Online learning method for predicting air environmental information used in agricultural robots

Yueting Wang[1,2,3], Minzan Li[2], Ronghua Ji[2], Minjuan Wang[2], Yao Zhang[2], Lihua Zheng[2*]

(1. *National Research Center of Intelligent Equipment for Agriculture, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China*;
2. *Key Laboratory of Smart Agriculture Systems, Ministry of Education, China Agricultural University, Beijing 100083, China*;
3. *Key Laboratory of Agricultural Sensors, Ministry of Agriculture and Rural Affairs, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China*)

Abstract: Air environmental information plays an important role during plant growth and reproduction, and prompt and accurate prediction of atmospheric environmental data is helpful for agricultural robots to make timely decisions. In the interest of efficiency, an online learning method for predicting air environmental information was presented in this work. This method combines the advantages of convolutional neural network (CNN) and experience replay technique: CNN is used to extract features from raw data and predict atmospheric environmental information, while experience replay technique can store environmental data over some time and update the hyperparameters of CNN. To validate the effects of this method, this online method was compared with three different predictive methods (including random forest, multi-layer perceptron, and support vector regression) using a public dataset (Jena). According to the results, a suitable sample sequence size (e.g., 16) has a smaller number of training sessions and stable results; a larger replay memory size (e.g., 200) can provide enough samples to capture useful features; and 6 d of historical information is the best setting for training predictor. Compared with traditional methods, the method proposed in this study is the only method that can be applied for various conditions.

Keywords: online learning method, convolutional neural network, real-time prediction, air environmental information

DOI: 10.25165/j.ijabe.20241705.7972

Citation: Wang Y T, Li M Z, Ji R H, Wang M J, Zhang Y, Zheng L H. Online learning method for predicting air environmental information used in agricultural robots. Int J Agric & Biol Eng, 2024; 17(5): 206–212.

1 Introduction

Prompt and accurate prediction of air environmental information helps agricultural robots to make timely decisions[1-3]. However, the most common prediction paradigm (i.e., offloading learning) has poor timeliness, and it requires large memory to store samples for training. The online learning method is another learning paradigm, which has been widely studied in several research fields[4]. The goal of online learning is to make a sequence of accurate predictions given knowledge of correct answers to previous prediction tasks and possibly additional available information[5,6]. Online learning applications of air environmental information prediction are few, partly because it is difficult to capture suitable features and predict for continuously changing data, and partly due to lack of timely feedback to update the predictive model.

As for feature extraction and prediction, many well-established methods can be used for predicting air environmental information. These methods can be roughly divided into two categories: conventional models and deep-learning models. Conventional models include random forest (RF)[7], support vector regression (SVR)[8], etc., in which fitting ability is limited by the base model (e.g., decision tree (DT)) or kernel function (e.g., radial basis function kernel (RBF)). As for deep-learning models, multi-layer perception (MLP) can be seen as the basic deep-learning model, which can freely fit different parameters by linear combinations[9]. Recurrent neural networks (RNNs, a typical deep-learning method) constitute a class of neural networks that exhibit state-of-the-art performances for modeling sequential data[10]. However, they require more storage for data and require more computing resources to process[11]. Convolutional neural network (CNN) is another classical deep-learning method[12]. CNN is able to process ambient data faster than RNN because CNN is better at parallel computing than RNN[13]. As for timely feedback for updating the predictive model, the existing strategy used in the online learning method is difficult to apply to air environmental information prediction. This is because the randomness of air environmental data cannot be anticipated. For these questions, experience replay technique provides a pathway[14,15]. At each time step, air environmental data is stored in experience replay memory, and samples are split and sampled from replay memory in each training cycle to drive learning and update the weights of the predictor. Although replay memory size is limited, access is quite fast, which can be used for the online learning method.

Based on the above description, an online learning method for predicting air environmental data was developed, which can be used for agricultural robots to make timely decisions. The main

Received date: 2022-10-16 **Accepted date:** 2023-06-04

Biographies: Yueting Wang, PhD, research interest: agricultural Information, smart agriculture, Email: wyt_study@163.com; **Minzan Li**, PhD, Professor, research interest: agricultural information, smart agriculture, Email: limz@cau.edu.cn; **Ronghua Ji**, PhD, Associate Professor, research interest: agricultural information, Email: jessic1212@cau.edu.cn; **Minjuan Wang**, PhD, Associate Professor, research interest: image processing, Email: minjuan@cau.edu.cn; **Yao Zhang**, PhD, Associate Professor, research interest: agricultural information, Email: zhangyao@cau.edu.cn.

***Corresponding author: Lihua Zheng**, PhD, Professor, research interest: agricultural information. College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China. Email: zhenglh@cau.edu.cn.

contributions of this work are as follows: 1) The framework of this online learning method was introduced and more details were given about the key component (i.e., predictor). 2) The optimal working conditions were investigated and many experiments were carried out. Finally, the other methods (including RF, MLP, and SVR) were compared with the proposed method's results, indicating that this online learning method can adapt to different air environmental data in different collected conditions.

2 Materials and methods

2.1 Dataset description

A weather dataset recorded at the Weather Station at the Max Planck Institute for Biogeochemistry in Jena, Germany was selected. With this dataset, 14 different ambient variables (e.g., air temperature, atmospheric pressure, humidity, etc.) were recorded every 10 min. You can obtain more information at www.bgc-jena.mpg.de/wetter. This study only used three types of air environmental data (including air temperature, humidity, and atmospheric pressure), which were recorded from 2009-2016, and more details are listed in Table 1.

2.2 Structure of the online method for predicting air environmental data

The structure of the online method for predicting air environmental data is shown in Figure 1. The input (i.e., air environmental data collected at different times) is constructed into different sample sequences (i.e., sample sequence$_t$, the length is L_s). The last record of each sample sequence is selected as a predictive variable (V_p) and others are selected as training variables (V_t). On the one hand, a sample sequence can be used to predict air environmental data at time t (output is O_t) by predictor (at time T). On the other hand, every sample sequence is stored in replay memory. Replay memory size (M_s) is set manually, and old data will be removed when storage capacity is full. The sample acquisition period (P) in replay memory can be computed by Equation (1).

$$P = S_i \cdot M_s \cdot L_s \tag{1}$$

where, S_i is the sampling interval of raw ambient data. When the training interval is reached, samples in replay memory are divided into training samples and testing samples (the ratio is 8:2 in this work) which are used for hyperparameters updating and evaluation, respectively. Then, a new state of predictor (at time T+1) is obtained and it is used to substitute the old predictor. Note that, T and t are two different parameters.

Table 1 Description of ambient dataset

Total samples	Interval	Air temperature/°C			Humidity/(% RH)			Atmospheric pressure/Pa		
		Min	Max	Mean	Min	Max	Mean	Min	Max	Mean
420 451	10 min	−23.01	37.28	9.45	12.95	100	76.01	95	6377	1356

Note: The statistical properties of ambient data are calculated including minimum values (Min), maximum values (Max), and mean values (Mean).

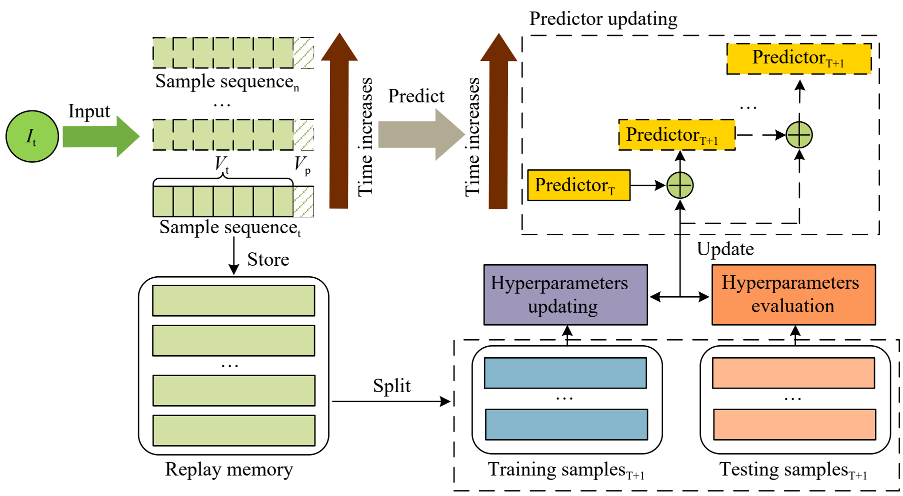


Figure 1 Structure of this online learning method to predict air environmental data

The predictor (shown in Figure 2) is the most important component of this method. The input is raw air environmental data (in light blue), target output is predicted data (in red), and intermediate output is reconstructed data (in light orange). Different convolutional/un-convolutional layers and pooling/up-sampling layers are distinguished by different colors. Feature reconstruction is used to provide more features for predicting, and a fully connected layer is a linear regression model. In order to extract the right features, a CNN-based auto-encoder is introduced as a predictor that can take into account air environmental information between adjacent periods[16]. The CNN-based auto-encoder includes two parts: encoder and decoder. In the encoding phase, raw data is expressed according to multi-layer convolutional and pooling operations, and each neuron in the output of the encoder covers a receptive field of many original data. In the decoding phase, un-convolutional and up-sampling operations were carried out to recover air environmental data by encoded features. The quality of encoded features was evaluated by the mean squared error (MSE, Equation (2)) of raw data and recovered data.

$$\mathrm{MSE}(\bar{y}, y) = \frac{\sum_{i=1}^{n} (y_i - \bar{y}_i)^2}{n} \tag{2}$$

where, y_i and $\bar{y}_i$ are measured values and reconstructed values, respectively; n is the number of samples in the training set; i is the ith sample.

For better performance, feature reconstruction is carried out to obtain more features for predicting air environmental data[17]. In this work, the convolutional operation was used as a specific operation of feature reconstruction[18]. The fully connected layer is a 2-layer artificial neural network to predict air environmental data based on features outputted by the flattened layer. Adam optimizer is used to search the local minimum of the objective function. All hyperparameters are listed in Table 2. Kernel/pooling size are the values of different convolutional and pooling layers in part of auto-encoder, and operations of convolution/un-convolution use the same hyperparameters. Kernel size R is used in feature reconstruction. Hidden layer 1 means the number of neurons within the 1st hidden layer, and hidden layer 2 means the number of neurons within the 2nd hidden layer.

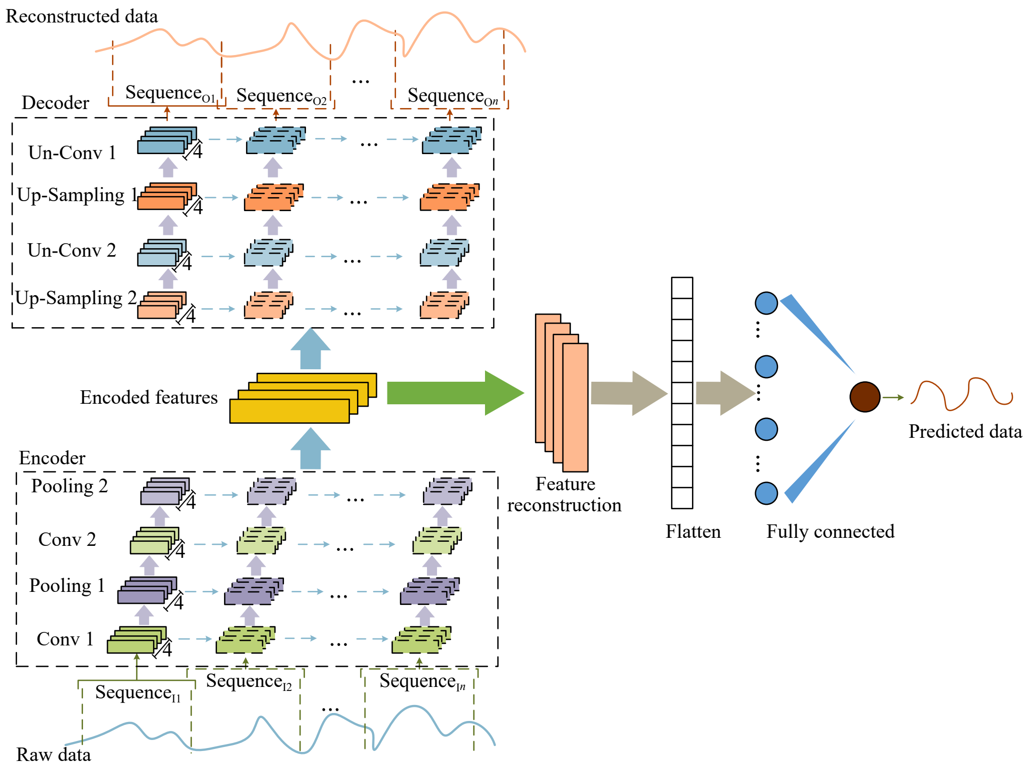


Figure 2 Schematic of predictor used in this online learning method

Table 2 Hyperparameter settings in predictor

Hyperparameter	Kernel size 1	Kernel size 2	Kernel size R	Hidden layer 1	Hidden layer 2	Learning rate
Value	5	3	3	12	8	0.01

2.3 Design of experiments

To validate the performance of this online learning method, three predictive models (including random forest (RF) regression, support vector regression (SVR), and multi-layer perception (MLP)) are selected to compare performance. RF consists of a set of many un-pruned ensembles of regression trees, which are composed of root nodes, branch nodes, and leaf nodes. Regression trees are generated based on bootstrap sampling from original training data, and bootstrap resampling of data for training each tree increases diversity between trees[19]. Random forest regression is widely used in prediction with the characteristics described above[20,21]. SVR is the application of a support vector in a regression function. SVR derives a function on the basis of training data to predict numerical values. SVR can be seen as an intrinsically non-linear prediction method because it projects datasets characterized by the presence of non-linear structure-property relationships in original feature spaces into higher-dimensional space representations where a linear regression function can be fitted[8]. MLP is an artificial neural network that models complex functions, which consist of three or more layers of nodes[22]. Compared with traditional methods, MLP does not require any prior assumptions regarding the distribution of training data, avoiding the influence of data distribution on performance.

Results of the different prediction models were evaluated by the coefficient of determination (R^2), root mean squared error of prediction (RMSEP), and the residual prediction deviation (RPD). Their calculations are shown in Equations (3)-(6).

$$R^2(y,\hat{y}) = 1 - \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2} \tag{3}$$

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n}} \tag{4}$$

$$\text{RPD} = \frac{s_y}{\text{RMSEP}} \tag{5}$$

$$s_y = \sqrt{\frac{\sum_{i=1}^{N}(y_i-\bar{y})^2}{N}} \tag{6}$$

where, s_y is the standard deviation of the observed values, which is calculated using Equation (6). In which, $\bar{y}$ is the arithmetic mean of y_i; N is the number of samples in the test set.

3 Results and discussion

3.1 Performance of this online learning method with different settings

3.1.1 Impact of different sequence sizes on this online learning method

Sample sequence size is one of the most important factors for performance. To analyze the effect of sample sequence size, this experiment was carried out to compare performance with different sample sequence sizes (as shown in Table 3). There were 7 experimental groups with different sample sequence sizes selected to predict air temperature with 9600 samples. Memory size was 100, training interval was 100, and number of epochs was 100.

Table 3 Performance of this online learning method with different sample sequence sizes

Sequence size	Indicators	Number of training sessions										
		1	2	3	4	5	6	7	8	9	10	11
8	R^2	0.66	0.97	0.98	0.95	0.97	0.99	0.97	0.99	0.98	0.99	0.99
	RMSEP	1.95	0.22	0.25	0.11	0.54	0.45	0.37	0.27	0.21	0.36	0.25
	RPD	1.73	5.92	7.75	4.59	5.62	10.59	5.43	15.42	7.15	8.65	17.92
10	R^2	0.89	0.96	0.99	0.99	0.99	0.98	0.99	0.99	--	--	--
	RMSEP	1.40	0.52	0.06	0.18	0.10	0.20	0.08	0.30	--	--	--
	RPD	3.07	5.14	10.03	20.04	21.94	7.38	17.28	10.69	--	--	--
12	R^2	0.95	0.96	0.96	0.81	0.98	0.96	0.98	--	--	--	--
	RMSEP	0.42	0.42	0.65	0.86	0.45	0.18	0.61	--	--	--	--
	RPD	4.64	5.21	4.88	2.32	8.19	5.19	6.61	--	--	--	--
16	R^2	0.97	0.99	0.98	0.99	0.99	--	--	--	--	--	--
	RMSEP	0.29	0.36	0.38	0.13	0.37	--	--	--	--	--	--
	RPD	6.29	8.92	6.52	14.81	13.69	--	--	--	--	--	--
20	R^2	0.99	0.99	0.99	--	--	--	--	--	--	--	--
	RMSEP	0.10	0.74	0.11	--	--	--	--	--	--	--	--
	RPD	12.91	8.28	16.65	--	--	--	--	--	--	--	--
24	R^2	0.98	0.99	0.99	--	--	--	--	--	--	--	--
	RMSEP	0.39	0.32	0.49	--	--	--	--	--	--	--	--
	RPD	6.84	10.82	10.65	--	--	--	--	--	--	--	--
30	R^2	0.98	0.98	--	--	--	--	--	--	--	--	--
	RMSEP	1.07	0.77	--	--	--	--	--	--	--	--	--
	RPD	6.54	6.38	--	--	--	--	--	--	--	--	--

Small sequence sizes (e.g., 8, 10, 12) could capture more features because they had a larger number of sample sequences (e.g., 1200, 960, 800). However, this case could not extract stable features, and it was relatively easy for performance to obtain a greater fluctuation (e.g., the first training session when the sample sequence size was 8 and 10, and the 4th training session when the sample sequence size was 12). As sample sequence size increased, performance became more stable (e.g., 24 and 30), but more details were lost (i.e., larger values of RMSEP). This is attributed to too large of a sample sequence to capture instantaneous changes in raw data.

3.1.2 Impact of different replay memory sizes on this online learning method

Replay memory size is another important factor for the performance of this online learning method, which directly controls the size of training and testing samples. This experiment was carried out to analyze the effect of replay memory size. Five groups of different memory sizes were selected to predict air temperature with 20 000 samples. The sample sequence size was 20, the training interval was 150, and the number of epochs was 100.

As shown in Figure 3, when the replay memory size was small (e.g., 50, 100, and 150), the predictor could be trained by new air temperature data. In addition to the first training session, many old records were used with a large memory size (e.g., 200 and 250). Small memory size (e.g., 50) was beneficial for fitting new features, but it could not provide enough samples to obtain a good model when air temperature data had large variations. With increasing replay memory size, performance improved significantly. When replay memory size and training interval were equal (i.e., 150), every sample sequence was trained. However, because of the overly large replay memory size, the predictor could not capture suitable features, which led to poor performance. When memory size was larger than the training interval (e.g., 200 and 250), some samples (10 and 50 samples, respectively) were trained repeatedly. Too many trained samples (e.g., when memory size was 250) improved the performance of the predictor, but this required more computing resources and more memory.

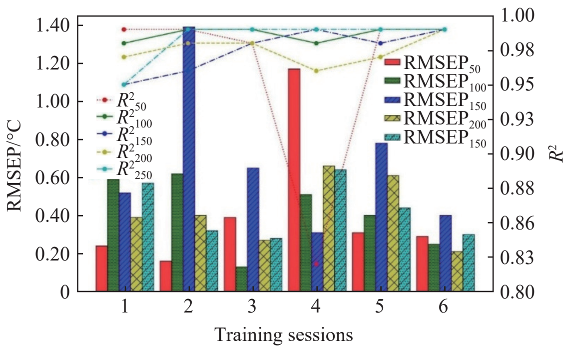


Figure 3 Performance of online learning method with different replay memory size

3.1.3 Impact of different historical information on this online learning method

To analyze the effect of historical information on replay memory, this experiment was carried out to compare the performance with different historical information (Table 4). For convenience, the sample sequence size was set to 7, in which 6 samples (i.e., 1 h) were selected as training variables and one sample was used as a predictive variable. With this operation, each sample sequence could provide features in the past 1 h (only considering training variables). Meanwhile, the replay memory size was set to 72, 96, 120, 144, 168, and 336, respectively. In this way, 6 experimental groups with different historical information (including 3, 4, 5, 6, 7, and 14 d) were obtained. There were 42 000 samples used for predicting air temperature, and the training interval was 400.

Table 4 Performance of this online learning method with different historical data

Historical interval/d	R^2				RMSEP/°C			
	Min	Max	Mean	Std	Min	Max	Mean	Std
3	0.81	0.99	0.97	0.0486	0.05	0.55	0.24	0.1384
4	0.90	0.99	0.97	0.0350	0.04	2.25	0.46	0.5515
5	0.94	0.99	0.98	0.0140	0.07	0.67	0.26	0.1638
6	0.98	0.99	0.99	0.0033	0.06	0.63	0.23	0.1574
7	0.96	0.99	0.99	0.0088	0.13	1.28	0.38	0.2663
14	0.97	0.99	0.99	0.0071	0.14	1.39	0.45	0.3067

Due to insufficient historical information, performance (especially values of R^2) with more days of historical data was better than those with fewer historical days. But, because of more samples and data fluctuation, values of RMSEP with more historical days were higher than those with fewer historical days. Based on a combination of R^2 and RMSEP, 6 d of historical information was suitable for training predictors. What needs special attention is that there was some abnormal record (i.e., the value of RMSEP was 2.25) with 4 d of historical information. This was caused by an inappropriate segmentation method, which was not avoidable because of random variations in air temperature.

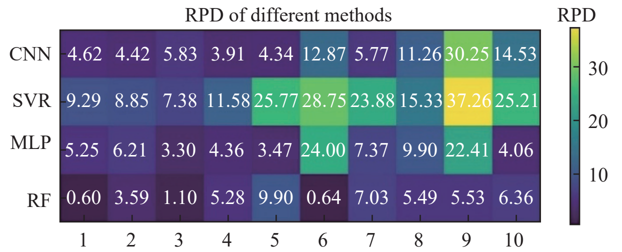
Overall, this online learning method can predict air environmental information in a suitable setting. To avoid the effect of random variation, longer sample sequences and larger replay memory size help extract effective and stable features, which can obtain good prediction performance. Historical information is the underlying factor, which is affected by sample interval, sample sequence size, and replay memory size. Sufficient historical data is beneficial for capturing stable features, but it will lose some sensitivity.

3.2 Comparison of predictive performance with different prediction methods

3.2.1 Impact of different training intervals on different prediction methods

Real-time air environmental information prediction is a major challenge, and the uncertainty of newly collected data requires the model to have a strong ability for feature extraction and prediction. Furthermore, a shorter training interval is another way to help prediction models for fitting raw data by subdivision of raw data. In order to analyze the impact of training intervals on different predictive models, different models (including RF, MLP, SVR, and this online learning method (labeled as CNN)) were built and compared with different training intervals. 15 120 samples were selected to carry out this experiment, the sample sequence size was 7 (6 samples used for feature extraction, and one sample used to predict), and the replay memory size was 72. Five different experimental groups were carried out, and training intervals were 144 (i.e., 6 d), 216 (i.e., 9 d), 288 (i.e., 12 d), 360 (i.e., 15 d), and 720 (i.e., 30 d).

During this experiment, some prediction models (including RF and SVR) could not work in some cases. The predictor presented in this work was the only one to be applied for all cases, and all values of RPD were higher than 2.0. Figure 4 shows values of RPD with 4 prediction models when the training interval was 216 (i.e., 9 d). SVR performed the most stably in this case, but it had a poor performance (value of RPD is 0.45) at the first training session when the training interval was 144. Moreover, the introduction of Grid-search took more time on the hyperparameter setting. RF performed the worst of these four models, and was not able to work at 3 periods (the 1st, 3rd, and 6th training sessions) in this case. In other cases, there were also 5 times where the value of RPD was lower than 1.5. Due to its lower ability of feature extraction, MLP showed poorer performance than the CNN-based method, but it could also obtain a stable model (all the values of RPD were higher than 1.5).



Note: CNN: Convolutional neural network; SVR: Support vector regression; MLP: Multi-layer perception; RF: Random forest. Same below.

Figure 4 Credibility of different predictive models

To compare the accuracy of different predictive models with different training intervals, average values of R^2 and RMSEP were computed and are shown in Figure 5. It should be specially explained that these cases, in which values of RPD were lower than 1.5, were seen as outliers and were removed. The CNN and MLP models, which both belong to the deep-learning method, performed better and became more stable with increasing training sessions. However, because of weaker feature expression ability, MLP had fluctuations of R^2 and higher RMSEP. SVR had the optimal performance, but it took much time to select hyperparameters. Meanwhile, its performance was limited with the kernel function (linear or radial basis function kernel), sometimes to the point of failure. As for RF, it could not infer values outside training values due to regression trees, and this determined poor performance.

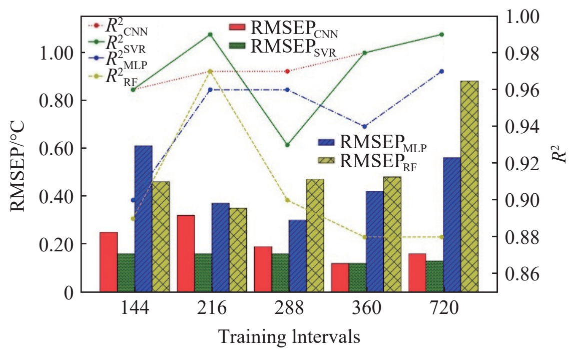


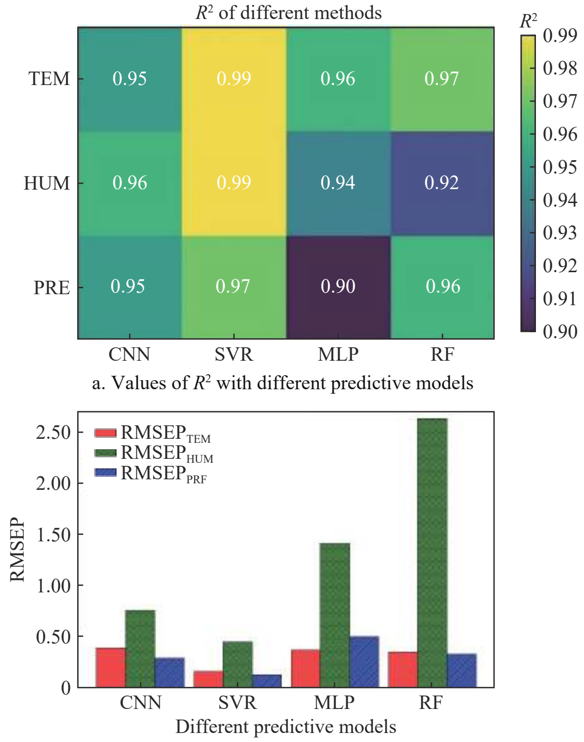
Figure 5 Predictive results obtained by four different models with different training intervals

3.2.2 Impact of different types of air environmental data on different prediction methods

Different types of air environmental data have different characteristics. Three different types of air environmental data (including air temperature, humidity, and atmospheric pressure) were selected to analyze the impact of different environmental information. Of these, the air temperature had positive and negative values, humidity was the most discrete (standard deviation is 16.48), and atmospheric pressure was relatively flat. During this experiment, 15 120 samples (i.e., 90 d) participated, 3 d of historical data were stored in replay memory (sample sequence size was 7, and replay memory size was 72), and training interval was 216 (i.e., 9 d). The predictive results (average of 10 training sessions) are shown in Figure 6.

Note that the cases in which values of RPD are lower than 1.5

were seen as outliers and were removed. For air temperature, three predictive results were removed (1st, 3rd, and 6th) in RF. For humidity, the predictive result obtained by the first training session was removed in SVR, three results (1st, 5th, and 8th) were removed in MLP, and two results were removed in RF. As for atmospheric pressure, MLP (4th, 5th, and 7th) and RF (1st, 3rd, and 6th) each had three results removed.



a. Values of R^2 with different predictive models

b. Values of RMSEP with different predictive models

Figure 6 Predictive results obtained by four predictive models with different types of air environmental data

The CNN-based model, presented in this work, was the only predictive model suitable for three ambient data types, in which all values of RPD were higher than 2.0. All other predictive models (SVR, MLP, and RF) had abnormal cases, which could not be applied. After removing outliers, SVR obtained the optimal predictive results, and the performances of MLP and RF were similar. This further illustrates that the ability of feature extraction and prediction cannot meet different ambient information requests with these predictive models, and they have practical limitations in real applications. Special attention needs to be given to the discrete distribution of ambient information (e.g., humidity), which greatly affects the values of RMSEP. This indicates that the characteristics of raw data directly affect the performance of predictive models[23].

4 Conclusions

Air environmental information plays an important role during crop growth, affecting photosynthesis, respiration, and transpiration. Predicting air environmental information is crucial in aiding agricultural robots to make timely decisions. To address this concern, an online learning method for predicting air environmental data is presented in this work. The following conclusions are accordingly achieved.

1) This online learning method has addressed challenges of air environmental data prediction: accurate feature collection and timely feedback to update predictive models. The convolutional neural network is used to extract useful features from continuously changing air environmental data, and the experience replay technique can provide timely feedback.

2) Due to the randomness of air environmental data, prior knowledge-based models (e.g., random forest, support vector regression) are unsuitable for real-time prediction. The randomness will lead to assumptions (i.e., base model or kernel function) of prior knowledge-based models that are no longer satisfied in real conditions. Of course, poor capability of feature extraction (e.g., multi-layer perception) will similarly affect prediction performance.

3) An appropriate amount of historical data helps extract effective and stable features of air environmental data, and it can improve the performance of this online learning method. However, too much historical data will compromise the fitting ability of prediction models, leading to the loss of more details. Meanwhile, the characteristics (e.g., discrete distribution) of raw data directly affect the performance of prediction models.

Online learning is a current hotspot of research, and it has important implications for the development of smart agriculture. In this work, we only considered the model structure and implementation. With the development of hardware and software, more and more deep-learning algorithms can be applied to embedded devices, and this will be the focus of our future studies.

Acknowledgements

This research was funded by the National Key Research and Development Program (Grant No. 2022YFD2002202), Beijing Innovation Consortium of Agriculture Research System (BAIC08-2024-FQ04), Key Laboratory of Agricultural Sensors, Ministry of Agriculture and Rural Affairs (Grant No. PT2024-46), China Postdoctoral Science Foundation (Grant No. BX20230048), and Postdoctoral Fund of Beijing Academy of Agriculture and Forestry Sciences (Grant No. 2023-ZZ-025).

[References]

[1] Hamrita T K, Conway R H. First order dynamics approaching of broiler chicken deep body temperature response to step changes in ambient temperature. Int J Agric & Biol Eng, 2017; 10(4): 13–21.

[2] Fu X, Shen W Z, Yin Y L, Zhang Y, Yan S C, Kou T L, et al. Remote monitoring system for livestock environmental information based on LoRa wireless ad hoc network technology. Int J Agric & Biol Eng, 2022; 15(4): 79–89.

[3] Boomgard-Zagrodnik J P, Brown D J. Machine learning imputation of missing Mesonet temperature observations. Computers and Electronics in Agriculture, 2022; 192: 106580.

[4] Wang Y T, Li M Z, Ji R H, Wang M J, Zhang Y, Zheng L H. A convolutional operation-based online computation offloading approach in wireless powered multi-access edge computing networks. Computers and Electronics in Agriculture, 2022; 197: 106967.

[5] Orabona F. A modern introduction to online learning. arXiv, In press. 2019; arXiv: 1912.13213.

[6] Ross S, Gordon G J, Bagnell J A. A reduction of imitation learning and structured prediction to no-regret online learning. Journal Of Machine Learning Research, Aistats, 2011; pp.627–635.

[7] Breiman L. Bagging predictors. Machine Learning, 1996; 24: 123–140.

[8] Rachman T. Support vector regression machines. Angew. Chemie Int. Ed. 2018; 6(11): 951–952.

[9] Tomassetti B, Verdecchia M, Giorgi F. NN5: A neural network based approach for the downscaling of precipitation fields - Model description and preliminary results. Journal of Hydrology, 2009; 367: 14–26.

[10] Deznabi I, Arabaci B, Koyuturk M, Tastan O. DeepKinZero: Zero-shot learning for predicting kinase-phosphosite associations involving understudied kinases. Bioinformatics, 2020; 36(12): 3652–3661.

[11] Shi W Z, Caballero J, Theis F. Huszar A, Aitken A, Ledig C, et al. Is the deconvolution layer the same as a convolutional layer? 2016; arXiv, In Press. arXiv: 1609.07009.

[12] Wang Y T, Li M Z, Ji R H, Wang M J, Zhang Y, Zheng L H. Novel

encoder for ambient data compression applied to microcontrollers in agricultural robots. Int J Agric & Biol Eng, 2022; 15(4): 197–204.

[13] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, et al. Attention is all you need. In: NIPS'17: Procedings of the 31st International Conference on Neural Information Processing Systems, 2017; pp.6000–6010. doi: 10.5555/3295222.3295349.

[14] Lillicrap T P, Hunt J J, Pritzel A, Heess N, Erez T, Tassa Y, et al. Continuous control with deep reinforcement learning. In: 4th Int. Conf. Learn. Represent. ICLR 2016 - Conf. Track Proc, 2016.

[15] Montague P R. Reinforcement learning: An introduction, by Sutton, R. S. and Barto, A. G. Trends in Cognitive Sciences, 1999; 3(9): 360.

[16] Jarrett K, Kavukcuoglu K, Ranzato M A, LeCun Y. What is the best multi-stage architecture for object recognition? In: 2009 IEEE 12th International Conference on Computer Vision, 2009; pp.2146–2153. doi: 10.1109/ICCV.2009.5459469.

[17] Batal I, Hauskrecht M. Constructing classification features using minimal predictive patterns. In: CIKM '10: Proceedings of the 19th ACM International Conference on Information, 2010; pp.869–878. doi: 10.1145/1871437.1871549.

[18] Wang Y T, Li M Z, Ji R H, Wang M J, Zhang Y, Zheng L H. Construction of complex features for predicting soil total nitrogen content based on convolution operations. Soil and Tillage Research, 2021; 213: 105109.

[19] Jin Z W, Shang J X, Zhu Q W, Ling C, Xie W, Qiang B H. RFRSF: Employee turnover prediction based on random forests and survival analysis. In: Web Information Systems Engineering - WISE 2020, 2020; pp.503–515. doi: 10.1007/978-3-030-62008-0_35.

[20] Gonzalez-Mora A F, Rousseau A N, Larios A D, Godbout S, Fournel S. Assessing environmental control strategies in cage-free aviary housing systems: Egg production analysis and Random Forest modeling. Computers and Electronics in Agriculture, 2022; 196: 106854.

[21] Wang Y T, Li M Z, Ji R H, Wang M J, Zheng L H. A deep learning-based method for screening soil total nitrogen characteristic wavelengths. Computers and Electronics in Agriculture, 2021; 187: 106228.

[22] Li F, Shirahama K, Nisar M A, Koping L, Grzegorzek M. Comparison of feature learning methods for human activity recognition using wearable sensors. Sensors, 2018; 18(2): 679.

[23] Hu J L, Lu J W, Tan Y-P, Zhou J. Deep transfer metric learning. IEEE Transactions on Image Processing, 2016; 25(12): 5576–5588.