

doi: 103969/j. issn. 0490-6756. 2016. 03. 006

# 基于相空间重构和支持向量机的小麦条锈病预测方法

刘 诚, 熊万丹, 付蜀智

(四川农业大学, 都江堰 611830)

**摘 要:** 小麦条锈病预测对小麦灾害的预防和控制具有十分重要的意义. 本文使用相空间重构和最小二乘支持向量机相结合的方法对成都市小麦条锈病发病率进行预测. 实验中使用成都地区的气象数据作为模型的输入参数, 发病率作为模型的输出参数, 对 20 例发病病例样本进行预测, 得到了比较好的预测效果. 实验中还与简单最小二乘支持向量机预测结果作了对比, 结果表明, 相空间重构和最小二乘支持向量机结合的模型预测准确率明显高于简单最小二乘支持向量机, 因此本模型对小麦条锈病预测是十分有效的.

**关键词:** 小麦条锈病; 相空间重构; 支持向量机

**中图分类号:** O29

**文献标识码:** A

**文章编号:** 0490-6756(2016)02-0270-05

## Method of wheat stripe rust predicting base on phase space reconstruction and support vector machine

LIU Cheng, XIONG Wan-Dan, FU Shu-Zhi

(Sichuan Agricultural University, Dujiangyan 611830, China)

**Abstract:** Wheat stripe rust prediction has very vital significance for the prevention and control of wheat disaster. This paper predicted the wheat stripe rust incidence in Chengdu using PSR and LSSVM. The Meteorological data was as input parameters, and the incidence rate was as the output parameter in the test. To predict the 20 cases samples, this paper got better predictive effect. With comparison, the method in this paper is much better than SVM. So the model in this paper is very effective for wheat stripe rust prediction.

**Key words:** Wheat stripe rust; Support vector machine; Phase space reconstruction

## 1 引 言

小麦条锈病是一种典型的叶部气传真菌病害, 是世界各小麦主产国最主要的病害之一, 也是长期影响我国小麦安全生产的严重生物灾害之一, 在我国的流行区域主要是四川、甘肃、重庆、云南、贵州等地. 近年来成都平原的气候不断变暖, 秋季多雨, 冬季多雾等都为小麦条锈病的发生和流行提供了条件, 使得成都市的小麦条锈病多次发生, 且危害严重, 在流行年份导致成都市小麦减产 10%~

20%, 特大流行年份减产高达 60% 以上, 甚至导致小麦绝收. 正是由于小麦条锈病对粮食安全和品质造成了巨大的危害, 因此预测小麦条锈病的发病率具有重要意义, 它不仅可以有效预防和控制小麦条锈病的发生, 还可以提高农业生产中的管理水平, 发展精准农业, 减少病害损失、提高农业的产量和品质.

为了对该病害及时准确预报, 前人已经有了大量的研究, 比如特尔菲法、相关回归分析, 时间序列分析等等, 但大多数是以线性分析为基础的数理统

收稿日期: 2014-3-24

基金项目: 国家自然科学基金(U1433129); 四川省教育厅青年基金(12ZB092)

作者简介: 刘诚(1982-), 男, 四川都江堰人, 讲师, 主要从事数学及其在生物学、经济学中的应用. E-mail: liucheng@sicau.edu.cn

计模型<sup>[1-8]</sup>.然而小麦条锈病的发生与流行本身十分复杂,表现出高度的非线性和多时间尺度特性,因此采用传统的预报模型并不能反映出预测过程中的不确定性和非线性.本文提出一种将相空间重构(PSR)和最小二乘支持向量机(LSSVM)结合的非线性时间序列预测建模方法,建立一个在小样本情况下的小麦条锈病预测模型.

## 2 基于 PSR-LSSVM 的小麦条锈病发病率预测模型

### 2.1 相空间重构算法原理

相空间重构是混沌时间序列预测的有效方法,可通过利用系统长期演化的任一变量时间序列来研究系统的混沌行为.从而解决了从实际的时间序列里提取非线性特征物理量的科学问题.在此基础上,可以利用事物发展的时间序列本身所计算出来的客观规律进行预测,此法可避免预测的人为主观性,从而提高预测的精度和可信度.

相空间重构理论是 Packard 等<sup>[3]</sup>人在 1980 年最先提出来的.之后, Takens 则从数学上为其奠定了可靠的基础<sup>[7]</sup>,提出通过选取最佳的时间延迟和嵌入维数,从单个变量时间序列中恢复出整个原相空间的状态轨迹.其基本原理如下:

假设时间序列为  $\{x_t | t = 1, 2, \dots, N\}$ , 引入时间延迟参数  $t$  和重构维数  $m$ , 得到:

$$X_t = \{x(t), x(t + \tau), x(t + 2\tau), \dots, x(t + (m - 1)\tau)\}.$$

根据 Takens 定理, 时间延迟参数和重构维数对于重构相空间, 更好地恢复动力学系统起着至关重要的作用. 目前, 关于这两个参数的选取方法很多, 本文选用 C-C 算法来对两个参数进行选取, C-C 算法容易操作, 计算量小, 对小数据组可靠, 具有较强的抗噪声能力. C-C 算法的原理如下:

对于单变量时间序列  $X = \{x(i) | i = 1, 2, \dots, N\}$ , 按照 Takens 定理进行延迟相空间重构, 有

$$X_i = \{x_i, x_{i+\tau}, \dots, x_{i+(m-1)\tau}\}, \\ i = 1, 2, \dots, M,$$

式中  $M = N - (m - 1)\tau$ ,  $M$  是重构  $m$  维相空间中的相点个数. 其对应的关联积分为:

$$C(m, M, r, t) = \frac{2}{M(M+1)} \sum_{1 \leq i < j \leq M} \theta(r - \|X(t_i) - X(t_j)\|) \quad (1)$$

(1) 式用于计算重构相空间中任意两点之间的距离  $d_{ij} = |X_i - X_j|$  小于搜索半径  $r$  的概率;  $\theta(x)$  是 Heaviside 函数, 当  $x < 0$  时,  $\theta(x) = 0$ , 当  $x > 0$  时,  $\theta(x) = 1$ .

为了描述非线性时间序列的相关性, 引入关联积分的 BDS 统计量:

$$S_1(m, n, r, t) = C(m, n, r, t) - C^m(1, n, r, t) \quad (2)$$

计算(2)式时需要把原始时间序列拆分成  $t$  个互不相交的子时间序列, 然后采用分块平均策略, 分别计算每个时间序列的检测统计量  $S_2(m, n, r, t)$ , 即当  $n \rightarrow \infty$  时, 有

$$S_2(m, r, t) = \frac{1}{t} \sum_{s=1}^t [C_s(m, r, t) - C_s^m(1, r, t)].$$

对于  $\{x_i | i = 1, 2, \dots, N\}$ , 分成  $t$  个互不相交的子序列, 则每个时间序列的子序列长度  $l = \frac{N}{t}$ , 此时

$$S(m, N, r, t) = \frac{1}{t} \sum_{s=1}^t [C_s(m, \frac{N}{t}, r, t) - C_s^m(1, \frac{N}{t}, r, t)].$$

令  $N \rightarrow \infty$ , 有

$$S(m, r, t) = \frac{1}{t} \sum_{s=1}^t [C_s(m, r, t) - C_s^m(1, r, t)] \quad (3)$$

根据(3)式, 有  $S(m, r, t)$  恒为零, 但由于实际序列是有限的, 所以实际得到的  $S(m, r, t)$  不为零, 这样局部最大时间间隔可取为  $S(m, r, t)$  的零点或对所有搜索半径差别最小的点, 即:

$$\Delta S(m, t) = \max\{S(m, r, t)\} - \min\{S(m, r, t)\}.$$

对所有的  $m$  和  $t$ ,  $S(m, r, t)$  的零交叉和  $\Delta S(m, t)$  的最小值几乎相同, 因此延迟时间  $\tau$  就选为第一次出现这些局部最优的时间.

同时, 计算所有  $S(m, r_j, t)$  的平均值, 即:

$$\bar{S}(t) = \frac{1}{mj} \sum_m \sum_j S(m, r_j, t),$$

其中  $j$  和  $m$  分别表示嵌入维数  $m$  和  $r_j$  的个数; 对平均值  $\Delta \bar{S}(t) = \frac{1}{m} \sum_m \Delta S(m, t)$  取统计量  $S_{\text{cor}}(t) = \Delta \bar{S}(t) - |\bar{S}(t)|$ , 其最小值作为延迟时间的最优值, 从而得到最佳嵌入维数为  $m = \frac{\tau_m}{\tau} + 1$ .

## 2.2 最小二乘支持向量机算法原理

设训练样本集为  $\{x_i, y_i\}, i = 1, 2, \dots, N, x_i \in \mathbf{R}^n$  为输入变量的值,  $Y_i \in \mathbf{R}$  为对应的输出变量, 对应的回归函数可以定义为:

$$f(X) = w \cdot \varphi(X) + b \quad (4)$$

其中,  $w$  为权向量,  $b$  为偏差. 根据线性模型理论, (4)式的拟合问题可描述为如下的最优化问题:

$$\begin{aligned} \min_{w, b, e} (\omega, e) &= \frac{1}{2} \omega^2 + \frac{1}{2} \gamma \sum_{i=1}^N e_i^2 \\ \text{s. t. } y_i &= w^T \varphi(x_i) + b + e_i, \\ i &= 1, 2, \dots, N \end{aligned} \quad (5)$$

其中  $w \in \mathbf{R}^n$  为权向量,  $e_i$  为误差,  $\gamma_i > 0$  为惩罚系数.

将(5)式模型变换到对偶空间, 引入 Lagrange 乘子  $a_i$ , 得到如下 Lagrange 函数

$$\begin{aligned} L(w, b, e, a) &= \frac{1}{2} \omega^2 + \\ &\frac{1}{2} \gamma \sum_{i=1}^N e_i^2 - \sum_{i=1}^N a_i (w^T \varphi(x_i) + b + e_i - y_i). \end{aligned}$$

根据 KKT(Karush-Kuhn-Tucker)条件可得:

$$\begin{cases} \frac{\partial L}{\partial w} = 0 \rightarrow w = \sum_{i=1}^N a_i \varphi(x_i), \\ \frac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^N a_i = 0, \\ \frac{\partial L}{\partial e_i} = 0 \rightarrow a_i = \gamma e_i, \\ \frac{\partial L}{\partial a_k} = 0 \rightarrow w^T \varphi(x_i) + b + e_i - y_i = 0. \end{cases}$$

消去变量  $w, e_i$  得到线性方程组:

$$\begin{bmatrix} 0 & I^T \\ I & \Omega + \gamma^{-1} I \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix},$$

其中,  $y = [y_1, y_2, \dots, y_N]^T$ ,  $a = [a_1, a_2, \dots, a_N]^T$ ,  $I = [1, 1, \dots, 1]^T$ ;  $\Omega$  为核矩阵, 其第  $i$  行第  $j$  列的元素为

$$\begin{aligned} \Omega_{ij} &= K(x_i, x_j) = \varphi(x_i)^T \varphi(x_j)^T, \\ i, j &= 1, 2, \dots, N, \end{aligned}$$

本文取

$$K(x_i, x_j) = \exp\left(-\frac{x_i - x_j^2}{\sigma^2}\right),$$

为高斯核函数. 容易验证  $\Omega$  满足 Mercer 条件, 即对任意  $\varphi(x) \neq 0$ ,  $\int \varphi(x) dx < \infty$ , 有

$$\iint K(x_i, y_j) \varphi(x_i) \varphi(y_j) dx_i dy_j \geq 0.$$

解该线性方程组, 可得 LSSVM 拟合模型为

$$y(X) = \sum_{i=1}^N a_i K(x_i, X) + b.$$

## 2.3 预测及结果分析

### (1) 预测原理

根据小麦灾害的不适定性, 首先采用相空间重构方法对小麦条锈病时间序列数据进行相空间重构, 再将重构后的数据作为 LSSVM 的训练数据输入, 其中 LSSVM 参数采用网格法优化调整.

为了验证 PSR-LSSVM 用于小麦条锈病发病率预测的可行性和有效性, 本文选取了成都市 1990 至 2009 年的小麦种植面积、发病面积、幼苗期气温、幼苗期雨日、分叶拔节期气温、孕穗抽穗期气温、孕穗抽穗期雨日, 共计 20 例样本. 以此数据为例, 对成都市小麦条锈病发病率进行预测, 并对预测结果进行分析. 在模型构建时, 以幼苗期气温、幼苗期雨日、分叶拔节期气温、孕穗抽穗期气温、孕穗抽穗期雨日为输入数据, 对应年的小麦发病率为预测值. 使用前 16 组数据训练样本, 后 4 组样本为验证样本, 即 1990 年至 2005 年的数据作为训练样本, 2006 年至 2009 年的数据作为验证样本.

### (2) 预测结果与误差分析

利用上述方法对成都市小麦条锈病发病率进行预测. 预测结果如图 1~3 所示. 其中图 1 为训练样本组预测结果, 图 2 为验证样本组预测结果, 图 3 为整个样本组的误差分布. 从预测结果可以看出, 由相空间重构和最小二乘支持向量机相结合的方法对小麦条锈病进行预测效果比较好, 且具有较强的泛化能力和鲁棒性.

为了进一步说明 PSR-LSSVM 方法的优越性, 本文将 PSR-LSSVM 方法预测精度和基本 LSSVM 方法进行比较(如图 4 和表 1 所示). 图 4 为两种方法预测误差对比分析图, 表 1 为两种方法实际数据与预测数据的相关性分析.

从图 4 和表 1 的分析中可以明显看出 PSR-LSSVM 方法的预测效果要明显优于基本 LSSVM 方法.

根据以上分析, 将 PSR-LSSVM 方法用于小麦条锈病预测具有可行性和一定的优越性. 同时也说明小麦条锈病的发生与流行是一个复杂的动力学问题, 如果不对原始数据进行相空间重构, 势必会影响预测的准确性.

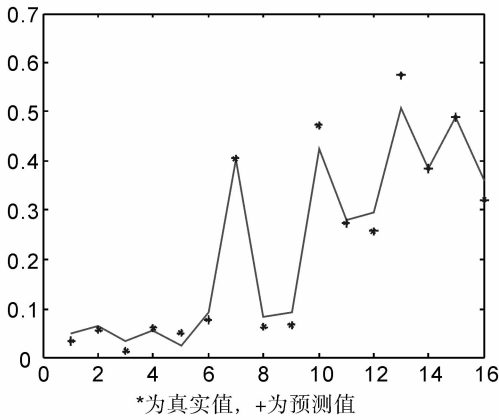


图 1 训练组预测结果

Fig. 1 The training group result based on PSR-LSSVM

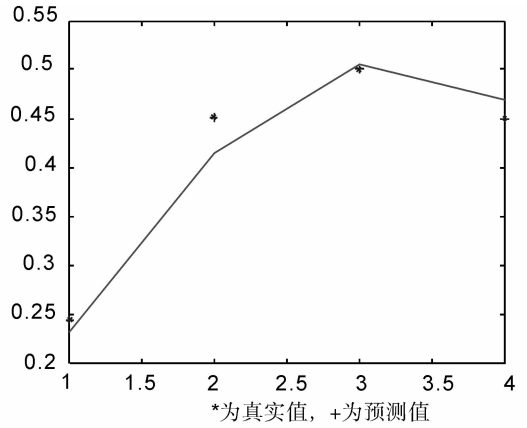


图 2 验证组预测结果

Fig. 2 The validation group result based on PSR-LSSVM

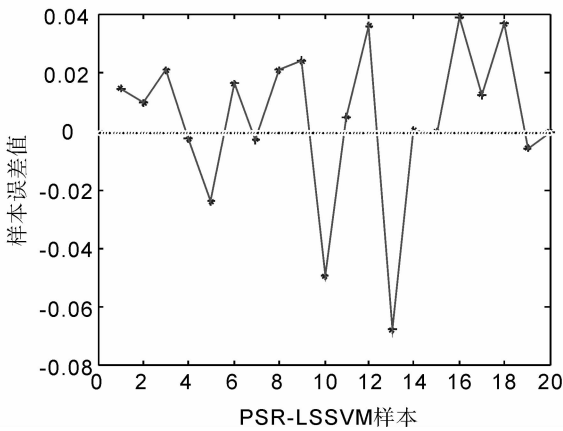


图 3 误差分布

Fig. 3 Error distribution based on PSR-LSSVM

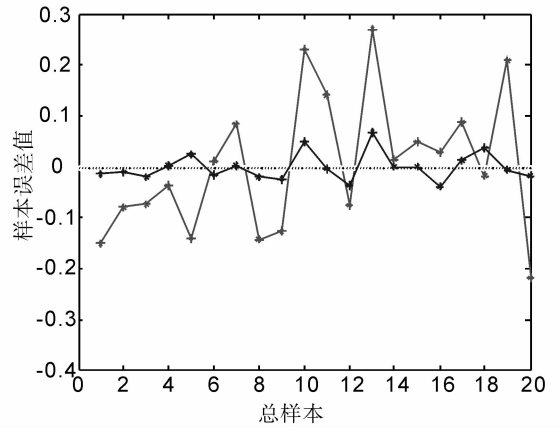


图 4 误差对比分析

Fig. 4 Error analysis comparison between LSSVM and PSR-LSSVM

表 1 LSSVM 与 PSR-LSSVM 误差和相关系数对比分析

Tab. 1 Error and correlation coefficient comparison between LSSVM and PSR-LSSVM

不同预测模型	LSSVM 模型		PSR-LSSVM 模型	
	训练样本	验证样本	训练样本	验证样本
误差分析				
均方误差(MSE)	0.0154	0.0201	0.0039	0.000057
平均绝对误差(MAE)	0.1004	0.1150	0.0461	0.0179
相关系数(R)	0.7739	0.5820	0.9718	0.9812

### 3 讨论

小麦条锈病的发生与流行是一个复杂的动力学系统. 本文提出了一种基于相空间重构和最小二乘支持向量机结合的小麦条锈病预测方法. 此算法首先考虑了小麦条锈病数据的混沌特征, 先对原始数据进行相空间重构, 再将重构数据作为最小二乘支持向量机的输入数据. 结果表明, 该方法在小麦

条锈病预测中具有很好的效果, 最后还将该方法的预测结果与基本 LSSVM 回归模型预测结果进行对比, 表面了基于相空间重构和最小二乘支持向量机相结合的预测方法的优越性. 但是由于样本数较少, 本文研究结果尚未达到统计学上的统计意义, 但在短期灾害预测中, PSR-LSSVM 方法是有效的. 基于自身的优点, 该方法对于其他灾害预测也是适用的.

## 参考文献:

- [1] 费永成, 王芬, 于成. 成都市小麦条锈病的发生与流行规律[J]. 贵州农业科学, 2011, 39(4): 100.
- [2] 易亮, 赵豫. 豫南小麦条锈病长·中·短期气象预报模型研究[J]. 安徽农业科学, 2008, 36(21): 9137.
- [3] Packard H, Crutch F P, Farmer D, Shaw S. Geometry from a time series[J]. Phys Rev Lett, 1980, 45: 712.
- [4] 汪可宁, 谢水仙, 刘孝坤, 等. 我国小麦条锈病防治研究的进展[J]. 中国农业科学, 1988, 21(2): 1.
- [5] Ustun B, Melssen W J. Determination of optimal support vector regression parameters by genetic algorithms and simplex optimization[J]. Analytical Chimica Acta, 2005, 54(4): 292.
- [6] 郑永康, 陈维荣, 戴朝华, 等. 小波支持向量机与相空间重构结合的短期负荷预测研究[J]. 继电器, 2008, 36(7): 29.
- [7] Takens F. Detecting strange attractors in fluid turbulence [J]. Lecture Notes in Mathematics, 1981, 28: 366.
- [8] 刘婷. 基于相空间重构和支持向量机的和弦识别[J]. 计算机与数字工程, 2010, 252(38): 139.