

局部支持向量回归在小麦蚜虫预测中的研究与应用

王秀美¹,牟少敏^{1*},时爱菊²,浩庆波¹

1. 271018
2. 271018

摘要:

1990~2013

82.69% 91.03%

199366 213108

196362 198780

80.77% 91.03%

关键词:

中图分类号: TP391

文献标识码: A

文章编号: 1000-2324(2016)01-0052-05

Research and Application of Local Support Vector Regression in Prediction of Wheat Aphid

WANG Xiu-mei¹, MU Shao-min^{1*}, SHI Ai-ju², HAO Qing-bo¹

1. College of Information Science and Engineering/Shandong Agricultural University, Taian 271018, China

2. College of Chemistry and Material Science/Shandong Agricultural University, Taian 271018, China

Abstract: Aiming at the accuracy of wheat aphid prediction is low, this paper proposed a short-term forecasting algorithm of wheat aphid based on local support vector regression. Firstly, feature selection was realized by correlation analysis. Secondly, the normalized processing of selected features was calculated. Finally, the short-term forecasting model was established and the prediction value of test sample was obtained by the established model. Experiments were conducted on the wheat aphid data and meteorological data of Yantai area from 1990 to 2013 year and contrast test was conducted by the standard support vector regression. The standard support vector regression achieved the Mean Square Error at 199366 in prediction and 213108 in back-substitution check, the accuracy at 80.77% in prediction and 91.03% in back-substitution check, while local support vector regression achieved the Mean Square Error at 196362 in prediction and 198780 in back-substitution check, the accuracy at 82.69% in prediction and 91.03% in back-substitution check. The results showed that local support vector regression has better performance in accuracy and generalization ability for the short-term prediction of wheat aphid.

Keywords: Local support vector regression; kernel function; correlation analysis; prediction; wheat aphid

[1]

[1,2]

[3]

[4]

2009

2

[5] 2011

(Support Vector Regression, SVR

[6]

Steinwart 2002

[7]

2006 ZHANG

[8]

[9]

[10]

收稿日期: 2014-07-12

修回日期: 2014-09-24

基金项目: (ZR2012FM024)

作者简介: (1992-), , , : E-mail: wxmsdau@163.com

*通讯作者: Author for correspondence. E-mail: msm@sdau.edu.cn

1 局部支持向量回归

$$(x_1, y_1), \dots, (x_n, y_n) \quad x_i \in R^n \quad y_i \in R$$

$$x_i \in N \quad y_i$$

$$y = f(x) \quad x \quad y$$

1.1 支持向量回归

[11] φ

$$f(x) = w \bullet \varphi(x) + b$$

$$L_\varepsilon(y) = \begin{cases} 0 & |f(x) - y| \leq \varepsilon \\ |f(x) - y| - \varepsilon & else \end{cases} \quad (1)$$

$$\min \|w\|^2 / 2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (2)$$

$$\begin{cases} y_i - f(x_i) \leq \varepsilon + \xi_i \\ f(x_i) - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \quad i = 1, 2, \dots, n \quad (3)$$

C n Lagrange KKT (Karush-Kuhn-Tueker) (2)~(3)

$$f(x) = w \bullet \varphi(x) + b = \sum_{i=1}^n (a_i - a_i^*) K(x_i, x) + b \quad (4)$$

$$f(x_i) - y_i = \varepsilon \quad b \quad 0 \leq a_i, a_i^* \leq C \quad x_i \quad y_i - f(x_i) = \varepsilon$$

$$K(x_i, x) = (\varphi(x_i) \bullet \varphi(x))$$

1.2 局部支持向量回归

[12] [13]

2007 Cheng

Localized Support Vector Machine LSVM LSVM LSVM $\sigma(x_i, x_j^*)$

x_i x_j^* σ LSVM

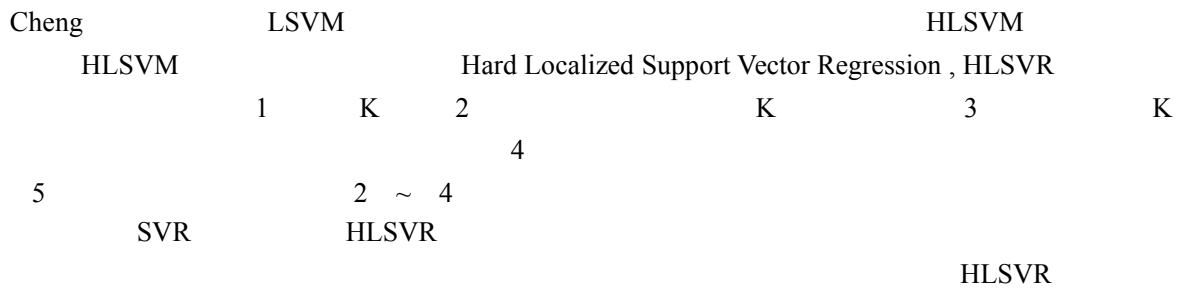
$\sigma \in [0, 1]$ LSVM SLSVM(Soft Localized Support Vector Machine

SLSVM) σ LSVM HLSVM(Hard Localized Support Vector Machine

HLSVM)

$$\sigma(x_i, x_j^*) = \begin{cases} 1 & x_i \in x_j^* \\ 0 & \end{cases} \quad (5)$$

x_j^* K



2 基于 HLSVR 的小麦蚜虫百株蚜量短期预测模型

HLSVR [2]

HLSVR

2.1 特征选择

[14]

Spearman

Pearson
Spearman

Pearson

Pearson : X_i Y Pearson $r_{X,Y}$

$$r_{X,Y} = \frac{S_{X,Y}}{\sqrt{S_{X,X_i}} \sqrt{S_{YY}}} \quad (6)$$

S_{X,X_i} S_{YY} $S_{X,Y}$ X_i Y

2.2 数据预处理

0 1

“ ”

$$x_{ij} = \frac{x_{ij} - x_i^{\min}}{x_i^{\max} - x_i^{\min}} \quad (7)$$

$$x_{ij} \quad j \quad i \quad x_i^{\max}, x_i^{\min} \quad i$$

[0,1]

3 小麦蚜虫百株蚜量预测实验及分析

3.1 数据来源

1990~2013

	1990~2007	1992~1994		2008~2013	
	2	3	4	5	5
		1		x19	1
	x1-x18	19			
2	r	P	P		

表 1 小麦蚜虫发生程度分级指标
Table 1 The classification index of wheat aphid occurrence degree

Amount of occurrence	1	2	3	4	5
(Y)	Y≤500	500 < Y ≤ 1500	1500 < Y ≤ 2500	2500 < Y ≤ 3500	Y > 3500

表 2 相关分析结果
Table 2 The result of correlation analysis

Variable	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10
r	0.00495	0.00495	0.00495	-0.03086	-0.12044	0.01201	0.23146	0.12303	-0.03104	0.0991
P	0.9602	0.9602	0.9602	0.7558	0.2233	0.9037	0.0181	0.2134	0.7545	0.3169
Variable	x11	x12	x13	x14	x15	x16	x17	x18	x19	
r	0.0991	0.0991	-0.11039	0.28725	-0.11984	0.13983	-0.07264	0.02879	0.79788	
P	0.3169	0.3169	0.2646	0.0031	0.2256	0.1569	0.4637	0.7717	<.0001	

0.5 2 x1~x4 x6 x9 x18 P

0.5 12

1990~2007	1992~1994	78	2008~2013	26
		104		

3.2 实验结果及分析

		RBF		RBF		[15]
	RBF					[16]
			C	ε	δ	
		K				K
3						

表 3 模型参数
Table 3 Model parameter

SVR			LSVR			
C	δ	ε	C	δ	ε	K
4096	9.77E-04	5	4096	9.77E-04	5	40

2008~2013 (Mean Square Error, MSE)

$$MSE = \sum_{i=1}^n (y_i - y'_i)^2 / n \tag{8}$$

$y_i \quad y'_i \quad n \quad MSE$

1990~2007
 2008~2013
 SVR

HLSVR

2008~2013

4
 SVR
 HLSVR

SVR HLSVR
 1990~2007

表 4 均方误差及发生程度准确率
Table 4 MSE and the accuracy of classification index

Model	MSE of aphids/100 plants		Accuracy of statistics	
	Prediction	Back substitution	Prediction	Back substitution
SVR	199366	213108	80.77%	91.03%
HLSVR	196362	198780	82.69%	91.03%

4
 1990~2007
 HLSVR SVR
 2008~2013
 HLSVR SVR HLSVR 26

4 结语

HLSVR

HLSVR SVR

HLSVR SVR

hadoop

参考文献

[1] [J]. ,2011(3):81-84

[2] [J]. ,2009(3):440-444

[3] [J]. ,2003(6):33-36

[4] [J]. ,2015(4):127-131

[5] [J]. ,2009(14):147-148

[6] [J]. ,2011(1):28-31

[7] STEINWART I. Support vector machines are universally consistent [J]. Journal of Complexity, 2002,18(3):768-791

[8] ZHANG H, BERG AC, MAIRE M, *et al.* SVM KNN: discriminative nearest neighbor classification for visual category recognition[C] //Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York:IEEE,2006:2126-2136

[9] HAIQIN YANG, KAIZHU HUANG, IRWIN KING, *et al.* Local support vector regression for time series prediction [J]. Neurocomputing, 2009,72(10-12):2659-2669

[10] [D]. : ,2012:26-35

[11] [D]. : ,2006:9-15

[12] [J]. ,2012(1):170-174,189

[14] CHENG H, TAN PN, JIN R. Localized support vector machine and its efficient algorithm[C]//Proc. of STAM International Conference on Data Mining 2007, Minneapolis, Minnesota, 2007:461-466

[15] [J]. ,2011(9):59-62

[16] [J]. ,2007(2):163-167

[17] [J]. ,2015(5):1287-1290,1297