Comparison of Price Forecasting Models based on Collective Intelligence

Yuanxin Li[1,a], Wei Chen[2,b], Zhi Wang[3,c]
[1] College of Engineering, Ocean University of China, Qingdao, China
[2] Mount Everest property insurance company limited by shares, Beijing, China
[3] College of Computer, University of Mississippi, USA
E-mail: [a]842708989@qq.com, [b]3094158007@qq.com, [c]wangz6@go.olemiss.edu

***Abstract*—The determination of commodity trading price is related to the fairness and efficiency of different parties on both sides. Within the framework of this collective-intelligence (group wisdom) scenario, this study compares the prediction methods in commodity price. After quantifying the commodity trading price in all the frequently used models under the collective intelligence basis, the radial basis network has the best performance, and the overall prediction is boosted nearly 20% than the traditional method. Due to non-time sensitive data samples, prediction feature importance may be contrary to the common sense. But this also indicates that the excess human opinions added to the valuation process are not always helpful. Instead, the real-time price will be more likely to tell which feature should have more importance, which is also the reflection of Collective Intelligence from another perspective.**

***Index Terms*—group wisdom, collective intelligence, commodity price, prediction, machine learning, marketing model**

I. Introduction

Throughout the entire development of economy and various industries, the most important business activities in the market economy is commodity trading. Trading activities can be divided into general trading and auction, and the transaction price deeply exemplifies the game theory between the parties of supply and demand, which constantly reflects the intrinsic value of goods. With the rapid development of the economy and the fast definition updating for commodities values, the original asset evaluation method has remained difficult to meet the needs of various emerging new scenarios. Under such a background, the problem of re-examining the prediction of commodity transaction prices has become very necessary.

The process of determining the transaction price is in fact a game process, which is the instantaneous knowledge generated out of collective intelligence that is shared or formed in different parties represented by the individual and collective efforts through cooperation and competition. And it only appears in the decision-making process reached by consensus. Collective IQ (Intelligence Quotient) is an indicator of group intelligence. It is the result of a combination of data and information knowledge extracted from organizations of software, to hardware and human participation. Technically, it can be defined as an emerging and combined property between people and information processing methods. In the application of Co-intelligence theory, advantages of unorganized decision-making are divided into three categories: cognition, cooperation and coordination. If the mechanism of mapping from commodity value to price can be clarified, the experience obtained in the mapping scenario and process can speed up the process of commodity price determination and improve the efficiency of the trading and transaction activity. This will not only save both the time of the buyer and seller, but is also promotes the rational and effective distribution for the resources from the transaction platform.

The current Co-intelligence based price forecasting model is formulated upon the hypothesis that the value of an item or a service is determined by its innate attributes or characteristics. In environmental economics, researchers also use the technology to evaluate or quantify the specific asset environment attributes. For example, the pricing of auto piloting cars requires not only the consideration of their own material costs, but also the consideration of adjacent environmental and human contributions, including the embedded software systems[1-3]. Maritime transport vessels not only need to consider equipment environments such as satellite navigation and positioning systems[4-9], it is also necessary to consider the long-distance environment variations such as the extreme sea conditions of typhoons and hurricanes during the traveling[10-20]. Similarly, the pricing of the health care for the human body also needs to be relied upon various sections in targeting choosing, mechanism elucidation, drugs design and synthesis as well as multiple clinical trials[21-34]. As can be seen from the aforementioned, the potential sources of benefits, production cost and operating cost are all attributed to the data generation and collection, software and hardware construction as well as experts' knowledge[35], and the security cost to prevent unexpected system breaches[36-40], all of which contribute into the final value of the commodity. Even for specific technique such as space drilling, expert's knowledge will add up to the final value[41-48]. The same is true even for wireless communication and networks[49-52] or its new application field of environment protection[53]. In this regard, the additive but often non-apparent environmental services cannot be ignored, which mainly refers to natural non-renewable land resources, water and air (including

the associated ecosystems and the bio-organisms that grow in them). As all human acclivities rely on these, it also means that the natural environment provides a very important use or benefit under the broader concept of "service". In various price quantifying scenarios, the auction price is very representative in that it can be expanded to generate more specific scenes.

II. ANALYSIS OF PRICING MODELS

The advantage of the Hedonic Price Model (HRM) is that it is highly intuitive. It directly solves the incremental benefits of monetization in a logical way using the market prices. The main disadvantage is its need for a fairly complex empirical process. Weighted K-nearest neighbor (weighted KNN) is a method suitable for price prediction in Co-intelligence scenarios54-57. It can be used to better reflect the price change that stems from the change of commodity characteristic variable that is measured by the distance between the features/ price predictors. Decision Tree (DT) is also a classification (prediction) and learning method. By learning the data input and output patterns in the training set, the data in the prediction set is classified (predicted), and whose outstanding feature is to reproduce the process of human decision making. For the regression tree (CRT) with numerical input variables, the data is first sorted in ascending order; then, in value from small to large, the median value of the adjacent value is used for grouping. The difference from the sample output in the group will be calculated as heterogeneity. The pruning process of the regression tree is usually performed by a combination of pre-pruning and post-pruning which is aim to limit the decision tree depth, and groupings of parent node and child nodes. For chi-squared automatic interaction detector (CHAID), as a decision tree algorithm, its input and output variables can be either categorical or numerical. The algorithm is very similar to CRT despite the difference that when the data is pre-processed, the numerical variables are processed by binning, and the categorical variables are combined to form a "superclass". The advantage of this is that the input variables will be filtered, which is beneficial since it simplifies the growth process of the decision tree.

III. MODEL COMPARISON RESULT ANALYSIS

The data source for this study comes from the papers on commodity price, some of which are shown in Table 1. The sample has a total of 3,730 observations and there are no missing values. In order to compare the quantitative effects of different research methods, the data set is randomly divided into training and prediction sets. The size for the training set is 3000, accounting for about 80% of the total data, with the prediction set sample size around 730, accounting for about 20%. The final indicator to be compared will be the mean absolute error and relative error of the predicted and true values of the 730 samples in the prediction set.

TABLE 1. PARTIAL DATA FROM THE MODELING ANALYSIS

price	speed	hd	ram	screen	cd	multi	premium	ads	trend
2534	66	730	8	17	1	0	1	162	22
1839	25	212	4	15	0	0	0	339	17
2594	66	107	4	14	1	0	1	298	8
2099	33	120	4	14	0	0	0	246	9
2744	50	340	8	17	0	0	1	275	12
1449	25	120	4	14	0	0	0	249	7
3090	66	528	16	15	0	0	1	216	13
2495	33	250	8	15	0	0	1	176	6
3244	66	452	16	14	0	0	1	249	7
2794	50	720	16	15	1	0	1	205	21

TABLE 2 DESCRIPTIVE STATISTICAL ANALYSIS OF DATA SETS

Descriptive Statistics					
	N	minimum	maximum	mean	Std.deviation
price	3730	999	5399	2257.29	621.754
speed	3730	25	100	54.40	20.957
hd	3730	80	2100	425.14	262.059
Ram	3730	2	32	8.15	5.341
screen	3730	14	17	14.76	.972
cd	3730	0	1	.50	.500
multi	3730	0	1	.18	.387
premium	3730	0	1	.89	.321
ads	3730	39	339	218.63	76.658
trend	3730	1	35	16.06	8.096
Valid N	3730				

Where price = PCs 486 PC price (USD), speed = running frequency (MHz), hd = hard disk size (MB), ram = memory size (MB), screen = screen size (inches), cd = dummy variable, for CD-ROM, its value is 1, multi =

dummy variable. If it contains multimedia kit (such as speaker, sound card), its value will be 1, premium = dummy variable. If the manufacturer is a high-end company (such as IBM, COMPAQ), the value will be 1, ads = Monthly price list quantity, trend = time trend month mark (from January 1993 to December 1995). Descriptive statistical analysis of all data is shown in Table 2.

The mean absolute error (MAE) and the mean relative error (MER) were used to evaluate the prediction results. From equations (1) and (2). It is obvious the smaller the mean absolute or relative error, the better the prediction effect (Table 3). The differences between the prediction effects of the models are also vividly shown in Fig. 1.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y - \hat{y}\right| \quad (1)$$

$$MER = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|y - \hat{y}\right|}{y} \quad (2)$$

TABLE 3 COMPARISON OF MAE AND MRE OF VARIOUS MODELS

	HRM	Weighted KNN	CRT	CHAID	MLP	RBF
MAE	214.8	201.0	293.1	298.9	189.4	447.4
MRE	9.51%	8.62%	12.93%	12.89%	8.42%	20.54%

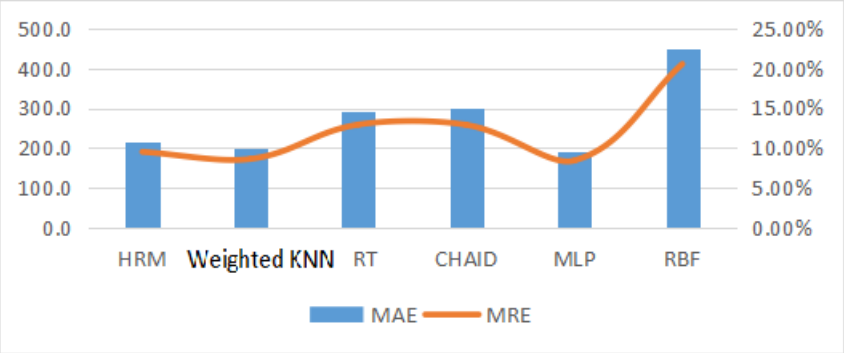

Figure 1 MAE and MRE in bar and line chart

It can be seen that the weighted K-nearest neighbor and the multi-layer perceptron have relatively better prediction results, with the mean relative errors 8.62% and 8.42%, respectively, which is 11.92% and 12.11% lower than the radial basis function network who has the worst prediction effect. The mean absolute error is 201.0 and 189.4, respectively, which is lower than the predicted radial basis function network by 246.4 and 258.0, respectively.

Table 4 gives the predictor importance for different models. Three different intensities of blue shades indicate that features that have varying weights in different models. The darker the color, the more important it is.

TABLE 4 IMPORTANCE OF PREDICTORS IN DIFFERENT MODELS

Models		Predictor/feature importance from high to low								
HRM	HRM	premium	multi	screen	cd	trend	ram	speed	hd	ads
KNN	Weight KNN	trend	premium	ram	speed	hd	screen	multi	cd	ads
DT	CRT	ram	trend	hd	speed	premium	ads	screen	cd	multi
	CHAID	\								
ANN	MLP	hd	ram	ads	trend	premium	speed	screen	cd	multi
	RBF	ram	screen	hd	multi	speed	trend	ads	cd	premium

IV. CONCLUSIONS

Predictors can help sellers finding the right-on spot for commodity price, and it also helps consumers to choose the right product based on the key factors in building up the final prices. When using a variety of different features to predict numerical data, the weighted K-nearest neighbor method has better performance than the decision tree. In the application, the above method will effectively guide the trading/transaction process with more accurate and controllable sales price. It is also very helpful to determine market behaviors for different parties concerning the price change and the estimated psychological purchase behavior. Additionally, and more specifically, 1. The features with the best predictive effect of the multi-layer perceptron in the upper one-third interval of the normalized importance are hd and ram; For the similar predictors from the second best weighted K-nearest neighbor, there are trend and premium. 2. Among these established models, the best ones are the multi-layer perceptron and the weighted K-nearest neighbor method.

Even they are from the same neural network category, there is difference between the multi-layer perceptron and the radial basis function network from our results. This indicates that in some data scenarios, there will be some extent superiority of multi-layer perceptrons over radial basis function networks, therefore it is somewhat very risky for the indiscriminate application of the two.

REFERENCES

[1] PONCE-LÓPEZ V, CHEN B, OLIU M, et al. ChaLearn LAP 2016: First Round Challenge on First Impressions - Dataset and Results. Computer Vision – ECCV 2016 Workshops. Springer, Cham, 2016: 400–418.

[2] Zhang SF, Shen W, Li DS, Zhang XW, Chen BY，Nondestructive ultrasonic testing in rod structure with a novel numerical Laplace based wavelet finite element method. Latin American Journal of Solids and Structures, 2018, 15(7):1-17, e48.

[3] ESCALANTE H J, PONCE-LÓPEZ V, WAN J, et al. ChaLearn Joint Contest on Multimedia Challenges Beyond Visual Analysis: An overview. 2016 23rd International Conference on Pattern Recognition (ICPR). 2016: 67–73.

[4] Liu, X.; He, Y.; Fu, H.; Chen, B.; Wang, M.; Wang, Z. How Environmental Protection Motivation Influences on Residents' Recycled Water Reuse Behaviors: A Case Study in Xi'an City. Water2018, 10, 1282.

[5] Song Jiang , Minjie Lian, CaiWu Lu ,Shunling Ruan ,Zhe Wang ,and Baiyu Chen ,.SVM-DS fusion based soft fault detection and diagnosis in solar water heaters, Energy Exploration & Exploitation, DOI: 10.1177/0144598718816604

[6] CHEN B, ESCALERA S, GUYON I, et al. Overcoming Calibration Problems in Pattern Labeling with Pairwise Ratings: Application to Personality Traits. Computer Vision – ECCV 2016 Workshops. Springer, Cham, 2016: 419–432.

[7] Liu, X.; He, Y.; Fu, H.; Chen, B.; Wang, M .Scientometric of Nearly Zero Energy Building Research:A Systematic Review from the Perspective of Co-Citation Analysis.Journal of Thermal Science.

[8] Wang L-P, Chen B, Zhang J-F, Chen Z. A new model for calculating the design wave height in typhoon-affected sea areas. Natural Hazards. Springer Netherlands; 2013;67: 129–143.

[9] Chen B, Yang Z, Huang S, Du X, Cui Z, Bhimani J, et al. Cyber-physical system enabled nearby traffic flow modelling for autonomous vehicles. 36th IEEE International Performance Computing and Communications Conference, Special Session on Cyber Physical Systems: Security, Computing, and Performance (IPCCC-CPS) IEEE. 2017.

[10] ANTHONY BARRS, CHEN BAIYU, How Emerging Technologies Could Transform Infrastructure. http://www.governing.com/gov-institute/voices/col-hyperlane-emerging-technologies-transform-infrastructure.html.

[11] CHEN B, WANG B. Location Selection of Logistics Center in e-Commerce Network Environments. American Journal of Neural Networks and Applications, Science Publishing Group, 2017, 3(4): 40-48.

[12] Wang L, Xu X, Liu G, Chen B, Chen Z. A new method to estimate wave height of specified return period. Chin J Oceanol Limnol. Science Press; 2017;35: 1002–1009.

[13] CHEN B, LIU G, WANG L. Predicting Joint Return Period Under Ocean Extremes Based on a Maximum Entropy Compound Distribution Model. International Journal of Energy and Environmental Science, 2017,2(6): 117-126. 2017.

[14] Wang L-P, Chen B-Y, Chen C, Chen Z-S, Liu G-L. Application of linear mean-square estimation in ocean engineering. China Ocean Eng. Chinese Ocean Engineering; 2016;30: 149–160.

[15] Baiyu Chen,Guilin Liu,Jianfang Zhagn et al., A calculation method of design wave height under the three factors of typhoon: China, ZL 2016 1 0972118.X. Patent, 2016-10-31.

[16] Liping Wang, Liu Guilin, Baiyu Chen et al., et al, Typhoon influence considered method for calculating combined return period of ocean extreme value: China, ZL 2010 1 0595807.6. Patent, 2013-03-20.

[17] Guilin Liu, Zheng Zhenjun，Liping Wang,et al., Power-type wave absorbing device and using method thereof: China, ZL 2015 1 0575336.5. Patent, 2017-11-03.

[18] Liping Wang, Liu Guilin, Baiyu Chen, et al., Typhoon based on the principle of maximum entropy waters affect the design wave height calculation method: China, ZL 2010 1 0595815.0[P]. Patent, 2015-08-19.

[19] LIU G, CHEN B, WANG L, et al. Wave Height Statistical Characteristic Analysis. Journal of oceanology and limnology. 2018, 36(4):1123-1136.

[20] CHEN Baiyu, LIU Guilin, WANG Liping, ZHANG Kuangyuan, ZHANG Shuaifang, Determination of Water Level Design for an Estuarine City. Journal of oceanology and limnology. 2019, https://doi.org/10.1007/s00343-019-8107-z

[21] Cheun Y, Kou Y, Stevenson B, Kim H-K, Koag MC, Lee S. Synthesis of C17-OH-north unit of ritterazine G via "Red-Ox" modifications of hecogenin acetate. Steroids. 2013;78: 639–643.

[22] Kou Y, Lee S. Unexpected opening of steroidal E-ring during hypoiodite-mediated oxidation. Tetrahedron Lett. 2013;54: 4106–4109.

[23] Kou Y, Others. Structural and kinetic study of N7-methyl, N7-benzyl and C8-chloro guanine lesions using human DNA polymerase β [Internet]. 2015. Available: https://repositories.lib.utexas.edu/handle/2152/46815

[24] S Lee, Y Kou, M Koag. Mechanism of alkylation and platination-induced mutagenesis. ENVIRONMENTAL AND MOLECULAR MUTAGENESIS 59, 107-107.

[25] Kou Y, Koag M-C, Lee S. N7 methylation alters hydrogen-bonding patterns of guanine in duplex DNA. J Am Chem Soc. 2015;137: 14067–14070.

[26] Cheun Y, Koag MC, Kou Y, Warnken Z, Lee S. Transetherification-mediated E-ring opening and stereoselective "Red-Ox" modification of furostan. Steroids. 2012;77: 276–281.

[27] Lei X, Kou Y, Fu Y, Rajashekar N, Shi H, Wu F, et al. The cancer mutation D83V induces an α-helix to β-strand conformation switch in MEF2B. J Mol Biol. 2018; doi:10.1016/j.jmb.2018.02.012

[28] Lei X, Shi H, Kou Y, Rajashekar N, Wu F, Sen C, et al. Crystal Structure of Apo MEF2B Reveals New Insights in DNA Binding and Cofactor Interaction. Biochemistry. 2018;57: 4047–4051.

[29] Koag M-C, Kou Y, Ouzon-Shubeita H, Lee S. Transition-state destabilization reveals how human DNA polymerase β proceeds across the chemically unstable lesion N7-methylguanine. Nucleic Acids Res. 2014;42: 8755–8766.

[30] Yang Y, Chen L, Pan H-Z, Kou Y, Xu C-M. Glycosylation modification of human prion protein provokes apoptosis in HeLa cells in vitro. BMB Rep. 2009;42: 331–337.

[31] Koag MC, Cheun Y, Kou Y, Ouzon-Shubeita H, Min K, Monzingo AF, et al. Synthesis and structure of 16,22-diketocholesterol bound to oxysterol-binding protein Osh4. Steroids. 2013;78: 938–944.
[32] Kou Y, Koag MC, Cheun Y, Shin A, Lee S. Application of hypoiodite-mediated aminyl radical cyclization to synthesis of solasodine acetate. Steroids. 2012;77: 1069–1074.
[33] Kou Y, Cheun Y, Koag MC, Lee S. Synthesis of 14',15'-dehydro-ritterazine Y via reductive and oxidative functionalizations of hecogenin acetate. Steroids. 2013;78: 304-311.
[34] Kou Y, Koag M-C, Lee S. Structural and Kinetic Studies of the Effect of Guanine N7 Alkylation and Metal Cofactors on DNA Replication. Biochemistry. 2018;57: 5105–5116.
[35] Baiyu Chen, Kuangyuan Zhang, Liping Wang, Song Jiang, Guilin Liu ,Generalized Extreme Value - Pareto Distribution Function and Its Applications in Ocean Engineering, Chinese Ocean Engineering; 2019;2
[36] Chen Y, Wang Z, Whalley D, Lu L. Remix: On-demand live randomization. In Proceedings of the Sixth ACM Conference on Data and Application Security and Privacy 2016 Mar 9 (pp. 50-61). ACM.
[37] Wang X, Qi Y, Wang Z, Chen Y, Zhou Y. Design and implementation of SecPod, a framework for virtualization-based security systems. IEEE Transactions on Dependable and Secure Computing. 2017 Mar 2.
[38] Wang X, Chen Y, Wang Z, Qi Y, Zhou Y. SecPod: a Framework for Virtualization-based Security Systems. In USENIX annual technical conference 2015 Jul 8 (pp. 347-360).
[39] Zhou Y, Wang X, Chen Y, Wang Z. Armlock: Hardware-based fault isolation for arm. In Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security 2014 Nov 3 (pp. 558-569). ACM.
[40] Chen Y, Zhang Y, Wang Z, Xia L, Bao C, Wei T. Adaptive android kernel live patching. In Proceedings of the 26th USENIX Security Symposium (USENIX Security 17) 2017 Aug.
[41] Deng Z Q, Tian Y, Tang D W, et al. (2013). Research on new structure coring bit for extraterrestrial bodies exploration. Chin. J. Mech. Eng., 49(19), 104–110
[42] TIAN YE, DENG ZONG-QUAN. Coring Bit with Enhanced Structural Parameters for Improved Lunar Soil Sampling and Reduced Mechanical Disturbance. Journal of aerospace engineering, American Society of Civil Engineers, 2016, 29(4): 04016015.
[43] TIAN Y, DENG Z-Q, TANG D-W, et al. Drilling Power Consumption Analysis of Coring Bit in Lunar Sample Mission. Journal of aerospace engineering, American Society of Civil Engineers, 2017, 30(5): 04017055.
[44] TIAN Y, SUN Z H, ZHANG L. Analysis of Structural Parameters of Flexible Spiral Conveyor Blades and Study of its Transmission Power. Ammonia Plant Safety and Related Facilities, 2012, 200: 455–458.
[45] TIAN Y, TANG D, DENG Z, et al. Drilling power consumption and soil conveying volume performances of lunar sampling auger. Chinese Journal of Mechanical Engineering, Chinese Mechanical Engineering Society, 2015, 28(3): 451–459.
[46] Tian, Y., Deng, Z. Q, and Tang, D. W. (2012). Structure parameters optimization and simulation experiment of auger in lunar soil drill-sampling device. Chin. J. Mech. Eng., 48(23), 10–15 (in Chinese).
[47] Ye Tian, Panpan Yuan, Fei Yang, et al. Research on the Principle of a New Flexible Screw Conveyor and Its Power Consumption. Applied Sciences. 2018, 8(7): 1038-1045.
[48] Tian, Y., Deng, Z.Q, Tang, D.W., et al. Power consumption of lunar subsurface coring driller and earth environment stimulant experiments. Journal of Jilin University Engineering and Technology Edition.2016
[49] Cui, P.; Tonnemacher, M.; Rajan, D.; Camp, J. WhiteCell: Energy-efficient use of unlicensed frequency bands for cellular offloading. In Proceedings of 2015 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN), Stockholm,Sweden, 29 Sept.-2 Oct. 2015;IEEE:Piscataway, NJ,2015; pp.188-199.
[50] Cui, P.; Liu, H.; He, J.; Altintas, O.; Vuyyuru, R. ; Rajan, D.; Camp, J. Leveraging diverse propagation and context for multi-modal vehicular applications. In Proceedings of 2013 IEEE 5th International Symposium on Wireless Vehicular Communications (WiVeC), 2 June – 3 June 2013; IEEE: Dresden, Germany
[51] Cui, P.; Liu, H.; Rajan, D.; Camp, J. A measurement study of white spaces across diverse population densities. . In Proceedings of), 2014 12th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), 12-16 May 2014, Hammamet, Tunisia
[52] Cui, P.; Chen, S.; Camp, J. GreenLoading: Using the Citizens Band Radio for Energy-Efficient Offloading of Shared Interests. In Proceedings of the 21st ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems.
[53] Song, J.; Feng, Q.; Wang, X.; Fu, H.; Jiang, W.; Chen, B. Spatial Association and Effect Evaluation of CO2 Emission in the Chengdu-Chongqing Urban Agglomeration: Quantitative Evidence from Social Network Analysis. Sustainability 2019, 1,1.
[54] Liu Z, Li H, Cao G. Quick estimation model for the concentration of indoor airborne culturable bacteria: an application of machine learning. International journal of environmental research and public health. 2017 Jul 30;14(8):857.
[55] Li H, Zhang Z, Liu Z. Application of artificial neural networks for catalysis: a review. Catalysts. 2017;7(10):306.
[56] Tong L, Chen Y, Wang Z, Qin Q. Diesel Engine Fault Diagnosis Based on D-S Evidence Theory. Science Technology and Engineering. 2010(15):3749-52.
[57] Li H, Yang S, Zhao W, Xu Z, Zhao S, Liu X. Prediction of the physicochemical properties of woody biomass using linear prediction and artificial neural networks. Energy Sources, Part A: Recovery, Utilization, and Environmental Effects. 2016 Jun 2;38(11):1569-73.