International Journal of Control and Automation
Vol. 8, No. 9 (2015), pp. 381-388
http://dx.doi.org/10.14257/ijca.2015.8.9.37

An Improved Prediction Model Based on Grey Clustering Analysis Method and its Application in Power Load Forecasting

Wang Ya

Institute of Information Engineering, Xuchang University
Xuchang, Henan Province, China
wangya_xcc@126.com

Abstract

Current grey clustering analysis methods have some defects. So, this paper proposes a prediction model based on improved grey clustering analysis. Firstly, it constructs the grey classical domain and the grey sector domain based on prediction subjects and data and according to relevant theory about grey clustering analysis. Secondly, it categorizes samples according to features of prediction subjects and confirms the analysis categories corresponding to the classical domain. Thirdly, based on the grey system theory, it constructs the grey correlation coefficient model and grey correlation degree model so as to obtain the weighed grey correlation degree. Thus, prediction subjects can be divided into proper category. Finally, power load forecasting in the power industry is taken as a case to prove that the model is reliable and has efficacy.

Keywords*: grey clustering analysis; grey prediction; classical domain; correlation degree; power load forecasting; model*

1. Introduction

Today's world is advancing towards a higher level of social productivity and a rapid development of science and technology. Prediction analysis, which depends on the analysis result to reduce uncertainties about the future, has received widely attention and been applied to many fields, in particular the engineering industry [1-3]. It gives instructions on decision and planning so as to avoid blindness. The prediction analysis on complex systems is defined as to use quantitative and qualitative methods to estimate, measure and analyze design features (either internal or external) or development state of the complex system according to data available [4-5]. At present, common methods of prediction analysis include qualitative analysis, mathematical models such as time-series model, and regression model and simulation methods such as dynamic demand system. These methods are tailored to different situations and have yielded fruitful results [6-9]. A new theory, the grey system theory, is worthy of noticing. It was first put forward by Professor Deng Julong. Grey system an emerging subject aimed at solving problems with many uncertainties based on mathematical theories. It is well applied to address samples with "clear extension but blur connotation" or situations where there are only a small scale of samples and a lack of enough information [10-15]. The prediction of a complex system usually involves grey information. Therefore, this paper proposes an improved prediction model based on grey system theory and proves the efficacy of the model through a case study.

2. Grey Sequence and Grey Correlation Coefficient

According to the grey system theory, sub-systems at different levels are subject to grey correlation analysis. The development state, correlation of indicators and dynamic change

ISSN: 2005-4297 IJCA

of the system are measured in a quantitative way. There are two key concepts, namely grey sequence and grey correlation coefficient.

2.1. Grey Sequence

The change of the state of the complex system has a time-series feature. n Correlation factors of the complex system consist of the sequence, named the comparative sequence V_i. Then m comparative sequences corresponding to m time-series state are:

$$\begin{aligned} V_1 &= \{v_1(1), v_1(2), \cdots, v_1(n)\} \\ V_2 &= \{v_2(1), v_2(2), \cdots, v_2(n)\} \\ &\cdots\cdots \\ V_m &= \{v_m(1), v_m(2), \cdots, v_m(n)\} \end{aligned} \tag{1}$$

In particular, the sequence that can reflect the behavior and state of the complex system is named reference sequence V_0. There is:

$$V_0 = \{v_0(1), v_0(2), \cdots, v_0(n)\} \tag{2}$$

2.2. Grey Correlation Coefficient

To show the correlation of the development state under different time-series, grey correlation coefficient ξ_i is used to describe how the correlation changes with the time and grey correlation degree r_i is used to describe how the correlation changes with the subjects.

Grey correlation coefficient ξ_{ij} can represent the correlation between the comparative sequence V_i and the reference sequence V_0 under different influence factors. There is:

$$\xi_{ij} = \frac{\min_i \min_j \left|\Delta\left(v_i(j) - v_0(j)\right)\right| + \beta \max_i \max_j \left|\Delta\left(v_i(j) - v_0(j)\right)\right|}{\left|\Delta\left(v_i(j) - v_0(j)\right)\right| + \beta \max_i \max_j \left|\Delta\left(v_i(j) - v_0(j)\right)\right|} \tag{3}$$

Where β refers to the resolution coefficient $\beta \in [0,1]$, usually $\beta = 0.5$. $\Delta\left(v_i(j) - v_0(j)\right)$ Refers to the gap of two values.

Grey correlation degree r_i refers to the closeness of two curves of the comparative sequence V_i and the reference sequence V_0. There is:

$$r_i = \frac{1}{n}\sum_{j=1}^{n} \xi_{ij} \tag{4}$$

3. An Improved Prediction Model Based on Grey Clustering Analysis

3.1 Grey Classical Domain and Grey Sector Domain

In the grey clustering analysis, the primary thing is to categorize different state sets P through clustering analysis. Feature indicators in each set have a certain range of value of quantity. Based on that, construct the grey classical domain and the grey sector domain. Suppose the classical domain V_j^i of the state set P_i about feature j is:

$$V_j^i = \left\langle v_j^i(a), v_j^i(b) \right\rangle, v_j^i(a) \le v_j^i(b) \tag{5}$$

Thus, the sector domain V_j^o of the state set P_i about indicator j is:

$$V_j^o = \left\langle v_j^o(a), v_j^o(b) \right\rangle = \left\langle min\left(v_j^i(a)\right), max\left(v_j^i(b)\right) \right\rangle \tag{6}$$

3.2. The Weight of Prediction Feature

There are many influence factors for the complex system. The importance of prediction features varies from one another. So it is necessary to allocate weight for these features. This paper adopts a comprehensive evaluation method to allocate weight. X Experts are invited to score on the importance of prediction features by a 1-9 scale. Based on the score, we can get the initial judgment matrix $\boldsymbol{B}$.

$$\boldsymbol{B} = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ b_{X1} & b_{X2} & \cdots & b_{Xn} \end{bmatrix} \tag{7}$$

So the weight w_j of predicator feature j is:

$$w_j = \sum_{i=1}^{X} b_{ij} / \sum_{j=1}^{n} \sum_{i=1}^{X} b_{ij} \tag{8}$$

3.3. Processing Prediction Features

Prediction features usually have different scales. To facilitate the prediction analysis, we need to nondimensionalize the classical domain of the prediction feature in order to unify the value scales.

For classical domain $U_j^i = \left\langle u_j^i(a), u_j^i(b) \right\rangle$ of prediction features of effective type, the unified value of quantity $V_j^i = \left\langle v_j^i(a), v_j^i(b) \right\rangle$ is:

$$V_j^i = \left\langle v_j^i(a), v_j^i(b) \right\rangle = \left\langle \frac{u_j^i(a) - \min\limits_{1\le i\le m}\left(u_j^i(a)\right)}{\max\limits_{1\le i\le m}\left(u_j^i(b)\right) - \min\limits_{1\le i\le m}\left(u_j^i(a)\right)}, \frac{u_j^i(b) - \min\limits_{1\le i\le m}\left(u_j^i(a)\right)}{\max\limits_{1\le i\le m}\left(u_j^i(b)\right) - \min\limits_{1\le i\le m}\left(u_j^i(a)\right)} \right\rangle \tag{9}$$

For classical domain $U_j^i = \left\langle u_j^i(a), u_j^i(b) \right\rangle$ of prediction features of effective type, the unified value of quantity $V_j^i = \left\langle v_j^i(a), v_j^i(b) \right\rangle$ is:

$$V_j^i = \left\langle v_j^i(a), v_j^i(b) \right\rangle = \left\langle \frac{\max\limits_{1\le i\le m}\left(u_j^i(b)\right) - u_j^i(a)}{\max\limits_{1\le i\le m}\left(u_j^i(b)\right) - \min\limits_{1\le i\le m}\left(u_j^i(a)\right)}, \frac{\max\limits_{1\le i\le m}\left(u_j^i(b)\right) - u_j^i(b)}{\max\limits_{1\le i\le m}\left(u_j^i(b)\right) - \min\limits_{1\le i\le m}\left(u_j^i(a)\right)} \right\rangle \tag{10}$$

After standardization, all values of quantity of classical domains fall between 0-1. In other word, there is $0 \le v_j^i(a) \le v_j^i(b) \le 1$, which is conductive to the grey clustering analysis of prediction feature of the complex system.

3.4. Grey Correlation Coefficient and Grey Correlation Degree

Suppose the sample sequence of the system to be tested is V_d:

$$V_d = \{v_d(1), v_d(2), \cdots, v_d(n)\} \tag{11}$$

According to abovementioned analysis, standardize the sample sequence V_d and obtain the grey correlation coefficients ξ_{ij}^d between sequence about prediction feature j and different classical domains $V_j^i = \left\langle v_j^i(a), v_j^i(b) \right\rangle$:

$$\xi_{ij}^d = \frac{\min_i \min_j d\left\langle v_d(j), V_j^i \right\rangle + \beta \max_i \max_j d\left\langle v_d(j), V_j^i \right\rangle}{d\left\langle v_d(j), V_j^i \right\rangle + \beta \max_i \max_j d\left\langle v_d(j), V_j^i \right\rangle} \tag{12}$$

Where $d\left\langle v_d(j), V_j^i \right\rangle$ refers to the distance between prediction feature j and different classical domains $V_j^i = \left\langle v_j^i(a), v_j^i(b) \right\rangle$:

$$d\left\langle v_d(j), V_j^i \right\rangle = \frac{\left|v_d(j) - v_j^i(a)\right| + \left|v_d(j) - v_j^i(b)\right|}{2} \tag{13}$$

After the grey correlation coefficients are obtained, we need to analyze the category of the system, namely the grey correlation degree r_i^d between sequence V_d about prediction feature j and different classical domains $V_j^i = \left\langle v_j^i(a), v_j^i(b) \right\rangle$:

$$r_i^d = \sum_{j=1}^{n} \left(w_j * \xi_{ij}^d \right) \tag{14}$$

Obtain the grey correlation degree r_i^d between the system and the category sets and select the maximum grey correlation degree r_s^d following the optimum principle. Its corresponding category is the state set in which the system falls:

$$r_s^d = \max\left(r_1^d, r_2^d, \cdots, r_m^d\right) \tag{15}$$

3.5. The Model and the Algorithm

According to abovementioned analysis, we first construct the grey classical domain and the grey sector domain by comparing prediction features on the basis of grey clustering analysis and we can get the grey correlation coefficient and grey correlation degree between the system and different categories. Then, according to the comprehensive grey correlation degree, we can figure out which state set the system belongs to and finally are able to get the prediction result.

To be more specific, the algorithm of the prediction model based on the improved grey clustering analysis is described as followings:

Step 1 Obtain time-series sequences on the basis of influence factors of the system

Step 2 Categorize the state according to prediction features of the system and confirm the state sets or the category set.

Step 3 Construct the grey classical domain and the grey sector domain of different category sets.

Step 4 Standardize prediction features of different types and scales and allocate weight to them.

Step 5 Obtain influence factors of the system and the grey correlation coefficients between influence factors and different category sets.

Step 6 Obtain the grey correlation degrees between the system and the category sets.

Step 7 Confirm to which category the system belongs based on grey correlation degree and get the prediction result of grey clustering analysis.

Step 8 Judge whether the result reaches the prediction accuracy. If not, detail the category and repeat Step 2-7.

4. Case Study of Power Loading Forecast

Power loading is crucial for regional economic development. It holds significance to predict power loading in different phase to ensure sufficient power supply as enough supply provides support for a rapid, reliable and stable development of regional economy. However, power loading is limited to many factors. An accurate prediction cannot be realized unless these factors are categorized. This paper proposes an improved prediction model for the complex system in order to realize effective power loading prediction. Through surveys and inspections, it is known that there is a relation between the power loading and the development state of the industry. Sequences of the latest 10 years are obtained after statistics analysis, as shown in Table 1.

Table 1. Power Loading and the Development of the Industry

Year	Influence factors-industrial value (100 million yuan)			Power loading / 100 million kWh
	Primary industry	Secondary industry	Tertiary industry	
2005	67.278	86.452	32.439	31.211
2006	80.767	100.382	41.926	38.635
2007	90.331	135.624	50.356	44.662
2008	95.467	158.687	58.735	47.934
2009	108.623	170.981	66.546	55.157
2010	121.938	195.234	74.325	67.433
2011	131.346	248.805	89.987	79.127
2012	148.639	298.556	123.768	99.223
2013	198.639	375.466	165.620	127.316

To better construct the grey classical domain and the grey sector domain, we use feature growth rate to describe characteristic quantity, as shown in Table 2.

Table 2. Power Loading and the Growth Rate of Development Value of the Industry

Year	Influence factors-industrial value(100 million yuan)			Power loading / 100 million kWh
	Primary industry	Secondary industry	Tertiary industry	
2006	0.2004	0.1611	0.2925	0.2379

2007	0.1184	0.3511	0.2011	0.1560
2008	0.0569	0.1701	0.1664	0.0733
2009	0.1378	0.0775	0.1330	0.1507
2010	0.1226	0.1418	0.1169	0.2226
2011	0.0771	0.2744	0.2107	0.1734
2012	0.1317	0.1999	0.3754	0.2540
2013	0.3364	0.2576	0.3381	0.2831

From Table 2, we can see the growth rate of power loading ranges between 0.070-0.300. If measured by such a growth rate $\eta(pow)$, there are four prediction categories of power loading, namely:

$$\begin{cases} LV_1 : 0.070 < \eta(pow) \le 0.100 \\ LV_2 : 0.100 < \eta(pow) \le 0.200 \\ LV_3 : 0.200 < \eta(pow) \le 0.250 \\ LV_4 : 0.250 < \eta(pow) \le 0.300 \end{cases} \quad (16)$$

Primary industry, secondary industry and tertiary industry serve as three prediction features. We can get the average prediction feature under each category. Construct the grey classical degree of each of the four categories, as is shown in Table 3.

Table 3. Grey Relational Coefficient of Indicators of Hardware Features

Category	Primary industry		Secondary industry		Tertiary industry	
	Average	Classical domain	Average		Average	Classical domain
LV_1	0.0569	0.027-0.087	0.1701	0.140-0.200	0.1664	0.136-0.196
LV_2	0.1111	0.081-0.141	0.2343	0.204-0.264	0.1816	0.152-0.212
LV_3	0.1615	0.132-0.192	0.1515	0.122-0.182	0.2047	0.175-0.237
LV_4	0.1317	0.101-0.162	0.1999	0.170-0.230	0.3754	0.345-0.405

Classical domains of the growth rate of the three industries under different categories are taken as the comparative sequence. The growth rate of the three industries is taken as the reference sequence. We can obtain the grey correlation coefficient of prediction features, as shown in Table 4.

Table 4. Grey Correlation Coefficient of the System to be Tested

Category	Grey correlation coefficient		
	Primary industry		Primary industry
LV_1	0.389	0.717	0.524
LV_2	0.447	1.000	0.551
LV_3	0.518	0.663	0.597
LV_4	0.473	0.826	0.921

With the weight of hardware features and software features taken into account, we can get the comprehensive weighed grey relational degree of each indicator, namely, $\sigma_1 = 0.786$, $\sigma_2 = 504$, $\sigma_3 = 503$. By comparing the three, we can judge that object I is the optimal one and conductive to later research.

Table 5. Grey Correlation Coefficient of the System to be Tested

Category	Grey correlation degree	Prediction feature	Weight
LV_1	0.554		
LV_2	0.694	Primary industry	0.35
LV_3	0.596	Secondary industry	0.40
LV_4	0.726	Tertiary industry	0.25

From Table 5, we can see the system belongs to category LV_4, with the range of growth rate at $0.250 < \eta(pow) \le 0.300$. The power loading falls into $\langle 124.028, 128.990 \rangle$. The actual power loading value is 127.316, as shown in Table 1. Its growth rate is 0.2831, as shown in Table 2. As $127.316 \in \langle 124.028, 128.990 \rangle$ and $0.2831 \in \langle 0.250, 0.300 \rangle$, which indicates that the actual value is in line with the prediction. Thus, the efficacy of the model gets proved.

5. Conclusion

This paper studies the prediction about complicated systems and proposes an improved prediction model based on grey clustering analysis. It tests the model with a case of power loading prediction and proves that the model can fulfill the purpose. In this model, grey classical domain and grey sector domain of prediction features are constructed according to grey data sequence. The grey correlation coefficient and grey correlation degree between prediction features and different categories of classical domain are computed and the weighed grey correlation degree is obtained. Thus, the prediction of grey clustering analysis is available. The model has clear physical definitions. It is accurate and reliable in calculation, and easy to achieve on the computer. This model is able to provide support for the predication of intelligent and complicated systems.

Acknowledgments

This paper is supported by Key State Project of the Research in Science Technology of the Education Department of Henan Province (Project Name: Research in Mechanism of Syncretism and Self-control of Heterogeneous Internet of Things. Project Number: 14A510018).

References

[1] L. Xiaobai, C. Xiuling and L. Rongling, “Forecasting method for aeroengine performance parameters”, Journal of Beijing University of Aeronautics and Astronautics, vol.34, no. 3, **(2008)**, pp.253-256.

[2] Y. Xin and F. Hua, “Prediction of coal and gas outburst based on data fusion and case-based reasoning”, Journal of Southeast University(Natural Science Edition), vol.41, **(2011)**, pp.59-63.

[3] L. Fang, “Application of Grey Model and Neural Network in Grain Production Prediction”, Computer Simulation, vol.29, no.4, **(2012)**, pp.225-228+267.

[4] L. Nihuan, H. Zhihong, M. Yingying and L. Wangyun, “Review of the short-term load forecasting methods of electric power system”, Power System Protection and Control, vol.39, no.1, **(2011)**, pp.147-152.
[5] S. Qiang and Y. Jiguang, “Review on fault prognostic methods based on uncertainty”, Control and Decision, vol.5, **(2014)**, pp.769-778.
[8] W. Li, “Water bloom prediction and factor analysis based on multidimensional time series analysis”, CIESC Jorunal, vol.64, no.12, **(2013)**, pp.4649-4655.
[10] G. Wei, “An Improved GM(1,1)Forecast Model Based on the Markov Theory”, Computer Engineering and Science, vol.33, no.2, **(2011)**, pp.159-163.
[11] T.-c. Wang, A.-j. Yang and S.-s. Zhong, “Multi-Attribute Extension Fuzzy Optimized Decision-Making Model of Scheme Design”, Tehnički vjesnik/Technical Gazette, vol.21, no.2, **(2014)**, pp.239-247.
[12] T.-F. Chuang and Y.-H. Chang, “Comparison of physical characteristics between Rana latouchtii and Rana adenopleura using grey system theory and Artificial Neural Network”, Ecological Engineering, vol.68, no.7, **(2014)**, pp.223-232.
[13] K.-C. Wang, “A hybrid Kansei engineering design expert system based on grey system theory and support vector regression”, Expert Systems with Applications, vol.38, no.7, **(2011)**, pp.8738-8750.
[14] J. J. Guo, J. Y. Wu and R. Z. Wang, “A new approach to energy consumption prediction of domestic heat pump water heater based on grey system theory”, Energy and Buildings, vol.43, no.6, **(2011)**, pp.1273-1279.
[15] S. H. Hashemi, A. Karimi and M. Tavana, “An integrated green supplier selection approach with analytic network process and improved Grey relational analysis”, International Journal of Production Economics, vol.159, no., pp.178-191.

