The Study on Livestock Production Prediction in Heilongjiang Province Based on Support Vector Machine

Tang Yang
1.School of Economics and Management,Northeast Agricultural University. 2. Investment and Insurance Department, Harbin Finance University.
Harbin, China
tangyang0717@163.com

Li Cuixia
1.School of Economics and Management,Northeast Agricultural University.
Harbin, China
Licuixia0451@163.com

***Abstract*—This paper uses the support vector machine (SVM) algorithm to study the prediction of livestock production in Heilongjiang province, forms the sample set with the 1985-2010 data in Heilongjiang province, and set up the SVM model between factors and livestock production. Use SVM on the input and output data for training and learning, approximate the implied function relationship by historical data, complete the mapping of the new data series, in order to complete the livestock production prediction for future years, and compare the prediction effects with other methods. The results show that, the prediction accuracy of livestock production of the SVM model is superior to other prediction methods.**

Keywords*- *livestock production, support vector machine, prediction

I. INTRODUCTION

Accurate prediction of livestock production in Heilongjiang province can provide decision-making basis for the government, so as to effectively guide the livestock production and sales, so the livestock industry in Heilongjiang province can develop towards a healthy and orderly direction.

At present, the prediction research for livestock production is very little. However, the affecting factors of livestock production are intricate and complicated, and there exits the characteristic of the uncertainty and non-linear between livestock production and a number of their affecting factors, the traditional forecasting methods to solve such highly nonlinear problems has great limitations, therefore, based on the analysis of affecting factors of livestock production the models we established are more able to reflect the internal relations and development trend of the livestock production changes.

Support Vector Machine(SVM) is a new, very effective machine learning method based on statistical learning theory, is to specifically solve the study rule of small sample problem, it avoids the traditional process from the induction to the deduction, achieves an efficient "transductive inference" from the training sample to prediction sample, greatly simplifies the usual classification and regression problems[1].

The livestock production's historical data is limited, is the typical small sampled data; simultaneously the livestock production is also a complex nonlinear system, in view of

II. EMPIRICAL ANALYSIS OF LIVESTOCK PRODUCTION PREDICTION BASED ON SVM

A. *Collection of Sample Data Set*

This paper according to the availability of statistical data and the applicability of modeling, selects the progress of science and technology, natural environment, economic environment and livestock production structure as the primary factors to do the livestock production prediction.

Therefore, select rice sowing area X1, mechanical power X2, irrigation area X3, fertilizer rate X4 and hazard area X5 as explanatory variables for livestock production Y.

Select 1985 - 2010 statistics as total samples, and the total sample is divided into two parts: 1985 - 2006 total 22 samples as the fitting and training sample, 2007 - 2010 total 4 samples as prediction inspection samples. Specific data of Model can be seen in table I:

TABLE I. LIVESTOCK PRODUCTION MODEL SAMPLE DATA

	Y	X_1	X_2	X_3	X_4	X_5
1985	21.5	721.6	949.5	679	42.1	3028
1986	22.5	571.5	935.2	720	47.7	1537
1987	23.0	741.2	1093.5	766	52.2	2220
1988	29.4	688.6	1105.2	740	54.4	3056
1989	34.0	726.2	1162.8	777	66.1	3408
1990	49.3	742	1173.4	1078	76.6	680
1991	55.9	742.7	1179.5	1118	83.8	2450
1992	57.1	734.8	1172.6	1157	88.5	1468
1993	64.5	755.8	1185.3	1164	100.2	1663
1994	106.0	750.1	1190	1015	108.5	2146
1995	134.3	750	1226.1	1095	108.9	1030
1996	151.5	779.6	1254.8	1335	115.1	997
1997	168.7	799.5	1285.4	1607	121.8	1698
1998	184.5	808.3	1454.5	1648	125.9	2056
1999	165.9	809.9	1559.7	1966	126.3	650
2000	175.7	785.2	1613.8	2032	121.6	2300
2001	224.6	795.7	1648.29	2090	123.2	2388
2002	252.1	783.3	1741.8	2185	129.7	1967
2003	294.2	786.3	1807.74	2112	125.7	4160
2004	400.7	821.6	1952.2	2282	143.8	1087
2005	461.2	988.9	2234	2394	150.9	873
2006	448.7	1052.6	2570.62	2648	162.2	2247

Published by Atlantis Press, Paris, France.

2007	585.0	1082.1	2785.3	2950	175.2	3187
2008	813.1	1098.8	3018.36	3123	180.7	1345
2009	870.2	1313.3	3401.27	3406	198.9	3130
2010	965.8	1354.9	3736.29	3875	214.9	974

Source: China statistical yearbook over the years

B. *Establishment of SVM Prediction Model*

This paper in the process of establishment of SVM prediction model and prediction mainly uses MATLAB (R2010a) to unify the Libsvm-2.89 toolbox to complete.

- Data preprocessing

In order to avoid the numerical calculus's difficulty during training, improve the convergence speed and prediction accuracy, need to carry on the normalized pretreatment to the data, using a normalized mapping as follows:

$$f : x \to y = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

In this formula:

$$x, y \in R^n, x_{\min} = \min(x), x_{\max} = \max(x)$$

The normalized result makes the original data to be regulated to the [0, 1] range.

- Selection of kernel function

Kernel function of support vector machine determines the performance of support vector machine regression. According to learning algorithm analysis and simulation experimental comparison of the different kernel Function of support vector machine, it is concluded that in regression estimation support vector machine which is based on Radial Basic Function (RBF), its performance is better than based on other kernel function of support vector machine[3].

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

In this formula:

σ >0, it is the parameter of kernel function.

- Determination of optimal parameters

As is known to all, support vector machine regression generalization performance depends on the parameters selected situations, such as penalty coefficient C, insensitive loss coefficient ε, and parameter σ of RBF kernel function, etc.

This paper selects PSO method to determine the parameters. Constantly adjust parameters, finally select optimal parameter $C = 1.1525$ 、 $\sigma = 0.25396$ 、 $\varepsilon = 0.002328$, obtain SVM forecasting model. Specific parameter choice result diagram is shown in figure 1.

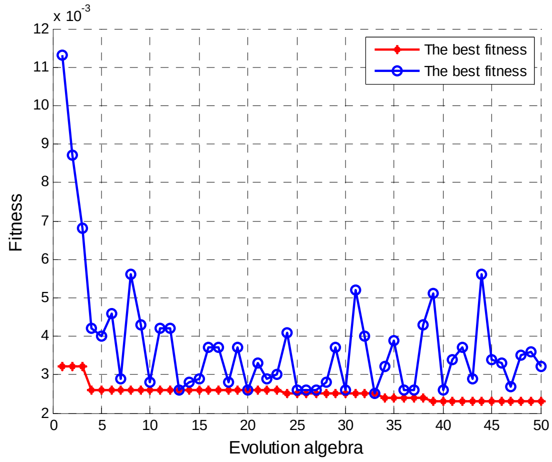

Figure 1. Contour map of parameter selection

C. *Model Training and Forecast*

Select 1985 - 2006 total 23 samples as the fitting and training sample, 2007- 2010 total 4 samples as prediction inspection samples. According to the above prediction model, the fitting and training result of livestock production from 1985 to 2006 is shown in table II and the predicted results of livestock production from 2007 to 2010 is shown in table III.

TABLE II. FITTING RESULTS OF SVM MODEL

	Actual value	Predictive value	Absolute error	Relative error (%)
1985	21.5	12.29173	(9.2)	-0.42829
1986	22.5	13.61264	(8.9)	-0.39499
1987	23.0	24.25395	1.3	0.054519
1988	29.4	30.45717	1.1	0.035958
1989	34.0	53.04243	19.0	0.560071
1990	49.3	88.70329	39.4	0.799255
1991	55.9	74.18904	18.3	0.327174
1992	57.1	87.32185	30.2	0.529279
1993	64.5	102.4437	37.9	0.588274
1994	106.0	102.8952	(3.1)	-0.02929
1995	134.3	125.4164	(8.9)	-0.06615
1996	151.5	150.9624	(0.5)	-0.00355
1997	168.7	167.5324	(1.2)	-0.00692
1998	184.5	188.0872	3.6	0.019443
1999	165.9	239.4054	73.5	0.44307
2000	175.7	214.9967	39.3	0.223658
2001	224.6	225.321	0.7	0.00321
2002	252.1	250.7457	(1.4)	-0.00537
2003	294.2	260.8733	(33.3)	-0.11328
2004	400.7	315.6629	(85.0)	-0.21222
2005	461.2	388.8869	(72.3)	-0.15679
2006	448.7	447.7833	(0.9)	-0.00198

As can be seen from Table II, SVM model fitting effect is still quite good, and the relative error of less than 15% account for 50%.

TABLE III. PREDICTION RESULTS OF SVM MODEL

	Actual value	Predictive value	Absolute error	Relative error (%)
2007	585.0	501.5252	(83.4)	-0.14263
2008	813.1	555.4559	(257.6)	-0.31687
2009	870.2	617.2645	(252.9)	-0.29066
2010	965.8	684.8275	(281.0)	-0.29092

As can be seen from Table III, the SVM model can carry on the forecast by the high precision to the future data, have the good forecast performance. Specifically speaking, the SVM model can reflect well the complex nonlinear relationship between the livestock production and its influencing factor, in the livestock production forecast domain has the good application prospect.

III. COMPARISON OF DIFFERENT MODELS TO PREDICT THE LIVESTOCK PRODUCTION

A. Predicted Results of Different Models

In order to confirm the livestock production forecast model based on SVM strong predictive ability, this paper selects cubic exponential smoothing model and grey prediction model separately as the prediction comparison, each model uses the same training sample and the forecast sample, the concrete training and forecast situation is as shown in figure 2, figure 3, figure 4 below.

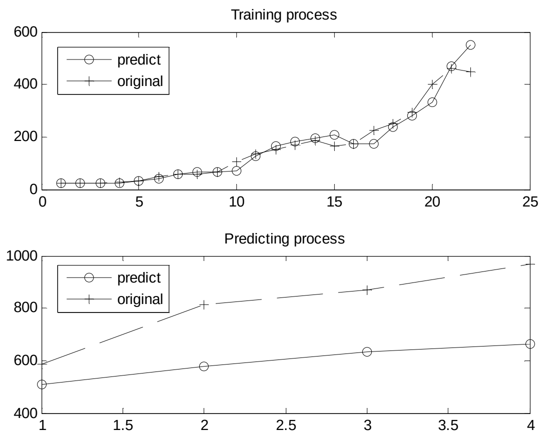

Figure 2. Prediction results of cubic exponential smoothing model

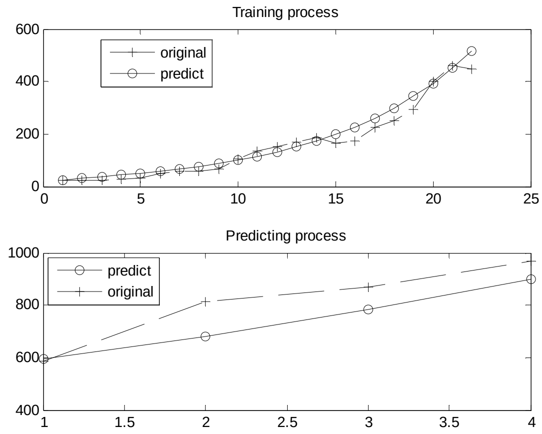

Figure 3. Prediction results of grey prediction model

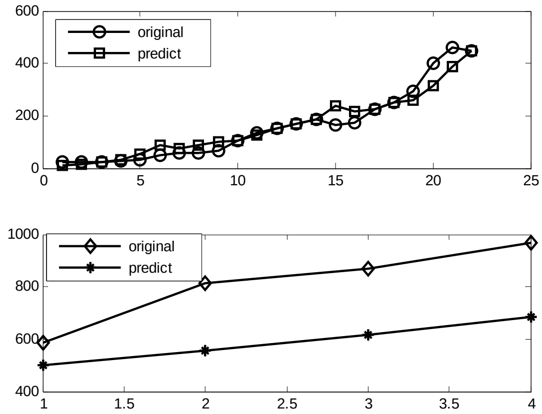

Figure 4. Prediction results of SVM model

B. Comparison and Evaluation of Predicted Results of Different Models

The model forecasting result analysis includes the absolute error analysis and the relative error analysis, the mean absolute error and the root mean squared error belong to the absolute error analysis, Theil inequality coefficient and the mean absolute percent error belong to the relative error analysis. Mean absolute error and root mean squared error have the consistent application of principles, choose one kind then. Therefore, this paper finally chooses the root mean squared error (RMSE), the mean absolute percent error (MAPE) and Theil inequality coefficient (TIC) three kinds of evaluating indicators to carry on the evaluation to the model forecasting result [4].

TABLE IV. COMPARISON OF PREDICTION ACCURACY OF DIFFERENT MODELS

model	RMSE	MAPE	TIC
SVM model	205.1052	20.0694	0.1407
grey prediction model	185.9467	18.7037	0.1548
cubic exponential smoothing model	232.5244	26.0286	0.1644

As can be seen from Table IV, linear model cubic exponential smoothing model prediction accuracy is not very high, the main reason is that it cannot capture the nonlinear characteristics of livestock production data; Grey prediction accuracy is not satisfactory, mainly because it excels in trends for short-term forecasting precision, but for the future slightly inadequate[5].

In all models, SVM model's fit and predictive ability of generalization are the best, this is because SVM based on structural risk minimization, can solve the small sample, nonlinear and other issues, and have the excellent generalization ability. Therefore, we believe that SVM model predicts livestock production more accurately than the other two models.

IV. CONCLUSIONS

This paper analyzed the major affecting factor of livestock production, based on this constructed the SVM model to carry on the forecast research of livestock production in the situation of having less sampled data, this model had reflected well the complex nonlinear relationship between livestock production and its influencing factors, obtained the high forecast precision[6].

Meanwhile, in order to further specify this SVM model's advantages to forecast the livestock production, compare the prediction accuracy with the cubic exponential smoothing model and gray prediction model, the results show that, the SVM model's forecast precision is highest, has the good application prospect in the livestock production's forecast domain.

In addition, because the livestock production's influencing factor is numerous, moreover is in unceasingly in the dynamic change, therefore, to enhance the accuracy of forecasts, a number of factors need to be continuously screening and analysis, so as to find the livestock production's most appropriate explanatory variables.

Because the SVM theory and method have short historical development, in addition the livestock production fluctuation condition is complex, at present the combination of the two is also at the start stage, but the author believes that SVM theory has the prospect in the livestock production research aspect very much, hopes that the government department and the scientific research institution can at the right moment advance SVM in the livestock production forecast application, grasp the livestock production fluctuation tendency accurately promptly, provide the powerful help for the livestock production level's fast promotion.

ACKNOWLEDGMENT

The research project has funds supported by 1. National natural science foundation "the research on benign competition and cooperation development mechanism of breeding and grain production based on grain security in northeast region" (71173035).2. Heilongjiang province colleges and universities in the new century excellent talent training plan "the research on China's dairy quality safety guarantee system"（1252-NCET-001）.3. Heilongjiang province key project "the research on stable development of animal husbandry Based on circular economy in Heilongjiang Province"(GC10D206).

LiCuixia*- Corresponding Author

REFERENCES

[1] Vapnik V. The Nature of Statistical Learning Theory (Second Edition) [M]. New York: Springer Verlag, 1999.225-260.

[2] Zhang Xuegong. Introduction to Statistical Learning Theory and Support Vector Machines [J].Acta Automatica Sinica, 2000, 26(1):32-42.

[3] Zhao Jiwen, Liu Yongbin, Kong Fanrang. Application of SVM with Genetic Algorithms Optimizing Kernel Parameters in Linear Motor Model[J]. Journal of System Simulation, 2006, 18 (12): 3547 – 3553.

[4] Yi Danhui. Data analysis and EViews application [M].Bei Jing: China Renmin University Press, 2008(10): 53 – 54.

[5] M.Ldmna, Introduction to pattern Recognition, Journal of ACM，1984.6:12-13.

[6] Divide. Chapter 12. Informational Today Inc.: Medford, New Jersey.