Research on Industry Prosperity Index Forecast based on Hidden Markov Model

Junlong Xu, Haiyang Li *, Yuyang Chen, Qi Chen

Shanghai Institute of Measurement and Testing Technology, Shanghai, China

* lihy@simt.com.cn

Abstract. The research is mainly based on the electricity consumption information of 800 key energy-consuming enterprises. By cleaning and sorting the data, high-frequency structured electricity consumption data is obtained, which is based on the production theory in economics. Using complex network models and Hidden Markov Algorithm constructs an industry prosperity index based on power operation data by industry, so as to objectively reflects the operation of various industries and plays a predictive and early warning role in economy.

Keywords: Complex Network Models; Hidden Markov Algorithm; Electricity Consumption Information; Power Operation Data; Economy.

1. Research Background

The energy industry is a pillar industry in the national economy. With the rapid development and popularization of computer and information technology, a large amount of meaningful data has been accumulated in the process of energy production, storage and use. If these data are used in conjunction with appropriate weather data, environmental data, and economic data, they can provide a favorable basis for analyzing user behavior, grasping the fluctuation rules of user behavior, obtaining urban energy usage, and improving environmental pollution. The energy measurement data of an enterprise can provide many aspects of information for the entire society and economic development.

Through the sorting, analysis and mining of high-frequency energy measurement data, and the use of machine learning methods, the correlation between energy consumption and climate, social and economic factors is obtained, and the industrial industry prosperity index prediction model is established for multi-dimensional analysis. These data on the one hand, it reflects the information of the power system, which helps to understand the energy supply and transmission pressure of this region; on the other hand, it also helps to understand the electricity consumption behavior of users, so that their production can be analyzed, and then it can be extended to environmental issues related to the national economy and people's livelihood. This is of great significance both in theory and in reality.

2. Research Status at Home and Abroad

The prosperity index is a quantitative indicator that reflects the operating conditions of various industries and is used to reflect changes in the economic prosperity of the industry. Accurately predicting the industry prosperity index is very important to the development of production activities and macroeconomic regulation. The internationally popular method of measuring economic prosperity is the Composite Index, which uses a country's industrial growth level as a reference and selects some macro-statistical data, which is divided into a leading indicator group, a consistent indicator group, and a lagging indicator group. The indicator group is used to construct an indicator system for economic prosperity analysis to analyze and predict the turning points of business cycle fluctuations and economic changes.

During the operation of the national economy, the development of the industry economy exhibits a certain cyclical phenomenon, from prosperity to depression, or from contraction to expansion. This is the result of the interaction of a series of economic factors and policy environments. While, describing the current development trends of various industries and grasping the trends of the

industrial economy are of great practical significance for promoting the stable and healthy development of the industrial economy. As an indicator of economic trends, on the one hand, GDP is vulnerable to price and other factors; on the other hand, the accounting system of GDP is complex, subjective and untimely. Therefore, GDP lacks objectivity and real-timeliness in reflecting economic conditions. As the “digital growth” of Chinese economy is largely affected by administrative-led investment stimulus, compared with the other two indicators, power consumption is less affected by administrative-led investment, and reflects the economic situation more objectively and truthfully. In addition, power consumption is also closely related to modern industrial production, which has advantages of Synchronization of production and use and Convenience and accuracy in measurement. Therefore, the construction of the industrial production prosperity index by using the big data of the enterprise power consumption has four aspects of significance for the operation of the national economy:

- (1) Accurately reflect the activity of Chinese industrial production and the operating rate of factories;
- (2) A barometer reflecting economic operation;
- (3) An important frame of reference for judging the quality of statistical indicators such as GDP;
- (4) Grasp the trend of the industrial economy and promote the stable and healthy development of the industrial economy;

In summary, it is a good guide and reference for formulating the economic policies of the government, and formulating development goals of the enterprises, and personal consumption and investment.

3. Model Building

The study uses the daily frozen electricity of more than 800 key energy-consuming enterprises and the corresponding sub-industry codes as data samples. The time span of the data samples is from 2017 to 2018, with a total of 726 days. At the same time, the data samples also includes observational data of daily frequency weather in the same area, including average temperature, rainfall, wind speed, air pressure, and humidity. Due to the strong correlation between the highest temperature and the lowest temperature, only the high temperature data were selected for analysis.

In order to take into account production factors such as industry structure and its upstream and downstream relationships, the complex network model is used for analysis, which is a network describing relationship between industry-related connections and evolution. For the dynamic industry structure, this study sets a certain length of window and scroll length, selects the data within the window length to construct the current complex network model, And by moving the data window backward by the corresponding scroll length to obtain the data in the new window to construct a new phase of the complex network model, repeat the process to obtain dynamic complex network results.

Each node in the network represents one industry, and the size of the node is expressed by the growth rate of the average daily frozen electricity of the industry. For prosperity index of each industry, it must be a stable time data series without working day cycles or seasonal cycles. According to the requirement of time data series should to be weakly stable, the data series must have a constant mean value, and the autocorrelation coefficient of any order tends to zero. That is, the data series gradually tends to a white noise data series. The unit square root test method is used to test the stationarity of the data series.

$$(1 - \rho L)x_t = \varepsilon_t \quad (1)$$

$$E[\varepsilon] = 0, V(\varepsilon) = \sigma < \infty, Cov(\varepsilon, \varepsilon) = \mu < \infty \quad (2)$$

Where L is the lag operator and ε_t is a white noise data series. If $|\rho| < 1$, the time data series is a stationary series data. This method is used to test the stationarity of time series data of electricity consumption growth rates in all industries.

Pearson's correlation coefficient can be used to measure the degree of influence between various industries. The formula is as follows:

$$r_{xy} = \frac{\sum_{t=1}^T (x_t - \bar{x})(y_t - \bar{y})}{\sqrt{\sum_{t=1}^T (x_t - \bar{x})^2} \sqrt{\sum_{t=1}^T (y_t - \bar{y})^2}} \quad (3)$$

Among them, x_t and y_t are the time series data of the two industries, and T is the data length. If the degree of similarity r_{xy} is non-zero at the 95% confidence level, it is recorded as the weight (edge) of the connection between the two nodes. Even if there is no correlation between industries, the correlation coefficient should be a value with a small absolute value and not zero. Therefore, the correlation coefficient must be greater than the set threshold to be considered as an inter-industry correlation, which can exclude those Industries that are weakly or even unrelated. Through the pairwise correlation analysis of all industries, the degree of interconnection among all industries can be obtained.

After determining the correlation between the industries, it is also necessary to determine whether the driving relationship between the industries exists and the time required to affect the transmission. The Minimization Akaike Information Amount Criterion (AIC) can be used to determine the number of periods in which the interaction between industries occurs, that is, if the leading industry's business conditions change, how long will the change be transmitted to the affected industry? Its formula as follows:

$$AIC = \ln \hat{\sigma}_k^2 + \frac{n+2k}{n} \quad (4)$$

Among them, k is the number of regression variables, n is the size of the sample, $\hat{\sigma}_k^2 = SSE_k/n$, SSE is the residual sum of squares of the regression.

At the same time, it is based on the autoregressive Granger causality test model to test whether the driving relationship exists, that is, whether the driving relationship between the two industries is statistically related, and the expression is:

$$dy_t = \alpha_0 + \sum_{i=1}^p \alpha_i dy_{t-i} + \mu_t \quad (5)$$

$$dy_t = \alpha_0 + \sum_{i=1}^p \alpha_i dy_{t-i} + \sum_{j=1}^q \beta_j dx_{t-j} + \varepsilon_t \quad (6)$$

Among them, the residual μ_t, ε_t are normal distributions with an average value of 0 and a constant variance, p and q are the order of lag, y_t represents the target industry, and x_t represents the leading industry, where q is derived from the Minimization Akaike Information Amount Criterion. The alternative hypothesis is that the coefficient β_j is not all zero. If the hypothesis is true, then x_t is the reason for the change in y_t , which is determined by the F test. This study conducted the judgment of the number of lagging periods between all industries and the Granger causality test, and determined the driving relationship, the number of transmission periods, and the degree of correlation between all industries.

4. Using Hidden Markov Model to Predict Industry Prosperity Index

After the screening of the complex network, we can obtain some independent industries that are related to the electricity consumption of the central industry, Combine the meteorological data and use the hidden Markov model to estimate the electricity consumption growth of the core industry based on the development of the industry.

Excluding the influence of external factors, the electricity consumption growth rate is defined as state 1 when the adjusted electricity consumption growth rate is negative, and state 2 is defined when the adjusted electricity consumption growth rate is positive. Obviously, the original electricity consumption growth rate is divided into a high growth rate part and a low growth rate part. The state of high growth rate is the prosperity state, the production intensity value is higher than 100, and the low growth rate state is the recession state, and the production intensity value is lower than 100.

Figure 1. Original electricity consumption growth and state division

The adjusted growth rate of electricity consumption can reflect the current prosperity of this industry. In the state transition model, state 2 corresponds to the situation where the growth rate of electricity consumption is positive. To assign a value for the electricity consumption, a state greater than 0 is defined as a prosperous state, and a value greater than 100 is assigned to the prosperous index, and a state less than 0 is defined as a depressed state, and a value less than 100 is assigned to the prosperous index. The industry prosperity index from January 2017 to December 2018 is shown in the figure below:

Figure 2. Daily Production Prosperity Index and Original Electricity Consumption

By analyzing the daily production prosperity index, it can be seen that the prosperity and unprosperity state states are interchangeable. The prosperity state is dominant for a certain period of time, and the unprosperity state will be switched in the next time period. In addition, it can be seen that the industry prosperity index obtained through the model and algorithm is different from the original electricity consumption. This is because we have controlled production and non-production factors, and whether an industry's production is prosperous is only related to production factors, so non-production factors are eliminated. Since the prosperity is the state of economic operation within a certain time range, we can use month and quarter as the time periods to add up the daily prosperity value.

5. Analysis of the Validity of the Prosperity Index

Figure 3. Prosperity Index of the Automobile Manufacturing Industry

The automobile manufacturing industry is a capital-intensive and technology-intensive industry. Its upstream docks with parts, steel, rubber raw materials enterprises and production equipment manufacturing enterprises, etc. The midstream is automobile manufacturing, and the source of profit is automobile sales, mainly including auto shows and 4S Direct sales models such as stores, auto trading markets, channel distribution and other distribution models, supplemented by e-commerce platforms. The downstream of automobile companies meets the demand side, mainly including the demand for mining, road transportation, special purpose vehicles, etc., The demand for vehicles from enterprises and institutions or individual families, and the extended demand for automobile maintenance and services.

Therefore, the industry prosperity index will provide standardized measurement indicators for the production status of various industries, so that government can grasp the production dynamics of various industries in a timely manner, so as to formulate corresponding policies and respond in a timely manner. Play a forecasting and early warning role in the economic operation. It has real social benefits.

References

- [1] Schumacher C , Forecasting German GDP using alternative factor models based on large datasets. *J Forecast* 26(4) (2007):271–302.
- [2] M. Bonaiuto, F. Fornara, S. Ariccio, U. G. Cancellieri, L. Rahimi, Perceived Residential Environment Quality Indicators (PREQIs) relevance for UN-HABITAT City Prosperity Index (CPI), *Habitat International* 45 (2015) 53-63.
- [3] Ba'n'bura M, Modugno M, Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data. *J Appl Econom* 29(1) (2014):133–160.
- [4] G. Sands, New economy jobs and economic health, prosperity in Canada's midsize urban areas, *Habitat International* 45 (2015) 15-19.
- [5] C. Wong, A framework for 'City Prosperity Index': Linking indicators, analysis and policy, *Habitat International* 45 (2015) 3-9.
- [6] K. Mori, T. Yamashita, Methodological framework of sustainability assessment in City Sustainability Index (CSI): A concept of constraint and maximization Indicators, *Habitat International* 45 (2015) 10-14.
- [7] C. Scotti, Surprise and uncertainty indexes: real time aggregation of real-activity macro-surprises. *Journal of Monetary Economics*, 82 (2016) 1-19.