Final Prediction of Product Quality in Batch Process based on Bidirectional Neural Network Algorithm

Wanshui Li1,*, Xiang Wang2, Qianyuan Feng3

1University College London
2School of Science, Xi’an Jiaotong-Liverpool University, Suzhou, Jiangsu, 215000
3School of Automation, Shenyang Aerospace University, Shenyang, Liaoning, 110135

*Corresponding author: wanshuili541728@foxmail.com

Abstract. Based on the analysis of time series characteristics of production process based on common batch process end point quality prediction a predictive model based on bidirectional gated loop neural network is proposed to predict final product quality for unequal interval batch processes. Based on the requirement of forecasting value in actual production, loss function adapted to batch process is constructed, which makes model meet forecasting requirement under guarantee prediction precision, thus obtaining greater production benefit. Two-way gated loop (GRU) neural networks with different loss functions are compared with multi-directional partial least squares (MPLS) neural networks support vector regression SVR and gated cyclic unit neural networks. Results show that bidirectional gated loop neural networks have stronger applicability and higher accuracy.

Keywords: Depth learning; intermittent process; bidirectional cyclic neural networks; loss function; time series.

1. Introduction

As an important production method, batch process is mainly used to produce low output and high added value products. However, due to the uncertainty of raw material composition, frequent change of product varieties and complex technological process, the product quality of batch process fluctuates greatly [1-2]. On the one hand, due to the lack of on-line sensors or on-line measurement of product quality indicators in most batch production processes, analytical values can only be obtained through laboratory analysis, which is difficult to meet the needs of real-time on-line control of quality indicators. On the other hand, it is difficult to get the mechanism model of batch process, so the data-based model has become a research hotspot [3-4].

Recurrent Neural networks (RNN) have shown a greater adaptability in sequential data analysis through the introduction of temporal concepts into the design of Network structures, providing better solutions for extracting time series features. With the increase of the processing time series length, the use of conventional activation function will cause the gradient disappearance during network training, which will lead to the deficiency of RNN network prediction accuracy. However, its variants, such as Long ShortTerm Memory (LSTM), gated loop unit (GRU) and bidirectional operation of the model, make up for the deficiency of RNN by adding some threshold gates. These methods have been widely
used in multiple sequential learning problems. In industry, paper [5] proposed a framework of deeply heterogeneous GRU model for tool wear prediction. Paper [6] used the automatic encoder method based on bidirectional circulation neural network to predict the life of turbofan engine. Paper [7] used LSTM to extract features of the long time series of each stage in the intermittent process, so as to obtain the quality-related comprehensive hidden features. To solve the above problems, the loss function of BiGRU is improved and used to predict the quality index of batch process. The bidirectional circulation neural network can integrate the sequential information of batch data before and after and fully exploit the sequential characteristics of batch to batch caused by the uncertainty of raw materials. By imposing different penalties on different predicted values, the improved loss function makes the predicted values meet the requirements of actual production and improves the practicability of the algorithm.

2. BiGRU based batch process final product quality prediction

2.1. Data representation of intermittent processes
Batch processes for repeat production, compared with pure time-series data, the data also contains a batch process in batch, thus general with 3 d data $X$, said matrix $Y$ said the final product quality variables, which I said batch batch processes, $J$ said the dimensions of the process variables, $K$ said hits every batch processes, as shown in figure 1.

2.2. Circulating neural network
The advantage of cyclic neural network lies in the processing of sequential data, which is embodied in the fact that the output of the current moment in the network depends on the network's memory of previous information. In other words, the input of the hidden layer is no longer the output of the input layer at the current moment, but also the output of the hidden layer at the previous moment. There are three main structures of common circulating neural networks, including two many-to-many structures and one many-to-one structure, as shown in figure. 2.

Fig.1 Data in batch process

Fig.2 RNN structure
There are two kinds of many-to-many structures in Figure 2. The first type requires the same length of input sequence and output sequence, so the application scope of this structure is relatively narrow. In the second structure, the input and output are unequal sequences, and it is an Encoder-decoder structure, which is often used in machine translation, speech recognition and other fields. However, many-to-one structure is often used to deal with sequence classification or regression problems, and only one output result is needed in the end. Therefore, in this paper, a many-to-one cyclic neural network structure is selected, and the specific network prediction structure is shown in Figure 3.

![Fig.3 RNN network prediction structure diagram](image)

RNN's single-layer network structure and network parameter solving algorithm based on BPTT (Back Propagation Through Time) determine that gradient disappearance or gradient explosion will occur during the training process of RNN, so it is difficult to learn network parameters and long-term dependence.

### 2.3. Gated circulation unit

Gated circulation unit was proposed by Cho. in 2014 and applied in the field of machine translation. Similar to LSTM, GRU reduces gradient dispersion by introducing a special gate structure, so that errors can be propagated over a long distance, and thus it has the ability of long-term memory. The difference is that GRU reduces the LSTM gate structure to update gate and reset gate. The model is simpler, with fewer network parameters and faster convergence. The update gate in GRU replaces the forgetting gate and output gate in LSTM to control which information in the current information needs to flow into the candidate state. The specific formula of renewal gate is shown in Equation (1).

\[ z_t = \text{sigmoid}(W_z[h_{t-1}, x_t] + b_z) \]  \hspace{1cm} (1)

The reset gate determines the effect of the \( h_{T-1} \) output from the hidden layer unit on the candidate state at the previous time. The specific formula is shown in Equation (2).

\[ r_t = \text{sigmoid}(W_r[h_{t-1}, x_t] + b_r) \]  \hspace{1cm} (2)

### 2.4. Bidirectional circulation neural network

Schuster proposed bidirectional circulation neural network (BiRNN) in 1997. Compared with the ordinary one-way RNN information which can only be propagated in the positive time direction, BiRNN has two sets of hidden layer information, one is the input sequence in the positive time direction, the other is the input sequence in the negative time direction. Therefore, BiRNN can better
capture information in time series and is widely used in speech recognition, machine translation, emotion classification and other aspects. The specific structure is shown in Figure 4.

![Fig.4 Structure of BiRNN](image)

Tradational algorithms are widely used in the establishment of intra-batch models. Meanwhile, the traditional inter-batch model cannot fully mine the time-series characteristic information between batches, while RNN performs well in the time-series data modeling. Considering the problems of long relied on RNN, while LSTM and 6 Locke nan, such as: two-way door control cycle unit of the neural network based on the intermittent process of the final product quality can predict 827 GRU helped make up for this problem, GRU helped compared with less LSTM network parameters at the same time, to facilitate the training of the model, this article USES the BiGRU, as the main part of the prediction model, which can extract the feature information, improve the precision of the model.

3. Case Analysis

3.1. Experimental data set
Resin production is a multi-stage batch process, which mainly consists of two stages, polycondensation and drying and dehydration. Each stage has specific control objectives, different dominant variables and process characteristics. In the condensation polymerization stage, phenols, aldehydes and catalysts were added to the reactor in turn for condensation polymerization. After a period of time, they were switched to the drying and dehydration stage. In the drying and dehydration stage, vacuum is built in the reaction kettle. The resin is dehydrated under a certain degree of vacuum. After a certain period of time, the vacuum is broken and the material is taken, and the production process ends. Phenols as reactants are themselves a mixture, stored in the tundish, with the passage of time, the composition of the inside also slowly change, specific can be analyzed by chromatography, but generally not analyzed, even if the analysis, it is difficult to analyze all the substances. Therefore, the uncertainty of raw material composition brings great difficulty to accurately predict the quality of the final product in the process. Meanwhile, such uncertainty is accompanied by the change of time, that is, the process batch itself has certain timing. In this paper, the softening point of the resin was predicted based on the production process of the resin.

3.2. Data acquisition and processing
Data of process variables in 400 batches of production were collected, including the flow rate of phenols and aldehydes, as well as the pressure, temperature and quality of the reactor. The resin softening point measured off-line in the laboratory was used as the predictive variable. Because the operation process of the resin is usually fixed, according to the operating conditions for different operation stages in the process of production, quality and pressure of reaction kettle temperature, start and end values differ long process of the original data for each batch of data extraction process, and according to the material flow for each batch of the quality of the reactants. According to the maximum information coefficient, the characteristics of the process data and the reactant mass were selected, and
the four variables with the strongest correlation with the resin softening point were selected, namely, the addition amount of aldehyde, the temperature value at the beginning of the reaction, the temperature value at the end of the reaction and the vacuum degree, and the input variable $X_{400}^{4}$ of the model was obtained. In order to avoid the influence of the dimension of each variable on the model prediction results, the maximum and minimum value method is used to normalize the samples.

4. Establishment of the model

Using the resin production process data, the model parameters were set by the method of program self-cycle optimization, and the main parameters included the size of time window, network structure and training times. By means of the limited variable method, the parameter was changed one at a time and the RMSE value was monitored, and the optimal parameter was determined by comparative analysis. In order to reduce the influence of randomness in model training, the average value was obtained after repeated operation for 10 times. Since BiGRU neural network needs to input time series data, the data of 400 batches is continuously sampled with sliding window as 5. The quality index corresponding to the last batch of data of each sliding window is taken as the predicted value of the model, and 396 samples are obtained.

In order to better mine the uncertainty existing in the data, the first 350 samples are the training data set, and the last 46 samples are the test data set. The final prediction model structure of this paper is shown in Figure 5. The number of hidden layer neurons of the BiGRU neural network was set as 80, and the network parameters were trained with Adam optimizer. The EPOCH was 100, and the batch_size was 20.

![Fig.5 Model structure](image)

5. Analysis of results

As can be seen from Figure 6, MPLS had the worst prediction effect, while THE prediction results of SVR and NN were similar, and the prediction results of some batches had a large deviation, which was slightly worse than the GRU model. GRU helped the model prediction effect than the traditional algorithm has greatly promoted, instructions on the raw material of intermittent process uncertainty feature extraction, GRU helped with more 6 Locke nan, such as: two-way door control cycle unit of the neural network based on batch processes the end product quality forecast 829 strong memory, dig to the temporal aspects of characteristics, also shows that the intermittent process of the final product quality index has closely associated with the change of raw material composition. The GRU fitting degree is generally good, but the performance of some batches is not ideal. However, BiGRU comprehensively considers the forward timing sequence characteristics and the reverse timing sequence characteristics to obtain more in-depth global characteristics, and the prediction results are better than GRU.
Fig. 6 Softening point prediction results of six algorithms

6. Conclusion
In order to solve the problem of batch process modeling with uncertain raw materials by traditional algorithms, a product quality prediction method based on BiGRU model for batch process (Vol. 46, Journal of East China University of Science and Technology (Natural Science edition) was proposed in this paper, and the method was verified in the quality prediction of softening point of certain resin. Compared with MPLS, NN, SVR and GRU algorithms, the BiGRU model-based batch process product quality prediction method achieves better prediction results than the traditional algorithm, which verifies that the bidirectional gated circulation unit neural network has better prediction ability for industrial batch process data. In order to improve system capacity rate and predict the results of practicability, this paper adopts BiGRU - ALF model, based on the predicted results of different deviation, different penalty value, makes the forecast results have greater safety margin, for batch process modeling research provides a new train of thought, also for field application of deep learning in batch processes provides some guidance.

Conferences
[1] CORBETT B, MHASKAR P. Subspace identification for data-driven modeling and quality control of batch processes [J]. AIChE Journal, 2016, 62(5): 1581-1601.
[2] WANG D. Robust data-driven modeling approach for realtime final product quality prediction in batch process operation [J]. IEEE Transactions on Industrial Informatics, 2011, 7(2): 371-377.
[3] ZHANG Q, MA Y, GU M, et al. End-to-end Chinese dialects identification in short utterances using CNNBiGRU[C]//2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC). USA: IEEE, 2019: 340-344.
[4] ZHAO Z, LIANG B, WANG X, et al. Remaining useful life prediction of aircraft engine based on degradation pattern learning [J]. Reliability Engineering System Safety, 2017, 164: 74-83.
[5] NOMIKOS P, MACGREGOR J F. Multi-way partial least squares in monitoring batch processes [J]. Chemometrics and Intelligent Laboratory Systems, 1995, 30(1): 97-108.
[6] WANG J, YAN J, LI C, et al. Deep heterogeneous GRU model for predictive analytics in smart manufacturing: Application to tool wear prediction [J]. Computers in Industry, 2019, 111: 1-14.
[7] JIN H, CHEN X, YANG J, et al. Multi-model adaptive soft sensor modeling method using local learning and online support vector regression for nonlinear time-variant batch processes [J]. Chemical Engineering Science, 2015, 131: 282-303.