EnPress Publisher

Prediction of growth trend and application of Synechococcus PCC7002 in industrial culture based on MTS-Mixers

Sichao Hu

Hunan University of Science and Technology, Xiangtan 411100, China.

Abstract: Industrial closed culture of Marine microalgae requires higher environmental parameters in the whole process.Synechococcus PCC7002 was selected as the culture object in the pilot stage of the culture, and a variety of parameters in the culture process were collected and a data set was established. A growth prediction model of synechococcus PCC7002 based on LSTM neural network was constructed. Multiple environmental parameters, such as temperature, pH, light intensity, air input and dissolved oxygen value, were used as input parameters of the model. After repeated training and learning, turbidity value (i.e., cell concentration) of the growth state of microalgae was obtained. The turbidity value of quantifiable microalgae growth conditions and the growth trend of PCC7002 were obtained after the coupled training of the main environmental factors in the microalgae culture process. Then, the MTS-Mixers algorithm was integrated to predict the external environment prediction curve required for the growth of microalgae in the stable and exponential stages in the culture process.

Keywords: LSTM; MTS-Mixers; Industrial training; Synechococcus PCC7002

1.Introduction

With the explosion of population, the problem of food shortage has been difficult to solve effectively by traditional industrial and agricultural means. In recent years, people have turned their research direction to biotechnology, hoping to develop organisms with rich nutrient content and short culture cycle into one of the raw materials for nutrient intake through biotechnology means. Due to the advantages of fast splitting speed, high photosynthetic efficiency and short production cycle, microalgae organisms have shown wide application prospects and huge demand in animal feed and food industry. Among them, Synechococcus sp. strain PCC7002 and other algae have the potential to act as nutritive substitute materials due to their high protein, low carbohydrate water and low fat characteristics, which have gradually attracted the attention of scientists.

Microalgae are widely distributed in nature, it is difficult to achieve enrichment, and it is not worth the loss to harvest microalgae raw materials by artificial fishing. The photobioreactor is the core equipment for the industrialization and large-scale cultivation of microalgae. At the same time, the growth process of microalgae can be divided into delayed period, exponential period, stable period and decline period according to the growth characteristics. The lag period is characterized by slow growth of cell number; The exponential stage is characterized by the close relationship between the number of viable bacteria and the total number of bacteria, the shortest generation time of cell division, and the largest growth rate, which is the key stage of microalgae culture. The stable period is characterized by the growth of microalgae has reached the peak, which is the best period to harvest products. The decline stage was characterized by a very low proportion of the live bacteria in the total bacteria number, and a large area of microalgae died. In the whole culture process, the lag period and exponential period have direct influence on the output. When the culture environment is suitable for the growth of microalgae, the lag period is shortened, and the exponential cell generation is shortened and the growth rate is accelerated. According to the growth process of microalgae and the demand of industrial culture, a photobioreactor which can be intelligently controlled was designed to improve the culture efficiency.

With the rapid development of deep learning, some neural networks can predict the trend of data within a certain period of time according to the demand because of their nonlinear fitting and ultra-high training operation ability. Based on Long short-term memory (LSTM) and multiple time series prediction model (MTS-Mixers), this paper predicted and analyzed the cultivation process of polycoccus PCC7002 in the pilot stage, and obtained the optimal curve of various environmental parameters suitable for polycoccus cultivation in this equipment. LSTM is a kind of temporal Recurrent Neural Network, which comes from solving the long-term dependence problem of recurrent neural network (RNN). Compared with RNN, which only has a hidden state ht(hidden state),LSTM adds a passing state and a cell state ct(cell state) in the optimization process. ct acts as a “processor” to determine whether the information is useful. Adam optimization algorithm effectively solves the problem of gradient descent. In this paper, LSTM can be introduced to solve the problem of predicting the growth trend of microalgae by environmental variables during the culture process. MTS-Mixers is a general framework for multivariate time series prediction proposed by Huawei in 2023. It is not only used to improve the performance of time series prediction by combining attention mechanism, but also to solve the problem of the lack of explanation of attention mechanism in the actual use of Transformer, which is used to capture the long-term correlation of time series prediction. It also solves the problem that the data in the process of acquisition is affected by the sampling frequency and the sensor.

In the process of microalgae culture, a large amount of data will be generated, mainly including the temperature, pH, dissolved oxygen value, light and turbidity of the algae solution in the culture environment, and the turbidity in the reactor. The process environmental parameter database after successful culture of polycoccus can be established, and the model can be trained and established to predict the growth trend of polycoccus with high accuracy. The department confirmed the scientific nature of the culture process with LSTM. However, due to the real-time monitoring and control of the temperature, pH, light intensity and dissolved oxygen value inside the reactor in a wide range during the culture process, the synechococcus is not in the best growth environment at every moment due to its biological characteristics such as growth, division and cell activity. Therefore, the optimization curves of various environmental factors in the stable period and exponential period of polychlococcus in the process of polychlococcus culture can be predicted by the turbidity change of polychlococcus in turn through the fusion of MTS-Mixers.

In view of the above studies, a microalgae growth prediction model was proposed in this paper in combination with the corresponding deep learning framework and the data set of the successful cultivation of PCC7002 synechococcus, and the optimized environmental parameters of the synechococcus in the stable and exponential stages of the culture process were predicted by the changes in the cell concentration of the synechococcus. First, the environmental data of a single successful culture of synechococcus were collected, including temperature, pH, light intensity, air input and dissolved oxygen value, and the culture process was repeated several times to obtain the data set. Secondly, a multivariable temporal recurrent neural network prediction model with five environmental parameters as input parameters was established to predict the growth of synechococcus in the short term. Finally, the fusion MTS-Mixers algorithm predicted the external environment prediction curve of polycoccus in the stable and exponential stages of the culture process through the cell concentration changes, providing technical support for improving the success rate and yield of industrial cultured microalgae products.

2. Data and models

In this study, a microalgae growth prediction model was proposed based on the multi-variable temporal recurrent neural network and the data set of the successful cultivation of PCC7002 in the pilot stage. Five environmental parameters, including temperature, pH, light intensity, dissolved oxygen value and air input, which have major effects on microalgae growth, were taken as input parameters. The turbidity value (i.e., cell concentration) of the growth state of microalgae was used as the output variable. The construction of the model is divided into three stages: (1) data collection; (2) Data preparation and preprocessing; (3) Neural network training.

2.1 Strain pre-culture and data collection

The data set used in this study is the real-time environmental parameters recorded in the reactor during the self-cultured synechococcus PCC7002. Polycoccus was inoculated with BG11 (Rippka, 1988) solid medium plate (1.5% AGAR). When algal colonies appeared on the plate, single algal colonies were selected and inoculated in 50mL liquid BG11 medium (250mL conical bottle). 75 μmol photons m-2s-1 (4000K white light) in 1% (v/v) CO2 at 37℃.

2.2 Data Preparation

After the strain was placed in a 1L culture bottle for secondary expansion, the algal liquid was poured into a closed tube photobioreactor for further culture, and the growth of the polycoccus was recorded. In order to reduce the complexity of the model, the environmental parameters were divided into elements, and a time series model was established to predict the turbidity value of microalgae. The input variables suitable for the artificial neural network included the temperature inside the reactor, pH, light intensity, air input and dissolved oxygen value, and the cell concentration at a certain time was defined as the output variable. This dataset converts the time read into a timestamp, and the parameters for this study include five input parameters and one output parameter (Table 1).

Table 1 Input and output parameters of neural network

	temperature	pH	light	air_in	O2	turbidity
count	6012.000000	6012.000000	6012.000000	6012.000000	6012.000000	6012.000000
mean	30.524251	8.129087	102.700432	1.996770	11.777792	312.283081
std	8.772746	0.090406	151.761402	0.015042	6.509991	204.550519
min	17.900000	7.910000	0.000000	1.899100	5.800000	1.130127
25%	23.100000	8.060000	0.000000	1.987200	6.347000	182.438259
50%	29.800000	8.120000	1.000000	1.996400	7.005000	317.564981
75%	36.800000	8.210000	146.000000	2.006000	20.000000	429.218871
max	48.800000	8.400000	501.000000	2.075800	20.000000	806.312219

2.3 Neural network training

2.3.1 Prediction of growth process of Polycoccus by LSTM

LSTM is a multivariable temporal recurrent neural network optimized to alleviate the gradient vanishing and forgetting problems of RNN for excessively long sequences. Its output mode is shown in Figure 1. LSTM adds a passing state in the optimization process. Compared with RNN, which only has a passing state ht(hidden state),LSTM adds a cell state ct(cell state) on this basis. With the increase of ct-1, ct not only changes slowly with the change of c, but ht and ht-1 will show great differences with the rapid change of h.

The internal structure of LSTM is trained by the current input value xt and the ht-1 passed from the previous state value through the network to obtain four states, which are transmitted to the next state through the forget stage, the selection stage and the output stage. In the forget stage, the input from the previous node is selectively forgotten, and the discarded part of ct-1 in the previous state is controlled by the calculated zf. In the select stage, the input information is filtered, the current input content is represented by z calculated in the previous step, and the signal is controlled by zi. The ct transmitted to the next state is obtained by combining the two, namely:

$$c^t = z^f \odot c^{t-1} + z^i \odot z \qquad (1)$$

The output phase uses zo to determine the output that will be regarded as the current state, and the output yt is often changed by the transfer state ht:

$$h^t = z^o \odot \tanh(c^t) \qquad (2)$$

$$y^t = \sigma(W'h^t) \qquad (3)$$

The neural network model is built by using all parameter training data, and the optimal neural network architecture is obtained by integrating training set, verification set and test set. In this experiment, 5 input parameters temperature, pH, light, air_in, O2 and 1 output parameter turbidity were used. After data set partitioning and training for several times, the data set was divided into the training set and the verification set in the ratio of 4:1. Adam optimization function was used and the number of prediction steps was set to 200 and the number of iterations was set to 50 to get the corresponding curve. The LSTM algorithm model is shown in Figure 1.

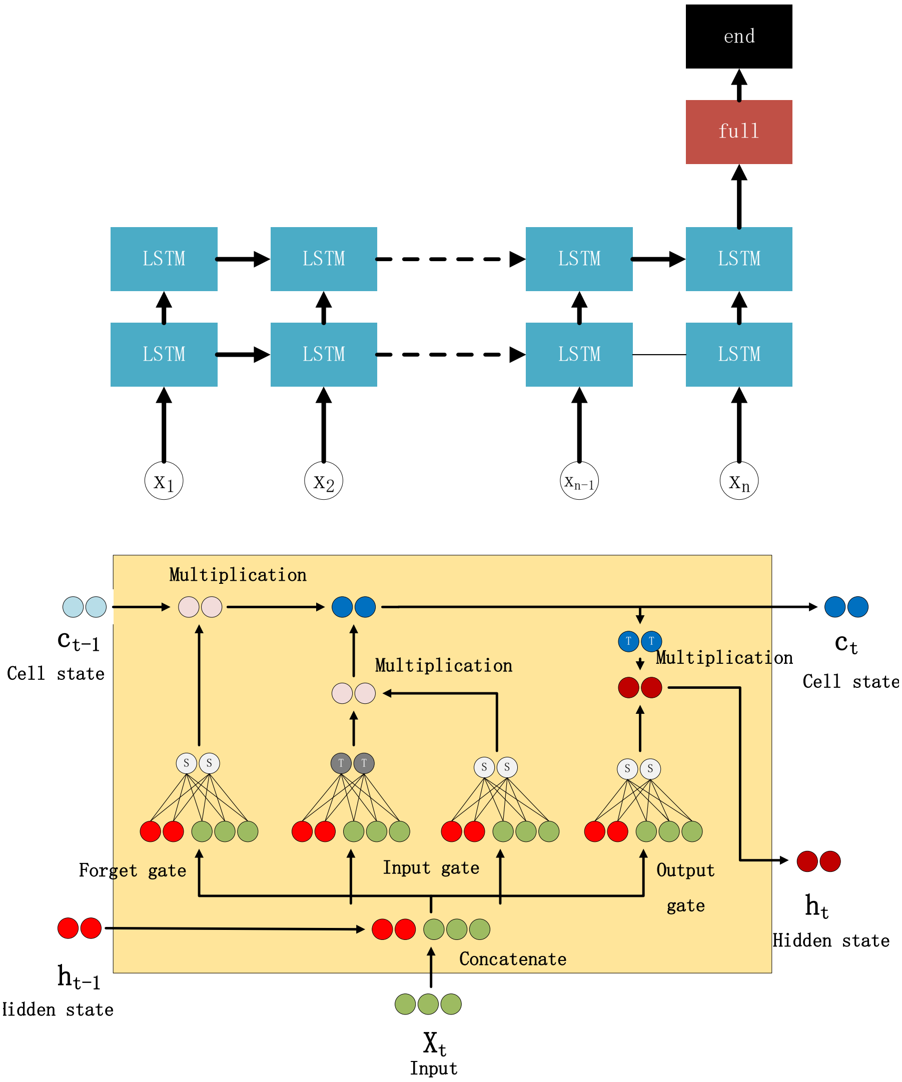


Figure 1 LSTM algorithm structure (top), LSTM internal structure (bottom)

The turbidity prediction of polycoccus was obtained through training, as shown in Figure 2, where the horizontal coordinate is the time stamp and the interval is 1 minute. The ordinate is the turbidity value used to characterize the concentration of synechococcus cells. The left figure is the comparison between the actual value of the trained data set and the predicted value, and the right figure is the comparison between the actual value of the trained test set and the predicted value. After model training and curve comparison, it can be seen that the actual value is highly consistent with the predicted value.

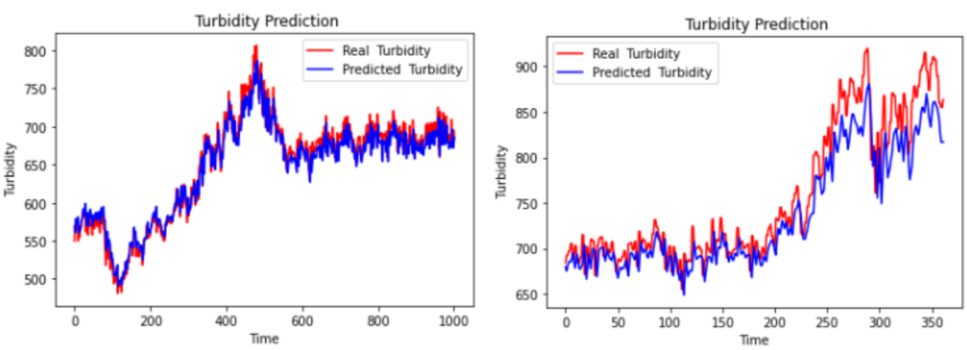


Figure 2 Turbidity prediction

2.3.2 Prediction of environmental parameters of synechococcus culture by MTS-Mixers algorithm

On the condition that the prediction curve of the growth trend of polycoccus was highly consistent with the actual value, the rationality of the culture process was confirmed. On this basis, the MTS-Mixers algorithm was integrated to predict various environmental parameters of the culture process. The structure of MTS-Mixers algorithm is shown in Figure 3. In the actual industrial culture process, what has a direct impact on the yield of microalgae is the growth of microalgae in the lag period and exponential period. When the culture environment is suitable for the growth of microalgae, the lag period is shortened, and the exponential cell generation is shortened, and the growth rate is accelerated, and a large number of products can be harvested in a short time through industrial culture means. The hysteresis period and exponential period of synechococcus were analyzed respectively.

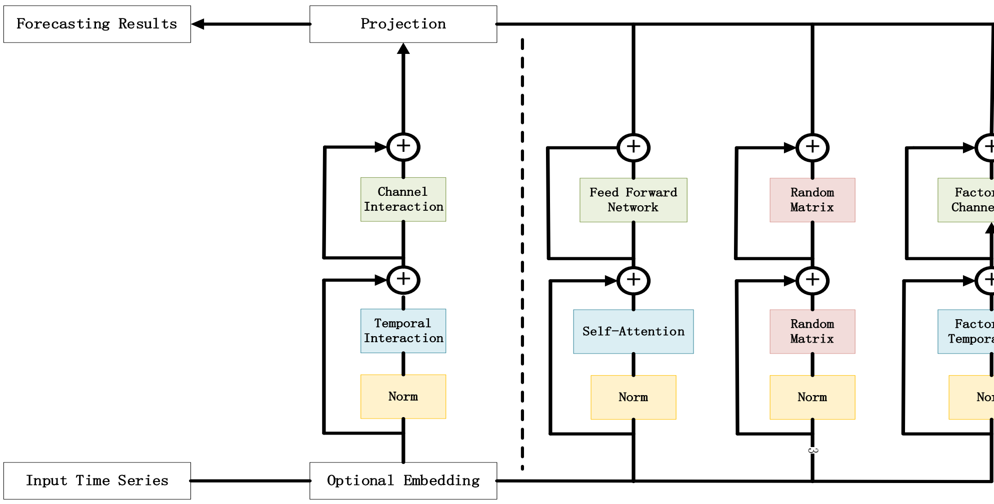
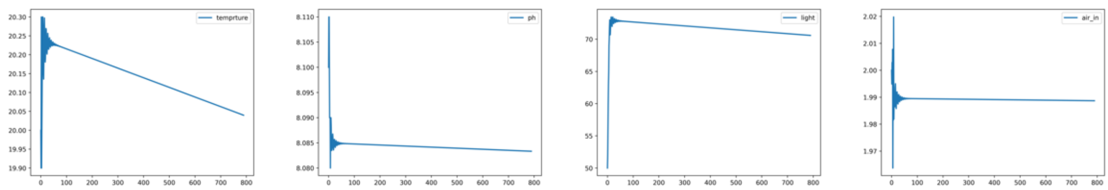


Figure 3 MTS-Mixers algorithm structure

When synechococcus is in the lag period, its growth is determined by the length of time. By excerpting the turbidity changes of synechococcus during the lag period, the optimization curve of the predicted environmental parameters is shown in Figure 4. Where the horizontal coordinate is the time stamp of the delayed period of the synechococcus, and the vertical coordinate is the corresponding eigenvalue.



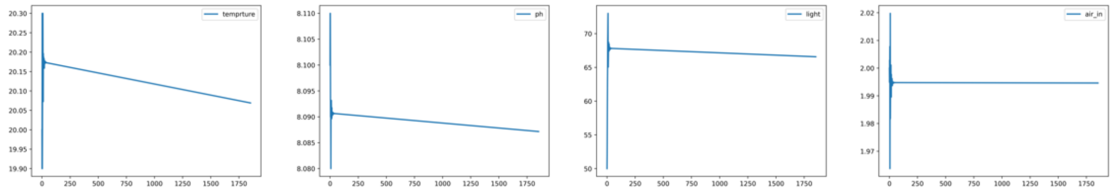
Temperature pH Light Intensity Air Input

Figure. 4 Prediction of environmental parameters of demurrage period

Through the analysis of the predicted data of the lag period, the predicted data fluctuated weakly, which can be approximately regarded as linear changes. As shown in the figure, during the growth process of the lag period, the temperature should be controlled within the range of 20.0-20.2℃, the pH value should be controlled at 8.08, the light intensity should be controlled at 70W/m2, and the air input should be controlled at 1.99Ln/m.

When polycoccus is in exponential stage, its growth is determined by multiplication algebra $n = 3.322(\lg x_2 - \lg x_1)$, growth rate constant $R = 3.322(\lg x_2 - \lg x_1)/(t_2 - t_1)$ and generation time $G = (t_2 - t_1)/3.322(\lg x_2 - \lg x_1)$.

By excerpting the turbidity changes of synechococcus in the exponential period, the optimization curve of the predicted environmental parameters is shown in Figure 5. Where the horizontal coordinate is the time stamp of the exponential phase of synechococcus, and the vertical coordinate is the corresponding eigenvalue.



Temperature pH Light Intensity Air Input

Figure 5 Prediction of environmental parameters in the index period

Through the analysis of the predicted data in the index period, the predicted data has a weak fluctuation range, which can be approximately regarded as linear change. As shown in the figure, during the growth process of the lag period, the temperature should be controlled within the range of 20.0-20.2℃, the pH value should be controlled within the range of 8.08-8.09, and the light intensity should be controlled within 67W/m2. air input should be controlled between 1.99-2.00Ln/m.

3. Conclusion

The typical growth curve of microorganisms is usually derived by appropriate culture in a laboratory setting and periodic measurement of cell content, with few precedents in pilot stage and industrial production culture environments. In the laboratory environment, the lag time of synechococcus depends on the volume of synechococcus, and the exponential period is between 1.5-2.3 hours. In the process of cultivating polycoccus, the exponential generation time of the equipment is mostly about 1.9 hours, indicating that under the premise of increasing the cultivation capacity in the culture process, the normal growth of polycoccus can be achieved through intelligent control and the high yield of polycoccus raw materials can be obtained through industrial production. In this paper, a large capacity photobioreactor was designed as a vessel for data collection. The fusion LSTM predicted the growth curve of polycoccus by taking environmental factors in the culture process as input parameters, which confirmed the authenticity and scientificity of the data set. The fusion MTS-Mixers predicted the optimization curve of various environmental factors in the culture process through cell concentration in the culture process. For the industrial production of high-value photoorganisms such as polychlococcus, the situation of large-scale death of organisms caused by poor culture environment during industrial production of photoorganisms was solved, and the technical support was provided for improving the survival rate and yield of photoorganisms.

References

[1] Zhilan Chen，Yun Tian，Chenhong Zhu， et al. Sensitive detection of oxidative DNA damage in cyanobacterial cells using supercoiling-sensitive quantitative PCR[J]. Chemosphere， 2018， 211: 164-172.

[2] Ryan L. Clark，Laura L. McGinley，Hugh M. Purdy， et al. Light-optimized growth of cyanobacterial cultures: Growth phases and productivity of biomass and secreted molecules in light-limited batch growth[J]. Metabolic Engineering， 2018， 47: 230-242.

[3] Zhe Li，Zhongwen Rao，Lujia Pan， et al. MTS-Mixers: Multivariate Time Series Forecasting via Factorized Temporal and Channel Mixing[J]. Machine Learning， 2023: 1-14.

[4] Noguchi R, Ahamed T, Rani D S, et al. Artificial neural networks model for estimating growth of polyculture microalgae in an open raceway pond[J]. Biosystems Engineering, 2019, 177: 122-129.

[5] Hochreiter, S, and J. Schmidhuber. “Long short-term memory.” Neural Computation 9.8(1997):1735-1780.

[6] Gers F A, Schmidhuber J, Cummins F. Learning to forget: Continual prediction with LSTM[J]. 1999.