Prediction of chilling damage risk in maize growth period based on probabilistic neural network approach

Chunqiao Mi\(^{1,2,3}\), Changhua Zhao\(^{1,2,3,*}\), Qingyou Deng\(^{4}\), Xiaowu Deng\(^{1,2,3}\)

(1. School of Computer Science and Engineering, Huaihua University, Huaihua 418000, Hunan, China; 2. Key Laboratory of Intelligent Control Technology for Wuling-Mountain Ecological Agriculture, Huaihua 418000, Hunan, China; 3. Key Laboratory of Wuling-Mountain Health Big Data Intelligent Processing and Application in Hunan Province Universities, Huaihua 418000, Hunan, China; 4. Human Resource Department, Huaihua University, Huaihua 418000, Hunan, China)

Abstract: Low temperature chilling damage is one of the most serious disasters in maize production, which is a typical non-linear complex issue with numerous influencing factors and strong uncertainty. How to predict it is not only a hot theoretical research topic, but also an urgent practical problem to be solved. However, most of the current researches are focusing on post-disaster static descriptive assessment rather than pre-disaster dynamic predictive analysis, resulting in the problems such as no indicative result and low accuracy. In this study, the satisfaction rate of environmental accumulated temperature for maize production was used to measure the chilling damage risk, and a model for maize chilling damage risk prediction based on probabilistic neural network was constructed. The model was composed of input layer, pattern layer, summation layer and output layer. The obtained results showed that the prediction accuracy for the most serious risk level was as high as 0.91, and the rates of the Type I Error and Type II Error made by the model were 0.1 and 0.09, respectively. This indicated that the model employed was promising with good performance. The results of this research are of both theoretical significance for providing a new reference method of pre-disaster prediction to study maize chilling disaster risk and practical significance for reducing maize production risk and ensuring yield safety.

Keywords: maize chilling damage, risk prediction, probabilistic neural network

DOI: 10.25165/j.ijabe.20211402.5732


Introduction

Maize needs a certain accumulated temperature in the whole growth period to mature normally. Heat deficiency caused by low temperature will cause slow growth and delay development of maize. Maize chilling damage refers to the disaster caused by low temperature and calorific deficiency during the process of maize growth, which can result in delayed growth and production reduction, even no production due to the low temperature and frost hazard in the later growth period. Low temperature chilling damage is one of the most serious disasters in maize production; especially in Northeast China, the chilling damage caused by low temperature in growing season is the biggest meteorological disaster affecting agricultural production\(^{[1]}\). More specifically, in most parts of Northeast China, due to the low temperature and insufficient accumulated temperature in the growing season, maize usually gets slow growth and delayed maturity; in some severe years, maize even could not mature normally before autumn frost, which would result in maize yield reduction by more than 15%\(^{[2]}\). Studies have shown that from now on to 2044, the risk of maize chilling damage in the eastern Songnen Plain and the eastern Sanjiang Plain will still increase\(^{[3]}\). Therefore, it is of great significance to study the occurrence regularity and risk prediction method of maize chilling damage to prevent low temperature related disasters and ensure maize production safety\(^{[4]}\), which can also provide an important basis for scientific maize production planning and disaster prevention and mitigation\(^{[5]}\).

At present, the relevant study of chilling damage is both a research hotspot in the field of agro-meteorology disaster prediction and an urgent practical problem to be solved in relevant agriculture production and disaster prevention departments. Especially with the context of future climate change, more and more attention has been paid to chilling damage risk assessment and prediction by scholars and policy makers in various countries. In foreign countries, relevant studies began in the late 1980s, mainly focusing on the establishment of assessment indices and system method. In China, relevant studies started in the 1990s, focusing on the exploration of risk analysis method and technology\(^{[6]}\). Among all existing related studies, the specific methods used can be summarized into three categories. (1) The first category is index-based evaluation method, which mainly focuses on selection, optimization and weight calculation of disaster risk indicators and then evaluating disaster risk level by constructing specific mathematical models. Typical implementation method includes multi-factor analysis\(^{[7]}\), analytic hierarchy process, fuzzy comprehensive evaluation, principal component analysis, expert scoring method, historical comparison method and Delphi method, etc. For example, Ma et al.\(^{[5]}\) established a comprehensive climate-disaster risk assessment model for maize...
chilling damage, which was composed of cold climate risk index and agricultural production structure factors such as maize yield and area ratio. Gao et al.\cite{9,10} constructed a chilling damage index in maize growth stage based on heat index, and systematically analyzed the temporal and spatial distribution and periodic characteristics of chilling damage at different growth stages of Maize in Northeast China in recent 50 years. Zhu et al.\cite{11} took accumulated equivalent temperature, negative accumulated equivalent temperature and heat matching index as indicators to study the identification method for cold damage year based on rice biological characteristics, circadian temperature rhythm and temperature coefficient. Ma et al.\cite{12} executed a low-temperature treatment at different altitudes on the northern slope of Changbai Mountain to reveal the effect of cold accumulated temperature in different phases of flowering on Japonica rice seed setting rate and to define the weather indexes of rice chilling injury. (2) The second category is the probability-based analysis method, which usually uses mathematical statistics method to analyze the sample data to obtain the statistical regularity of disaster risk, and then carries on disaster risk assessment. Typical analysis method includes regression model, time series model, clustering analysis, probability density function estimation, information diffusion theory, etc. For example, some scholars have carried out risk assessment of low temperature and frost in agriculture based on information diffusion theory\cite{13-15}. According to the principle of crop growth and development, Lin et al.\cite{16} counted the number of frost days of different disaster grades at different growth stages of winter wheat and calculated the probability of frost occurrence for different grades during the whole growth period, and then the risk degree of frost disaster of winter wheat was evaluated. (3) The third category is scenario-simulation-based method, which can be divided into two perspectives. One is to establish a crop growth model under disaster risk scenarios by describing the dynamic growth process of crops, and then dynamically assess the meteorological disaster risk. For example, Ma et al.\cite{17} applied the improved dynamic model of maize growth, development and dry matter accumulation, and established a risk assessment model of maize chilling damage. Cai et al.\cite{18} studied the drought and cold disaster during the two growth stages of Maize germinating-tasselling and tasselling-maturing, and dynamically evaluated the risk of drought and chilling damage of Maize in Northeast China. Another perspective is to use crop models to simulate the growth process of crops, and assess the risk of agro-meteorological disasters by setting different model parameters. For example, Wang et al.\cite{19} and Zhang et al.\cite{20} assessed the risk of maize cold damage in Northeast China and the risk of late frost in Huanghuaizhi region by introducing WOFOST (world food study) model.

To some extent, these related studies all have revealed the formation mechanism, influencing factors, calculation methods and techniques about maize chilling disaster. However, there are only a few studies that carry out dynamic risk analysis using crop simulation models for the whole or critical period of crop growth, while most of the current researches are usually focusing on post-disaster static descriptive assessment rather than pre-disaster dynamic predictive analysis. In addition, most of the methods used are traditional mathematical statistics methods, while machine learning algorithms based on data mining are very rare. Although the calculation process of traditional method is simple, there are some issues to be improved, such as unstable accuracy and no indicative result. The risk of maize chilling disaster is a typical non-linear complex problem with strong uncertainty. On the one hand, it is related to the fluctuation of relevant agro-meteorological factors. On the other hand, it is also related to the imprecise calculated results by the unsuitable method used to understand and evaluate the disaster risk. Probabilistic neural network (PNN) is a parallel algorithm based on the Bayesian minimum risk criterion. PNN places Bayesian estimation in a feed forward neural network and makes Bayesian decision according to the nonparametric estimation of probability density function. It has the characteristics of both fast operation and high accuracy\cite{21}. Based on the advantages of PNN in solving complex non-linear problems, this study attempts to predict low-temperature chilling damage risk in maize growth period with PNN, which can provide a new reference method to study the risk of maize chilling disaster and has very good practical significance for reducing the risk of maize production loss and ensuring yield safety.

2 Materials and methods

2.1 Description of source data

In this study, the satisfaction rate of accumulated temperature during maize growth period, which referred to the satisfaction degree of accumulated temperature of the environment to that required by maize growth, was used to measure the risk of low temperature chilling damage in maize production. The original data used are 50 years’ meteorological daily data from 235 national basic meteorology stations. Among them, 77 (one-third of the total) stations were randomly selected as testing samples, and the remaining two-thirds, namely 158 stations, were used as training samples. Each station contains five basic information, namely station ID, station name, latitude, longitude and altitude, as shown in Table 1. In addition, each station also contains 50 years’ daily meteorological data of 12 indices, one of which used in this study is daily average temperature as shown in Table 2. The stations distributed in 10 provinces in China’s main maize production areas, namely the Northeast China and Huanghuaihai regions, as shown in Figure 1. All the original data are from the China Meteorological Science Data Sharing Service Network. In the practice of maize production, only when the accumulated temperature of the target planting environment is stably greater than the accumulated temperature required by maize growth can maize crop mature normally. Therefore, in the next sections, the characteristic values of accumulated temperature will be calculated.

| Table 1 Examples of basic information of meteorology stations |
|---------------------------|------------------|------------------|------------------|------------------|
| Station ID | Station name | Latitude/(°) | Longitude/(°) | Altitude/m |
| 50468 | Heihe | 50.25000 | 127.45000 | 166 |
| 50514 | Manzhouli | 49.56667 | 117.43333 | 662 |
| 50656 | Be‘lan | 48.28333 | 126.51667 | 270 |

Note: The table omits the basic information of other stations.

| Table 2 Examples of daily average temperature data of meteorology stations |
|---------------------------|------------------|------------------|------------------|------------------|
| Station ID | Year | Month | Day | Daily average temperature/°C |
| 50136 | 2010 | 6 | 1 | 15.8 |
| 50136 | 2010 | 6 | 2 | 17.7 |
| 50136 | 2010 | 6 | 3 | 20.4 |

Note: Only one station is selected in the table, and the data of three days is presented. The original data are from 235 stations with 18 250 d.
Figure 1  Spatial distribution of meteorology station samples

2.2 Calculation of accumulated temperature characteristics

Because only the temperature stably higher than 0°C is significant for maize growth, in this study the accumulated temperature of maize planting environment is measured by the cumulative sum of the daily average temperatures which are stable above 0°C at the corresponding meteorology station. Therefore, for each meteorology station, the corresponding environmental accumulated temperature characteristics can be calculated by Equations (1)-(3),

\[ t_i = \sum_{j=B}^{E} d_j \]

\[ \mu = \frac{1}{50} \sum_{i=1}^{50} t_i \]

\[ \sigma = \sqrt{\frac{\sum_{i=1}^{50} (t_i - \mu)^2}{50 - 1}} \]

where, \( i \) is the year; \( j \) is the date; \( B \) and \( E \) represent the start and end dates respectively when the daily average temperature is stable above 0°C; \( d_j \) is the daily average temperature on date \( j \) of year \( i \), °C; \( t_i \) is the environmental accumulated temperature in year \( i \), °C; \( \mu \) and \( \sigma \) are the mean and standard deviation of environmental accumulated temperature during recent 50 years. The examples of calculated statistics (\( \mu \) and \( \sigma \)) of environmental accumulated temperature at all stations are shown in Table 3.

<table>
<thead>
<tr>
<th>Station ID</th>
<th>Mean (( \mu ))</th>
<th>Standard deviation (( \sigma ))</th>
<th>Satisfaction rate (( P ))</th>
<th>Risk level (( Y ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>58436</td>
<td>5678.56</td>
<td>193.25</td>
<td>1.0000000</td>
<td>0</td>
</tr>
<tr>
<td>58531</td>
<td>5975.34</td>
<td>209.25</td>
<td>1.000000</td>
<td>0</td>
</tr>
<tr>
<td>50844</td>
<td>3385.69</td>
<td>155.47</td>
<td>0.993446</td>
<td>1</td>
</tr>
<tr>
<td>53231</td>
<td>3216.11</td>
<td>167.18</td>
<td>0.901939</td>
<td>1</td>
</tr>
<tr>
<td>50742</td>
<td>3063.84</td>
<td>158.87</td>
<td>0.656098</td>
<td>2</td>
</tr>
<tr>
<td>50468</td>
<td>2682.53</td>
<td>143.73</td>
<td>0.013594</td>
<td>2</td>
</tr>
</tbody>
</table>

Note: the table omits the calculated environmental accumulated temperature statistics and maize chilling damage risk information of other stations.

2.3 Test of accumulated temperature probability distribution

In order to calculate the satisfaction rate of environmental accumulated temperature for maize production, it is necessary to first determine the probability distribution type of accumulated temperature. According to experience, the daily average temperature (the average value of the average temperature of many periods in a day) should follow the normal distribution. The environmental accumulated temperature can be expressed as a linear combination of the daily average temperature of each day in the growth period, so it can be inferred that the environmental accumulated temperature should also follow the normal distribution. In this study, a combination of qualitative and quantitative methods is used to test and determine the probability distribution type of environmental accumulated temperature.

Randomly take one station as an example, firstly, in qualitative analysis, from Figure 2, it can be seen that the environmental accumulated temperatures are approximately symmetrically distributed with respect to the mean value. Besides, from the normal Q-Q diagram in Figure 3, it can be clearly seen that each point is very close to the solid line. Therefore, it can be qualitatively said that the environmental accumulated temperature follows the normal distribution. Secondly, in quantitative analysis, the Kolmogorov-Smirnov test and Shapiro-Wilk test were used to determine whether the environmental accumulated temperature followed the normal distribution. All calculations were implemented in the SPSS17.0 software. When the significance level is set as 0.05, the obtained P-values of Kolmogorov-Smirnov and Shapiro-Wilk tests were 0.2 and 0.5, respectively; both were greater than the significance level 0.05, which quantitatively further showed that the normal distribution can be used to simulate the distribution of environmental accumulated temperature. Therefore, in this study, the normal distribution was used to calculate the environmental accumulated temperature satisfaction rate in the following section.

Figure 2  Histogram of environmental accumulated temperature of one example meteorology station named Changchun

Figure 3  Normal Q-Q plot of environmental accumulated temperature of one example meteorology station named Changchun

2.4 Calculation of chilling damage risk

After knowing that the environmental accumulated temperature follows the normal distribution, for each meteorology station, the probability that the environmental accumulated...
temperature satisfies maize growth in production can be calculated by Equation (4),

\[
P = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt
\]

where, \( H \) is the accumulated temperature required by maize growth, °C; \( P \) is the accumulated temperature satisfaction rate of the target planting environment corresponding to the meteorology station. In this study, in order to better control the risk, \( H = 3000 \)°C was taken conservatively according to the average value that most maize varieties need for accumulated temperature in the whole growth period.

Then, the risk level \( Y \) can be determined by Equation (5) based on the field expert experiences,

\[
Y = \begin{cases} 
0 , & \text{if } P = 1 \\
1 , & \text{if } 1 > P \geq 0.9 \\
2 , & \text{if } 0.9 > P \geq 0 
\end{cases}
\]

where, 0 represents no risk, 1 indicates low risk, 2 means serious risk, and level 0 is the basis of risk level, corresponding to the risk level information in Table 3.

2.5 Chilling damage risk prediction based on PNN model

Probabilistic neural network \([22]\), proposed by D. F. Specht in 1989, can be regarded as a radial basis function neural network, which combines density function estimation and Bayesian decision theory, having the characteristics of simple structure and wide applications \([23-25]\). The topological structure of probabilistic neural network used in this study is shown in Figure 4, which consists of four layers: input layer, pattern layer, summation layer and output layer.

![Figure 4](https://via.placeholder.com/150)

**Figure 4** Topology of the constructed PNN model

Suppose the feature dimension of each sample be \( L \) in this study \( L = 3 \) (latitude, longitude, altitude); the number of testing sample be \( M \), in this study \( M = 77 \); the number of training sample be \( N \), in this study \( N = 154 \); the number of risk levels is \( S \), in this study \( S = 3 \) (risk level 0, 1 and 2). In addition, let the testing sample set be \( X = \{X_i\} = \{X_0\} \), where \( X_0 \) is the \( i \)th testing sample, \( X_0 \) is the \( i \)th feature of \( X_i \), \( 1 \leq i \leq M \); training sample set be \( T = \{T_i\} = \{T_n\} \), where \( T_n \) is the \( n \)th training sample, \( T_n \) is the \( n \)th feature of \( T_n \), \( 1 \leq n \leq N \); Risk level set is \( C = \{C_i\} \), where \( C_i \) is the \( i \)th risk level, \( 1 \leq i \leq S \).

Then the number of neurons in the input layer is equal to the sample feature dimension, namely there are three neurons in input layer in this study as shown in Figure 4. The number of neurons in the pattern layer is equal to the number of training samples. Pattern layer is essentially radial base layer, where each neuron corresponds to a radial base center (also corresponding to a training sample). So for any input vector (testing sample) \( X_0 \), the output of the \( i \)th neuron in pattern layer is determined by Equation (6),

\[
G_i = \exp\left(-\frac{||X_i - T_i||^2}{2\sigma^2}\right)
\]

where, \( G_i \) is the probability density of the \( i \)th neuron in pattern layer; \( ||X_i - T_i|| \) is the distance between the \( i \)th testing sample and the \( i \)th training sample; \( \sigma \) is a smoothing factor which is a hyperparameter of the model and is usually a constant value between 0, 1).

The number of neurons in the summation layer is equal to the number of risk levels, namely there are three summation neurons in this study as shown in Figure 4. The summation layer calculates the weighted summation output of the neurons belonging to the same risk level in pattern layer as shown in Equation (7), where \( V_n \) is the output value of the \( n \)th neuron in summation layer corresponding to \( X_0 \) and \( N_n \) is the number of neurons belonging to the \( n \)th risk level in pattern layer.

\[
V_n = \sum_{i=1}^{N_n} G_i
\]

The number of neurons in the output layer is 1. Output layer firstly normalizes the output of the summation layer to get the probabilities that the input sample \( X_0 \) belongs to different risk levels as shown in Equation (8), and then judges the final predicted risk level that the input test sample belongs to as shown in Equation (9).

\[
\hat{Y} = \text{arg max}(V_n')
\]

Considering that the value of different features of input variable varied greatly, in order to eliminate the impact of inconsistent measurement units, for all samples each feature value was pre-processed by standardization method before being fed into the model. Furthermore, in order to compare the effects of different standardization methods on prediction result, two different standardization methods that are shown as Equations (10)-(11) were adopted in data pre-processing stage, where \( \text{MIN}(\cdot) \) and \( \text{MAX}(\cdot) \) were the minimization function and maximization function, respectively; \( \text{MEAN}(\cdot) \) and \( \text{STD}(\cdot) \) were the mean value and standard deviation value, respectively. So all the values inputted into the model were actually the standardized values, namely \( X' \). Here, Equation (10) was defined as MAX-MIN method and Equation (11) as MEAN-STD method.

\[
X' = \frac{X - \text{MIN}(X)}{\text{MAX}(X) - \text{MIN}(X)}
\]

\[
X' = \frac{X - \text{MEAN}(X)}{\text{STD}(X)}
\]

Finally, the accuracy of classification predicted by the model is determined by Equation (12), where \( M \) is the total number of testing samples; \( S \) is the number of risk levels, and \( T_n \) is the number of samples correctly classified into the \( n \)th risk level.
3 Results and discussion

All data processing and model calculation in this study were implemented in Python 3.7. Firstly, in order to determine the optimal value of the hyperparameter $\sigma$, the value of 0.05 was taken as step and investigated the change of the model accuracy when $\sigma$ varied in the interval $(0, 1)$ with different step values. The difference caused by MAX-MIN pre-processing method and MEAN-STD pre-processing method with different values of $\sigma$ is shown in Figure 5, from which it can be seen that the accuracy of MEAN-STD pre-processing method is generally higher than that of MAX-MIN method. By MEAN-STD pre-processing method, the accuracy of the model is basically stable with little fluctuation, and it also has good precision performance for different $\sigma$ values. However, the accuracy by MAX-MIN pre-processing method varies greatly when $\sigma$ is greater than 0.25, and the overall performance is relatively low and unstable. Therefore, in this experiment, the MEAN-STD method was finally chosen for pre-processing the original data before inputted into the model. In this case, the relatively optimal value of the hyperparameter $\sigma$ is 0.8 corresponding to the highest precision, which can be clearly seen from Figure 5.

When the optimal pre-processing method was determined as MEAN-STD and the optimal value of hyperparameter $\sigma$ was chosen as 0.8, the final predicted classification results of the model could be obtained as shown in Figure 6, and the prediction details of each risk level could be got as shown in Table 4. From Figure 6, it can be seen that for risk level 0 and risk level 2, the number of predicted results from the model is less than the actual sample number, while for risk level 1, the number of predicted results is more than that of observation. This is a phenomenon of classifying the cases in the two end levels into the middle level, which shows that the prediction trend of the model is somewhat conservative.

According to Table 4, it can be calculated and obtained that the overall accuracy of the model for the three risk levels is 0.81 ($(27+14+21)/77$). This performance is generally good in the field of agricultural disaster research where the influencing factors are very complex and usually with strong uncertainty. Besides, for the risk level 2 (the most serious), the accuracy of the model is as high as 0.91 (21/23), which is a very good performance in this field. So the prediction performance of the model is acceptable especially for cases with serious risk levels.

\[
\text{precision} = \sum_{\text{level}} \frac{T_p}{M} \quad (12)
\]

4 Conclusions

Maize chilling damage by low temperature is a serious disaster in maize production. Prediction of its risk is a typical non-linear complex issue. Probabilistic neural network is a kind of classification neural network with simple structure, easy algorithm and good performance. It can use linear learning algorithm to achieve the function of non-linear approximation, allows increasing and good performance. It can use linear learning algorithm to achieve the function of non-linear approximation, allows increasing and good performance. What’s more, according to Table 4, it can be obtained that the ratio of Type I Error made by the model (classifying the non-risky samples as risky ones or the less risky samples as more serious ones, shown as the bold and underlined numbers in the upper right corner of Table 4) is 0.1 ($(7+0+1)/77$). The ratio of the Type II Error made by the model (classifying the risky samples as risk-free or risk-less ones, shown as the bold and italic numbers in the lower left corner of Table 4) is 0.09 ($(5+0+2)/77$). Considering the impact of the two types of errors on maize production, the second type of error is more concerned. That is to say, it is much more dangerous to mistake a risky sample as a risk-less one than to mistake a risk-less one as a risky one. The rate of second type error of the model is smaller than that of the first type error, so the model is of practical value in predicting the risk of maize chilling damage, which is helpful for risk control in agriculture production. In addition, according to Table 4, through further analysis, it can be known that the model does not produce the phenomenon of over-level prediction errors, namely the sample of level 0 is not mistakenly predicted to be level 2 and the sample of level 2 is also not mistakenly predicted to be level 0. This implies that the model has good application value.

Through empirical analysis and data calculation, the following
concrete conclusions were obtained: Firstly, the environmental accumulated temperature during maize growth period, which was used to calculate maize chilling damage risk, followed the normal distribution, and the risk of chilling damage could be calculated by probability model of normal distribution. Secondly, to predict the risk of maize chilling damage by probabilistic neural network, the original data were standardized by different pre-processing methods, and the accuracy of MEAN-STD method was higher than that of MAX-MIN method. Thirdly, the optimal value of the model hyperparameter, namely smoothing coefficient $\sigma$, was obtained as 0.8 by trial-and-error method. Fourthly, the overall accuracy of the model was 0.81, and especially the prediction accuracy of the risk level 2 (the most serious level) was as high as 0.91. Lastly, the rates of the Type I Error and Type II Error made by the model were 0.1 and 0.09, respectively. The model had a smaller Type II Error and did not produce over-level prediction errors.

In summary, all these finds indicate that the risk prediction model of maize chilling damage in growth period based on probabilistic neural network has very good prediction performance and credibility, which is conducive to chilling damage risk control in maize production. It has good academic and theoretical reference significance for transforming post-disaster evaluation to pre-disaster prediction and expanding new ideas for the risk study of maize chilling disaster. The model is also of great practical significance for supporting scientific maize planting plans, making early disaster prevention and mitigation measures, and ensuring the safety of our grain production. In the next step, the real-time assessment and dynamic prediction of maize chilling damage risk based on adaptive probabilistic neural network algorithm will be further studied under the background of big data and the context of global climate change.

Acknowledgements

This study is supported by the general program of Humanities and Social Sciences Research of the Ministry of Education of China (Grant No. 19YJC880064), the Hunan Provincial Education Department (Grant No. 19B447), and the Hunan Provincial Natural Science Foundation (Grant No.2017J33252). This work is also supported in part by the Huaihua University Double First-Class initiative Applied Characteristic Discipline of Control Science and Engineering.

References