Study on the Forecast of Air Passenger Flow Based on SVM Regression Algorithm

Yan Ke-wu
School of Economics and Management
Jiangsu Teachers University of Technology
Changzhou 213001, China
yankewu@nuaa.edu.cn

***Abstract*—The forecast of air passenger flow plays an important role in the management of airline, but the traditional forecast methods can't guarantee the generalization capability when they face a large-scale, multi-dimension, nonlinear and non-normal distribution time series data. To improve the forecast ability of air passenger flow, the SVM regression algorithm is introduced in this paper. By selecting appropriate parameters and kernel function, compared with the other two forecast methods, we find that the result obtained by SVM regression algorithm shows the least error among the mentioned three methods.**

Keywords-Forecast; Support vector machine (SVM) regression algorithm; Air passenger flow

I. INTRODUCTION

In the process of airline's management, not only the operations management but also the FAP (fleet assignment program) needs to forecast the air passenger flow. The accurate forecast of air passenger flow plays an important role in the revenue management of airline; so many scholars have been studying on it. There are some traditional forecast techniques. Such as the analysis of time series, logistic regression analysis, grey theory, the combination forecasting, and so on [1]. Most of the methods above concentrate the analysis on the relation regression model and the time sequence model. They may give us an approximate tendency of passenger following. But when facing a large-scale, multi-dimension, nonlinear and non-normal distribution time series data. it can't guarantee the generalization capability of a model. To a certain extent, the artificial intelligence algorithm of neural networks has overcome the flaw of the conventional methods, it not only has the ability of generalization and nonlinear map but also the strong robustness and higher forecast precision. This method mainly depends on the principle of experiences risk minimum [2] and is easy to converge on the local minima when sample quantity is limited. On the other hand, when the study sample size is large, it's easy to fall into the dimension disaster, causes the generalization-ability drop and difficult to determine the model structure, these insufficiencies have limited the method's application in reality.

The support vector machines regression model has solved the above problem successfully. As a new algorithm on the foundation of structure risk minimum principle, the SVM algorithm has the superiority than other algorithm based on experience risk minimum principle. Applying the support vector machines in the function regression estimation form the SVM regression question, which might be thought to solve a convex quadratic programming[3,4] and can obtain the globally optimal solution theoretically. The algorithm transforms the actual problem to the high-dimensional space through the non-linear conversion and carries on the linear regression in the high-dimensional space to realize non-linear regression in the original space. The algorithm nature had guaranteed the SVM regression model to have the good generalization capability. At the same time, it solves the dimension problem ingeniously. The complexity of algorithm has nothing to do with the sample dimension. The SVM regression algorithm structures regression function according to the part of training sample, it doesn't need the prior information of the sample or the regression function structure.

II. SUPPORT VECTOR MACHINES REGRESSION MODEL

A. Linear support vector machines regression model

Establishing a n-dimension training sample.

$$\mathrm{D}=\{(x_k,y_k)|k=1,2,\cdots,n \quad x_k\in R \ ;y_k\in R\}$$

is the training sample data sets. x_k is a n dimension input-data and y_k is the output-data. The regression function can be expressed as the following linear equation:

$$f(x)=w\cdot x+b \qquad (1)$$

Firstly, we should try to get an optimal w and b to make the loss function

$$R_{emp}(w,b)=\frac{1}{n}\sum_{i=1}^{n}\left|y_i-(w\cdot x_i+b)\right|_{\varepsilon}$$

least. Supposing all the training data can be fitted with linear function as the formula (1) thus may result in the following formula:

$$\begin{aligned}&\min\frac{1}{2}\left\|w^2\right\|\\&s.t.\begin{cases}w\cdot x_i+b-y_i\le\varepsilon\\ y_i-w\cdot x_i-b\le\varepsilon\end{cases}\quad i=1,2,\cdots,n\end{aligned}\qquad (2)$$

Considering the error, relaxation factor ξ_i,ξ_i^* is introduced, and then the formula (2) evolves into formula (3)

$$\begin{aligned}&\min\frac{1}{2}\left\|w^2\right\|+c\sum_{i=1}^{n}(\xi_i+\xi_i^*)\\&s.t.\begin{cases}w\cdot x_i+b-y_i\le\varepsilon\\ y_i-w\cdot x_i-b\le\varepsilon\end{cases}\quad i=1,2,\cdots,n\end{aligned}\qquad (3)$$

978-0-7695-3604-0/09 $25.00 © 2009 IEEE
DOI 10.1109/DBTA.2009.33

In the formula (3), $c>0$ express the disciplinal degree for a sample that surpasses error ε . ξ_i, ξ_i^* is the upper-bound and lower bound. $\frac{1}{2}\|w^2\|$ makes the function smoother to improve the generalization ability. $c\sum_{i=1}^{n}(\xi_i+\xi_i^*)$ can reduce the error and the coefficient c make the two factors compromise.

The Lagrange function is introduced to this convex quadratic programming:

$$L(w,b,\alpha,\alpha^*)=\frac{1}{2}w\cdot w+c\sum_{i=1}^{n}(\xi_i+\xi_i^*)-\sum_{i=1}^{n}\alpha_i[\varepsilon+\xi_i+y_i$$

$$-(w\cdot x_i+b)]-\sum_{i=1}^{n}\alpha_i^*[\varepsilon+\xi_i^*-y_i+(w\cdot x_i+b)] \quad (4)$$

According to the saddle point theorem, the equation above can be transformed as the following dual problem:

$$\min w(\alpha,\alpha^*)=\frac{1}{2}\sum_{i,j=1}^{n}(\alpha_i-\alpha_i^*)(\alpha_j-\alpha_j^*)(x_i\cdot x_j)+$$

$$\sum_{i=1}^{n}y_i(\alpha_i-\alpha_i^*)-\sum_{i=1}^{n}\varepsilon(\alpha_i+\alpha_i^*)$$

$$s.t.\begin{cases}\sum_{i=1}^{n}(\alpha_i-\alpha_i^*)=0\\ 0\le\alpha_i \quad \alpha_i^*\le c\end{cases} \quad (5)$$

According to the formula (5) we can get the Lagrange coefficient α_i, α_i^* that make the objective function minimum, then the optimal weight will be a linear combination between α_i, α_i^* expressed as $w^*=\sum_{i=1}^{n}(\alpha_i-\alpha_i^*)x_i$, the regression function can be expressed as following:

$$y=f(x)=w^*\cdot x+b^*=\sum_{i=1}^{n}(\alpha_i-\alpha_i^*)x_i\cdot x+b^* \quad (6)$$

B. Nonlinear vector machine regression model

Regarding the nonlinear problem, we map data to a high dimension space $\mathbf{H}$ [5] through nonlinear mapping. With Φ transform $R^d\to H$. Through mapping x into $\Phi(x)$, then make it regress in high dimensions and introduce kernel function $\mathrm{k}(x,x_i)$, we can change the linear model into nonlinear model as formula (7)

$$y=f(x,x_i)=\sum_{i=1}^{n}(\alpha_i-\alpha_i^*)\mathrm{k}(x,x_i)+\mathrm{b} \quad (7)$$

When taking linear insensitive loss function as risk assessment model, we get the objective function of the quadratic programming Lagrange:
Its dual form is:

$$L(w,b,\alpha,\alpha^*)=\frac{1}{2}w\cdot w+c\sum_{i=1}^{n}(\xi_i+\xi_i^*)-\sum_{i=1}^{n}\alpha_i[\varepsilon+\xi_i+y_i$$

$$-(w\cdot\Phi(x_i)+b)]-\sum_{i=1}^{n}\alpha_i^*[\varepsilon+\xi_i^*-y_i+(w\cdot\Phi(x_i)+b)] \quad (8)$$

Through solving the model above , we can obtain the α_i and α_i^* applying formula (7) in forecast.

C. The choice of kernel function in the process of airline leg flow forecast.

In the process of establishing the support vector machines model, the choice of kernel function is direct related the model's properties, the choice of kernel function is an important problem for SVM which has been studying. Each kernel function has the data distribution type belong itself, regarding the different data sets, the different kernel function has different performances. Choosing the appropriate kernel function is easy to obtain the better forecast effect.

There are mainly three kinds kernel function at present [6,7,8]:

（1）polynomial kernel function:

$$\mathrm{k}(x,y)=(\mathrm{xy}+1)^{\mathrm{d}} \qquad d=1,2\cdots$$

（2）Radial basis kernel function (RBF):

$$\mathrm{k}(x,y)=\exp(-\frac{\|x-y\|^2}{2\sigma^2})$$

（3）Sigmoid kernel function：

$$\mathrm{k}(x,y)=\mathrm{anh}(\mathrm{b}(\mathrm{xy})\text{-}\mathrm{c})^{\mathrm{d}}$$

Until now, we can't give the appropriate method theoretically to choose the optimal kernel function and has to seek help from experiment's way. Aiming to find which kernel function is more suit to the air passenger flow forecast, we choose an airline's leg sample data of the past 24 years altogether. Comparing the forecast results which come from the different kernel function with the actual value, the average relative error is shown as table Ⅰ.

TABLE I. THE PREDICTION ERROR COMPARISON OF DIFFERENT KERNEL FUNCTION

Kernel function	c	σ	d	error
polynomial kernel function	30		5	6.3
	97		5	5.3
	450		5	4.6
	1500		5	4.3
	3500		5	5.6

RBF	30	5		15.2
	67	5		13.3
	150	5		12.8
	1500	5		11.2
	3500	5		12.3
Sigmoid kernel function	30			22.1
	90			26.8
	450			16.5
	1500			24.6
	3500			29.6

From the table 1, we find that the forecast result of polynomial kernel function is better, so we apply it to the forecast of air passenger flow. Regarding the choice of parameter and penalty factor of kernel function, though we haven't the unification theorical instruction at present, we may choose the appropriate parameter by experience as well as the computed results contrast of different parameter.

III. THE EMPIRICAL RESEARCH OF AIRLINE LEG PASSENGER FLOW FORECAST BASED ON SUPPORT VECTOR MACHINES REGRESSION ALGORITHM

A. Sample processing

In this article, we take the every year's passengers flow from 1982 to 2005 as our data sets ,as shown in table II.

TABLE II. AN AIRLINE LEG'S PASSENGERS FLOW FROM 1982 TO 2005.

No.	1	2	3	4	5	6	7	8
year	1982	1983	1984	1985	1986	1987	1988	1989
Passenger flow	2875	3209	3435	4578	5225	5967	6242	6827
No.	9	10	11	12	13	14	15	16
year	1990	1991	1992	1993	1994	1995	1996	1997
Passenger flow	7640	8290	8500	9066	8744	9225	9378	9893
No.	17	18	19	20	21	22	23	24
year	1998	1999	2000	2001	2002	2003	2004	2005
Passenger flow	11133	12551	14477	15693	17379	19204	10680	12044

B. Forecasting results and comparative analysis.

In view of time series of passenger flow in table 1, we forecast the fourth year data on the foundation of the past three years' data. Carrying on the training and simulation test with MATLAB6.5 and comparing the forecast data achieved from SVM regression algorithm with the results obtain from BPANN or the linear regression algorithm, the kernel function of SVM regression algorithm is $k(x_i, y_i)=(x_i \cdot y_i+1)^4$, the optimized parameter choose by intersection method. Regarding BPANN, build three layers neural networks, training forward network with the fast BP algorithm, the inputs neuron is 3, the Implying neuron is 4, output 1 neuron, training epochs is 1500, the expected error is 0.001, the rate of beginner study is 0.01. Finally using the linear regression to forecast the passengers flow time series in this airline leg. The simulation results is shown as tableIII.

TABLE III. THE FORECASTING RESULT CONTRAST OF DIFFERENT FORECAST METHODS

year	Real passenger flow	The forecast value of SVM	The forecast value of Linear Regression	The forecast value of BPANN	Forecast relative error of SVM %	Forecast relative error of Linear Regression %	Forecast relative error of BPANN %
1985	4578	4569	3730	4348	-0.21	-18.53	-5.03
1986	5225	5261	5181	5058	0.67	-0.85	-3.21
1987	5967	5957	5946	6146	-0.16	-0.34	3.01
1988	6242	6488	6609	6545	3.94	5.88	4.86
1989	6827	6817	6771	6981	-0.14	-0.82	2.26
1990	7640	7224	7350	7258	-5.45	-3.8	-5
1991	8290	7984	8351	8121	-3.69	0.74	-2.04
1992	8500	8690	9037	8934	2.23	6.32	5.11
1993	9066	8978	9069	9275	-0.97	0.03	2.31
1994	8744	9434	9649	9581	7.89	10.35	9.57
1995	9225	9338	9128	9729	1.22	-1.05	5.46
1996	9378	9551	9641	9602	1.84	2.8	2.39
1997	9893	9903	9916	10185	0.1	0.23	2.95
1998	11133	10344	10476	10478	-7.09	-5.9	-5.88

1999	12551	11642	12095	11607	-7.24	-3.63	-7.52
2000	14477	13439	13824	13213	-7.17	-4.51	-8.73
2001	15693	16210	16259	16182	3.294	3.605	3.119
2002	17379	17832	17903	17818	2.609	3.013	2.527
2003	19204	18808	15342	20312	-2.06	-20.11	5.77
2004	10680	10170	9658	10509	-4.78	-9.57	-1.6
2005	12044	11962	11583	12085	-0.68	-3.83	0.34

Carrying the result which obtains from the three forecast technique fitting with the real result , we get figure 1.

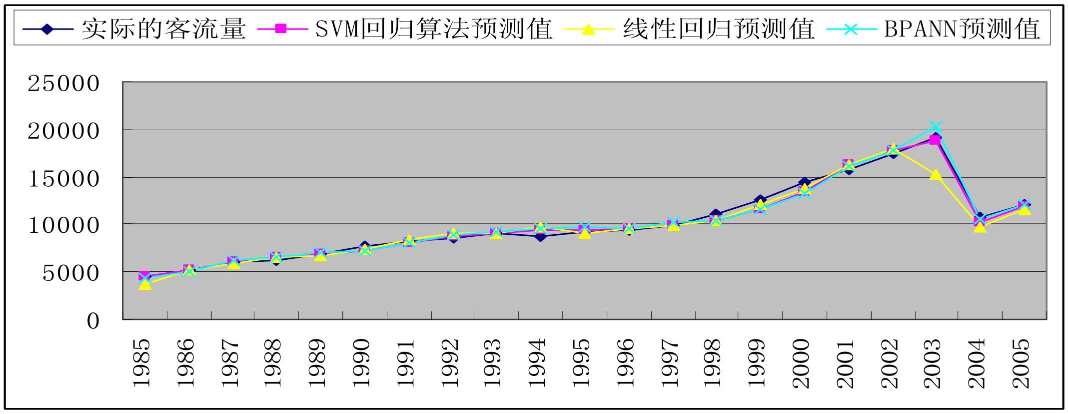

Figure 1. Fitting contrast line chart

From figure 1, we find that the fitting effect based on SVM regression is best among the three methods, the linear regression algorithm's effect is worst. This show that the aviation passenger flow forecast is not a linear question problem.

IV. Conclusion

The SVM regression algorithm is applied in the aviation passenger flow forecast in this article. Through comparing SVM with the linear regression algorithm and BPANN, we discovered that the SVM model with polynomial kernel function has the best forecast effect. But, in the specific model, there is no rule to follow in the selection of the kernel function and the parameter until now. Therefore, how to find one kind of feasible theory and the method in choice of the kernel function and the parameter is the key to increase the forecast precision.

References

[1] Zhai Yan-juan, “The study on airline’s FAP in hub-and-spoke network”, Chinese commercial aviation university. (2006) pp.18–26

[2] Xia Guo-en.(2008),Model of customer churn prediction on support vector machine,Systems Engineering-Theory & Practice, vol.01,2008. pp.71-77 .

[3] Steve .R. Gunn, “Support vector machines for classification and regression”, University of Southampton,1998 pp.152-156

[4] Suykens JAK, J Vandewalle, Least squares support vectors machine classifiers, Neural processing Letters,1999,9(3) pp.293-300.

[5] Li Guo-zheng, The introductory of support vector machine,Electronics industry publishing house, pp.56-78,2003

[6] Hang Wen-qiang, Application of Support Vector Machine in Airline Passenger Volume Forecast, Computer Engineering, vol.7,2005, pp.254.

[7] Liu Guang-li, Yang Zhi-min, A New Support Vector Regression Model,Journal of Jishou University(Natural Science Edition**.** 2002 ，vol.72（3）pp.28-32.

[8] Nello Cristianini, England, John Shawe-Taylor, An introduction to support vector machines and other kennel-baded learning method Beijing:Mechanical industry publishing house,2005