

Soft Sensor for Melt Index Prediction Based on Long Short-Term Memory Network^{*}

Min Jun Song^{*} Sungkyu Kim^{**} Seung Hwan Oh^{**}
Pil Sung Jo^{**} Jong Min Lee^{*}

^{*} School of Chemical and Biological Engineering, Seoul National
University, Seoul, Korea (e-mail: smj08217@snu.ac.kr,
jongmin@snu.ac.kr).

^{**} Big Data Team, LG Chem, Ltd, Seoul, Korea (e-mail:
paul.kim@lgchem.com, oshznim@lgchem.com, pilsungjo@lgchem.com)

Abstract:

This paper presents a soft sensor model for melt index (MI) prediction in an industrial polymerization process based on long short-term memory (LSTM) network. MI is one of the important specifications that determine the quality and grade of thermoplastic polymers. However, lack of online measurement of MI makes it difficult to monitor and control the quality of polymer products. Thus, there has been a great effort to build accurate soft sensor models to predict MI with easy-to-measure process variables by using black-box modeling approaches. However, real chemical processes have strong nonlinearity and complicated temporal correlations between the process and quality variables, which is very challenging for traditional static black-box models to handle. Recently, LSTM network that is an advanced form of recurrent neural network (RNN) has shown great advantages in capturing and modeling the long-term dynamic nature of complex industrial processes. We develop an LSTM-based MI prediction model for an industrial styrene-acrylonitrile (SAN) polymerization process in Korea. The developed model provides the most accurate predictions compared to other soft sensor models based on partial least squares (PLS), support vector machines (SVM), Gaussian process regression (GPR), and feedforward artificial neural network (ANN).

Keywords: Soft Sensor, Polymerization, Melt Index, Long Short-Term Memory

1. INTRODUCTION

Thermoplastic polymers such as polystyrene (PS), polyethylene (PE), and polypropylene (PP) are essential chemical materials that are widely used in various commercial products. The melt index (MI) of a thermoplastic polymer is a quality variable that indicates flow and mechanical properties of the polymer. Because the quality and grade of a polymer is directly affected by MI, MI is used to determine whether a product is out of specification or not. Therefore, it is necessary to monitor and control a polymerization process to keep the MI of the product within an acceptable range.

However, unlike process variables such as temperature and pressure are measured online, online measurement of MI is not available in industrial polymerization processes. Instead, an offline laboratory analysis for MI is available but it is costly and time-consuming. In most polymerization processes, MI is measured only once every two or four hours, which makes it very challenging to monitor the polymer quality. Therefore, there have been extensive studies to build soft sensor models to predict MI as quickly and accurately as possible.

There are two major challenges in modeling industrial chemical processes: strong nonlinearity and temporal correlations between variables. Since a number of chemical reactions and phase equilibria are involved in a chemical process, it is very time-consuming to build an accurate first-principle model that requires the prior knowledge of the process and mathematical equations, such as mass and energy balances. Instead, with the rapid growth of computer and sensor technology, data-driven modeling methods have been extensively studied and applied to industrial chemical processes.

In data-driven models, the modeling is based on historical data of easy-to-measure variables and their relationships to difficult-to-measure variables. The latent variable methods such as principal component regression (PCR) and partial least squares (PLS) are the most commonly used data-driven modeling methods (Kourti (2005), Kadlec et al. (2009), Han et al. (2005), Ahmed et al. (2013)). These methods reconstruct raw measurement data into a lower-dimensional space, and find the linear relationship between newly constructed variables. Additionally, to address the nonlinearity and temporal correlations, various extended versions of principal component analysis (PCA) and PLS have been developed and applied to chemical processes (Ku et al. (1995), Dong and Qin (2018)).

^{*} This research was financially supported by LG Chem, Ltd.

Recently, there has been a growing interest in predicting MI using nonlinear data-driven modeling approaches that are able to capture and model the nonlinearity of complex industrial polymerization processes. Han et al. (2005), Park et al. (2010), Liu et al. (2013), and Zhang and Liu (2013) developed soft sensor models based on support vector machines (SVMs) to predict MI of commercial polymerization processes. Furthermore, Gaussian process regression (GPR) models (Liu and Gao (2015), Chan and Chen (2017), Liu et al. (2017)) and artificial neural network (ANN) models (Liu and Zhao (2012), Li et al. (2012), Xu and Liu (2014)) were developed for industrial polymerization processes.

While the nonlinear modeling approaches mentioned above demonstrated accurate MI predictions, the prediction performance can be improved by further investigating and modeling temporal correlations and dynamic behavior of processes. Heo and Lee (2018) showed that the modeling performance of ANN was increased by augmenting input data with lagged data to account for process dynamics. Additionally, recurrent neural networks (RNNs) that have recurrent connections inside the layers have been developed and demonstrated improved performance for time series data of dynamic systems. However, RNNs have difficulties in modeling long sequences of data because of the vanishing gradient (Bengio et al. (1994)).

Thus, an improved version of RNN, long short-term memory (LSTM) network, was developed by addressing the long-term dependency problem and showed great performances in modeling and predicting sequential data such as speech recognition (Sundermeyer et al. (2012), Graves et al. (2013)) and video representation (Srivastava et al. (2015)). In recent years, there has been a growing effort to build a soft sensor model based on LSTM for industrial chemical processes such as a hydrocracking process (Yuan et al. (2020)) and a sulfur recovery unit (Ke et al. (2017)). In this work, we propose an LSTM-based soft sensor model for prediction of MI in an industrial styrene-acrylonitrile (SAN) polymerization process in Korea. The prediction performance of the proposed model is compared with other data-driven models, including PLS, SVM, ANN, and GPR.

2. METHODOLOGY

2.1 Recurrent Neural Network

RNN is a subset of neural networks and is mainly used to describe the temporal dynamic behaviors of dynamic

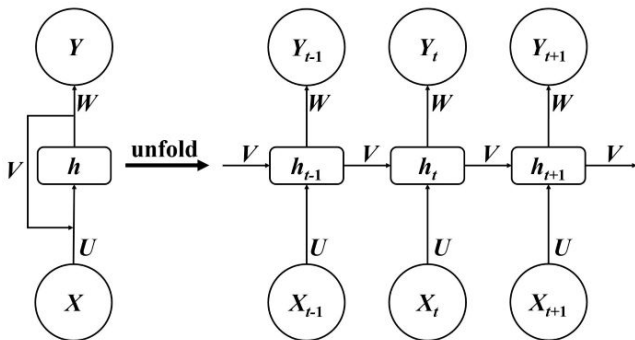


Fig. 1. Structure of a recurrent neural network

systems. Unlike traditional feedforward neural networks, RNN has recurrent connections inside its layers. Thus, the hidden state h_t at time t is determined by the input X_t and the former hidden state h_{t-1} . Fig. 1 shows the basic structure of RNN where the left structure can be unfolded into the right structure. The hidden state h_t and output vector Y_t at time t are calculated as

$$h_t = \sigma_h(UX_t + Vh_{t-1} + b) \quad (1)$$

$$Y_t = \sigma_Y(W h_t + c) \quad (2)$$

where σ_h and σ_Y are activation functions for the hidden and output layer; b and c are the corresponding bias terms; U , V , and W are the corresponding weight matrices. However, it is difficult for the standard RNN to learn from long sequences because of the gradient vanishing problem (Pascanu et al. (2013)).

2.2 Long Short-Term Memory Network

LSTM is developed as an improved version of the standard RNN by addressing the vanishing gradient problem. While the standard RNN has only one connection between the former and present hidden states, the basic LSTM unit has three gates; the input, forget, and output gate. An LSTM unit calculates the cell state, c_t , and hidden state, h_t , at time t from the input vector X_t and previous states, c_{t-1} and h_{t-1} through the three gates.

The input gate i_t and output gate o_t determine what information from the input X_t and the previous hidden state h_{t-1} should be remembered and passed into the new cell state c_t and hidden state h_t , respectively. The forget gate f_t determines how much information from the previous cell state should be remembered to calculate new states.

Fig. 2 shows the structure of a basic LSTM unit. The three gates are calculated as

$$f_t = \sigma(W_{fx}X_t + W_{fh}h_{t-1} + b_f) \quad (3)$$

$$i_t = \sigma(W_{ix}X_t + W_{ih}h_{t-1} + b_i) \quad (4)$$

$$o_t = \sigma(W_{ox}X_t + W_{oh}h_{t-1} + b_o) \quad (5)$$

where W and b represent the corresponding weight and bias matrices. Then the cell and hidden state are updated as

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_{cx}X_t + W_{ch}h_{t-1} + b_c) \quad (6)$$

$$h_t = o_t \odot \tanh(c_t) \quad (7)$$

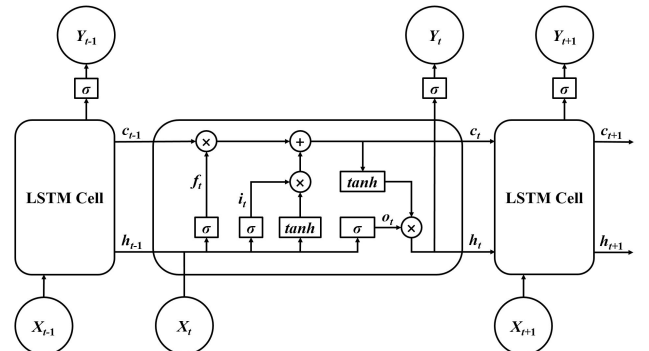


Fig. 2. Structure of a basic LSTM unit

where \odot represents element-wise multiplication. Finally, the output vector from the LSTM unit is calculated as

$$Y_t = \sigma(W_y h_t + b_y) \quad (8)$$

3. SOFT SENSOR FOR MELT INDEX PREDICTION

In this paper, an LSTM-based soft sensor model for MI prediction in an industrial multiple-grade SAN polymerization process is developed and its prediction performance is compared with other soft sensor models based on PLS, SVM, GPR, and ANN.

3.1 Process Description

The SAN polymerization process of interest is currently operated in Korea. Fig. 3 depicts the simple schematic diagram of the target process. The process consists of two continuous reactors in series, followed by two devolatilizers and a pelletizer. Styrene, acrylonitrile, α -methylstyrene, and the initiator are mixed with monomers and solvents recovered from the devolatilizers, then fed into the first reactor. The polymerization reaction occurs in a liquid phase achieving monomer conversions of approximately 40 and 65 percent in the first and second reactor, respectively. The reaction mixture is then dried in the devolatilizers from which the residual monomers and solvents are vaporized and recovered. Finally, the polymer products are pelletized in the pelletizer and transported to be stored in silos. As a result, two distinct grades of the SAN polymer products with different values of MI are produced. Operating conditions such as reaction temperature and monomer flow rate vary from grade to grade.

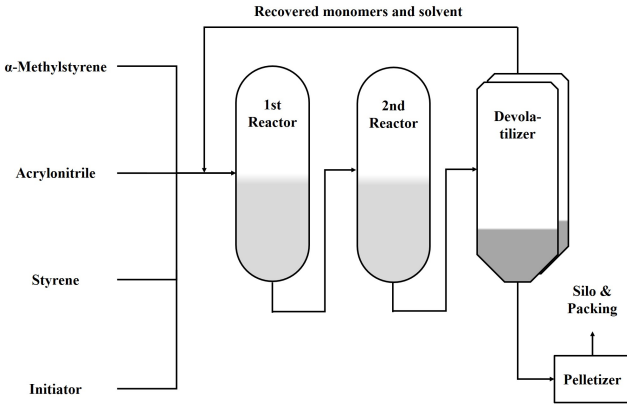


Fig. 3. Schematic of SAN polymerization process

3.2 Dataset Preparation

There are two types of variables in measurement data obtained from the SAN polymerization process. First, the process variables, a total of 27 variables including the temperatures and pressures of two reactors, are measured online every hour by the sensors in the process. The quality variables such as MI and color of the polymer product, the other type of the measurement variables, are measured less frequently than the process variables and measured with an offline laboratory analysis.

The data were collected for a period of approximately seventeen months from January 2, 2020 to May 23, 2021

including approximately four months of process shutdown and eight grade changeovers. A total of 2285 labeled quality measurement samples are available. 80 percent of the samples are selected as the training dataset. The remaining 20 percent of the samples are divided equally into the validation and testing dataset.

Before training soft sensor models, the historical data are preprocessed as follows. First, both the process measurements X and MI measurements y are normalized to z-scores \tilde{X} and \tilde{y} which have zero mean and unit variance as in (9).

$$z = \frac{x - \bar{x}}{\sigma_x} \quad (9)$$

where \bar{x} and σ_x are the mean and standard deviation of the variable x , respectively.

PCA is then applied to transform the normalized process variables \tilde{X} to linearly independent variables to remove the multicollinearity between the process variables and reduce computational load. PCA is a statistical dimension reduction technique that is widely used in process monitoring and modeling. PCA finds the latent variables called principal components (PCs) such that the variability of PCs is maximized by the orthogonal linear transformation of raw data. As shown in Fig. 4, the first 10 PCs explain approximately 94 percent of the total variance in raw process measurements. Therefore, only 10 PCs with the highest variances are used as model inputs rather than training a model with all 27 PCs, which reduces the computational requirements while retaining as much information in raw data as possible.

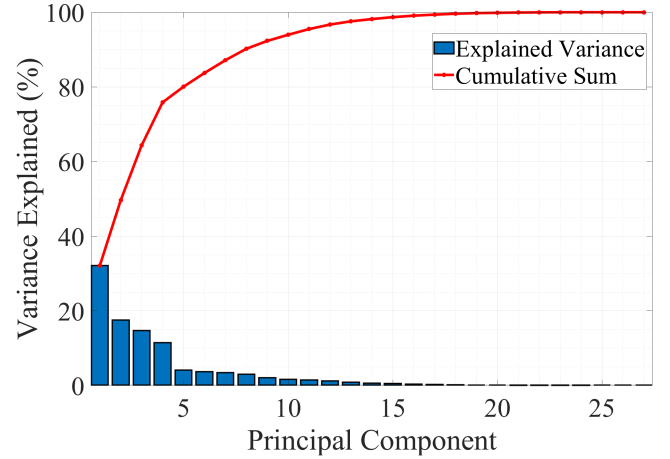


Fig. 4. Scree plot of variances explained by PCs

3.3 LSTM-Based Soft Sensor Modeling

The network of the proposed LSTM-based soft sensor is designed as follows. First, the sequence data of the process variables is preprocessed as in the previous subsection into a new sequence consisting of PCs that are linearly independent to each other. It is then sequentially fed to the input layer that is the first layer of the network. The last layer of the network is a fully-connected layer with a linear activation function whose output is \hat{y}_k , which is the predicted value of MI at time t_k .

The activation functions of the input, forget and output gate of the LSTM layers are sigmoid functions. For training of the LSTM network, the ADAM optimizer with the back propagation through time (BPTT) method is used because the ADAM optimizer has advantages in computational efficiency and memory requirements over stochastic gradient descent algorithm and the root mean square propagation (RMSProp) algorithm (Kingma and Ba (2014)). Before the training starts, the input and recurrent weights of the LSTM layers are initialized with Glorot initialization and orthogonal initialization, respectively. The learning rate is initially set as 0.1 and drops by a factor of 0.99 after every 10 epoch of training. The loss function for model training is the mean square error (MSE) that is calculated as

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (10)$$

where n is the number of samples, y_i and \hat{y}_i are the measured and predicted value of MI, respectively. To avoid over-fitting, MSE of the trained model on validation dataset is monitored every 10 epoch. When the validation loss ceases to decrease, the training stops and the early stop model with the lowest validation loss is obtained.

4. RESULTS AND DISCUSSIONS

An LSTM model with three hidden layers of 10 neurons was trained for MI prediction in the SAN polymerization process described in Section 3.1. Since an LSTM network takes a sequence as its input, MI is predicted using the sequence consisting of four hours of process measurement data. As shown in Fig. 5, the LSTM model was trained for 240 epochs. Both the training and validation loss decreased rapidly at the beginning of the training. The validation loss reached the lowest value at the 140-th epoch, after which the validation loss ceased to decrease while the training loss continued to decrease. Therefore, the early stop model with the lowest validation loss and better generalization performance was finally obtained in order to avoid over-fitting.

For the comparison of the prediction performance, soft sensor models based on PLS, SVM, GPR, and ANN were trained. To account for dynamic behavior of the process, the input data was augmented with four hours of time lagged data. Thus, a total of 108 input variables were used for soft sensor modeling. PLS is a traditional data-based modeling method which aims to find the linear combination of the process variables that maximizes covariance

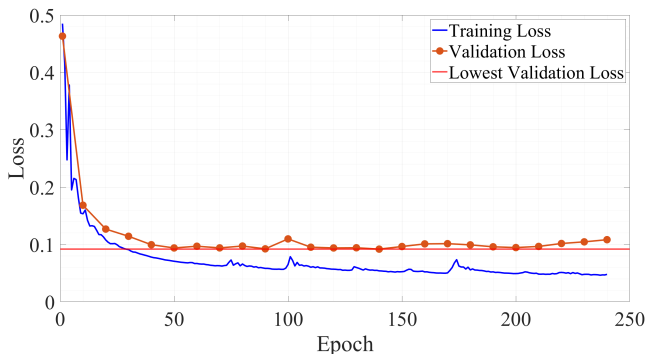


Fig. 5. Training procedure of the LSTM model

between the process variables and MI. The PLS model was trained using the SIMPLS algorithm developed by De Jong (1993). The PCA preprocessing step was omitted in PLS modeling. Instead, MI was predicted using 27 PLS components. The nonlinear SVM model was trained by solving a convex quadratic optimization problem constructed with Gaussian kernel function. The kernel function for the GPR model was an exponential kernel and quasi-Newton optimizer was used in training. The network structure of the ANN model was a feedforward multilayer perceptron with 3 hidden fully-connected layers of 10 neurons, which is the same as the number of hidden layers and neurons of the LSTM model. The ANN model was trained using the ADAM optimizer and the early stop model with the lowest validation loss was obtained.

The prediction performance of a soft sensor model was evaluated with four statistical indices; root mean square error (RMSE), coefficient of determination (R^2), mean absolute percentage error (MAPE), and Theil's inequality coefficient (TIC). These performance indices are defined as follows:

$$\text{RMSE} = \sqrt{\text{MSE}} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (11)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (12)$$

$$\text{MAPE} = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (13)$$

$$\text{TIC} = \frac{\sqrt{\sum_{i=1}^n (y_i - \hat{y}_i)^2}}{\sqrt{\sum_{i=1}^n y_i^2} + \sqrt{\sum_{i=1}^n \hat{y}_i^2}} \quad (14)$$

where $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$. The RMSE and MAPE are measures of the prediction accuracy of a soft sensor model, while the R^2 indicates the reliability of a model. The TIC indicates how well the predicted sequence matches to the measured sequence. For two identical sequences, the TIC value equals zero.

Table 1. MI prediction results of all models on the testing dataset

Method	RMSE	R^2	MAPE	TIC
PLS	0.3361	0.7144	2.749	0.01792
SVM	0.3049	0.7649	2.410	0.01627
ANN	0.2666	0.8204	2.275	0.01421
GPR	0.2653	0.8221	2.200	0.01414
LSTM	0.2583	0.8314	2.179	0.01379

Table 1 summarizes the prediction performance of the soft sensor models on testing dataset. Furthermore, the predicted values of MI of the PLS, SVM, ANN, GPR, and LSTM model on the testing dataset are shown in Fig. 6. As illustrated in Table 1 and Fig. 6, the LSTM model shows the best prediction performance among all soft sensor models. On the other hand, the PLS model shows the worst prediction performance. Since PLS is unable to explain the nonlinear relationships existent in complex chemical processes, there are large deviations between the predicted and measured MI values than other models. Although the nonlinear SVM model shows better prediction performance than the PLS model, there are still large deviations

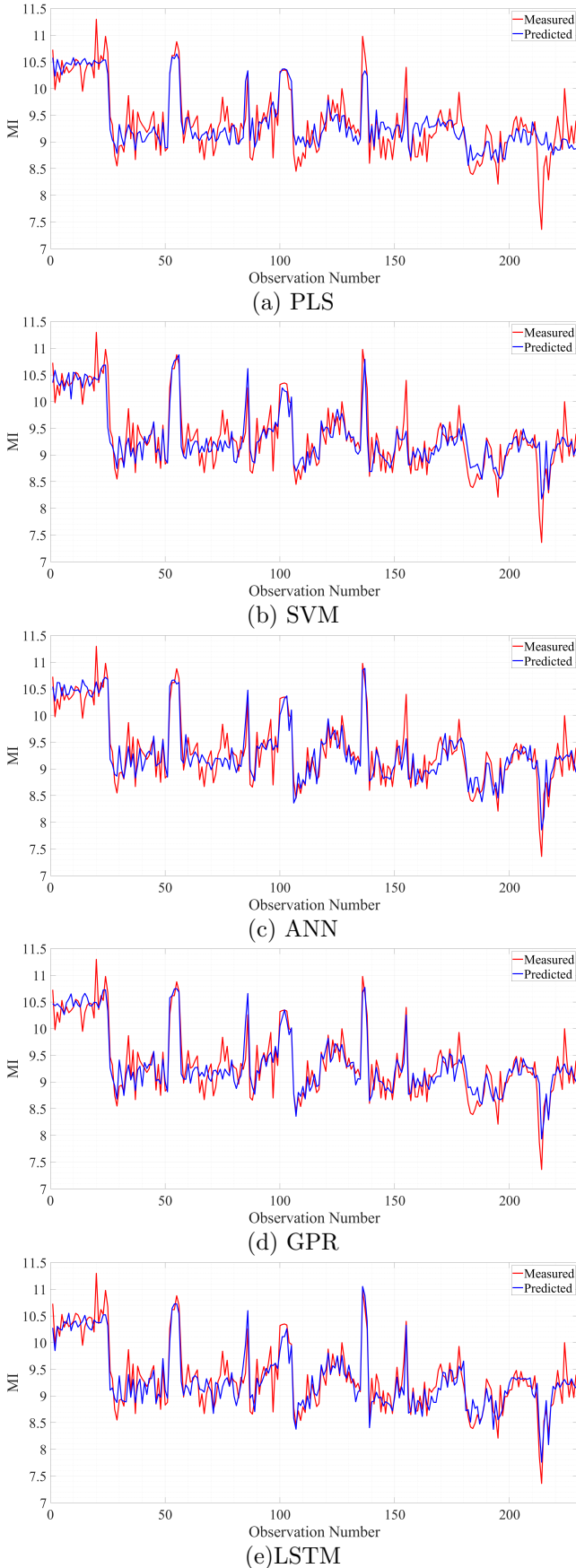


Fig. 6. MI prediction results of all models: (a) PLS (b) SVM (c) ANN (d) GPR (e) LSTM

between the predicted and measured MI values when the process is in grade change-over and unsteady state. The LSTM model captures the nonlinearity as well as the long-term dynamics of the target process with its recurrent network. As a result, the LSTM model outperformed other soft sensor models in all 4 statistical performance indices. As shown in Fig. 6, the predicted values of MI of the LSTM model show better matches with the measured values than other models, even when MI changes rapidly. For instance, the monomer and initiator flow rates fluctuated abruptly between the testing data of sample number 212 - 214, which resulted in sharp decrease of MI. Additionally, the LSTM model shows good prediction performance during grade changeovers where MI changes from one range to another in a short time.

Table 2. CPU time required for model training

Method	Training epoch	CPU time (s)	CPU time per epoch (s)
PLS	-	0.021	-
SVM	-	0.246	-
ANN	300	13.227	0.044
GPR	-	4.829	-
LSTM	240	21.879	0.091

Table 2 summarizes the required CPU time for training each model. The CPU time was measured on a Intel Core i7-8700 CPU @ 3.20 GHz. The PLS and SVM models required less than one second for training. On the other hand, training of the ANN and LSTM models required 13.227 and 21.879 seconds, which was much longer than the other three models. Additionally, because an LSTM network consists of more training parameters than a feed-forward neural network, the LSTM model took longer CPU time per epoch than the ANN model. Although both the LSTM and ANN models consist of three hidden layers of 10 neurons, a total of 2531 parameters were trained for the LSTM model while only 641 parameters were required for training the ANN model. Additionally, A large number of parameters makes training of an LSTM network more sensitive to parameter initialization.

In short, the proposed LSTM-based soft sensor model outperformed other machine learning models in predicting MI of polymer products from the industrial SAN process. To address the high computation load required for training an LSTM network, PCA was applied in data preprocessing step to reduce the dimension of the data.

5. CONCLUSION

In this paper, we developed an LSTM-based soft sensor model for MI prediction in an industrial polymerization process and compared its prediction performance with other black-box soft sensor models based on PLS, SVM, GPR, and ANN. The soft sensor models were trained and tested with the measurement data obtained from a commercial SAN polymerization process in Korea. The measurement data was normalized and then preprocessed by applying PCA to remove multicollinearity and reduce computational load. With the ability to capture the nonlinearity and long-term dynamic behavior of the process, the LSTM model accomplished the best prediction performance among all soft sensor models. The LSTM model

predicts MI accurately even when the process is in abrupt changes while other soft sensor models show large deviations between the predicted and measured MI values. The results suggest that an LSTM network is more effective in modeling complex chemical processes than traditional black-box modeling methods.

For further research, there are several suggestions that aid in more accurate prediction of MI using LSTM network. First, the improvement of the prediction performance can be achieved by the optimization of the network structure and hyperparameters. Second, the modeling performance of a black-box model is only guaranteed for the data in the training region. Thus, the chemical and thermodynamic knowledge of polymerization processes can be combined to build a hybrid soft sensor model to improve generalization ability.

ACKNOWLEDGEMENTS

This research was financially supported by LG Chem, Ltd.

REFERENCES

- Ahmed, F., Kim, L., and Yeo, Y. (2013). 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*, 30(1), 11-19.
- Bengio, Y., Simard, P., and Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks*, 5(2), 157-166.
- Chan, L., and Chen, J. (2017). Melt index prediction with a mixture of Gaussian process regression with embedded clustering and variable selections. *Journal of applied polymer science*, 134(40), 45237.
- De Jong, S. (1993). SIMPLS: an alternative approach to partial least squares regression. *Chemometrics and intelligent laboratory systems*, 18(3), 251-263.
- Dong, Y., and Qin, S. J. (2018). Regression on dynamic PLS structures for supervised learning of dynamic data. *Journal of process control*, 68(1), 64-72.
- Graves, A., Jaitly, N., and Mohamed, A. (2013). Hybrid speech recognition with deep bidirectional LSTM. In *2013 IEEE workshop on automatic speech recognition and understanding*, IEEE, 273-278.
- Han, I., Han, C., and Chung, C. (2005). Melt index modeling with support vector machines, partial least squares, and artificial neural networks. *Journal of applied polymer science*, 95(4), 967-974.
- Heo, S., and Lee, J. (2018). Fault detection and classification using artificial neural networks. *IFAC-PapersOnLine*, 51(18), 470-475.
- Kadlec, P., Gabrys, B., and Strandt, S. (2009). Data-driven soft sensors in the process industry. *Computers and chemical engineering*, 33(4), 795-814.
- Ke, W., Huang, D., Yang, F., and Jiang, Y. (2017). Soft sensor development and applications based on LSTM in deep neural networks. In *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, IEEE, 1-6.
- Kingma, D., and Ba, J. (2014). ADAM: a method for stochastic optimization. *arXiv preprint*, arXiv:1412.6980.
- Kourti, T. (2005). Application of latent variable methods to process control and multivariate statistical process control in industry. *International journal of adaptive control and signal processing*, 19(4), 213-246.
- Ku, W., Storer, R. H., and Georgakis, C. (1995). Disturbance detection and isolation by dynamic principal component analysis. *Chemometrics and intelligent laboratory systems*, 30(1), 179-196.
- Li, J., Liu, X., Jiang, H., and Xiao, Y. (2012). Melt index prediction by adaptively aggregated RBF neural networks trained with novel ACO algorithm. *Journal of applied polymer science*, 125(5), 943-951.
- Liu, X., and Zhao, C. (2012). Melt index prediction based on fuzzy neural networks and PSO algorithm with online correction strategy. *AIChE journal*, 58(4), 1194-1202.
- Liu, Y., Gao, Z., and Chen, J. (2013). Development of soft-sensors for online quality prediction of sequential-reactor-multi-grade industrial processes. *Chemical engineering science*, 102, 602-612.
- Liu, Y., and Gao, Z. (2015). Industrial melt index prediction with the ensemble anti-outlier just-in-time Gaussian process regression modeling method. *Journal of applied polymer science*, 132(22), 41958.
- Liu, Y., Liang, Y., and Gao, Z. (2017). Industrial polyethylene melt index prediction using ensemble manifold learning-based local model. *Journal of applied polymer science*, 134(29), 45094.
- Park, T., Kim, T., and Yeo, Y. (2010). 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*, 27(6), 1662-1668.
- Pascanu, R., Mikolov, T., and Bengio, Y. (2013). On the difficulty of training recurrent neural networks. In *International conference on machine learning*, PMLR, 1310-1318.
- Srivastava, N., Mansimov, E., and Salakhutdinov, R. (2015). Unsupervised learning of video representations using LSTMs. In *International conference on machine learning*, PMLR, 843-852.
- Sundermeyer, M., Schlüter, R., and Ney, H. (2012). LSTM neural networks for language modeling. In *Thirteenth annual conference of the international speech communication association*.
- Xu, S., and Liu, X. (2014). Melt index prediction by fuzzy functions with dynamic fuzzy neural networks. *Neurocomputing*, 142, 291-298.
- Yuan, X., Li, L., Shardt, Y., Wang, Y., and Yang, C. (2020). Deep learning with spatiotemporal attention-based LSTM for industrial soft sensor model development. *IEEE Transactions on industrial electronics*, 68(5), 4404-4414.
- Zhang, M., and Liu, X. (2013). Melt index prediction by fuzzy functions and weighted least squares support vector machines. *Chemical engineering and technology*, 36(9), 1577-1584.