International Journal of Electronics Communication and Computer Engineering
Volume 6, Issue 5, ISSN (Online): 2249–071X, ISSN (Print): 2278–4209

Energy Consumption Development Based on Gene Expression Programming model in Hebei Province

Hui-fang CHENG, Jian-li ZHAO, Hai-hua Qing

***Abstract* – This paper explorers the causal relationship between energy consumption and economic growth using data of Hebei Province. The study covers the period of 1980 to 2011. We have applied gene expression programming to analyses how they evolved together. Two ways adopted in the forecast, the one is based on the total energy consumption, and the other is based on energy intensity per GDP. Respectively, we build the GEP models for evolution. Two groups of predicted value and the actual value of the data fitting is good. It is found that the model can fit the variation trend of energy consumption and energy intensity per GDP by simulation. The model can reflect the actual status of the energy consumption and economic relations, which can provide scientific basis for policy makers to make a decision. The impact of economic growth and energy consumption is found to be positive.**

***Keywords* – Energy Consumption, Gene Express Programing, Evolution and Analysis.**

I. Introduction

Energy and environment have been the common concerned problem for humankind. It is also an important problem during social and economic development so it is significant to study about the relationship between energy consumption and economic growth. Energy consumption and economic growth is influenced by macroeconomic situations, policy regulation, economic development patterns, technological progress, industrial structure, energy structure and so on in a long span of the study period. Therefore, modeling energy consumption with good accuracy has become even more complex in such a complex environment. So a model can self-learning, without assumes parameters is more naturally than a fixed parameter one used by the existing energy economics literature(for example, see [1]Y Zhang, 2011; [2] Pedroni, 2004; [3] Ozturk, 2010; [4] Payne, 2010), and which play a vital role in understanding the dynamic relationship between energy consumption and economic growth, formulating reasonable strategies and policies for energy and environment and avoid costly mistakes. And because of this energy consumption models are developed specific to the area of Hebei Province.

In recent years, evolutionary algorithms have been used in various energy research because of their ability to self-learn and solve complicated problems [5-12]. By using genetic algorithms, Ceylan and Ozturk [12] have conducted an estimation study up to the year 2025 in Turkey's long-term energy demand forecasting. Ozturk and Ceylan presented some linear and nonlinear models based on GA to estimate electricity consumption in Turkish industrial sector. Based on their findings, the use of import and export variables leads to the improvement of the results [3]. Particle Swarm Optimization (PSO) and Genetic Algorithms (GAs) are applied to attain parameters of the models. Those models have been used in order to forecast and analyze energy in the Iranian metal industry[7]. In a similar study, Canyurt and Ozturk forecast fossil fuel consumption demand using GA [6]. Lee and Tong combine residual modification with genetic programming sign estimation to minimize gray forecasting errors [11]. As a promising variant of genetic programming, gene expression programming (GEP) is a kind of general adaptive random search algorithm, without any prior knowledge, do not understand things internal mechanism, only under the condition of the experimental data develop more accurate formula [13]. Therefore it has very good mapping ability to manage complicated nonlinear problems with more input and uncertainty. And researchers have made use of this advantage in predicting [13-14] , data mining [15-16]and have made very good progress. Therefore, based on gene expression programming algorithm to energy consumption evolution model is established.

The rest of the paper is organized as follows: Section 2 gives energy consumption status and data sources of study area, Section 3 presents GEP model and simulation, Section 4 reports the empirical analysis of results and finally, Section 5 concludes the study.

II. Energy Consumption Current Situation and Data Sources

In recent years, China has a wide range, long haze pollution weather affects hundreds of millions of people, Beijing-Tianjin-Hebei has become "heavy disaster area". Energy market in Hebei province has a rapidly developing due to heavy industrial structure, rapid urbanization and enormous population. However, energy sources in Hebei are quite scarce. In present situation, authorities is necessary to accurately reproduce the relationship between economic growth and energy consumption by using an accurate model.

The data used in this study are taken from the Hebei Statistical Yearbook, and cover 1998–2011. The variables used are total energy consumption (measured in 10000 tons of SCE), and Gross domestic product (measured in one hundred million yuan) in Hebei province. The specific district selected for the study and the timeframe was dictated by data availability.

Data in table I are calculated at current prices, in order to make the data comparable, gross domestic product data used in the study is for the conversion of data at comparable prices, further energy consumption data conversion for energy intensity per GDP in Hebei province.

International Journal of Electronics Communication and Computer Engineering
Volume 6, Issue 5, ISSN (Online): 2249–071X, ISSN (Print): 2278–4209

Table I: Primary consumption in Hebei province from 1998-2011

Time	Total Energy Consumption (10000 tons of SCE)	Energy Intensity Per GDP
1980	3120.50	14.23
1981	3627.80	16.30
1982	3929.05	15.6256
1983	4185.78	14.78
1984	4475.00	13.47
1985	4548.85	11.465
1986	5079.52	11.63
1987	5516.81	10.57
1988	5962.40	8.50
1989	6169.26	7.50
1990	6124.22	6.83
1991	6471.93	6.037
1992	6866.29	5.37
1993	7861.92	4.65
1994	8168.62	3.73
1995	8892.41	3.12
1996	8938.47	2.59
1997	9033.01	2.28
1998	9151.12	2.15
1999	9379.27	2.08
2000	11195.71	2.22
2001	12114.29	2.20
2002	13404.53	2.23
2003	15297.89	2.21
2004	17347.79	2.04
2005	24321.87	2.43
2006	21794.09	1.90
2007	23585.13	1.73
2008	24321.87	1.52
2009	25418.79	1.47
2010	27531.11	1.35
2011	29498.29	1.20

Once the data are ready, the next step is to apply model to learn the different influence for the given input data.

III. Gene Expression Programming Evolution Model and the Simulation Experiment for Energy Consumption and Economic Development

In addition to the traditional way of training the relationship between total energy consumption and economic growth, we applied gene expression programming algorithm (a machine learning mechanism) to our system. Gene Expression Programming (GEP), first introduced by (Ferreira 2001)[18], is an evolutionary algorithm that evolves computer programs and predicts mathematical models from experimental data. GEP begins with a random population of candidate solutions in the form of chromosomes. The chromosomes are evaluated based on a fitness function and selected by fitness to reproduce with modification via genetic operations. The new generation of solutions goes through the same process until the stop condition is satisfied. The fittest individual serves as the final solution. GEP has been used to solve symbolic regression, sequence induction, and classification problems efficiently ([19] Ferreira 2002; [20] Zhou 2003).

A. Build selection environment

On the basis of the research GEP schema theory[21], our GEP model consists of a well-balanced set in order to avoid the creation of models that can only solve a partial, and often marginal, aspect of the overall problem.

The individuals of gene expression programming are encoded in linear chromosomes (19 Ferreira, 2002)that store expressions in breadth-first form. GEP genes are composed of a head and a tail. The head may contains both function and terminal symbols. The tail may contain terminal symbols only. For each problem the head length (denoted h) is chosen by the user. The tail length (denoted by t) is evaluated by:

$t = (nmax-1)h + 1,$

Where, nmax is the number of arguments of the function with more arguments.

Function sets are chosen and shown in table 2, then we learn nmax =2. We choose h=7, then we get t=8, and the length of the gene is 7+8=15. Number of genes=4. Then the length of a chromosome is 60. Parameters for the energy consumption problem is shown in table 3; Max. fitness :1000. During the experiments sampling data of 31 cases are used, 12 -30cases were used for training, the rest cases for testing.

Table II: function set for the energy consumption problem

Representation	Definition	Representation	Definition
+	x+y	x3	x^3
-	x-y	Exp	e^x
*	x*y	log	logx
/	x/y	Abs	|x|
sin	Sin(x)	sqrt	$x^{1/2}$
Tan	Tan(x)	x2	x^2
Cos	Cos(x)	Not	(1-x)

Table III: Parameters for the energy consumption problem

Parameter	Value	Parameter	Value
Number of generations	10000	Population size	30
Number of fitness cases	30 (Table 1)	Function set	(Table 2)
Terminal set	(-10,10)	Head length	7
Number of genes	4	Linking function	+
Mutation	0.044	One-point recombination	0.3
Inversion	0.1	two-point recombination	0.3
IS transposition	0.1	Gene transposition	0.1

RIS transposition	0.1	Precision	0.0001
Fitness function	Eq(1)	Gene recombination	0.1

B. The choice of fitness function

Fitness function is applied to every training individual and produces a score accordingly. Then all individuals in the same training are ranked according to their scores, and individuals with top scores have greater probability to be selected to the next generation. When we speak of the fitness of an individual, on the one hand, it is always relative to a particular environment and, in the other, it is also relative to the measure (the fitness function) we are using to evaluate them. Consequently, the success of a problem not only depends on the way the fitness function is designed but also on the quality of the selection environment [19 p65].

Then the square of multiple correlation coefficient is used as the fitness to evaluate all the individuals and is expressed by equation (1).

R2=1-SSE/SST (1)

$$SSE=\sum_{j=1}^{m}\left(y_j-\hat{y}_j\right)^2,\ SST=\sum_{j=1}^{m}\left(y_j-\bar{y}_j\right)^2,$$

Refer to "(1),", every individual's forecast is denoted by $\hat{y}_j$ and the mean value of the generation by $\bar{y}_j$.

As it stands, R2 cannot be used directly to measure the fitness of the evolved models, for fitness proportionate selection, the fitness must increase with efficiency. Thus, the following equation is used to evaluate the fitness of an individual i:

Fitness = R2*1000 (2)

C. The simulation of energy consumption and economic evolution model

In the training stage, two set of training data (1998-2010 data set of Total Energy consumption and energy intensity per GDP) are provided, then they are ranked and evolved with their scores by GEP using GeneXproTools. In the testing stage, another two data set (1980-2011 data set of Total Energy consumption and energy intensity per GDP) are supplied to test the similarity between the twice evolutionary processes.

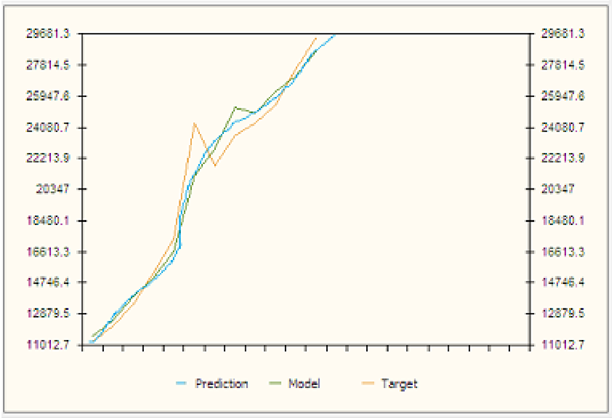


Fig.1. The evolution curve of Total Energy consumption with genetic algorithm 1998-2010

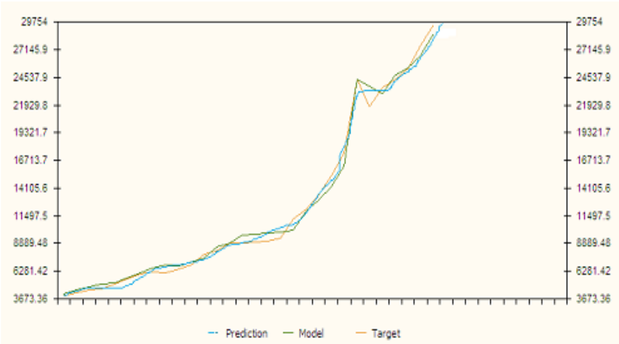


Fig.2. The evolution curve of Total Energy consumption with genetic algorithm 1980-2011

MSE: 1.41E+06 RAE: .162353244221607 RSE: .038472241401113 R-square: .961546954402695 Fitness: 836.020148479529
RMSE: 1,186.6784304253 MAE: 894.824066338133 RRSE: .196143420488969 Corr. Coef.: .980585006209403 Calc. Errors: 0

	Target	Model	Residual
1	11195.71	11520.4773660396	324.767366039552
2	12114.29	12511.12424297	396.834242970013
3	13404.53	13908.8624912061	504.332491206087
4	15297.89	15006.375545189	291.514454810951
5	17347.79	16590.86484529	756.925154709988
6	24321.87	21153.1649728492	3168.70502715076
7	21794.09	22803.6131372351	1009.52313723506
8	23585.13	25286.284177404	1701.154177404
9	24321.87	24936.5648013538	614.69480135806
10	25418.79	26220.1719315455	801.38193154549
11	27531.11	27130.4604394069	400.649560593087
12	29498.29	28730.8835489614	767.406451038558

Fig.3 Performance comparison in terms of GEP model validation parameters (Total Energy consumption with genetic algorithm 1998-2010)

International Journal of Electronics Communication and Computer Engineering
Volume 6, Issue 5, ISSN (Online): 2249–071X, ISSN (Print): 2278–4209

MSE: 4.14E+05 RAE: .075316172480515 RSE: .006648857127103 R-square: .993894148047110 Fitness: 924.607055765426
RMSE: 643.207853638357 MAE: 503.746010235929 RRSE: .081540524447070 Corr. Coef.: .996942399563340 Calc. Errors: 0

	Target	Model	Residual
1	3929.05	4120.15938738482	191.109387384817
2	4185.78	4462.009406088	276.229406088497
3	4475	4744.34856478806	269.34856478806
4	4548.85	5042.29896235858	493.448962358578
5	5079.52	5145.16769713734	65.6476971373377
6	5516.81	5660.92021284591	144.1102128459
7	5962.4	6118.53960730176	156.139607301764
8	6169.26	6599.453861409	430.193861409002
9	6124.22	6834.23973217707	710.019732177072
10	6471.93	6801.1149376001	329.184937600105
11	6866.29	7139.72783965699	273.437839656995
12	7861.92	7544.58141062903	317.338589370972
13	8168.62	8542.99163120867	374.371631208674
14	8892.41	8905.38232047254	12.9723204725433
15	8938.47	9632.8128926065	694.342892606499
16	9033.01	9722.9725495706	689.9625495706
17	9151.12	9821.02644865015	669.906448650148
18	9379.27	9926.36569792061	547.095697920606
19	11195.71	10175.6145061759	1020.09549382407
20	12114.29	11957.6830944455	156.606905554541
21	13404.53	12953.0696401829	451.460359817058
22	[illegible]	[illegible]	[illegible]

Fig.4 Performance comparison in terms of GEP model validation parameters
(Total Energy consumption with genetic algorithm 1980-2011)

Fig.5. The evolution curve ofenergy intensity per GDP with genetic algorithm 1998-2010

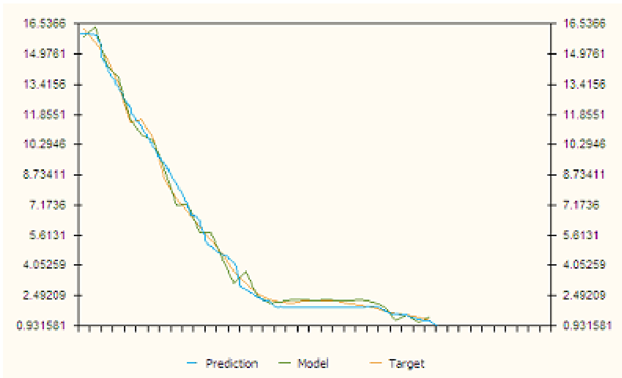


Fig.6. The evolution curve of energy intensity per GDP with genetic algorithm 1980-2011

MSE: .010975084856090 RAE: .190301504828520 RSE: .071533875548901 R-square: .929167784078993 Fitness: 788.980669711263
RMSE: .104762039194024 MAE: .066764111277339 RRSE: .267458175326351 Corr. Coef.: .963933495672286 Calc. Errors: 0

	Target	Model	Residual
1	2.22	2.20549593110587	1.45040688941318E-02
2	2.2	2.24989174415465	4.98917441546456E-02
3	2.23	2.23816466685507	8.16466685507189E-03
4	2.21	2.21749568968956	7.49568968955883E-03
5	2.04	2.22987981745984	.189879817459843
6	2.43	2.15613981662509	.273860183374914
7	1.9	1.83358408988801	.066415910111989
8	1.73	1.81177937453294	8.17793745329405E-02
9	1.52	1.60005529940256	8.00552994025581E-02
10	1.47	1.47937644469418	9.37644469417753E-03
11	1.35	1.35573730822296	5.73730822295748E-03
12	1.2	1.21400882793528	1.40088279352828E-02

Fig.7. Performance comparison in terms of GEP model validation parameters
(Energy Intensity Per GDP with genetic algorithm 1998-2010)

MSE:	.008536831253517	RAE:	.061614473262084	RSE:	.003640538082953	R-square:	.997360965649963	Fitness:	943.096507079666
RMSE:	.092394974178885	MAE:	.073565342115179	RRSE:	.060336871670258	Corr. Coef.:	.998679611111573	Calc. Errors:	0

	Target	Model	Residual
1	6.832550512	7.05172310660551	.219172594605506
2	6.036853937	6.14641985307448	.109565916074482
3	5.370582714	5.27865512275606	9.19275912439419E-02
4	4.649712569	4.75456222536982	.104849656369816
5	3.734243357	3.79590262074523	6.16592637452347E-02
6	3.120669446	3.1630359367493	4.23664907492967E-02
7	2.588632395	2.65116737872042	6.25349837204232E-02
8	2.28465165	2.28573180788789	1.08015788789384E-03
9	2.150164121	2.02012704400232	.130037076997684
10	2.077730446	2.1099344520278	3.22040060277997E-02
11	2.219627039	2.24981073215916	3.01836931591648E-02
12	2.195905507	2.31246179929763	.116556292297632
13	2.22730186	2.2682767135865	4.09748535864995E-02
14	2.210265887	2.1707658442261	3.95000427738954E-02
15	2.046301856	2.07682687802226	3.05250220222617E-02
16	1.981199767	2.00814571532933	2.69459483293308E-02
17	1.900492692	1.8526205363439	4.78721556560955E-02
18	1.733267829	1.68144587521865	5.18219537813498E-02
19	1.518980488	1.61214898347087	9.31684954708698E-02
20	1.474794436	1.50248086685088	2.76864308508837E-02
21	1.349944053	1.39571949893055	4.57754459305515E-02
22	1.203237836	1.41526729125333	.212029455253328

Fig.8. Performance comparison in terms of GEP model validation parameters
(energy intensity per GDP with genetic algorithm 1980-2011)

The simulation results are shown in Fig.1-Fig. 8, where it becomes obvious that the GEP algorithm fits these values well. In addition, the validation indexes RMSE, MSE, RAE, MAE,RSE, RRSE and R2 were compared with a part of experimental data and the estimated values for the test set given in Fig.3, Fig.4and Fig.7, Fig.8. The results also show that the greater the amount of given data, the model is more powerful. As given in Fig.1-Fig.8, the proposed model for the period1980–2011 have the best fitness than the propose model fortheperiod1998–2010.

IV. Conclusion

GEP model has attractive features, such as the ability to self-learning without any prior knowledge, do not understand things internal mechanism, only under the condition of the experimental data develop more accurate model. However, a proper set of parameters for specific problem is essential for good performance and efficiency of GEP model in estimation real numbers. GEP model is developed by getting the relationship between energy consumption and economic development, such as energy intensity per GDP. When a robust GEP model with high fitting and reliability was constructed, statistical data of different period used in the training stage are shown to be accurately estimated. The experiment results show that the proposed method is feasible and available.

Compared with the whole economic development speed and level in Hebei province, the energy consumption growth rate maintained a positive relationship. Namely as the growth of the economy, energy consumption will increase, and there exists a long-term consistency. Industrial proportion is larger in the economy in Hebei province. Industry is one of the major energy-consuming sectors using energy as a production factor for developing the economy. The increase of the energy input will lead to the increase of economic output, by the same token, when the economy expanded; lead to demand for energy elements will increase. Therefore, Hebei is in urgent need of development of green energy, developing low-carbon economy, the sustainable development of the scientific road. Actually it is so.

Acknowledgment

This research has been made possible through financial support by youth fund in Hebei province department of education (No. 20130020).

References

[1] Y Zhang. Interpreting the dynamic nexus between energy consumption and economic growth: Empirical evidence from Russia Int. Energy Policy. Volume 39, Issue 5, May 2011, Pages 2265–2272. http://www.sciencedirect.com/science/article/pii/S0301421511000346. (2011)

[2] Pedroni.P. Panel cointegration: asymptotic and finite samples properties of pooled time series tests with an application to the ppp hypothesis. Econometric Theory 20, 597-625. https://ideas.repec.org/a/cup/etheor/v20y2004i03p597-625_20.html, (2004)

[3] Ozturk, I.. A literature survey on energy-growth nexus. Energy Policy 38 340-349. http://wenku.baidu.com/link?url=w52BuDaTjBggeliK267s7ZTkQq8X3Me2-zAMCJCMqaBUQ7PXNAVugjgDjTfBsSNM0ZELlqGORYMCd0-Adm9pZh6knJXrfDITcskhS4g0HSW , (2010).

[4] Payne, J.E. Survey of the international evidence on the causal relationship between energy consumption and growth.J.Eco. Stud.37,53-95. http://www.docin.com/p-650544320.html. (2010)

[5] Piltan M, Shiri H, Ghaderi S.F. Energy demand forecasting in Iranian metal industry using linear and nonlinear models based on evolutionary algorithms. Energy Conversion and Management 58 (2012) 1–9. www.elsevier.com/locate/enconman

[6] Canyurt OE, Ozturk HK. Application of genetic algorithm (GA) technique on demand estimation of fossil fuels in Turkey. Energy Policy (2008); 36(7): 2562–9. https://ideas.repec.org/a/eee/enepol/v36y2008i7p2562-2569.html

[7] Azadeh A, Ghaderi SF, Tarverdian S, Saberi M. Integration of artificial neural networks and genetic algorithm to predict electrical energy consumption. Appl Math Compute (2007); 186(2):1731–41. http://link.springer.com/chapter/10.1007%2F11427391_123

[8] AlRashidi MR, EL-Naggar KM. Long term electric load forecasting based on particle swarm optimization. Appl Energy (2010);87(1):320–6. https://ideas.repec.org/a/eee/appene/v87y2010i1p320-326.html

[9] E. Assareh, M. A Behrang, A. Ghanbarzdeh. Forecasting energy demand in Iran using Genetic Algorithm(GA) and Particle Swarm Optimization (PSO) Methods. Energy Sources, Part B: Economics, Planning, and Policy, vol. 7, no. 4, pp. 411-422, (2012). http://academic.research.microsoft.com/Publication/57705171/forecasting-energy-demand-in-iran-using-genetic-algorithm-ga-and-particle-swarm-optimization

[10] Canyurt OE, Ozturk HK, Hepbasli A, Utlu Z. Estimating the Turkish residential–commercial energy output based on genetic algorithm (GA) approaches. Energy Policy (2005);33(8):1011–9. https://ideas.repec.org/a/eee/enepol/v33y2005i8p1011-1019.html

[11] Lee Y-S, Tong L-I. Forecasting energy consumption using a grey model improved by incorporating genetic programming. Energy Convers Manage (2011);52(1):147–52. http://www.sciencedirect.com/science/article/pii/S0196890410002748.

[12] Ceylan H, Ozturk HK. Estimating energy demand of Turkey based on economic indicators using genetic algorithm approach. Energy Convers Manage (2004);45:2525–37. http://www.sciencedirect.com/science/article/pii/S0196890403003479.

[13] Li Y, Chen X, He D .J. Research on GEP algorithm and its applications in foodstuff yield prediction from Shaanxi province[c]//Advanced Computer Theory and Engineering. (2008). ICACTE08. International Conference on. IEEE.2008:690-693.

[14] S SS Hosseini, A H Gandomi. Short-term load forecasting of power systems by gene expression programming. Neural Computing and Applications. (2012).21(2):377-389. http://link.springer.com/article/10.1007%2Fs00521-010-0444-y.

[15] S Xue, J Wu. Gene expression programming based on symbiotic evolutionary algorithm. International Conference on Artificial Intelligence, Management Science and Electronic Commerce - AIMSEC, (2011).

[16] K A Eldrandaly, A Negm. Performance Evaluation of Gene Expression Programming for Hydraulic Data Mining. The International Arab Journal of Information Technology, (2008).5(2): 126-131.

[17] Data source: Hebei province Statistical Yearbook, 2000-2014, China statistical publishing house

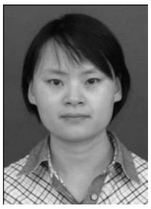
[18] Ferreira, C. Gene expression programming: anew adaptive algorithm for solving problems. Complex Systems, (2001).13(2): 87-129. http://www.gene-expression-programming.com/webpapers/GEPfirst.pdf

[19] C. Ferreira,. Gene expression programming mathematical modeling by an artificial intelligence. (2nd Ed.) (2006, Jul 20). http://link.springer.com/chapter/10.1007%2F978-3-642-12842-4_33.

[20] Zhou, C., Xiao, W., Tirpak, T. M., and Nelson, P. C. Evolving classification rules with gene expression programming. IEEE Transactions on Evolutionary Computation, (2003).7(6):519 – 531. http://d.wanfangdata.com.cn/NSTLQK_NSTL_QKJJ0213245933.aspx.

[21] H. Cheng, J. Zhao, H. Wang. Research of exact schema theorems of programming. Journal of computer and Information systems. submitted for publication

AUTHOR'S PROFILE



Huifang Cheng

was born in Wei Zhuang Village, XingTai City, in 1980. She received the B.S. in Hebei Normal University, ShiJiaZhuang, in 2002, and M.S. degrees in computer application technology from Hebei University of Engineering, Handan, China, in 2009. From 2002 to 2007, she was an Assistant with information and electricity-engineering institute, Hebei University of Engineering. Since 2007, he has been a lecturer with information and electricity-engineering institute, Hebei University of Engineering. She is the author of three books, more than 15 articles. Her research interests include evolutionary algorithm and data mining and applications.

**This paper need color printing.*