Application of Projection Pursuit Regression in the Forecast of the Percentage of Damaged Seeds of Leguminivora Glycinivorella

Dan Xu and Hualing Deng

Science College Northeast Agricultural University,Harbin,150030,P.R. China

Abstract

In the paper, we build the forecast model of the percentage of damaged seeds of leguminivora glycinivorella by the way of projection pursuit regression, because the percentage of damaged seeds of leguminivora glycinivorella is affected by several factors, such as the cardinal number that leguminivora glycinivorella live through the winter and weather conditions and so on, and it is a nonlinear and abnormal system. The historical fitting rate and the forecast fitting rate are both satisfactory. So projection pursuit regression can be a new method of the insect pest forecasting.

Keywords leguminivora glycinivorella(Mats.), the projection pursuit regression (PPR), the percentage of damaged seeds

1 Introduction

Leguminivora glycinivorella is one of the major pests in soybean producing areas of Heilongjiang Province. Because of the boring of its larvae, the soybean are of inferior quality and it made the yield from soybean have a great loss. The annual percentage of damaged seeds almost changes from 10% to 20% or indeed 30% to 40%, and it fluctuates greatly depending upon the area,the year and the variety of soybean. A long-term study by experts about Leguminivora glycinivorella proves that, its percentage of damaged seeds for the current year. are mostly affected by five factors[1-2]. The first three factors are the amount of damaged seeds the average air temperature and the average rainfall in the last twenty-days of September of the last year and the average air temperature and the average rainfall of July for the current year. For the insect pest disaster systems exhibiting complex properties of irregularity, difference, diversity, sudden change and stochasticity, so the result in classical methods of multivariate statistical analysis theory are unfit for the research in this field[3-6].We can use the technique of projection pursuit regression to analyze the nonlinear and abnormal insect pest system. Based on the linear projection of datum, the technique of projection pursuit regression is used to find the nonlinear structure from linear projection, so it can solve problems of nonlinearity and abnormality to some extent[7]. In this paper, a projection pursuit regression model on percentage of damaged seeds forecast of Leguminivora glycinivorella was proposed.

2 Principle of PPR Model

In essence, projection pursuit regression is a method to process and analyze multivariate data, it converts multivariate data to low-dimensional data through linear combination by computer technology, and analyze the data structures on low-dimensional data, so as to achieve for statistical purposes easily[8]. Given the projection pursuit regression techniques can overcome the dimension disaster and can also mine and process varieties of non-normal and non-linear information in raw data adequately, so the projection pursuit regression techniques is now being applied successfully in drought forecasting, flood forecasting, air pollution concentrations forecasting and another research fields[9].

In order to avoid the conflict that the non-linear situation can not reflect by the linear regression, projection pursuit regression model takes the method of regression function approximated by the sum of a series of Ridge function. The mathematical expression of projection pursuit regression is :

$$Y = \sum_{m=1}^{M} b_m g_m(\sum_{j=1}^{p} a_{mj}^T X), (m = 1, 2, ..., M, j = 1, 2, ..., p) \tag{1}$$

Where Y is the dependent variable, M is the number of subfunction which approximate to the regression function, g_m is the m-th smooth Ridge function, b_m is the weighed value which expresses the contribution of the m-th Ridge function to the output value, a_{mj} is the j-th component of the m-th projection direction, and p is the dimension of the input space; where $\sum\limits_{j=1}^{p} a_j^2 = 1$ in formula(1).

We construct a projection pursuit regression model based on the Hermite polynomial, that is, we use alterable order orthogonal Hermite polynomial to fit one-dimensional Ridge function. Its mathematical expression:

$$h_r(z) = (r!)^{\frac{1}{2}} \pi^{\frac{1}{4}} 2^{-\frac{r-1}{2}} H_r(z)\phi(z), (-\infty < z < \infty) \tag{2}$$

where $r!$ denotes the factorial of r ,$z = a^T X$,φ is Gauss function in standard form, $H_r(z)$ is Hermite polynomial which show in recurrence form, as $H_0(z) = 1 H_1(z) = 2z \ H_r(z) = 2(zH_{r-1}(z) - (r-1)H_{r-2}(z))$. Now the projection pursuit regression model in formula (1) is changed as follow:

$$f(X) = \sum_{i=1}^{m} \sum_{j=1}^{R} c_{ij} h_{ij}(a_i^T X) \tag{3}$$

where R is the order of polynomial, c is coefficient of polynomial, h denotes orthogonal Hermite polynomial which calculated by formula (2).

The steps to construct the projection pursuit regression model based on Hermite polynomial are as follows[10]:

(1) Setting the dependent variables $y_i(i = 1, 2, ..., n)$ and the variables $x_1, x_2, ...x_p$, we can construct the data table, that is $X = (x_1, x_2, ...x_p)_{n\times p}$ and $Y = (y)_{n\times 1}$, between the dependent variables and the variables by observing n sample points. Compute the projection values by formula (4).

$$z_i = \sum_{j=1}^{p} a_j x_{ij}, (i = 1, 2, ..., n; j = 1, 2, ..., p) \tag{4}$$

where $a_j(j = 1, 2, ..., p)$ is the projection direction, x_{ij} has been normalized already.

(2)Using the orthogonal Hermite polynomial to fit the scatters points (z, y).Now the projection pursuit regression model based on Hermite polynomial is

$$\hat{y} = \sum_{i=1}^{n}\sum_{j=1}^{r} c_{ij} h_{ij}(z), (i = 1, 2, ..., n; j = 1, 2, ..., r) \tag{5}$$

Where r is the order of polynomial, c is coefficient of polynomial which we can obtained through the method of least square, and h denotes the orthogonal Hermite polynomial.

(3) Optimize the projection target function. To Optimize the projection direction and the coefficient of polynomial simultaneously. We can estimate the best value of a and c by solving the minimization problem of projection target function.

$$minQ(a, c) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y})^2 \tag{6}$$

$$s.t\sum_{j=1}^{p} a^2(j) = 1 \tag{7}$$

(4) Calculate the first fitting residuals $r_1 = y - \hat{y}$,if r_1 is smaller than a user specified threshold then stop, and output the value of a and c, else go to step (5).

(5) Replace y by r_1 and go to step (1) for the next Ridge function until meet the specified threshold.

3 Examples

3.1 Materials

Table 1 The comparison between the fitting values of percentage of damaged seeds and the observed values

year	x_1	x_2	x_3	x_4	x_5	y	$\hat{y}$	relative error %
1991	16	124	459	217	1535	8.6	8.18	4.85
1992	68.15	138	241	232	787	7.68	7.49	2.53
1993	42.68	120	321	231	710	10.15	10.45	-2.97
1994	97.5	137	668	246	4118	9.71	9.69	0.12
1995	68.5	145	263	229	1305	14.7	14.5	1.38
1996	86.1	144	142	232	1029	10.22	10.38	-1.56
1997	54.2	132	28	250	561	4.5	4.95	-9.97
1998	3.1	128	139	236	1314	5.2	5.08	2.37
1999	143	152	432	254	740	36.1	36.14	-0.12
2000	210.6	118	79	248	757	13.7	13.69	0.017

Table 2 The comparison between the forecast values of percentage of damaged seeds and the observed values

year	x_1	x_2	x_3	x_4	x_5	y	$\hat{y}'$	relative error %
2001	166.9	161	23	213	739	31.1	29.83	-4.10
2002	77.2	143	141	231	1031	10.3	9.91	-3.82
2003	87.1	147	265	227	1307	14.6	14.67	1.84
2004	48.9	121	319	232	709	10.1	9.95	-1.49
2005	50.9	144	262	228	1306	15.4	15.89	3.18

In this paper, we study the relation between the percentage of damaged seeds of leguminivora glycinivorella $y(\%)$(for the current year) and five estimation factors in Shuangcheng county, Heilongjiang province form 1991 to 2005, which x_1 denotes the amount of damaged seeds for the last year, $x_2(°C)$ and x_3(mm) denote the average air temperature and rainfall in the last twenty-days of September for the last year respectively, $x_4(°C)$ and x_5(mm) denote the average air temperature and rainfall of July for the current year respectively. And the datum is listed in table 1 and table 2. We build the forecast model of the percentage of damaged seeds by projection pursuit regression with datum of 1991 to 2000 and keep the datum of 2001 to 2005 to test the forecast results.

3.2 Method

Replace the variables in formula (4)-(7) by the dependent variable series $y(i)(i = 1, 2, ..., 10)$ and the estimation variable(factor) series $x_1(i)$, $x_2(i)$, $x_3(i)$, $x_4(i)$, $x_5(i)(i = 1, 2, ..., 10)$, one Ridge function was used in fitting to the model, and set the order of the Hermite polynomial 4, then we can got the value of a, c by using the method of accelerating genetic algorithm based on real coding, which model the Biomimetic method of selecting the superior and eliminating the inferior and the information exchange mechanism of internal chromosome.

3.3 Results and Analysis

From the computer program, we get the results:

Value of projection target function is 0.0587;

Projection direction vector $a = (-0.4428, 0.7863, 0.1509, 0.0909, -0.3933)$;

Coefficient of polynomial $c = (291.68, -39.40, 271.01, -18.23)$

Through the results from the forecast model in table 2 we can see that the average relative error, $\frac{y_i - \hat{y}_i}{y_i}$(where $\hat{y}_i$ is the fitting values, y_i is the observed values), is 2.59% to the modeling sample from 1991 to 2000. And the average relative error is 2.89% to the sample from 1991 to 2000. So the forecast precision of the model is satisfactory.

4 Some Discussion

We build pest forecast model by projection pursuit regression with only the original observed datum of the estimation factors that closely related to the dependent variable, without any processing datum for classification , so the method is easy to operate.

For the method of projection pursuit regression have the characteristic of processing high dimensional datum ,so it enable us to make use of more estimation factors that related to the dependent variable and obtain adequate information from the datum.

The pest forecast model built by projection pursuit regression can process data effectively whatever the the datum show normal distribution or not and the relation between estimation factors and dependent variable is linear or not. We don't need to make any assumption for the distribution of samples. Therefore the results from the model have stability, high accuracy. However, it is rare to apply projection pursuit regression to pest forecasting ,so we hope more statist's in-depth exploration and research.

Acknowledgements

This research has been supported by the Heilongjiang Postdoctoral Foundation of China (LBHZ05033).

References

[1] Zhonglan Li. (1987), "Sdudies on forecasting the percentage of damaged seeds of leguminivora glycinivorella", *Journal of Entormological Knowledge*, Vol.4, pp.215-218.

[2] Yuebo Gao, Zongzhi Lu, and Ya-jie Sun. (2005), "Studies on forecasting the occurrence of soybean moth (Leguminivora glycinivorella) and its Application", *Journal of Jilin Agricultural Sciences*, Vol.30, pp.18-20.

[3] Xiaoyi Zhang. (1995), "The theory foundation of insect pest monitoring and forecasting", *Journal of Entormological Knowledge*, Vol.32, pp.55-60.

[4] Jiancong Chen, Zhihong Guan and Hu Chen. (2008), "Fuzzy association analysis of complex system in correlation", *An International Journal of Advances in Systems Sciences and Applications*, Vol.2, pp.251-257.

[5] Anming Wang, Xiaogen Li, TongJiang, Zhiquan Huang, Zhangming Li. (2007), "Stability analysis of Jjijiariver landslide and its study of information system of monitoring and forecasting", *An International Journal of Advances in Systems Sciences and Applications*, Vol.3, pp.382-390.

[6] Heli Yang, Kun Wu. (2006), "Evaluation and establishment of project renewal", *An International Journal of Advances in Systems Sciences and Applications*, Vol.3, pp.446-452.

[7] Friedman J H, Tukey J W. (1974), "A projection pursuit regresssion algorithm for exploratory data analysis", *Journal of IEEE Trans. Computers C*, Vol.23, pp.881-889.

[8] Friedman J H, Stuetzle W. (1981), "Projection pursuit regression", *Journal of J.Amer.Statis. Assoc1*, Vol.76, pp.817-823.

[9] Qiang Fu. (2006), "Data processing method and its application in agriculture", Science press, Vol.8, pp.283-325.

[10] Qiang Fu, Xiaoyong Zhao. (2006), "The principle and application of projection pursuit model", Science Press, Vol.6, pp.131-148.

Corresponding author

Author can be contacted at: xd_neau@126.com