similarity measure**	**RMSE** $(\times \mathbf{10^3})$	**MAPE (%)**	**TIC** $(\times \mathbf{10^3})$	$\boldsymbol{R^2}$
SVM	1	Euclidean distance	2.521	0.1321	1.2792	0.6179
SVM	18	Euclidean distance	0.5751	0.02568	0.2919	0.9801
SVM	18	DTW $(\delta = 9)$	0.4556	0.02020	0.2312	0.9875
SVM	18	DTW $(\delta = \infty)$	0.4635	0.02043	0.2353	0.9871
GPR	1	Euclidean distance	2.348	0.1348	1.192	0.6686
GPR	18	Euclidean distance	0.3407	0.01478	0.1729	0.9930
GPR	18	DTW $(\delta = 9)$	0.3124	0.01348	0.1585	0.9941
GPR	18	DTW $(\delta = \infty)$	0.3072	0.01339	0.1559	0.9943

Chapter 5. Concluding Remarks

5.1. Conclusions

This thesis proposes novel modeling approaches to improve prediction accuracy and generalizability for multi-grade and time-varying processes. The prediction of product qualities and reactor conditions from easy-to-measure process variables is required to process monitoring and quality control. However, industrial plants suffer from frequent dynamics changes derived from multi-grade operation and process drifts, posing challenges in accurate prediction. The modeling approaches proposed in this thesis contribute to more accurate prediction of time-varying processes, while achieving improved generalization ability.

First, a hybrid modeling approach that combines mechanistic modeling and data-driven machine learning modeling is proposed for accurate prediction of the quality of polymer products of commercial plants. The proposed hybrid model consists of two submodels in series. The first part of the hybrid model is a white-box submodel which is based on the mechanistic modeling approach. Polymerization-related variables including molecular weight and polymerization rate are calculated using process knowledge such as polymerization mechanisms and material and energy balances. The white-box submodel outputs a total of 17 variables that cannot be measured in real plant but significantly affect the quality of the polymer product, which aids in more accurate and robust quality prediction. The second part of the hybrid model is a black-box submodel which predicts MI from process variables and the white-box submodel's outputs. A machine learning regression model is trained using latent variables obtained by applying PCA to the raw variables. Additionally, their hyperparameters are optimized using Bayesian optimization. The proposed hybrid model demonstrates more accurate MI for a commercial SAN polymerization process compared to conventional mechanistic models and data-only machine

learning models. Additionally, the results on the unsteady state data indicate that the proposed hybrid model achieves improved generalizability by combining prior process knowledge into soft sensor modeling.

Second, a JITL modeling approach combined with DTW is proposed for processes with consistently changing dynamics. A global data-driven soft sensor model's prediction performance drastically reduced when the operating condition deviates from the training region. Therefore, a soft sensor model able to address changes in the data distribution is required for the robust prediction in industrial chemical processes. The proposed DTW-based JITL model focuses on training a local model for every new query data using time series. First, in order to account for temporal dynamic behaviors of the process, the raw data variables are augmented into the data sequence, which is used as a model input. Next, the similarities between the query data sequence and the data sequences in the historical database are calculated using DTW. DTW accurately calculates distance between two different data sequences even when they are distorted. Based on the similarity calculation results, the most similar data sequences are selected for local prediction model training. Lastly, the prediction for the query data is made using the trained local model which is discarded after the prediction is made. The proposed modeling approach achieves more accurate prediction performance and fast adaptation ability for time-varying dynamic systems compared to the conventional JITL models based on the Euclidean distance. Additionally, the optimality of model parameters is examined in the thesis to give direction to more improved modeling performance.

5.2. Future Works

The two modeling approaches proposed in this thesis demonstrates improved prediction performance and generalizability for an actual polymerization process and simulation case studies. Nevertheless, there are still room for further improvement in the proposed methods. In this last part of the thesis, several possible directions for further research are suggested.

In Chapter 3, the serial hybrid modeling approach is developed for a specific thermoplastic polymerization process. In order to apply the proposed modeling method to various types of industrial plants, the modeling performance of the hybrid model should be validated for other chemical processes. The white-box submodel in this thesis focuses on the bulk radical polymerization reaction in continuous serial reactors. However, changes in reaction conditions, such as reactor type, temperature, and pressure, require revision of the modeling procedure for the white-box submodel. Therefore, a generalized mechanistic model should be developed so that the proposed hybrid model can be applied to various types of industrial processes.

Next, there are suggestions for improving the performance of the DTW-based JITL modeling approach. First of all, the just-in-time learning method has limited modeling performance for processes where similar inputs and operating conditions yield different outputs. For instance, fed-batch processes have product qualities that are cumulative over the various reaction phases, while most process inputs remain unchanged. Therefore, in order to improve the modeling performance of the proposed modeling approach, the past measurement data may be augmented into the model inputs. By considering recent measurements, a just-in-time learning model can be more accurate in dividing similar process conditions. Additionally, batch indices can be considered when applying the proposed modeling approach to batch processes where batches close to one another have high similarities.

Lastly, the two proposed modeling methods may be combined to provide

synergistic effect to further improve modeling performance. Most chemical processes inherently exhibit dynamic behaviors. Additionally, multi-grade operation and process drifts are common in actual plants. Therefore, by considering the time-varying nature of not only the process variables but also the white-box output variables that can be calculated using mechanistic modeling, both the prediction accuracy and generalizability are expected to be improved. Additionally, GPR models have demonstrated good prediction performance when applied to both proposed modeling approaches. Unlike other popular regression models, GPR models provide not only predicted values but also prediction uncertainties. The prediction performance of the proposed models may be improved by utilizing the uncertainty information obtained from GPR models. For example, ensemble approach can be implemented where multiple GPR models are trained and their predictions are combined.

Bibliography

[1] M. Ohshima and M. Tanigaki, "Quality control of polymer production processes," *Journal of Process Control,* vol. 10, no. 2-3, pp. 135-148, 2000.

[2] T. Kourti, "Application of latent variable methods to process control and multivariate statistical process control in industry," *International Journal of Adaptive Control and Signal Processing,* vol. 19, no. 4, pp. 213-246, 2005.

[3] P. Kadlec, B. Gabrys and S. Strandt, "Data-driven soft sensors in the process industry," *Computers & Chemical Engineering,* vol. 33, no. 4, pp. 795-814, 2009.

[4] F. Ahmed, L.-H. Kim and Y.-K. Yeo, "Statistical data modeling based on partial least squares: Application to melt index predictions in high density polyethylene processes to achieve energy-saving operation," *Korean Journal of Chemical Engineering,* vol. 30, no. 1, pp. 11-19, 2013.

[5] I.-S. Han, C. Han and C.-B. Chung, "Melt index modeling with support vector machines, partial least squares, and artificial neural networks," *Journal of Applied Polymer Science,* vol. 95, no. 4, pp. 967-974, 2005.

[6] T. C. Park, T. Y. Kim and Y. K. Yeo, "Prediction of the melt flow index using partial least squares and support vector regression in high-density polyethylene (HDPE) process," *Korean Journal of Chemical Engineering,* vol. 27, no. 6, pp. 1662-1668, 2010.

[7] W. Yan, H. Shao and X. Wang, "Soft sensing modeling based on support vector machine and Bayesian model selection," *Computers & Chemical Engineering,* vol. 28, no. 8, pp. 1489-1498, 2004.

[8] J. Shi and X. Liu, "Melt index prediction by weighted least squares

support vector machines," *Journal of Applied Polymer Science,* vol. 101, no. 1, pp. 285-289, 2006.

[9] W. Wang and X. Liu, "Melt index prediction by least squares support vector machines with an adaptive mutation fruit fly optimization algorithm," *Chemometrics and Intelligent Laboratory Systems,* vol. 141, pp. 79-87, 2015.

[10] M. Zhang and X. Liu, "A real-time model based on optimized least squares support vector machine for industrial polypropylene melt index prediction," *Journal of Chemometrics,* vol. 30, no. 6, pp. 324-331, 2016.

[11] Z. Ge, T. Chen and Z. Song, "Quality prediction for polypropylene production process based on CLGPR model," *Control Engineering Practice,* vol. 19, no. 5, pp. 423-432, 2011.

[12] Y. Liu and Z. Gao, "Industrial melt index prediction with the ensemble anti-outlier just-in-time Gaussian process regression modeling method," *Journal of Applied Polymer Science,* vol. 132, no. 22, p. 41958, 2015.

[13] Y. Liu, Y. Liang and Z. Gao, "Industrial polyethylene melt index prediction using ensemble manifold learning-based local model," *Journal of Applied Polymer Science,* vol. 134, no. 29, p. 45094, 2017.

[14] L. L. T. Chan and J. Chen, "Melt index prediction with a mixture of Gaussian process regression with embedded clustering and variable selections," *Journal of Applied Polymer Science,* vol. 134, no. 40, p. 45237, 2017.

[15] L. L. T. Chan and J. Chen, "Gaussian process model based multi-source labeled data transfer learning for reducing cost of modeling target chemical processes with unlabeled data," *Control Engineering Practice,* vol. 117, p. 104941, 2021.

[16] C. Zhao, x. Liu and F. Ding, "Melt Index Prediction Based on Adaptive

Particle Swarm Optimization Algorithm-Optimized Radial Basis Function Neural Networks," *Chemical Engineering & Technology,* vol. 33, no. 11, pp. 1909-1916, 2010.

[17] J. Li and X. Liu, "Melt index prediction by RBF neural network optimized with an adaptive new ant colony optimization algorithm," *Journal of Applied Polymer Science,* vol. 119, no. 5, pp. 3093-3100, 2011.

[18] J. Li, X. Liu, H. Jiang and Y. Xiao, "Melt index prediction by adaptively aggregated RBF neural networks trained with novel ACO algorithm," *Journal of Applied Polymer Science,* vol. 125, no. 2, pp. 943-951, 2012.

[19] J. Li and X. Liu, "Melt index prediction by RBF neural network optimized with an MPSO-SA hybrid algorithm," *Neurocomputing,* vol. 74, no. 5, pp. 735-740, 2011.

[20] M. Huang, X. Liu and J. Li, "Melt index prediction by RBF neural network with an ICO-VSA hybrid optimization algorithm," *Journal of Applied Polymer Science,* vol. 126, no. 2, pp. 519-526, 2012.

[21] Z. Zhang, T. Wang and X. Liu, "Melt index prediction by aggregated RBF neural networks trained with chaotic theory," *Neurocomputing,* vol. 131, pp. 368-376, 2014.

[22] X. Liu and C. Zhao, "Melt index prediction based on fuzzy neural networks and PSO algorithm with online correction strategy," *AIChE Journal,* vol. 58, no. 4, pp. 1194-1202, 2012.

[23] S. Xu and X. Liu, "Melt index prediction by fuzzy functions with dynamic fuzzy neural networks," *Neurocomputing,* vol. 142, pp. 291-298, 2014.

[24] W. Wang, M. Zhang and X. Liu, "Improved fruit fly optimization algorithm optimized wavelet neural network for statistical data modeling for industrial polypropylene melt index prediction," *Journal of*

Chemometrics, vol. 29, no. 9, pp. 506-513, 2015.

[25] M. Zhang, X. Liu, Z. Zhang, G. Gong, G. Zhang, C. Zhang, X. Chen and H. Zhang, "A novel modeling approach and its application in polymer quality index prediction," *Transactions of the Institute of Measurement and Control,* vol. 41, no. 7, pp. 2005-2015, 2019.

[26] C.-H. Zhu and J. Zhang, "Developing soft sensors for polymer melt index in an industrial polymerization process using deep belief networks," *International Journal of Automation and Computing,* vol. 17, no. 1, pp. 44-54, 2020.

[27] H. Wu, Y. Han, J. Jin and Z. Geng, "Novel deep learning based on data fusion integrating correlation analysis for soft sensor modeling," *Industrial & Engineering Chemistry Research,* vol. 60, no. 27, pp. 10001-10010, 2021.

[28] W. Li, H. H. Yue, S. Valle-Cervantes and J. S. Qin, "Recursive PCA for adaptive process monitoring," *Journal of Process Control,* vol. 10, no. 5, pp. 471-486, 2000.

[29] H. D. Jin, Y.-H. Lee, G. Lee and C. Han, "Robust recursive principal component analysis modeling for adaptive monitoring," *Industrial & Engineering Chemistry Research,* vol. 45, no. 2, pp. 696-703, 2006.

[30] X. Wang, U. Kruger and G. W. Irwin, "Process Monitoring Approach Using Fast Moving Window PCA," *Industrial & Engineering Chemistry Research,* vol. 44, no. 15, pp. 5691-5702, 2005.

[31] J. S. Qin, "Recursive PLS algorithms for adaptive data modeling," *Computers & Chemical Engineering,* vol. 22, no. 4-5, pp. 503-514, 1998.

[32] B. S. Dayal and J. F. MacGregor, "Recursive exponentially weighted PLS and its applications to adaptive control and prediction," *Journal of Process control,* vol. 7, no. 3, pp. 169-179, 1997.

[33] F. Ahmed, S. Nazir and Y. K. Yeo, "A recursive PLS-based soft sensor for prediction of the melt index during grade change operations in HDPE plant," *Korean Journal of Chemical Engineering,* vol. 26, no. 1, pp. 14-20, 2009.

[34] L. Xie, J. Zeng and C. Gao, "Novel just-in-time learning-based soft sensor utilizing non-Gaussian information," *IEEE Transactions on Control Systems Technology,* vol. 22, no. 1, pp. 360-368, 2013.

[35] H. Kaneko and K. Funatsu, "Ensemble locally weighted partial least squares as a just-in-time modeling method," *AIChE Journal,* vol. 62, no. 3, pp. 717-725, 2016.

[36] J. Liu, T. Liu and J. Chen, "Quality prediction for multi-grade processes by just-in-time latent variable modeling with integration of common and special features," *Chemical Engineering Science,* vol. 191, pp. 31-41, 2018.

[37] J. Liu, J. Hou and J. Chen, "Dual-layer feature extraction based soft sensor methods and applications to industrial polyethylene processes," *Computers & Chemical Engineering,* vol. 154, p. 107469, 2021.

[38] Y. Liu, Z. Gao, P. Li and H. Wang, "Just-in-time kernel learning with adaptive parameter selection for soft sensor modeling of batch processes," *Industrial & Engineering Chemistry Research,* vol. 51, no. 11, pp. 4313-4327, 2012.

[39] K. Yang, H. Jin, X. Chen, J. Dai, L. Wang and D. Zhang, "Soft sensor development for online quality prediction of industrial batch rubber mixing process using ensemble just-in-time Gaussian process regression models," *Chemometrics and Intelligent Laboratory Systems,* vol. 155, pp. 170-182, 2016.

[40] X. Yuan, J. Zhou, Y. Wang and C. Yang, "Multi-similarity measurement

driven ensemble just-in-time learning for soft sensing of industrial processes," *Journal of Chemometrics,* vol. 32, no. 9, p. e3040, 2018.

[41] Y. Wu, D. Liu, X. Yuan and Y. Wang, "A just-in-time fine-tuning framework for deep learning of SAE in adaptive data-driven modeling of time-varying industrial processes," *IEEE Sensors Journal,* vol. 21, no. 3, pp. 3497-3505, 2020.

[42] J. Wang, K. Qiu, R. Wang, X. Zhou and Y. Guo, "Development of Soft Sensor Based on Sequential Kernel Fuzzy Partitioning and Just-in-Time Relevance Vector Machine for Multiphase Batch Processes," *IEEE Transactions on Instrumentation and Measurement,* vol. 70, pp. 1-10, 2021.

[43] J. Zheng, F. Shen and L. Ye, "Improved Mahalanobis Distance Based JITL-LSTM Soft Sensor for Multiphase Batch Processes," *IEEE Access,* vol. 9, pp. 72172-72182, 2021.

[44] X. Yuan, Z. Ge, B. Huang, Z. Song and Y. Wang, "Semisupervised JITL framework for nonlinear industrial soft sensing based on locally semisupervised weighted PCR," *IEEE Transactions on Industrial Informatics,* vol. 13, no. 2, pp. 532-541, 2016.

[45] F. Guo and B. Huang, "A mutual information-based Variational Autoencoder for robust JIT soft sensing with abnormal observations," *Chemometrics and Intelligent Laboratory Systems,* vol. 204, p. 104118, 2020.

[46] F. Guo, W. Bai and B. Huang, "Output-relevant variational autoencoder for just-in-time soft sensor modeling with missing data," *Journal of Process Control,* vol. 92, pp. 90-97, 2020.

[47] F. Guo, B. Wei and B. Huang, "A just-in-time modeling approach for multimode soft sensor based on Gaussian mixture variational

autoencoder," *Computers & Chemical Engineering,* vol. 146, p. 107230, 2021.

[48] A. Kholmatov and B. Yanikoglu, "Identity authentication using improved online signature verification method," *Pattern Recognition Letters,* vol. 26, no. 15, pp. 2400-2408, 2005.

[49] N. Gillian, B. Knapp and S. O'modhrain, "Recognition Of Multivariate Temporal Musical Gestures Using N-Dimensional Dynamic Time Warping," in *Nime*, Oslo, 2011.

[50] X. Meng, H. Fu, L. Peng, G. Liu, Y. Yu, Z. Wang and E. Chen, "D-LSTM: short-term road traffic speed prediction model based on GPS positioning data," *IEEE Transactions on Intelligent Transportation Systems,* vol. 23, no. 3, pp. 2021-2030, 2020.

[51] K. Q. Zhou, Y. Qin, B. P. L. Lau, C. Yuen and S. Adams, "Lithium-ion Battery State of Health Estimation based on Cycle Synchronization using Dynamic Time Warping," in *IECON 2021--47th Annual Conference of the IEEE Industrial Electronics Society*, 2021.

[52] Y. Si, Z. Chen, J. Sun, D. Zhang and P. Qian, "A data-driven fault detection framework using mahalanobis distance based dynamic time warping," *IEEE Access,* vol. 8, pp. 108359-108370, 2020.

[53] A. Kassidas, J. F. MacGregor and P. A. Taylor, "Synchronization of batch trajectories using dynamic time warping," *AIChE Journal,* vol. 44, no. 4, pp. 864-875, 1998.

[54] H. Wold, "Causal flows with latent variables: partings of the ways in the light of NIPALS modelling," *European Economic Review,* vol. 5, no. 1, pp. 67-86, 1974.

[55] S. De Jong, "SIMPLS: an alternative approach to partial least squares regression," *Chemometrics and Intelligent Laboratory Systems,* vol. 18,

no. 3, pp. 251-263, 1993.

[56] P. Geladi and B. R. Kowalski, "Partial least-squares regression: a tutorial," *Analytica Chimica Acta,* vol. 185, pp. 1-17, 1986.

[57] V. Vapnik, *Estimation of dependences based on empirical data*, Springer Science & Business Media, 2006.

[58] B. E. Boser, I. M. Guyon and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, 1992.

[59] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning,* vol. 20, no. 3, pp. 273-297, 1995.

[60] I. Guyon, V. Vapnik, B. Boser, L. Bottou and S. A. Solla, "Structural risk minimization for character recognition," *Advances in Neural Information Processing Systems,* vol. 4, 1991.

[61] E. Osuna, R. Freund and F. Girosi, "An improved training algorithm for support vector machines," in *Neural Networks for Signal Processing VII. Proceedings of the 1997 IEEE Signal Processing Society Workshop*, 1997.

[62] J. Platt, "Sequential minimal optimization: A fast algorithm for training support vector machines," 1998.

[63] G. W. Flake and S. Lawrence, "Efficient SVM regression training with SMO," *Machine Learning,* vol. 46, no. 1, pp. 271-290, 2002.

[64] C. Williams and C. Rasmussen, "Gaussian processes for regression," *Advances in Neural Information Processing Systems,* vol. 8, 1995.

[65] C. K. Williams and C. E. Rasmussen, *Gaussian processes for machine learning*, MIT press Cambridge, MA, 2006.

[66] A. Shenoy, S. Chattopadhyay and V. Nadkarni, "From melt flow index to rheogram," *Rheologica Acta,* vol. 22, no. 1, pp. 90-101, 1983.

[67] T. Bremner, A. Rudin and D. Cook, "Melt flow index values and molecular weight distributions of commercial thermoplastics," *Journal of Applied Polymer Science,* vol. 41, no. 7-8, pp. 1617-1627, 1990.

[68] S. S. Bafna and A.-m. Beall, "A design of experiments study on the factors affecting variability in the melt index measurement," *Journal of Applied Polymer Science,* vol. 65, no. 2, pp. 277-288, 1997.

[69] J. R. Richards and J. P. Congalidis, "Measurement and control of polymerization reactors," *Computers & Chemical Engineering,* vol. 30, no. 10-12, pp. 1447-1463, 2006.

[70] K. B. McAuley and J. F. MacGregor, "On-line inference of polymer properties in an industrial polyethylene reactor," *AIChE Journal,* vol. 37, no. 6, pp. 825-835, 1991.

[71] T. Y. Kim and Y. K. Yeo, "Development of polyethylene melt index inferential model," *Korean Journal of Chemical Engineering,* vol. 27, no. 6, pp. 1669-1674, 2010.

[72] G. H. Varshouee, A. Heydarinasab, A. Vaziri and B. Roozbahani, "Predicting molecular weight distribution, melt flow index, and bulk density in a polypropylene reactor via a validated mathematical model," *Theoretical Foundations of Chemical Engineering,* vol. 55, no. 1, pp. 153-166, 2021.

[73] J. van Gisbergen and J. den Doelder, "Processability predictions for mechanically recycled blends of linear polymers," *Journal of Polymer Engineering,* vol. 40, no. 9, pp. 771-781, 2020.

[74] Y. Sun, Y. Wang, X. Liu, C. Yang, Z. Zhang, W. Gui, X. Chen and B. Zhu, "A novel Bayesian inference soft sensor for real-time statistic learning modeling for industrial polypropylene melt index prediction," *Journal of Applied Polymer Science,* vol. 134, no. 40, p. 45384, 2017.

[75] N. F. Jumari and K. M. Yusof, "Comparison of Melt Flow Index of Propylene Polymerisation in Loop Reactors using First Principles and Artificial Neural Network Models," *Chemical Engineering Transactions,* vol. 56, pp. 163-168, 2017.

[76] A. Singh, P. Majumder and U. K. Bera, "Prediction of polypropylene business strategy for a petrochemical plant using a technique for order preference by similarity to an ideal solution-based artificial neural network," *Polymer Engineering & Science,* vol. 62, no. 4, pp. 1096-1113, 2022.

[77] M. von Stosch, R. Oliveira, J. Peres and S. F. de Azevedo, "Hybrid semi-parametric modeling in process systems engineering: Past, present and future," *Computers & Chemical Engineering,* vol. 60, pp. 86-101, 2014.

[78] H.-T. Su, N. Bhat, P. A. Minderman and T. J. McAvoy, "Integrating neural networks with first principles models for dynamic modeling," in *Dynamics and Control of Chemical Reactors, Distillation Columns and Batch Processes*, 1993.

[79] D. C. Psichogios and L. H. Ungar, "A hybrid neural network-first principles approach to process modeling," *AIChE Journal,* vol. 38, no. 10, pp. 1499-1511, 1992.

[80] M. L. Thompson and M. A. Kramer, "Modeling chemical processes using prior knowledge and neural networks," *AIChE Journal,* vol. 40, no. 8, pp. 1328-1340, 1994.

[81] H. Te Braake, H. van Can and H. Verbruggen, "Semi-mechanistic modeling of chemical processes with neural networks," *Engineering Applications of Artificial Intelligence,* vol. 11, no. 4, pp. 507-515, 1998.

[82] N. Bhutani, G. Rangaiah and A. Ray, "First-principles, data-based, and hybrid modeling and optimization of an industrial hydrocracking unit,"

Industrial & Engineering Chemistry Research, vol. 45, no. 23, pp. 7807-7816, 2006.

[83] F. S. Mjalli and A. S. Ibrehem, "Optimal hybrid modeling approach for polymerization reactors using parameter estimation techniques," *Chemical Engineering Research and Design,* vol. 89, no. 7, pp. 1078-1087, 2011.

[84] S. H. Ju, S. Kim, S. H. Oh, P. S. Jo and J. M. Lee, "Dynamic Modeling of Modified Styrene-Acrylonitrile Process," *IFAC-PapersOnLine,* vol. 55, no. 7, pp. 272-277, 2022.

[85] C.-C. Chen, "A segment-based local composition model for the Gibbs energy of polymer solutions," *Fluid Phase Equilibria,* vol. 83, pp. 301-312, 1993.

[86] A. Fredenslund, R. L. Jones and J. M. Prausnitz, "Group-contribution estimation of activity coefficients in nonideal liquid mixtures," *AIChE Journal,* vol. 21, no. 6, pp. 1086-1099, 1975.

[87] F. J. Karol, G. L. Brown and J. M. Davison, "Chromocene-based catalysts for ethylene polymerization: Kinetic parameters," *Journal of Polymer Science: Polymer Chemistry Edition,* vol. 11, no. 2, pp. 413-424, 1973.

[88] M. Ogawa, M. Ohshima, K. Morinaga and F. Watanabe, "Quality inferential control of an industrial high density polyethylene process," *Journal of Process Control,* vol. 9, no. 1, pp. 51-59, 1999.

[89] Y. Liu and J. Chen, "Integrated soft sensor using just-in-time support vector regression and probabilistic analysis for quality prediction of multi-grade processes," *Journal of Process Control,* vol. 23, no. 6, pp. 793-804, 2013.

[90] Y. Liu, Z. Gao and J. Chen, "Development of soft-sensors for online quality prediction of sequential-reactor-multi-grade industrial processes,"

Chemical Engineering Science, vol. 102, pp. 602-612, 2013.

[91] S. J. Qin, H. Yue and R. Dunia, "Self-validating inferential sensors with application to air emission monitoring," *Industrial & Engineering Chemistry Research,* vol. 36, no. 5, pp. 1675-1685, 1997.

[92] Z. Ge, "Quality prediction and analysis for large-scale processes based on multi-level principal component modeling strategy," *Control Engineering Practice,* vol. 31, pp. 9-23, 2014.

[93] M. K. Hartnett, G. Lightbody and G. W. Irwin, "Dynamic inferential estimation using principal components regression (PCR)," *Chemometrics and Intelligent Laboratory Systems,* vol. 40, no. 2, pp. 215-224, 1998.

[94] J. Yu, "Multiway Gaussian mixture model based adaptive kernel partial least squares regression method for soft sensor estimation and reliable quality prediction of nonlinear multiphase batch processes," *Industrial & Engineering Chemistry Research,* vol. 51, no. 40, pp. 13227-13237, 2012.

[95] W. Shao and X. Tian, "Adaptive soft sensor for quality prediction of chemical processes based on selective ensemble of local partial least squares models," *Chemical Engineering Research and Design,* vol. 95, pp. 113-132, 2015.

[96] Y. Matsuyama, S. Kim and S. Hasebe, "Robust parameter tuning method of LW-PLS and verification of its effectiveness by twelve industrial processes," *Computers & Chemical Engineering,* vol. 146, p. 107224, 2021.

[97] S. Park and C. Han, "A nonlinear soft sensor based on multivariate smoothing procedure for quality estimation in distillation columns," *Computers & Chemical Engineering,* vol. 24, no. 2-7, pp. 871-877, 2000.

[98] H. Jin, X. Chen, J. Yang, H. Zhang, L. Wang and L. Wu, "Multi-model adaptive soft sensor modeling method using local learning and online

support vector regression for nonlinear time-variant batch processes," *Chemical Engineering Science,* vol. 131, pp. 282-303, 2015.

[99] J. Yu, "Online quality prediction of nonlinear and non-Gaussian chemical processes with shifting dynamics using finite mixture model based Gaussian process regression approach," *Chemical Engineering Science,* vol. 82, pp. 22-30, 2012.

[100] R. Grbic, D. Sliskovic and P. Kadlec, "Adaptive soft sensor for online prediction and process monitoring based on a mixture of Gaussian process models," *Computers & Chemical Engineering,* vol. 58, pp. 84-97, 2013.

[101] Y. Liu, T. Chen and J. Chen, "Auto-switch Gaussian process regression-based probabilistic soft sensors for industrial multigrade processes with transitions," *Industrial & Engineering Chemistry Research,* vol. 54, no. 18, pp. 5037-5047, 2015.

[102] J. C. B. Gozaga, L. A. C. Meleiro, C. Kiang and R. Maciel Filho, "ANN-based soft-sensor for real-time process monitoring and control of an industrial polymerization process," *Computers & Chemical Engineering,* vol. 33, no. 1, pp. 43-49, 2009.

[103] A. J. De Assis and R. Maciel Filho, "Soft sensors development for on-line bioreactor state estimation," *Computers & Chemical Engineering,* vol. 24, no. 2-7, pp. 1099-1103, 2000.

[104] X. Yuan, L. Li and Y. Wang, "Nonlinear dynamic soft sensor modeling with supervised long short-term memory network," *IEEE Transactions on Industrial Informatics,* vol. 16, no. 5, pp. 3168-3176, 2019.

[105] X. Yuan, L. Li, Y. A. Shardt, Y. Wang and C. Yang, "Deep learning with spatiotemporal attention-based LSTM for industrial soft sensor model development," *IEEE Transactions on Industrial Electronics,* vol. 68, no. 5, pp. 4404-4414, 2020.

[106] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Transactions on Acoustics, Speech, and Signal Processing,* vol. 26, no. 1, pp. 43-49, 1978.

[107] H. Ding, G. Trajcevski, P. Scheuermann, X. Wang and E. Keogh, "Querying and mining of time series data: experimental comparison of representations and distance measures," *Proceedings of the VLDB Endowment,* vol. 1, no. 2, pp. 1542-1552, 2008.

[108] W. Ku, R. H. Storer and C. Georgakis, "Disturbance detection and isolation by dynamic principal component analysis," *Chemometrics and Intelligent Laboratory Systems,* vol. 30, no. 1, pp. 179-196, 1995.

[109] S. Heo and J. H. Lee, "Fault detection and classification using artificial neural networks," *IFAC-PapersOnLine,* vol. 51, no. 18, pp. 470-475, 2018.

[110] E. P. Nahas, M. A. Henson and D. E. Seborg, "Nonlinear internal model control strategy for neural network models," *Computers & Chemical Engineering,* vol. 16, no. 12, pp. 1039-1057, 1992.

초 록

실시간 온라인 품질 측정이 불가능한 많은 수의 산업 화학 공정에서는 운전 중 반응기의 상태 및 화학 제품의 성질과 품질을 예측하는 것이 매우 중요하다. 불량 제품 생산 감소 및 공정 효율 증대를 위해서 더욱 정확하고 보다 빠른 예측이 필수적이다. 따라서 제일 원리 혹은 기계 학습 회귀 방법론 등을 활용하여 산업 화학 공정의 정확하고 빠른 소프트 센서를 구축하기 위한 다양한 연구들이 지속적으로 수행되어왔다. 그러나 실제 산업에서 운전되는 화학 공정의 품질 예측을 어렵게 하는 몇 가지 문제점들이 존재한다.

첫째, 화학 공정의 공정 동역학은 다양한 화학 반응 및 에너지와 물질 전달, 그리고 상전이 등의 복잡한 상호 작용의 결과로써 예측 모델도 이러한 높은 수준의 공정 비선형성을 다룰 수 있어야 한다. 또한 공정의 높은 비선형성은 구축된 소프트 센서 모델의 해석을 더욱 어렵게 만들어 공정 모니터링 및 제어의 난이도가 높아진다. 둘째, 산업 화학 공정은 운전 중 동역학이 변화하는 시변 (time-varying) 공정인 경우가 많다. 고분자 중합 공정을 비롯하여 많은 수의 화학 공정들은 계속해서 변화하는 시장의 수요에 맞추기 위해 다양한 운전 조건에서 여러 제품을 번갈아 생산한다. 그러나 공정 초기 안정화 단계 및 등급 변경 등 비정상 상태에서 공정이 운전될 때에는 긴 안정화 시간 및 과도한 오버슈트 (overshoot) 등의 문제에 취약하다. 또한 대부분의 회분식 공정에서는 시간이 지남에 따라 반응기 내 화학 물질들의 농도가 변화하면서 지배적인 반응이 바뀌게 되어 공정은 이전과 다른 양상으로 진행된다. 실제 산업 현장에서 운전되는 공정에서는 위와 같이 공정의 특성으로 인한 동역학의 변화뿐만 아니라 촉매 비활성화 및 반응기 내 각종 오염으로 인해 의

도치 않은 공정 드리프트 (process drift)의 발생으로 인하여 공정 동역학이 변화하기도 한다. 이러한 공정 드리프트는 모델과 공정 간의 괴리를 야기하고 이로 인하여 구축된 소프트 센서 모델의 예측과 실제 품질 측정 값 간의 큰 편차가 발생할 수 있다. 따라서 더욱 정확하고 강건한 예측을 수행하기 위해서는 앞서 언급한 문제점들이 모델링 과정에서 적절하게 해결되어야 한다.

이러한 관점을 기반으로 본 논문은 산업 현장에서의 화학 공정의 모델링 과정에서 수반되는 문제점들을 해결하여 개선된 예측 정확도 및 강건함을 지닌 새로운 소프트 센서 개발 방법론들을 제안한다. 첫째, 실제 운전 중인 상업용 고분자 중합 공정에서 생산된 고분자 제품의 실시간 품질 측정을 위한 하이브리드 소프트 센서 모델링 방법론을 제안한다. 기존의 제일 원리 기반의 예측 모델 및 데이터만을 활용한 기계 학습 기반 예측 모델과 비교하였을 때, 제안한 하이브리드 모델은 반응 메커니즘과 같은 공정에 대한 사전 지식과 인공 신경망과 같은 데이터 기반 회귀 방법론을 결합함으로써 더욱 정확한 품질 예측을 수행할 수 있다. 또한 제안한 하이브리드 기법을 적용하여 얻어진 예측 모델은 공정에 대한 물리적, 화학적 정보를 통해 데이터 기반 예측 모델의 단점인 외삽 성능을 보완하여 더욱 뛰어난 일반화 성능을 나타낸다. 대상 고분자 중합 공정에서 얻어진 등급 변경과 같은 비정상 상태의 운전 데이터에 대하여 하이브리드 모델은 단순한 형태의 데이터만을 활용한 기계 학습 모델들에 비해 더 강건한 예측 성능을 보인다.

둘째, 동적 시간 왜곡 (dynamic time warping) 이 결합된 시계열 기반의 적시 학습 (just-in-time learning) 소프트 센서 모델링 기법을 제안한다. 다등급 공정 혹은 시변 공정에서 얻어진 공정 데이터는 여러 종류의 운전 조건이 혼합되어 다봉분포 (multimodal distribution)을 나타낸다. 이러한 형

태의 데이터는 단일한 전역 예측 모델만으로는 학습하기 어렵고 정확한 예측을 기대하기 어렵다. 적시 학습 모델링 기법에서는 모든 학습 데이터를 활용하여 단 하나의 전역 모델을 구축하는 대신 품질 예측이 필요한 현재 시간 스텝에서의 샘플과 유사한 샘플들만 추려내어 지역 모델을 학습한다. 따라서 적시 학습 기법으로 구축된 소프트 센서 모델은 시간에 따른 공정의 변화에도 적응할 수 있다는 장점이 있다. 또한 제안하는 모델링 기법에서는 특정 시간에서의 데이터 포인트 대신 시계열 형태의 데이터를 입력받음으로써 공정 데이터의 동적 양상을 고려한다. 적시 학습 기법에서는 어떠한 방식으로 샘플 간의 유사도를 계산하는지에 따라서 예측 성능이 크게 영향을 받는다. 따라서 시계열 간의 유사도를 보다 정확히 계산하기 위해서 동적 시간 왜곡을 제안한 모델링 기법에 적용한다. 동적 시간 왜곡은 데이터 배열 간의 정렬을 위해 널리 쓰이는 기법 중 하나로 배열을 늘이거나 압축하여 서로 다른 배열 간의 유사도를 정확히 계산할 수 있다.

제안한 모델링 방법론은 적시 학습 기법과 동적 시간 왜곡을 결합함으로써 다등급 및 시변 화학 공정에서의 소프트 센서 개발에 대해 다음 두 가지의 주요 기여를 달성한다. 첫째, 시계열을 모델의 입력으로 활용함으로써 복잡한 동역학을 지닌 공정에 대해 더 정확한 예측을 수행할 수 있다. 둘째, 테스트 시계열과 데이터베이스 내 시계열 간의 유사도 계산에 동적 시간 왜곡을 활용함으로써 유사한 샘플들을 더 정확하게 추려낼 수 있다.

또한 모델의 파라미터가 모델의 예측 성능에 미치는 영향을 연구한다. 민감도 분석 결과, 최적의 입력 길이는 공정의 입력 및 출력 변수 간의 상호 상관에 의해 결정되며 동적 시간 왜곡의 경로 너비를 제한함으로써 더 높은 예측 정확도와 계산량 감소 효과를 얻을 수 있다. 따라서 본 연

구에서는 최적화된 하이퍼파라미터 기반의 동적 시간 왜곡 기반이 결합된 적시 학습 소프트 센서 모델을 제안하며, 이 때 모델링 성능은 입출력 변수 간의 상호 상관 계수를 미리 계산하고 왜곡 경로의 너비를 제한하여 최적화된다. 제안한 모델링 방법론은 다등급 공정의 시뮬레이션 사례 연구들을 통해 개선된 예측 성능을 입증한다.

요악하면 본 논문은 높은 비선형성 및 다등급 혹은 시변 특성을 지닌 산업 화학 공정에 대해 개선된 예측 정확도와 일반화 성능을 지닌 새로운 소프트 센서 개발 방법론들을 제시한다. 시뮬레이션 및 실제 고분자 중합 공정의 데이터를 활용하여 기존 방법론들 대비 제안한 모델링 방법론의 우수한 예측 성능을 입증한다.

주요어: 소프트 센서, 기계 학습, 하이브리드 모델, 적시 학습, 동적 시간 왜곡

학번: 2017-28864