Systems & Control Transactions

Research Article - Peer Reviewed Conference Proceeding
ESCAPE 35 - European Symposium on Computer Aided Process Engineering
Ghent, Belgium. 6-9 July 2025
Jan F.M. Van Impe, Grégoire Léonard, Satyajeet S. Bhonsale,
Monika E. Polańska, Filip Logist (Eds.)

PSE PRESS

Predicting Final Properties in Ibuprofen Production with Variable Batch Durations

Kuan-Che Huang, David Shan-Hill Wong, Yuan Yao*

National Tsing Hua University, Department of Chemical Engineering, Hsinchu 300044, Taiwan
* Corresponding Author: yyao@mx.nthu.edu.tw

ABSTRACT

This study addresses the challenge of predicting final properties in batch processes with highly uneven durations, using the ibuprofen production process as a case study. Novel methodologies are proposed and compared against traditional regression algorithms, which rely on batch trajectory synchronization as a pre-processing step. The performance of each method is evaluated using established metrics. The data for this study were generated using Aspen Plus V12 simulation software, focused on batch reactors. To handle the unequal-length trajectories in batch processes, this research constructs a dual-transformer deep neural network with multi-head attention and layer normalization mechanism to extract shared information from the high-dimensional, uneven-length manipulated variable profiles into latent space, generating equal-dimensional latent codes. As an alternative strategy for representation learning, a dual-autoencoder framework is also employed to achieve equal-dimensional representations. The representation vectors are then used as inputs for downstream deep learning models to predict the target variables, achieving an accuracy with an R^2 score exceeding 0.9.

Keywords: Uneven durations, Batch process, Transformer, Autoencoder, Representation learning

INTRODUCTION

In the chemical industry, batch processing is a commonly used production method [1]. Throughout an batch process, the quality of the final product is a key indicator of interest to both industry and academia. However, variations between batches often lead to unequal processing times across different batches. Due to the requirement for consistent input in building a soft sensor model for quality prediction, there are traditionally two methods for handling unequal-length batch trajectories: truncation and time-series alignment [2]. However, these two methods often come with certain issues. For the truncation methods, cutting off process-related operational trajectories can result in the loss of important information, which may affect the accuracy of subsequent soft sensor modeling. As for the time warping methods, significant differences in time lengths can cause trajectory distortion, and compression in certain segments may introduce noise. This situation can also lead to wrong models.

In this research, novel methodologies are proposed and compared against traditional regression algorithms [3], which rely on batch trajectory synchronization as a pre-processing step. In batch processes, handling three-dimensional data containing batch, time, and process variables is typically done using batch-wise unfolding, which transforms the data into a two-dimensional matrix. Due to the multivariate nature of chemical plant systems, Partial Least Squares (PLS) regression is commonly employed for modelling. This framework is referred to as Multiway Partial Least Squares (MPLS) [4]. Even though MPLS is a well-established method for handling batch process modelling, combining time warping with batch-wise unfolding can still result in the curse of dimensionality and model instability, especially when the data sampling size is too small, leading to overfitting [5].

This study aims to address the issue of excessively unequal production times between batches in quality prediction by employing a strategy based on Transformer and Autoencoder neural network architectures. Additionally, a case study on the batch production of Ibuprofen was conducted to demonstrate the approach. Ibuprofen is a pain reliever and anti-inflammatory drug commonly

used for headaches, toothaches, and minor pain. The data for this study were generated using Aspen Plus V12 simulation software [6], focused on batch reactors. To maintain process stability, it is important to prevent significant changes in the trajectories of operational variables within a given time unit. We implemented a statistical sampling approach, employing Latin Hypercube Sampling (LHS) to select samples within acceptable process boundaries [7]. The completion of each batch was determined based on the final isobutylbenzene conversion rate. A total of 1,000 simulation runs were performed, and the generated data were utilized to build a neural network model. The target variables for prediction are: (1) the isobutylbenzene conversion rate, and (2) the accumulated mass of ibuprofen.

The entire model architecture proposed in this study is built on a dual-transformer framework [8], integrating multi-head attention and layer normalization mechanisms. This approach captures shared information from the high-dimensional, uneven-length manipulated variable profiles, mapping them into latent space to produce latent codes of equal dimensions. An alternative feature extraction strategy involves utilizing a dual-autoencoder framework [9], which is also employed to obtain equal-dimensional representations. The downstream network structure consists of two fully connected neural networks, each designed to predict one of the two target variables. Additionally, the Neural Network Scalable Vector Graphics (NNSVG) tool is employed for visualizing the model structure [10].

PRELIMINARIES

Autoencoder

Autoencoder is a neural network framework based on an encoder and decoder. Through dimensional transformations in the intermediate layers of the neural network, the goal is to reconstruct data that is identical to the input. The reconstruction error, which serves as the loss function, can be expressed as follows:

$$L(x, x') = \| \mathrm{x} - \mathrm{x}' \|^2 \quad (1)$$

Similar to Principal Component Analysis, Autoencoder provides the capability for dimensional reduction of high-dimensional data. The difference lies in that the activation function in an autoencoder can provide nonlinear treatment to extract latent data. For latent vectors, the target is often to have the low-dimensional data represent the original high-dimensional data through the shared information extracted by the Autoencoder. Reducing the dimensionality can improve the computational efficiency of subsequent model development and help avoid the curse of dimensionality. Furthermore, the low-dimensional representation of data, regarded as important information, can also serve a filtering function to do the denoise.

Transformer

The Transformer architecture utilizes attention mechanisms and layer normalization, eliminating the need for convolution or recurrence. There are various methods for data normalization. The advantage of using layer normalization in Transformers lies in its ability to handle three-dimension input data often associated with variable-length sequences, such as in translation tasks. When the length of sequences varies significantly, applying layer normalization to individual samples ensures that the mean and variance remain relatively stable. On the other hand, if batch normalization were used for mini-batch tasks, the insufficient padding part could affect the mean and variance, leading to fluctuations. When dealing with overly long new input sequences, the global mean and variance recorded by Batch Normalization may significantly affect the normalization of this unseen input.

In the Transformer model, multi-head attention is used, which involves multiple linear layers that project into lower-dimensional spaces. By utilizing multiple instances of scaled dot-product attention, the model learns projection parameters to identify different patterns in the sequence input. In scaled dot-product attention, the query and key undergo dot product and scaling, followed by the soft-max function to produce weight values between 0 and 1. Attention, in fact, can be regarded as a function, where the output is a weighted sum of the values. The weights are determined by the query's search over the keys corresponding to different values through the compatibility function, which can be viewed as a measure of similarity.

Latin Hypercube Sampling

There are numerous approaches to statistical sampling, such as the memoryless Monte Carlo sampling. In this study, Latin Hypercube Sampling (LHS), a memory-based method, was chosen. The goal of using this approach is to capture the population data distribution of the original process variables with the least amount of sampling. Additionally, over time, LHS data can help avoid excessive fluctuations in the process variables of the chemical plant during simulation outputs.

With a limited sample size, random or Monte Carlo sampling may result in sample clustering. In contrast, LHS employs stratified sampling, ensuring a well-distributed sample by dividing the range into equal intervals and selecting points within each.

DATASET AND SOFTWARE

This study used the batch production of ibuprofen as a case study to demonstrate that using feature extrac-

tion with latent codes of equal dimensions through Autoencoder and Transformer, expressed as feature learning for representing the original trajectories of chemical plant processes, combined with a neural network predictor, performs better than the approach of dynamic time warping combined with MPLS.

The dataset was generated from the Aspen Plus V12 simulation software, with the main reactor being a batch reactor, based on two feed streams and a final product outlet stream. The reactants include isobutylbenzene, acetic anhydride, hydrogen, and carbon monoxide for synthesis, with the primary product being ibuprofen. The side reaction produces acetic acid, and the chemical reaction is as follows,

$$\boldsymbol{C_{10}H_{14} + C_4H_6O_3 + H_2 + CO \rightarrow CH_3COOH + C_{13}H_{18}O_2} \quad (2)$$

The kinetic information for the chemical reaction involved in Ibuprofen production indicates that the consumption of isobutylbenzene follows a second-order reaction, while the other reactants exhibit first-order consumption kinetics. The formation of the products also follows a first-order reaction.

To accurately model phase equilibrium behavior in the system, we employed the Wilson thermodynamic model, which is expressed as:

$$\ln \gamma_i = 1 - \ln\left(\sum_j A_{ij} x_j\right) - \sum_j \frac{A_{ij} x_j}{\sum_k A_{jk} x_k} \quad (3)$$

where $lnA_{ij} = a_{ij} + \frac{b_{ij}}{T} + c_{ij} lnT + d_{ij} T + e_{ij}/T^2$.

The binary parameters $(a_{ij}, b_{ij}, c_{ij}, d_{ij}, e_{ij})$ were regressed using vapor-liquid equilibrium data from the Dortmund Databank, implemented via Aspen V12 software. In the pharmaceutical and chemical industries, the Wilson model is particularly advantageous due to its asymmetric interaction parameters, which enhance flexibility in capturing liquid-phase non-ideality, improving the accuracy of phase equilibrium predictions.

We used the conversion rate of isobutylbenzene as the stop criterion for each independent batch process. Due to different stop values and dynamic process operation profiles, the varying cycle times across batches were influenced. In this study, our goal was to do the regression two target variables: 1) the accumulated mass of ibuprofen and 2) the conversion rate of isobutylbenzene.

Using Python to drive Aspen software, we adjusted the two manipulated variables, temperature and pressure. The model training was based on Python 3.8.8, with the GPU being an NVIDIA GeForce RTX 2080 Ti. At each time step, the temperature and pressure points needed to obey with the range defined by LHS sampling while also avoiding large fluctuations in the variables, which was undesirable in chemical plant operations.

After data generation, we began by visualizing the issue of unequal lengths in batch data. As shown in Figure 1, we needed an effective solution for modeling with varying sequence lengths modeling. In 1,000 simulation experiments, the batch processing times varied significantly, with a maximum of 6158 seconds, a minimum of 2693 seconds, and a median of 3790 seconds. The subsequent comparison with the conventional methods employed the median as the representative template batch.

This study adjusted the dynamic process operations in the early stages of all batches, while the later stages reached a steady state following the Latin Hypercube Sampling method to generate data. We also visualize the variation of the two target variables across 1,000 independent batch simulations, shown in Figures 2 and 3.>

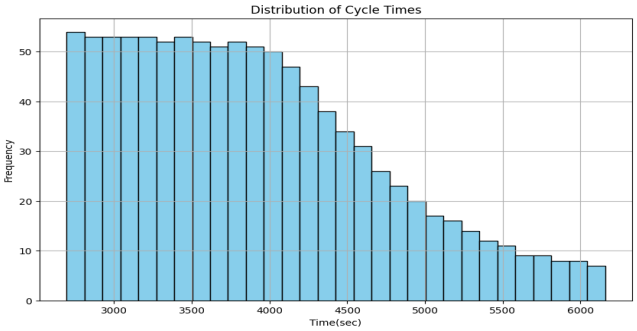

Figure 1. Batch time histogram for ibuprofen production

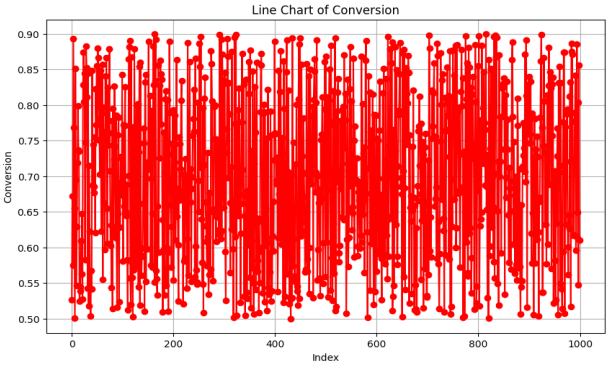

Figure 2. Trend of isobutylbenzene conversion rate

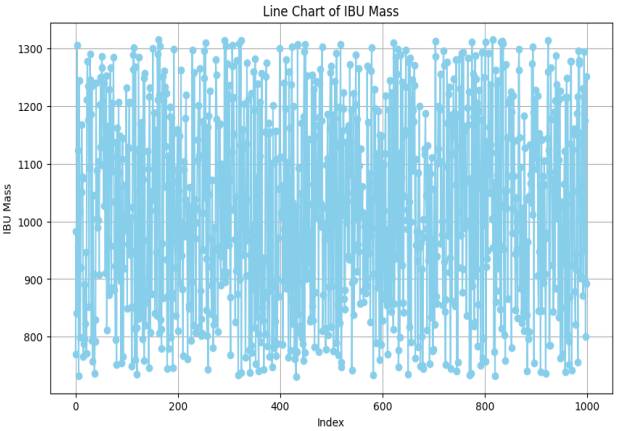

Figure 3. Trend of ibuprofen accumulated mass

METHODOLOGY AND RESULT DISCUSSIONS

This study employs two strategies to specifically address variations in process trajectories related to the manipulated variables of temperature and pressure, which could be extended to higher-dimensional systems with multiple process variables. The first strategy used in this study is the Autoencoder strategy, with the overall model architecture illustrated in the Figure 4. This model comprises a dual-autoencoder neural network, with two downstream predictors responsible for predicting the target variables: Ibuprofen accumulated mass and isobutylbenzene conversion rate, respectively. The overall loss function is a custom-defined loss function optimized through the Adam optimizer, which performs gradient descent to find the optimal weights. The total loss comprises the reconstruction loss of the two manipulated variables and the prediction loss of the two target variables.

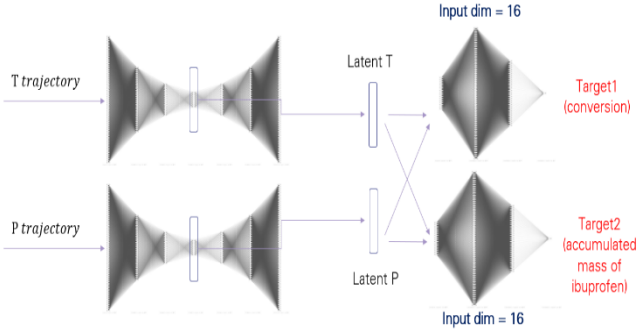

Figure 4. Framework of autoencoder strategy

The second strategy is the Transformer strategy, as shown in the model architecture Figure 5. Instead of using traditional recurrent neural networks to capture sequential relationships within the chemical engineering system, we utilize a multi-head attention mechanism. This mechanism applies multiple scaled dot-product attention layers and linear projections to identify trajectory similarities. To prevent the disproportionate influence of excessive padding on the global trajectory's mean and variance, layer normalization is applied. Furthermore, its scalability allows for adjusting the number of front-end Transformers to accommodate different chemical systems and corresponding manipulated variables.

The input layer for each autoencoder or Transformer basis consists of 62 parameters, where each corresponds to a single manipulated variable. The number 62 represents the sampling points for each manipulated variable in the longest production batch across all batches, with a time frequency of 100 seconds. For batches with fewer than 62 sampling points, padding is applied before entering the masking layer to ensure that these padded values are excluded from the loss function. Ultimately, the input data is compressed into an 8-dimensional latent space, and the two latent codes are merged into a 16-dimensional representation, which is then processed by the downstream predictor.

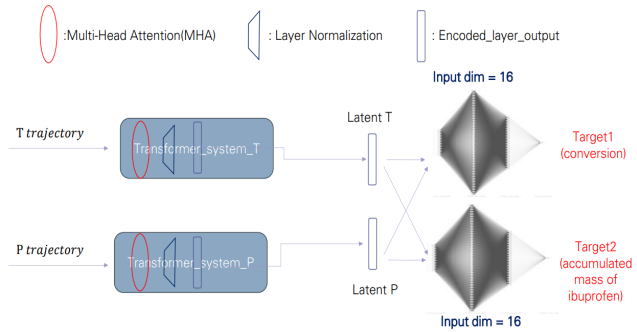

Figure 5. Transformer Method framework.

The conventional methods for comparison rely on using the golden batch as a template to align all candidate batches, followed by using dimensionality reduction machine learning algorithms (including MPLS and multi-way principal component regression (MPCR)) to predict the terminal properties of the process. The following Table 1 and Table 2 present performance comparisons.

Table 1: Ibuprofen accumulated mass prediction results

Ibuprofen accumulated mass performance metrics			
Subset	**Train**	**Validation**	**Test**
Method	$R^2/RMSE$	$R^2/RMSE$	$R^2/RMSE$
Autoencoder	0.99/7.6	0.99/7.6	0.99/7.5
Transformer	0.99/11.3	0.99/10.6	0.99/10.4
DTW+MPLS	0.02/167	-	-0.03/171
DTW+MPCR	0.01/167	-	-0.03/171

Table 2: Isobutylbenzene conversion rate prediction results

Isobutylbenzene conversion rate performance metrics			
Subset	**Train**	**Validation**	**Test**
Method	$R^2/RMSE$	$R^2/RMSE$	$R^2/RMSE$
Autoencoder	0.99/0.01	0.99/0.01	0.99/0.01
Transformer	0.99/0.01	0.99/0.01	0.99/0.01
DTW+MPLS	0.02/0.11	-	-0.03/0.1
DTW+MPCR	0.02/0.11	-	-0.03/0.1

The prediction results of the four methods on the test set, applied to the accumulated mass of Ibuprofen, are visualized in Figures 6 through 9. Among them, Figures 6 and 7 demonstrate that the proposed method in this study exhibits excellent predictive capability in addressing this a highly uneven-length batch production property prediction task.

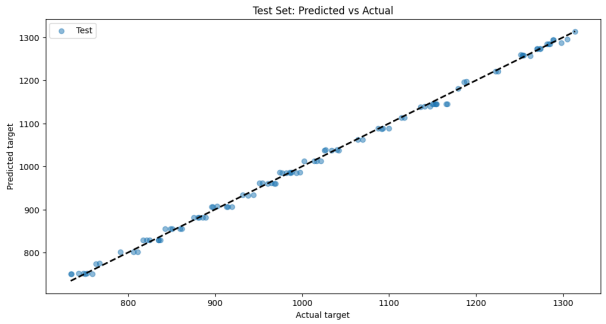

Figure 6. Ibuprofen accumulated mass (Autoencoder prediction results on the test set)

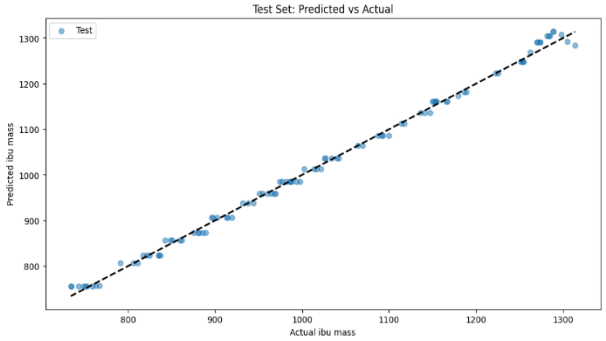

Figure 7. Ibuprofen accumulated mass (Transformer prediction results on the test set)

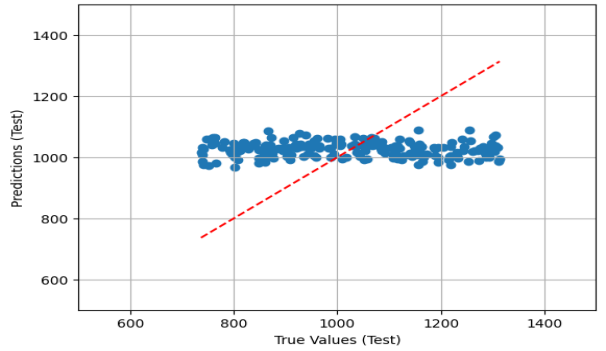

Figure 8. Ibuprofen accumulated mass (DTW+MPLS prediction results on the test set)

The prediction results of the four methods in test set can also be visualized in terms of the Isobutylbenzene conversion rate, as shown in the Figure 10 to Figure 13. The two strategies proposed in this study demonstrate better predictive performance compared to traditional trajectory synchronization and latent space dimensionality reduction machine learning methods. In the scatter plot, the prediction points are well-aligned near the ideal line, with R^2 for both target variables exceeding 0.9. This improvement is attributed to several factors: our neural network processes effectively handle the dynamics of trajectories with highly unequal lengths, avoiding the trajectory compression and introduction of error information that often result from warping. Additionally, both the autoencoder and transformer strategies incorporate activation functions, allowing for the handling of non-linear aspects of the system, which state-of-the-art methods lack in terms of non-linear learning capabilities.

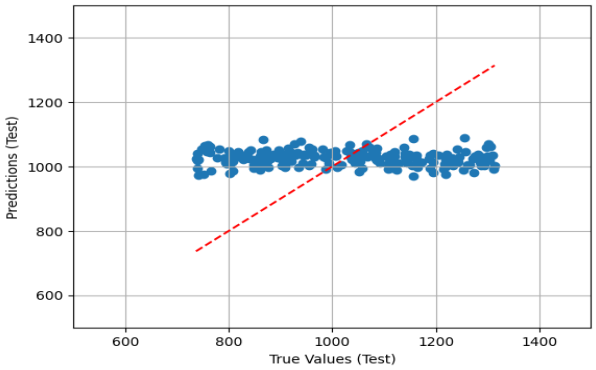

Figure 9. Ibuprofen accumulated mass (DTW+MPCR prediction results on the test set)

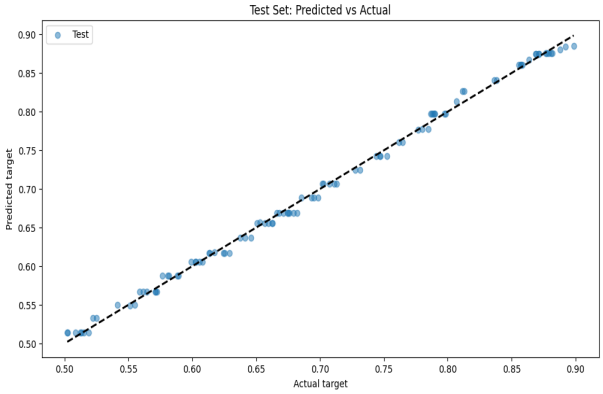

Figure 10. Isobutylbenzene conversion rate (Autoencoder prediction results on the test set)

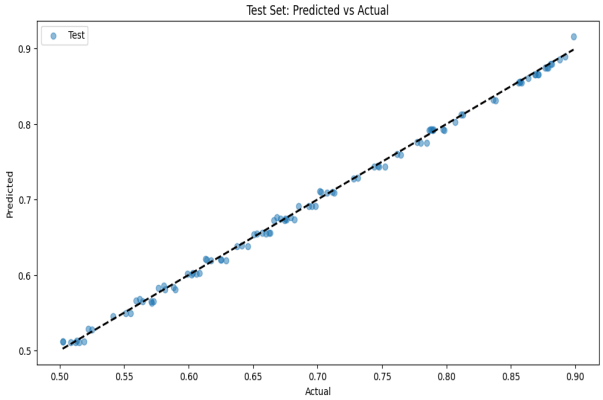

Figure 11. Isobutylbenzene conversion rate (Transformer prediction results on the test set)>

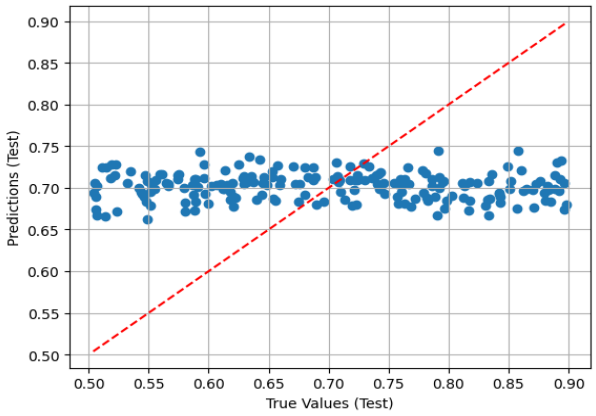

Figure 12. Isobutylbenzene conversion rate (DTW+MPLS prediction results on the test set)

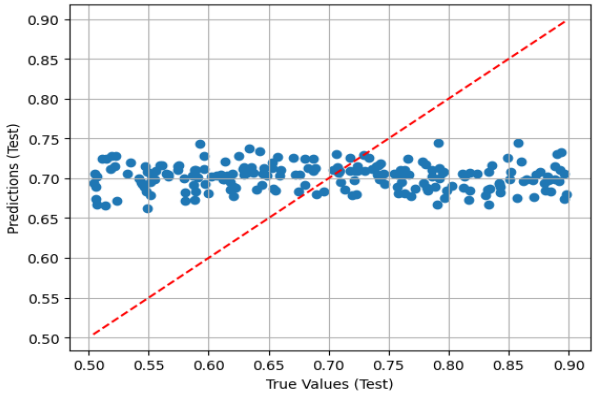

Figure 13. Isobutylbenzene conversion rate (DTW+MPCR prediction results on the test set)

CONCLUSIONS

This study proposes two innovative approaches to address the prediction problem of terminal properties in chemical processes with unequal batch lengths. The research contains two strategies, the dual-Transformer and dual-Autoencoder strategies. The neural network modules are determined by several process variables to extract latent space vectors from process trajectory data and input them into the downstream of the overall architecture for quality predictor modeling. In the case study of batch ibuprofen production, the R-squared values for both target variables exceed 0.9. This shows a significant difference compared to using only DTW combined with MPLS or MPCR.

ACKNOWLEDGEMENTS

This work was supported in part by the National Science and Technology Council, ROC, under Grant No. NSTC 113-2221-E-007-012-MY2.

REFERENCES

1. Yao Y, Gao F. A survey on multistage/multiphase statistical modeling methods for batch processes. *Annu Rev Control* 33:172-183 (2009). https://doi.org/10.1016/j.arcontrol.2009.08.001
2. Kassidas A, MacGregor JF, Taylor PA. Synchronization of batch trajectories using dynamic time warping. *AIChE J* 44:864-875 (1998). https://doi.org/10.1002/aic.690440412
3. Vigneau E, Devaux MF, Qannari EM, Robert P**.** Principal component regression, ridge regression and ridge principal component regression in spectroscopy calibration. *J Chemom* 11:239-249 (1997). https://doi.org/10.1002/(SICI)1099-128X(199705)11:3<239::AID-CEM470>3.0.CO;2-A
4. Nomikos P, MacGregor JF. Multi-way partial least squares in monitoring batch processes. *Chemom Intell Lab Syst* 30:97-108 (1995). https://doi.org/10.1016/0169-7439(95)00043-7
5. Sousa PF, Åberg KM. Can we beat overfitting?—A closer look at Cloarec's PLS algorithm. *J Chemom* 32:e3002 (2018). https://doi.org/10.1002/cem.3002
6. Aspen Technology, Inc. Aspen Plus User Guide (V12). Aspen Technology, Inc. (2017).
7. Iman RL. Latin Hypercube Sampling. John Wiley & Sons, Ltd. (2008).
8. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gómez AN, Kaiser Ł, Polosukhin I. Attention is all you need. *Adv Neural Inf Process Syst* (2017). https://doi.org/10.48550/arXiv.1706.03762
9. Vincent P, Larochelle H, Bengio Y, Manzagol PA. Extracting and composing robust features with denoising autoencoders. *Proc Int Conf Mach Learn* 25:1096-1103 (2008). https://doi.org/10.1145/1390156.1390294
10. LeNail A. NN-SVG: Publication-ready neural network architecture schematics. J Open Source Softw 4:747 (2019). https://doi.org/10.21105/joss.00747