Application of an Optimized SVR Model of Machine Learning

Zhikun Xu¹, Yabin Gao² and Yingying Jin³

College of Finance, Hebei Normal University of Science and Technology, Qinhuangdao, Hebei, P.R.China, 066004
¹xuvikp@126.com, ²522865482@qq.com, ³695735043@qq.com

Abstract

Machine learning is the core of artificial intelligence. It is a fundamental way to the computer intelligence. Support vector machine is one of the important methods in the field of machine learning. It has the advantages of global optimization and strong generalization ability. It has been successfully applied to face recognition, fault diagnosis, financial forecasting and other fields. In this paper, a novel SVR model is proposed to forecast GDP. In the model, the neighborhood rough set (NRS) is used to reduce the index set and the chaotic genetic algorithm (CGA) is adopted to parameters searching in SVR model. Then the novel model NRS-CGA-SVR is established to predict GDP of Anhui province. The results show that the proposed model is better than the other models presented in this paper on forecasting GDP.

Keywords: Prediction, Machine learning, SVR, GA, Rough set, Chaotic sequence

1. Introduction

Machine learning is an active field in the development of artificial intelligence. The purpose of the study of machine learning is to let the computer have the ability to acquire knowledge from the real world like human beings. At the same time, establish computational theory of learning and construct all sorts of learning system which are applied in various fields.

The gross domestic product (GDP) is one of the most important indicators to measure of national economic development. It is also the comprehensive reflection of the economic situation. GDP accounting has become the main methods of economic management departments of country and regional government to understand the functioning of the economy. It is also the main basis for the formulation of economic development strategy, planning, and a variety of macroeconomic policy. So it is very important to predict GDP in a scientific method.

At present, there are lots of methods to predict GDP. The main methods are regression method [1-3], time series prediction method [4-6], gray theory method [7-9], Markova prediction method [10-12] and neural network method [13-15], etc. In front of the four methods are the traditional forecasting methods. They mostly analyze the causal relationship between variables. Some problems which are multicollinearity and serial correlation exist in the practical application. The neural network algorithm is a kind of nonlinear optimization method in essence. It also has some shortcomings such as long convergence time and easy to fall into local minimum point.

Support vector machine (SVM) was proposed by Vapnik in 1995 [16]. It is based on structural risk minimization principle and can better solve over learning problem. Since then, pretty of Experts and scholars are focus on SVM. Bouzerdoum et al. [17] introduced a new
hybrid model which combines SARIMA and SVM for short-term power forecasting of a grid-connected photovoltaic plant. The comparative study showed that the developed hybrid model performs better than both the SARIMA and the SVM model. Li and Xiao [18] used the SVM model to predict GDP of Jiangmen and the satisfactory results had been achieved. Hu et al. [19] presented the SVM model to predict GDP of Jiangxi province and it also shows adaptability and accuracy of this model. Qiao et al. [20] proposed a Bayesian evidence framework to infer the LS-SVR model parameters. The results show that the Bayesian framework of an LS-SVR significantly improves the speed and accuracy of the forecast. Hong et al. [21] employed GA-SA algorithm to choose the suitable parameter combination for a SVR model and used the SVRGA–SA model to forecast Taiwanese 3G mobile phone demand. Li et al. [22] proposed a consensus least squares support vector regression (LS-SVR) method for calibrating the near-infrared (NIR) spectra and the predicted results show that consensus LS-SVR model is more robust and reliable than the conventional partial least squares (PLS) and LS-SVR methods. Zheng et al. [23] adapted an expectation-maximization (EM) algorithm based on two 2-level hierarchical-Bayes models to fast train multi-scale SVR. Mandal et al. [24] used rough set theory to reduce the length of experimental data as well as generate rules and used artificial bee colony (ABC) algorithm for the computational training of SVM. Peng et al. proposed a hybrid model based on rough set theory and SVM. RS theory was employed as an attribute reduction tool to identify the significant environmental parameters of a landslide and an SVM was used to predict landslide susceptibility. Compared with the general SVM, the hybrid model has superior prediction skill and higher reliability.

In this paper, the overall structure of the study is as follows: In Section 2, the standard SVR model and the principle of neighborhood rough set is introduced. The chaotic genetic algorithm is also presented in this section. In Section 3, we establish the NRS-CGA-SVR model to predict GDP of Anhui province. At last, the conclusion is given in Section 5.

2. Machine Learning

Machine learning is the extension of human cognitive behavior and is also an important branch of artificial intelligence. Machine learning uses some basic methods to understand the objective world and acquires all kinds of knowledge and skills from the study of human behavior. With the help of computer technology, it establishes all sorts of learning models, so as to give the learning ability to computer system. Figure 1 shows the basic idea about the machine learning system. Information may be collected from multiples sources.

![Figure 1. Leaning System](image-url)
One of the common theoretical bases is statistical from the existing machine learning method, including neural network. Vapnik [16] began to study machine learning problems in the finite sample from 1960s and formed a perfect theory system, named Statistical Learning Theory in 1990s. It achieved good results on the learnability, correctness, over learning and less learning and local optimum.

3. The optimized SVR model

3.1. The principle of Standard SVR Model

Set a sample set \( S = \{ (x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m) \} \). Through a nonlinear mapping \( \phi \), the data from sample set \( S \) are mapped into a high dimensional space \( F \). And the following function is used for linear regression in \( F \) space.

\[
f(x) = w \cdot \phi(x) + b, \quad \phi: R^n \rightarrow F, \quad w \in F
\]

(1)

Where \( b \) is threshold value. \( w \) is regression coefficient vector. The influence factors of \( w \) are the sum of empirical risk and the flatness of the high-dimensional space \( \|w\| \), that is

\[
R(w) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{m} \varepsilon(f(x_i) - y_i)
\]

(2)

\[
\varepsilon(f(x_i) - y_i) = \begin{cases} 0 & |f(x_i) - y_i| < \varepsilon \\ |f(x_i) - y_i| & |f(x_i) - y_i| \geq \varepsilon \\ \end{cases}
\]

(3)

Where \( m \) is the number of samples. \( \varepsilon(f(x_i) - y_i) \) is loss function. \( C \) is penalty factor.

In order to control the complexity of function, linear regression function should be brought to flat as far as possible. The relaxation factors \( \xi^+ \) and \( \xi^- \) are also introduced considering the regression error that beyond the precision. The relaxation factors are used to deal with the points which do not satisfy the Eq. (3). Based on structural risk minimization criterion of statistical learning theory, the variable of \( w \) and \( b \) are determined through support vector regression method by minimizing the objective function \( R(w) \):

\[
R(w, \xi, \xi^+) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{m} (\xi_i + \xi_i^+)
\]

\[
s.t. \begin{cases} y_i - w \cdot x_i - b \leq \varepsilon + \xi_i \\ w \cdot x_i + b - y_i \leq \varepsilon + \xi_i^+ \\ \xi_i, \xi_i^+ \geq 0 \end{cases}
\]

(4)

In the Eq.(4), \( \frac{1}{2} \|w\|^2 \) makes the regression function more flat and have better generalization ability. The second part of Eq.(4) is used to reduce the error. The penalty factor \( C \) is a constant number and \( C > 0 \). It is used to control the degree of punishment of samples beyond the error \( \varepsilon \). Then establish Lagrange’s equation:
\[
L(w, \xi^i, \xi^*i) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}(\xi^i + \xi^*i) - \sum_{i=1}^{m} \alpha_i[(e + \xi_i) + w \cdot \phi(x_i) + b] \\
- \sum_{i=1}^{m} \alpha_i^*[e + \xi_i^*] + y_i + w \cdot \phi(x_i) - b] - \sum_{i=1}^{m} (\lambda_i \xi_i + \lambda_i^* \xi_i^*)
\]

(5)

In order to make the Eq. (5) be minimum, function \(L(w, \xi^i, \xi^*i)\) of all parameters of partial derivative is equal to zero. Then the following dual optimization problem can be get.

\[
\min \{\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)[\phi(x_i) \cdot \phi(x_j)] + \sum_{i=1}^{m} \alpha_i(e - y_i) + \sum_{i=1}^{m} \alpha_i^*(e + y_i)\}
\]

(6)

\[
\sum_{i=1}^{m}(\alpha_i - \alpha_i^*) = 0
\]

\[
\alpha_i, \alpha_i^* \in [0, C]
\]

Then, the support vector regression problems can be summed up in quadratic programming problem. By solving the quadratic programming problem Eq. (6), we can get the function \(w\) which is described by the training sample points.

\[
w = \sum_{i=1}^{m}(\alpha_i - \alpha_i^*)\phi(x_i)
\]

(7)

Where \(\alpha_i\) and \(\alpha_i^*\) is the solution of \(L(w, \xi^i, \xi^*i)\). Then the regression function can be got. That is

\[
f(x) = \sum_{i=1}^{m}(\alpha_i - \alpha_i^*)k(x, x_i) + b
\]

(8)

Where \(k(x, x_i) = \phi(x) \cdot \phi(x_i)\) is the kernel function. Different SVR models can be got by selecting different kernel functions. In this paper, the radial basis function(RBF) is chosen in SVR model. The RBF is shown in bellow

\[
K(x, x_i) = \exp(-\gamma|x - x_i|^2)
\]

(9)

3.2. The Principle of Neighborhood Rough Set

Rough set is proposed in 1982 by professor Pawlak [26]. It uses the method of rough approximation from mass, desultorily, strong interference of data to dig out the potential and essential information. Therefore, the rough set is widely used in attribute reduction, decision rule extraction, classification prediction, etc. But classical rough set is only suitable for dealing with nominal variable. It needs to do discretization for numeric data through various methods, and discretization process inevitably brings effective information loss [27].

Lin [28] first put forward the concept of neighborhood model. The model adopted the neighborhood of space points to grain domain space. It considered the neighborhood as basic information particle and used it to describe the other concepts of the space. Neighborhood rough set model form a neighborhood \(\delta\) based on each point in real number space. The \(\delta\) neighborhood group constitutes the basic information particle which describes any concepts in the space. The basic description is as follows:

Set \(IS = \langle U, A, V, f \rangle\) is the information system. \(U = \{x_1, x_2, \cdots, x_n\}\) is said to be finite nonempty set. It is called discourse domain. The discourse domain consists of the sample
space set. $A$ is a set of attributes. Here it refers to the prediction index set. $V$ is numerical range. The information function is

$$f : U \times A \times V \rightarrow \mathbb{R}$$

It describes the relationship of mapping between samples and their attribute values. If $A = C \cup D$, where $C$ is condition attribute set. $D$ is decision attribute set. And $C \cap D = \emptyset$. Then $IS = \langle U, A, V, f \rangle$ is a decision table.

Let $x_i \in U$, $B \subseteq C$, the neighborhood of $x_i$ in attribute set $B$ can be defined as

$$\delta_B(x_i) = \{x_i | x_i \in U, d_B(x_i, x_2) \leq \delta\}$$

(10)

Where $d$ is the distance function. For $\forall x_1, x_2, x_i \in U$, $d$ satisfies the following relations.

$$
\begin{align*}
&d(x_1, x_2) \geq 0 \\
&d(x_1, x_2) = 0, \text{ if and only if } x_1 = x_2 \\
&d(x_1, x_2) = d(x_2, x_1) \\
&d(x_1, x_i) \leq d(x_1, x_2) + d(x_2, x_i)
\end{align*}
$$

(11)

When the sample set has $N$ attributes, the distance function can be defined as

$$d_B(x_i, x_2) = \sum_{i=1}^{N} f(x_i, a_i)$$

(12)

Where $f(x_i, a_i)$ is the value of sample $x_i$ in attribute $a_i$. It is worth noting that the distance function $d$ is used to numeric attribute set. But the neighborhood model is easy to extend the numeric attribute set to numeric data and the data which has symbols in distance calculation. For symbol type attribute $a_i$, set

$$
\begin{align*}
&f(x_i, a_i) - f(x_2, a_i) = 0, \text{ when } x_i \text{ and } x_2 \text{ have same value on } a_i \\
&f(x_i, a_i) - f(x_2, a_i) = 1, \text{ otherwise}
\end{align*}
$$

(13)

Then the lower approximation and upper approximation of neighborhood rough set are defined as

$$
\begin{align*}
&\underline{NX} = \{x_i | x_i \in U, d_B(x_i, x_2) \leq \delta\} \\
&\overline{NX} = \{x_i | x_i \in U, d_B(x_i, x_2) > \delta\}
\end{align*}
$$

(14)

The approximate boundary of $X$ is $BN(X) = \overline{NX} - \underline{NX}$.

For a neighborhood decision system $NDT = \langle U, C \cup D, V, f \rangle$, $U$ is divided into $N$ equivalence class by $D : X_{1}, X_{2}, \ldots, X_{N}$. The lower approximation, upper approximation and approximate boundary of decision attribute set $D$ about $B$ can be defined as

$$
\begin{align*}
&\underline{N_x}D = \bigcup_{i=1}^{N} N_x X_i \\
&\overline{N_x}D = \bigcup_{i=1}^{N} \overline{N_x} X_i
\end{align*}
$$

(15)

(16)
\[ BN(D) = N^b_D - N^a_D \]  \hspace{1cm} (17)

The lower approximation of decision attribute set D is also called decision positive domain, that is \( POS^b(D) \). The size of decision positive domain reflects the degree of separation of classification problem in certain attribute space. If the positive domain is greater, the overlapping region is less. Then the dependence of the decision attribute \( D \) to condition attribute \( B \) can be defined as

\[ \gamma^b_D = \frac{|POS^b(D)|}{|U|} \]  \hspace{1cm} (18)

Eq.(18) shows the radio of the samples which can be completely contains by one kind of decision account for the entire samples based on the description of condition attribute \( B \) in sample set.

### 3.3. The Chaotic Genetic Algorithm

Genetic algorithm is the search algorithm that combines the rule of survival of the fittest in biological process and random information exchange within the group based on the theory of natural selection and genetic theory. It can achieve parallel global search and has lots of advantages, such as simple, rapid and good robustness, etc. It provides a general framework for solving complex optimization problem. The problem does not depend on the specific field and the algorithm has very strong robustness in the types of the problem.

The chaos has the feature of randomness, ergodicity and regularity. It can’t repeat traversal of all state in a certain scope according to its own laws. It usually uses logistic iterative equation to describe chaotic phenomena.

\[ x^{i+1} = u(x^i - x^i) \]  \hspace{1cm} (19)

Where \( 0 < u \leq 4 \). The chaotic sequence \( X = \{x^1, x^2, \ldots, x^n\} \) can be got by Eq. (19). Fig2. shows the chaotic dynamics of the system where \( x_0 = 0.001, i = 300 \)

![Figure 2. Diversity of Logistic Function](image-url)
Chaos genetic algorithm is combined with the periodicity of chaotic phenomena and the inversion of the genetic algorithm, using the chaos mapping formula to transform the original population and gets the chaos population according to their own multiple iterations. Then use the new population to do genetic operations. Due to the use of the characteristics of the chaos, the chaos genetic algorithm has better performance on speed and accuracy of searching. Figure 3 shows the flowchart of chaotic genetic algorithm and the main steps of chaotic genetic algorithm are as follows.

Step 1: Chaotic population initialization. There are three parameters $A=(C,\gamma,\varepsilon)$ in a SVR model in this paper. So the individual $x^i$ has three components. $x^i$ is given $l$ different initial values. Set the ith iteration be $X^i$. Then adopt Eq.(20) to make $x^i$ locate in the interval $(0,1)$.

$$x^i = \frac{X^i - \text{Min}}{\text{Max} - \text{Min}}$$

(20)

Where $(\text{Min, Max})$ is the biggest interval of parameters. Then adopt the Eq.(19) to generate the next iteration chaotic variable $x^{i+1}$. Transform $x^{i+1}$ to $X^{i+1}$ by Eq.(21).

$$X^{i+1} = \text{Min} + x^{i+1}(\text{Max} - \text{Min})$$

(21)

After the transformation, encode individuals into binary format.

Step 2: Fitness evaluation. The fitness function is used to evaluate each individual. In this paper, the reciprocal of mean absolute percentage error (MAPE) is used as the fitness function. The function is shown in Eq.(22).

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{y_i - f_i}{y_i} \right| \times 100\%$$

(22)

where $y_i$ represents actual values, $f_i$ represents forecasting values, and $N$ is the forecasting periods.

Step 3: Selection. The purpose of selection is to choose a certain number of excellent individuals from current population. Here, the roulette wheel selection principle [29] is used to selection operation.

Step 4: Crossover operation. The selected individuals do match operation in the random position according to crossover probability $P_c$. In this paper, a single point crossover method is used.

Step 5: Annealing chaotic mutation. From crossover operation, we get population $X^i$. Then Eq.(23) is used to form the crossover chaotic variable space $x^i$.

$$x^i = \frac{X^i - \text{Min}}{\text{Max} - \text{Min}}$$

(23)

Where $(\text{Min, Max})$ is current solution space. The chaotic mutation variable is in the Eq. (24).

$$x^i = x^i + \delta x^i$$

(24)
Where $\delta$ is annealing operation. $x^i$ is the $i$th chaotic variable. Then the chaotic mutation variable is obtained in Eq. (25) which is mapped to the solution interval by the probability of mutation $p_m$

\[
X^i = \text{Min} + x^i (\text{Max} - \text{Min})
\]  

(25)

Step 6: Stop condition. If the generations reach the maximum number, then stop the algorithm. Otherwise, go back to Step 2.

Figure 3. The Flowchart of Chaotic Genetic Algorithm
4. Model Construct and Prediction

The NRS-CGA-SVR model is used to predict GDP of Anhui province. There are many factors that can affect GDP. In this paper, the index set which has influence on GDP is concluded in Table 1. Since the influence of GDP index is up to several dozen. These data often has a great deal of information overlap. They not only affect the SVM generalization ability, but also make the structure of SVR become complex. So the neighborhood RS method is used to reduce the index set so as to eliminate redundant attributes, reduce noise and improve the prediction accuracy. The new index set after attribute reduction is shown in Table 2.

Table 1. Index Set

<table>
<thead>
<tr>
<th>Variable</th>
<th>Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>X1</td>
<td>The proportion of first industrial population</td>
</tr>
<tr>
<td>X2</td>
<td>Total investment in fixed assets</td>
</tr>
<tr>
<td>X3</td>
<td>Fiscal expenditure</td>
</tr>
<tr>
<td>X4</td>
<td>Loan balance at the end of year</td>
</tr>
<tr>
<td>X5</td>
<td>Electricity consumption of whole society</td>
</tr>
<tr>
<td>X6</td>
<td>Turnover of freight traffic</td>
</tr>
<tr>
<td>X7</td>
<td>Business volume of post and telecommunications</td>
</tr>
<tr>
<td>X8</td>
<td>Total retail sales of social consumer goods</td>
</tr>
<tr>
<td>X9</td>
<td>CPI</td>
</tr>
<tr>
<td>X10</td>
<td>Urbanization rate</td>
</tr>
<tr>
<td>X11</td>
<td>Total population</td>
</tr>
<tr>
<td>X12</td>
<td>The RMB exchange rate against the dollar</td>
</tr>
<tr>
<td>X13</td>
<td>Steel production</td>
</tr>
<tr>
<td>X14</td>
<td>Chemical fiber production</td>
</tr>
<tr>
<td>X15</td>
<td>Grain production</td>
</tr>
<tr>
<td>X16</td>
<td>Pork production</td>
</tr>
<tr>
<td>X17</td>
<td>The per capita net income of rural residents</td>
</tr>
<tr>
<td>X18</td>
<td>The per capita real income of city residents</td>
</tr>
<tr>
<td>X19</td>
<td>Industrial added value</td>
</tr>
<tr>
<td>X20</td>
<td>High school students</td>
</tr>
<tr>
<td>X21</td>
<td>Total import and export volume</td>
</tr>
<tr>
<td>X22</td>
<td>The completed investment in real estate development</td>
</tr>
</tbody>
</table>
Table 2. The New Index Set After Attribute Reduction

<table>
<thead>
<tr>
<th>Variable</th>
<th>Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>X2</td>
<td>Total investment in fixed assets</td>
</tr>
<tr>
<td>X3</td>
<td>Fiscal expenditure</td>
</tr>
<tr>
<td>X5</td>
<td>Electricity consumption of whole society</td>
</tr>
<tr>
<td>X8</td>
<td>Total retail sales of social consumer goods</td>
</tr>
<tr>
<td>X13</td>
<td>Steel production</td>
</tr>
<tr>
<td>X17</td>
<td>The per capita net income of rural residents</td>
</tr>
<tr>
<td>X18</td>
<td>The per capita real income of city residents</td>
</tr>
<tr>
<td>X19</td>
<td>Industrial added value</td>
</tr>
<tr>
<td>X21</td>
<td>Total import and export volume</td>
</tr>
</tbody>
</table>

Then we use data from the new index set and GDP data to establish NRS-CGA-SVR model and adopt the model to predict the GDP of Anhui province. The Regression curve of NRS-CGA-SVR model is shown in Figure 4. From Figure 4, we can find intuitively that the regression and forecasting effect of the model are really good. Then NRS-CGA-SVR model is compared with CGA-SVR model, standard SVR model and BP neural network model. The comparison of predicting results are shown in Figure 5, Table 3 and Table 4.

![Figure 4. Regression Curve of NRS-CGA-SVR](image-url)
From Table 3, we can find that the NRS-CGA-SVR model do the best in both MAPE and RMSE. It shows that the model proposed in this paper has good predictive ability. In addition, Table 4 shows the relative error of different models. We can discover that in most of the years from 2007 to 2011, the NRS-CGA-SVR model has better prediction accuracy and the BP neural network model has the worst prediction accuracy of all these models. It demonstrates the superiority of the SVR model. It also indicates that the NRS and CGA methods can enhance the prediction accuracy.

Table 5. The Relative Error of Different Models

<table>
<thead>
<tr>
<th>Year</th>
<th>NRS-CGA-SVR</th>
<th>CGA-SVR</th>
<th>standard SVR</th>
<th>BP Neural Network</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007</td>
<td>-0.005233875</td>
<td>-0.00587745</td>
<td>-0.0079020176</td>
<td>-0.0089869884</td>
</tr>
<tr>
<td>2008</td>
<td>-0.0032483127</td>
<td>-0.0066318114</td>
<td>-0.0066034439</td>
<td>-0.0092617306</td>
</tr>
<tr>
<td>2009</td>
<td>-0.002099012</td>
<td>-0.0050802837</td>
<td>-0.0068991804</td>
<td>-0.0083194834</td>
</tr>
<tr>
<td>2010</td>
<td>0.0048824758</td>
<td>0.0040749652</td>
<td>-0.0082052676</td>
<td>-0.0116232733</td>
</tr>
<tr>
<td>2011</td>
<td>-0.0015777911</td>
<td>-0.0061500459</td>
<td>-0.009028455</td>
<td>-0.0126884302</td>
</tr>
</tbody>
</table>

5. Conclusion

In order to enhance the prediction accuracy, the NRS-CGA-SVR is proposed in this paper. The neighborhood rough set (NRS) is used to reduce the index set which has influence on GDP. The chaotic genetic algorithm (CGA) is adopted to parameters searching in SVR model.
Forecasting results show that the proposed NRS-CGA-SVR model is superior to other alternatives. By the way, this study still has many improvements. For example, there are other advanced optimization algorithms for parameters determination. The index set may be not comprehensive enough. There are some other methods that can be combined into SVR model.

References


Authors

Zhikun Xu, He received his bachelor's degree of Economics in Inner Mongolia Agricultural University, (1993) and master's degree of management in North China Electric Power University, (2010), now he work in hebei normal university of science and technology institute of finance and economics, Qinhuangdao, Hebei. He is a Master's Supervisor. Her major fields of study are practical economics.

Yabin Ga, She received her bachelor's degree of Science in TANGSHAN Normal University Tangshan, Hebei, (2004) and master's degree of Software Engineering in Beijing University of Technology, Beijing, (2012). Now she is a lecturer in Hebei Normal University of Science and Technology, Qinhuangdao, Hebei. Her major fields of study are education and teaching management, software engineering and education.

Yingying Jin, She received her bachelor's degree of education in Hebei Normal University, Baoding, Hebei (2001) and master's degree of education in Hebei Normal University, Baoding, Hebei. (2005), now she is a lecturer in Hebei Normal University of Science and Technology, Qinhuangdao, Hebei. Her major fields of study are practical economics.