International Journal of Research in Agricultural Sciences
Volume 5, Issue 2, ISSN (Online): 2348 – 3997

Plant Disease Forecasting Based on Wavelet Transformation and Support Vector Machine

Hong Wang[1*], Shanwen Zhang[2], Yu Shao[2] and Yunlong Zhang[2]
[1.] College of Science, Zhongyuan University of Technology, Zhengzhou, 450007, China.
[2.] SIAS International University, Zhengzhou University, 451150, China.
[*]Corresponding author email id: 921127911@qq.com

Date of publication (dd/mm/yyyy): 24/03/2018

***Abstract* – A forecasting method of plant disease based on wavelet transformation (WT) and Support Vector Machine (SVM) is introduced. The environment information data is essentially an unstationary time sequence, which can be decomposed into different frequency channels by WT and obtain the forecasting features. The disease can be forecasted by SVM. The average forecasting precision was over 86%. Experimental results on three common kinds of cucumber diseases show that the proposed method is more effective for plant disease forecasting.**

***Keywords* – Plant Disease Prediction; Environment Information; Wavelet Transformation (WT); SVM.**

I. Introduction

There are two of the main practical plant disease monitoring approaches including remote-sensing plant disease monitoring and on-farm plant disease monitoring. Remote-sensing plant disease monitoring uses remote-sensing technologies and devices such as satellites and aerial vehicles to extract the key parameters of agricultural environment information [1-6]. The remote-sensing technology is effective on regional-scale plant disease monitoring, but difficult to collect real-time environment information in farmlands. On-farm monitoring uses a variety of sensors deployed in farmlands as well as corresponding data processing and transmission equipment to dynamically acquire on-farm information [7-12]. IOT represents a vision in which the Internet extends into the real world embracing everyday objects [13, 14]. Physical items are no longer disconnected from the virtual world, but can be controlled remotely and can act as physical access points to Internet services. IOT makes computing truly ubiquitous a concept initially put forward by Mark Weiser in the early 1990s [2]. This development is opening up huge opportunities for both the economy and individuals. IOT is grounded in the belief that the steady advances in microelectronics, communications and information technology. In fact, due to their diminishing size, constantly falling price and declining energy consumption – processors, communications modules and other electronic components are being increasingly integrated into everyday objects today. Now, IOT technology based on the sensor nodes has been widely used in the field of agricultural environment monitoring in recent years. One logical development of the IOT is used to collect the environment information for forecasting plant diseases. However, when a regional-scale plant disease monitoring is carried out by using the IOT technology, there are some problems in the distribution of plant disease monitoring sites and the space representation of plant disease monitoring data. Aiming at the problems, an approach to evaluation of agricultural environment from on-farm sites to regional scales and the corresponding solution are proposed in this paper. Most of the factors of agricultural environment remain minor variation and appear to be uniform in the same ecological type region, except for extreme weather events or geographical conditions. Therefore, the agricultural environmental information acquired by an on-farm plant disease monitoring site could be regarded as the representation of the entire farm where the site located, and the agricultural environmental information of a representative farm could be also considered as the reflection of an entire ecological type region with similar geographical and environmental conditions. An IOT-based plant disease monitoring system has been put into application in the past five years which includes a total of 110 on-farm plant disease monitoring sites in representative regions of wheat belts across 14 provinces of China and a central plant disease monitoring platform. Based on the established central plant disease monitoring platform, a regional-scale plant disease monitoring system for agricultural environment is designed by integrating with the spatial analysis technology of Web GIS. In the system, three types of Web GIS-based functions were implemented, including site data positioning, regional-scale evaluation, and thematic mapping [15-17]. The RBF network is one novel effective forward neural network, it realizes through the nonlinear primary function's linear combination from space RN to the space RM nonlinear transformation, BFm and GM (1,1) combined model, especially qualify in nonlinear time series plant disease forecast [18]. Through an interactive online map, positioning, real-time display and query of on-farm sites information and collected data are realized in the site data positioning. The regional-scale evaluation is used for analysis of the collected agricultural environmental information, and decision-making on crop growth conditions and meteorological disasters from on-farm sites to regional scales, while the thematic mapping is used for automatic generation, dynamic display and query of regional-scale plant disease monitoring thematic maps of crop growth conditions and meteorological disasters based on the evaluation [19]. Wolf et al. [21] proposed a disease cycle approach to plant disease prediction method. Goodell et al. [22] proposed an integrated expert system for cotton production and management. Shi et al. [23] proposed an IOT application to monitoring plant disease and insect pests. The fusion analysis of on-farm IOT data and “spatial” Web GIS data is provided in accordance with the characteristics of monitored objects. In this paper, a plant disease forecasting method is proposed based on environment information, WT and SVM.

International Journal of Research in Agricultural Sciences
Volume 5, Issue 2, ISSN (Online): 2348 – 3997

II. Wavelet Transform (WT)

WT is a rapid development domain in modern mathematics and is widely applied in the signal analyzing, data forecasting and imagery processing [24]. In practical work, plant disease forecasting is extremely important, it plays important role in agriculture company's localization and market decision-making. Essentially speaking, the corporate management data, such as sales volume, profit and disbursement and so on all is one kind of time series. They has the same characteristic as the usual analysis signal, these data can be forecasted through the wavelet analyzing. But the corporate management data stochastic undulation is very big, such as in this article sales volume achievement, the stability is very bad, it belongs to the model a non - steady sequence kind, this time these traditional forecast method effects are not very good Therefore, so this article introduces the wavelet analysis method may decompose the signal wavelet to the different frequency channel. Because decomposes after the signal to be more unitary than on the frequency component the primary signal, and the wavelet decomposed to the signal has made smooth processing [25], so after decomposes, the time series stability is better than primitive time. After carrying wavelet decomposition on the some non-steady time series, it may be treat as the steady time series to process in the approximate significance, like this can use some traditional the forecast method to decompose after the time series to carry on the forecast, Moreover through the example proved this kind of forecast method effect is good. Usually, the wavelet decomposition and the heavy construction may realize through the Mallat algorithm. Supposes $\{\psi_{j,n,n\in z}\}$ $\{V_j\}_{j\in z}$ is $L^2(R)$ center more than criteria analyzes φ is the criterion function, $\{\psi_{j,n,n\in z}\}$ is the wavelet base [26], then influential the solution through the Mallat algorithm:

$$\begin{cases} c_k^{j+1} = \sum_{l\in Z} c_l^j \langle \varphi_{j+1,k}(x), \varphi_{j,l}(x) \rangle \\ d_k^{j+1} = \sum_{l\in Z} c_l^j \langle \psi_{j+1,k}(x), \varphi_{j,l}(x) \rangle \end{cases} \tag{1}$$

The heavy construction type is,

$$c_k^j = \sum_{l\in Z} c_l^{j+1} \langle \varphi_{j,k}(x), \varphi_{j+1,l}(x) \rangle + \sum_{l\in Z} d_l^{j+1} \langle \varphi_{j,k}(x), \psi_{j+1,l}(x) \rangle \tag{2}$$

Among them, $\{h_k\}_{k\in z} \in l^2(Z)$ is produced by type $\frac{1}{\sqrt{2}}\varphi\left[\frac{x}{2}\right] = \sum_k h_k^{\varphi}(x-k)$, May regard as the low pass filter coefficient; $\{g_k\}_{k\in z} \in l^2(Z)$ is produce by type g_k=$(-1)^k h_{1-k}^*$, regarding as high passes the filter coefficient, c_k^{j+1} is the $\frac{1}{\sqrt{2}}\varphi\left[\frac{x}{2}\right] = \sum_k h_k^{\varphi}(x-k)\, c_l^i$ approximate signal, d_k^{j+1} is the c_l^j detail signal.

III. Plant Disease Forecasting Based on WT

The environment information can be regards as one kind of time series, but it is non-steady time series, then namely makes $x_1, x_2, ..., x_n$ is non- steady time series, records is c^0. Carries on the data using the wavelet before the forecast, WT completes following two steps:

(a) Selects suitable wavelet criterion function to carry on the decomposition. In actual forecast process, may act according to the different question choice different wavelet female function, at the same time unifies the consideration different wavelet female function the different characteristic to forecast the value the influence, Compares various wavelets function processing signal through the analysis the result and compares with the theory result, determined the wavelet function quality with the error which elects. Table 1 has listed the decision wavelet function quality essential property and some commonly used wavelet function nature.

(b) Definite the wavelet decomposes suitable layer. The greatest criterion determination will be advantageous in grasps this forecast. The more criterion is big, the more computation work load is also bigger, the error also can increase, but, the ruler goes past is greatly more advantageous to from the deeper level clear signal trend analysis, it can cause the time series to be steadier, in the actual process, the greatest criterion J determination usually is to the signal under different criterion forecast error which studies carries on the contrast, will forecast in the erroneous root-mean-square value the smallest wavelet decomposition layer determination for the decomposition greatest criterion. When the time series data's quantity which will be forecasted is not very big, the decomposition layer generally is the 3~5 level.

Table 1 Commonly wavelet function's some nature

Wavelet function	Orthogonal	Tight structure	Structure length	Symmetry
Haar	Y	Y	1	Y
Daubechies	Y	Approximate	2N-1	Y
Symlets	Y	Approximate	2N-1	Y
Meyer	Y	Limited length	no	Y

After these two determinations, then uses the Mallat algorithm, selects a suitable wavelet female function, decomposes c^0 as follows to J wavelet and makes the wavelet decomposition to the primitive sequence, and uses vector form,

$$\begin{cases} c^{j+1} = Hc^j \\ d^{j+1} = Gd^j \end{cases} \qquad j = 0,\ 1\cdots \tag{3}$$

International Journal of Research in Agricultural Sciences
Volume 5, Issue 2, ISSN (Online): 2348 – 3997

Then, c^0 is decomposed into $d_1, d_2, \cdots d_J\ c_J$. To decomposes after using the Matlab software the signal to carry on the coefficient heavy construction, obtains each coefficient vector, $d_1, d_2, \cdots$ moreover $d_1 = \{d_{11}, d_{12}, \cdots \cdots \cdots \cdots$ so $x = d_1 + d_2 + \cdots c_J$. $d_1, d_2, \cdots$ is the first signal respectively, the second $, \cdots$ the Jth detail signal heavy construction result, $x_{J,i} = d_{1,i} + d_{2,i} + \cdots c_{J,i}$ is the heavy construction result of the Jth approaches the signal, so $x_{J,i} = d_{1,i} + d_{2,i} + \cdots c_{J,i}$.

Let $x_1, x_2, ..., x_l$ be a measured time series. The object data is to predict the value of x_{k+p} using all the observations until the instant *k*. For this purpose, a functional relationship that maps the vector $[x_1, x_2, ..., x_l]$ and the value x_{k+p} is to be constructed with the principal concern of minimizing the prediction error. The optimal prediction sequence $x'_{p+1}, x'_{p+2}, ...$ minimizes the expectation (or generalization risk).

$$C = \lim_{T \to \infty} \frac{1}{T} \sum_{k=1}^{T} E\{(x'_{p+1} - x_{p+1})^2 | x_k, x_{k-1}, ...\} \tag{4}$$

If the time series is random, it is given by $x'_{k+p} = E\{x_{k+p} | x_k, x_{k-1}, ...\}$, which cannot be computed since the information in the time series is limited to l measurements. The criterion to be minimized is the following empirical risk:

$$C_{emp} = \frac{1}{l} \sum_{i=1}^{l} (x_i - x'_i)^2 \tag{5}$$

All the expectations are omitted if the time series is deterministic. The relationship between x'_{k+p} and the sequence $x_k, x_{k+1}, ...$ are supposed to be nonlinear of an unknown nature.

On the other hand, two problems arise. The model's order choice: When the order is small, the training is simple but the information (about the past) is not enough to predict accurately. However, if the order is high, the training is hard and the approximation error decreases slowly. This is the curse of dimensionality.

According to wavelet transform theory, the data $x(n)$'s wavelet coefficient reconstruction can be replaced by the following finite sum.

$$x(n) = \sum_{j=0}^{J-1} \sum_{k \in Z} c_{j,k} \psi_{j,k}(n) \tag{6}$$

Eq. (6) means to project the input signal $x(n)$ into corresponding scale's orthogonal subspace so as to recur data series in different distinguishing level. The series $x_j(n) = \sum_{k \in Z} c_{j,k} \psi_{j,k}(n)$ is $x(n)$'s projective discrete form in wavelet subspace. The purpose of data forecast is to produce the discrete reconstruction of $x_j(n)$ $(j = 0, 1, \cdots J-1)$. $v_j(n)$ is the approximation of projection x_j:

$$v_j(n) = \sum_{k \in Z} \hat{c}_{j,k} \psi_{j,k}(n) \tag{7}$$

where, $\hat{c}_{j,k}$ is wavelet coefficient $c_{j,k}$'s discrete approximation.

$$\hat{c}_{j,k} = \sum_{l} x(l) \overline{\psi}_{j,k}(l) \tag{8}$$

Eq. (8) is put into Eq. (7), next result will be gained:

$$v_j(n) = \sum_{l} x(l) r_j(l, n) \tag{9}$$

where, $r_j(l, n) = \sum_{k \in Z} \overline{\psi}_{j,k}(l) \psi_{j,k}(n)$.

Eq. (9) is approximation Eq. of $x_j(n)$.

Under the supposition of time-steadiness and orthogonal, Eq. s (10) and (11) can be gained:

$$r_j(l, n) = r_j(l - n), \quad r_j(m) = r_0(2^j m) \tag{10}$$

Then

$$v_j(n) = \sum_{l} x(l) r_j(l - n) \tag{11}$$

The discussion above is about the signal demonstration pf discrete wavelet transformed. Next the discrete wavelet transforms adaptive LSM algorithm is to be deduced. If

$$\mathbf{V}(n) = [v_0(n), v_1(n), \cdots, v_{J-1}(n)]^T \tag{12}$$

$$\mathrm{x}(n) = [x(n), x(n-1),\ x(n-2), \cdots \quad V+1)]^T \tag{13}$$

$$[\mathbf{W}]_{jm} = r_j(m), \quad j = 0, 1, \cdots \quad \cdots \tag{14}$$

$$\mathbf{B}(n) = [b_0(n), b_1(n), \cdots, b_{J-1}(n)]^T \tag{15}$$

Then

$$\mathbf{V}(n) = \mathbf{W}\mathbf{x}(n) \tag{16}$$

Where, $\mathbf{W}$ is the wavelet transform $J \times N$ matrix, $\mathbf{V}(n)$ is the input signals after passing through discrete transform filter.

In the structure of LMS, filter $r_j(n)$ $(j = 0, 1, \cdots J-1)$ and multi-group delay line coefficient $b_j(n)$ $(j = 0, 1, \cdots J-1)$ constitute expected signal $d(n)$'s predictor. Such prediction is based on input $x(n)$'s J continuous linear combination. The adaptive filtering output signal is:

$$\begin{aligned} y(n) &= \mathbf{V}^T(n)\mathbf{B}(n) = \sum_{j=0}^{J-1} v_j(n) b_j(n) \\ &= \sum_{j=0}^{J-1}\sum_{i=0}^{N-1} x(n-i) r_j(i) b_j(n) = \sum_{i=0}^{N-1} \alpha_i x(n-i) \end{aligned} \tag{17}$$

where $\alpha_i = \sum_{j=0}^{J-1} r_j(i) b_j(n)$.

The adaptive filtering error signals are:

$$e(n) = d(n) - y(n) \tag{18}$$

Coefficient update Eq. is:

$$\mathbf{B}(n+1) = \mathbf{B}(n) + 2\mu e(n)\mathbf{V}(n) \tag{19}$$

The algorithm convergence condition is

$$0 < \mu < \frac{1}{\lambda_{V\max}} \tag{20}$$

To make R_{VV} is $\mathbf{V}(n)$'s self-correlation matrix, $\mathbf{R}_{\mathrm{V}d}$ is $\mathbf{V}(n)$ and expected signal $d(n)$'s cross correlation matrix. So $\lambda_{V\max}$ in Eq. (21) is R_{VV}'s maximum eigen-value.

$$\mathbf{B}(n+1) = \mathbf{B}(n) + 2\mu e(n)\mathbf{D}_w^{-1}\mathbf{V}(n) \tag{21}$$

IV. Experiments and Analysis

In cooperation with Northwest Agriculture and Forestry University, Shaanxi Yangling agricultural demonstration zone and Zhuque national Forest Park and other units, related to soil information, meteorological information and disease information networking sensor acquisition and utilization of agricultural crop diseases, in plant protection under the guidance of experts to build a plant disease management database. Collect related environmental information and diseases mainly include: soil information (region, soil temperature, soil moisture, soil moisture, soil salinity, soil, or even the soil pH value and microbial content), meteorological data (air temperature, air humidity, light intensity, photosynthetic active radiation, precipitation, rainfall, air pressure, wind speed and direction, the concentration of carbon dioxide etc.) and disease information (pesticides, disease types, seasonal incidence and disease grade). The environment information is formed as an environment information time series. We implement WT on this time series, and obtain a feature vector. The plant diseases can be forecasted by SVM.

The cucumber disease prediction problem, from the above 3 aspects were selected in recent 6 years, 3 kinds of common diseases (effect of cucumber downy mildew, leaf spot and anthracnose) 13 main factors: the occurrence and development of soil temperature, soil moisture, soil salinity, soil, microbial content, even if the air temperature, air humidity light intensity, precipitation, rainfall, carbon dioxide concentration, the use of pesticides and the incidence of the season. Because these factors are different in data dimension, scope, presentation, physical meaning and order of magnitude, they need to be quantified or discretized, standardized and normalized preprocessed to meet the comparability between different types of data. For non-digital raw data, it needs to be converted and hierarchical, and all the data are normalized.

Table 2. Forecasting correct rates and variances of three diseases of cucumber by three methods.

Disease	Forecasting rate（%）		
	SFSVM	BPNN	WT
Downy mildew	64.12 ± 3.28	73.44 ± 3.63	75.46 ± 3.48
Leaf spot	66.57 ± 3.74	71.38 ± 3.12	76.05 ± 3.21
Anthracnose	53.45 ± 3.62	71.71 ± 3.51	78.38 ± 3.54
Average	66.84	74.82	76.36

Form Table 2, the prediction of cucumber diseases based on WT and SVM is far higher than other two models, the main reasons are: the process of WT prediction model can imitate human thinking, learning and experience based on depth can occur from a large number of complex diseases of environment information mining classification and forecasting and, according to the characteristics of the information, judgment, reasoning, in order to get better prediction results. The prediction results show that the WT model has good characteristic learning performance in prediction of crop disease based on environmental information.

V. Conclusions

In practical work, forecasting of plant disease is extremely important, it plays important role in agriculture. Essentially speaking, the corporate environment information, such as soil temperature, soil moisture, soil moisture, soil salinity, has the same characteristic as the usual signal analysis, and these data can be used to forecast the plant diseases through the wavelet analyzing. In this paper, a plant disease forecasting method was proposed and the results showed that it is effective and feasible.

Acknowledgment

This work is partially supported by the scientific and technological research project of Henan provincial science and Technology Department (Nos. 182102210547, 182102210544 and 172102210109). The authors would like to thank all the editors and anonymous reviewers for their constructive advice.

References

[1] Bourke P M A. use of Weather Information in the Prediction of Plant Disease Epiphytotics. Annual Review of Phytopathology, 1970, 8(1):345-370.
[2] Wolf E D D, Isard S A. Disease Cycle Approach to Plant Disease

Prediction. Annual Review of Phytopathology, 2007, 45(45):203.
[3] Shi M W. Based on time series and RBF network plant disease forecasting. Procedia Engineering, 2011, 15:2384-2387.
[4] Siquan Hu, Haiou Wang, Chundong She, Junfeng Wang. AgOnt: Ontology for Agriculture Internet of Things.IFIP Advances in Information and Communication Technology, 344, 2011, 131-137.
[5] Welbourne E, Battle L, Cole G, et al. Building the Internet of Things using RFID. IEEE Internet Computing, 2009, 13(3):48-55.
[6] Thiago Teixeira, Sara Hachem, Valerie Issarny, Nikolaos Georgantas. Service Oriented Middleware for the Internet of Things: A Perspective. Service Wave, 2011, LNCS 6994: 220-229.
[7] Luigi Atzori, Antonio Iera, Giacomo Morabito. The Internet of Things: A Survey. Computer Networks, 2010(54): 2787-2805.
[8] Gao Lingwang, Chen Jiguang, Yu Xinwen, et al. Research and development of the expert system platform for forecast and prediction of agricultural pests. Transactions of the CSAE, 2006, 22(10): 154- 158. (in Chinese with English abstract)
[9] Lesovoi M, Smelyanets V, Aifaoui A, et al. Logical simulation of processes at forming of winter wheat germplasm with complex resistance to insect pests and pathogens, Simulation of breeding process of winter wheat. Archives of Phytopathology and Plant Protection, 1994, 29(2):153- 164.
[10] Goodell P B, Plant R E, Kerby T A, et al. CALEX/Cotton: an integrated expert system for cotton production and management. California Agriculture, 1990, 44(5):18-21.
[11] Shi Y, Wang Z, Wang X, et al. Internet of Things Application to Monitoring Plant Disease and Insect Pests. International Conference on Applied Science and Engineering Innovation, 2015: 31-34.
[12] Sannakki S, Rajpurohit V S, Sumira F, et al. A neural network approach for disease forecasting in grapes using weather parameters [C]. International Conference on Computing. IEEE, 2013:1-5.
[13] Siquan Hu, Haiou Wang, Chundong She, Junfeng Wang. AgOnt: Ontology for Agriculture Internet of Things. IFIP Advances in Information and Communication Technology, 344, 2011, 131-137.
[14] Welbourne E, Battle L, Cole G, et al. Building the Internet of things using RFID. IEEE Internet Computing, 2009, 13(3):48-55.
[15] Thiago Teixeira, Sara Hachem, Valerie Issarny, Nikolaos Georgantas. Service Oriented Middleware for the Internet of Things: A Perspective. Service Wave, 2011, LNCS 6994:220-229.
[16] Luigi Atzori, Antonio Iera, Giacomo Morabito. The Internet of Things: A Survey. Computer Networks, 2010(54): 2787-2805
[17] Atzori L, Iera A, Morabito G. The internet of things: A survey. Computer Networks, 2010, 54(15): 2787－2805.
[18] Shi M W. Based on time series and RBF network plant disease forecasting. Procedia Engineering, 15 (2011) 2384-2387.
[19] Sladojevic S, Arsenovic M, Anderla A, et al. Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification. Computational Intelligence and Neuroscience, 2016, 2016(6):1-11.
[20] Bourke P M A. Use of Weather Information in the Prediction of Plant Disease Epiphytotics. Annual Review of Phytopathology, 1970, 8(1):345-370.
[21] Wolf E D D, Isard S A. Disease Cycle Approach to Plant Disease Prediction. Annual Review of Phytopathology, 2007, 45(45):203.
[22] Goodell P B, Plant R E, Kerby T A, et al. CALEX/Cotton: an integrated expert system for cotton production and management. California Agriculture, 1990, 44(5):18-21.
[23] Shi Y, Wang Z, Wang X, et al. Internet of Things Application to Monitoring Plant Disease and Insect Pests. International Conference on Applied Science and Engineering Innovation, 2015:31-34.
[24] Zhouchen Li, Jianshu Luo. Wavelet analysis and its application. Beijing, Electronics industry publishing company, 2003, 462-465.
[25] Zhao kai, Zonghua Wang. Wavelet analysis and its analytical chemistry application. Beijing: Geological publishing house, 2000, 102-104.