Constructing the Industrial Prosperity Index Based on Big Data of Enterprise Electricity

Jun Wang¹, Yingjie Tian², Fang Jia¹ and Shuaishuai Zhang³*

¹ State Grid Shanghai Electric Power Company, Shanghai 200120, China.
² State Grid Shanghai Electric Power Research Institute, Shanghai 200437, China.
³ School of Economics, Fudan University, Shanghai 200433, China.
* Corresponding author’s e-mail: zhangss16@fudan.edu.cn.

Abstract. This paper is mainly based on the big electric power data of 5900 industrial enterprises above designated size in Shanghai. By combining a complex network and a hidden Markov model, the prosperity index of 164 medium-sized industries in Shanghai is constructed. Specifically, we use complex networks to describe the correlation between different industries, in order to mine the upstream and downstream drivers that affect industrial power consumption, and on this basis, consider the external factors that affect power consumption to establish a hidden Markov model that predicts changes in power consumption in the industry. Further, we use the state probability output by the Hidden Markov Model to define the industrial prosperity index, hoping that the index can fully reflect the economic operation of various industries in Shanghai and become a "barometer" and "wind vane" for economic development.

1. Introduction

Electric power represents an important energy source for industrial enterprises, which is regarded as one of the leading indicators to forecast economic operation. Compared with other leading indicators, electric power data is acquired more timely and accurately. As national economic accounting is vulnerable to human interference, the objectivity of electric power data becomes particularly important. Currently, Shanghai has completely adopted smart meters, making it easier and faster access to user information. Power consumption by both businesses and households can be acquired timely, greatly facilitating the value of electric power big data. In terms of the research on the prediction of economic operation through electric power data, one type of research is dedicated to the combination of power consumption with other economic variables. This method focuses on the analysis of macroeconomic trends through factor analysis, principal component analysis, neural network model, composite index, etc. (Chen et al, 2003; Tian, 2015; Wang, 2017; Chou et al, 2018). The other type of research advocates for processing the power consumption data through the seasonal adjustment model to assess the economic target from the perspective of electricity. (Liu et al, 2019). Besides, there are also some scholars dedicating to the study on the correlation between electric power consumption and economics. (Lin et al, 2003; Lin et al 2010.) However, the frequency of data acquisition for most of the studies is limited to monthly or annual records, which fails to take advantage of the real-time electric power data. In particular, during the outbreak of COVID-19 in 2020, daily electric power data is used to keep track of the fluctuations in power consumption by businesses. This move provides decision-makers with faster access to the impact of the epidemic on the social and economic operation. (Such as Zhejiang “Electric Power Index for Enterprise Resumption” launched by Zhejiang Province in February 2020). There are
quite many precedencies for the use of power consumption to predict economic operation. However, as most of the predictions are based on unadjusted power consumption data, it causes discrepancies in the monitoring of economic operation. For instance, such fluctuations as the reduction of power consumption on holidays and the increase of such figures due to an increase in temperature do not necessarily symbolize any trend of economic operation.

To address these issues, this paper sampled data from 5900 industrial enterprises above designated size in Shanghai, combining a complex network that includes the upstream and downstream drivers, and the factors of climate and holiday, the prosperity index of 164 medium-sized industries in Shanghai is constructed by adopting a hidden Markov model. Comparison is also made with the growth rate of the output value to reflect the effectiveness of the prosperity index to a certain extent.

2. Data
The name list of industrial enterprises above designated scale is based on the statistics released by the Shanghai Statistics Bureau in early 2018. The real-time power consumption of 5917 such industrial enterprises is obtained while the data regarding the other 2200 of such enterprises are failed to acquire due to various reasons. The number of enterprises above the industry scale and the statistics of enterprises with real-time power consumption that can be actually matched is shown in Figure 1. The period for the power consumption data dates from 1st August 2015 to 30th January 2019. The data losses or abnormal data during the transmission of intelligence meters are rectified through the certain algorithm.

Besides, industrial power consumption is not only relevant to the internal factors of the industry but also related to industrial structure, development of the upstream and downstream industries, and the external climate. Therefore, such factors shall be well controlled in the analysis of the industrial power consumption to define the industry prosperity. To describe the impact of the upstream and downstream industries, we establish a complicated network to acquire the changes of the downstream and upstream industries. As for the external climate, we control such factors as the average temperature, rainfall, wind speed, atmosphere, and humidity. Besides, as working days and holidays have strong implications for industrial production, we set up two virtual variables, weekends and legal holidays, to control the influence of these factors on power consumption.

Figure 1. Number of enterprises above scale and the total of samples in different industries.
3. The construction of Models

3.1. Complex Network Model of Inter-industry

High-frequency industrial power consumption data can be used to mine the correlations between the upstream and downstream of the industry chain. With the help of complex network algorithm, the upstream and downstream relationships between industries are depicted, so as to mine the upstream and downstream industrial chains between industries and master the intricate industrial structure in real time. Related method has been widely applied in research (Yao et al, 2016; Zhou et al, 2019). In the complex network models, the three major components of the directional complex network models include nodes, edges and directions in the network. Where each industry is a node, the connections between industries are determined by industrial power consumption subject to Granger causality test (Granger, 1969) and Pearson correlation coefficient, while lag periods affected by upstream and downstream industries are determined on the basis the AIC. We conducted lag period judgment and Granger causality test between the power consumption data of all industries, and determined the driving relationship, transmission time and correlation degree between any two industries.

For a dynamic industrial structure, we set a certain length of window and scroll length. The current-period inter-industry complex network can be established by selecting data within the window length. The next-period inter-industry complex network can be built by moving the data window backwards at the corresponding scroll length to obtain new window data. Such processes can be repeated to establish a dynamic complex network changing with time. We employ the QAP method to test relations between the complex network and actual industrial input and output network. The result shows that the correlations between two networks are stable, signifying the realistic significance of the inter-industry complex network. In addition, as correlations exist among all the industries, we adopt MST and PMFG (Prim et al, 1957; Aste et al, 2005) to trim the original network so that we can remove the redundant information from the network and keep the most important information there. The last issue of complex network base on MST and PMFG is shown in Figure 2.

Figure 2. The MST and PMFG network of inter-industry.

3.2. Hidden Markov Model

Electric power represents one of the most important production factors for an enterprise. Its changes over a short period of time signifies for expansion or contraction of production over a certain period. We categorize the factors that affect an enterprise’s fluctuation in power consumption into three types. Such factors include those relevant with the industry itself, those closely related to the fluctuation in power consumption by the upstream and downstream chain and the external variables such as weather, holiday, etc. By combining the complex network, controlling the impact of the industrial structure and integrating the external factors that may affect changes in power consumption, we establish the hidden Markov Model to predict changes in industrial power consumption. Assuming that the industry of analysis is linearly affected by other factors, we establish hidden Markov model under this condition, where the state $S = i$ function response is:
\[ dy_t(S = i) = c_i + \sum_{j=1}^{n} \alpha_{ij}(S = i)x_j + \sum_{k=1}^{n} \beta_{ik}(S = i) \times Out_{k,t} + \epsilon_t(i) \quad (1) \]

Where, \( dy_t \) stands for growth rate of power consumption by the central industry, \( x_j \) stands for the variables of leading industries, including \( N_{rela} \) relevant industry number, \( \overline{dy}_{rela} \) weighted average of power consumption growth rate, \( \overline{ind}_{rela} \) in-degree intensity weighted average and \( \overline{out}_{rela} \) out-degree intensity weighted average. Where the weight of the relevant industry is the correlation coefficient of the target industry, the number of transmission periods from the behaviour of the leading industry to the central industry is considered in the weighted electricity consumption. In case of any changes in the relations between the upstream and downstream in the industry structure, the correlation coefficient provided by the complex network and relevant industries will also change, so the dynamics of the industrial structure will be reflected in this index. \( Out_{k,t} \) represents all the external factors, including temperature, temperature squared, rainfall, wind speed, atmosphere, humidity, weekends, legal holidays, etc.. \( n \) stands for total of external factors. \( c_i, \alpha_{ij}, \beta_{ij} \) refers to the parameters to estimate.

The state of hidden Markov Model cannot be observed directly, but only through the vector sequence. Each probability vector represents a state by certain probability density distributions. We assume \( \epsilon_t(i) \) is a normal distribution with mean value of 0 and variance of \( \sigma_i^2 \) to demonstrate its conditional probability distribution, equations:

\[ p \left( dy_t(S = i) - dy_t(i), Out_{k,t}^{(0)} \right) \sim N(0, \sigma_i^2) \quad (2) \]

The foundation of prediction by hidden Markov Model is based on the belief that the current value is the superposition of each state value. Therefore, the transition probability between states needs to be determined before prediction. If the total number of states is \( N_s \), then the transition matrix is:

\[ \Gamma = \left( y_{ij} \right), \sum_{j=1}^{N_s} y_{ij} = 1, \quad 1 \leq i, j \leq N_s \quad (3) \]

Where, \( y_{ij} = p(S_t = i, S_{t+1} = j) \) is the probability of the system transition from one state \( i \) to another state \( j \). According to hidden Markov Model theory, it is assumed that future states depend only on the current state, not on the events that occurred before. If we know all the state probabilities and transition matrix at starting state, we can work out each state probability. The variation coefficients, transition matrix and state probabilities in hidden Markov Model can be worked through EM method.

3.3. Definition of Industrial Prosperity Index

The varying states of hidden Markov Model describe the production states of the industry. We assume that there are two states, “Prosperity” and “Recession” under hidden Markov Model. Then, the current fluctuations in the industrial power consumption is the superposition of the two states “Prosperity” and “Recession”. When the extent of Prosperity in the current industrial production is high, the probability for the Prosperity state in the hidden Markov Model is higher and vice versa. Therefore, the probability of Prosperity is consistent with the extent of Prosperity of such industry. By means of Hidden Markov Model output state probability, we define the prosperity index for the industrial production intensity and set the standard range from 0 to 200.

Given the analysis is modelled on medium-scale industries, we designate average power consumption of medium scale industries as the weight. With average weight of the medium-scale industries covered by the large-scale industries, we obtain the prosperity index of the large-scale industries. The equations for large scale industrial prosperity index:

\[ PIT = \sum_{i=1}^{n} Ele_t * \frac{PIT_i}{Ele} \quad (4) \]

Where, \( n \) is the total of medium-scale industries covered by the large scale industries; \( Ele \) is the average power consumption by the large-scale industry during the sampling period; \( Ele_t \) is the average power consumption in No i medium-scale industry during the sampling period, \( PIT_i \) is the Prosperity in the medium-scale industry.

4. Results and Discussion

The calculation results of industrial prosperity index is analysed by taking the mining, metallurgy and construction equipment manufacturing industry (industry code: 3510) as an example. There are two
states, “Prosperity” and “Recession”, the state at any period of industry is determined by the superposition of the two states. The value predicted by hidden Markov Model and the real value of fluctuation in industrial power consumption is shown in Figure 3 (it only shows fluctuation of the last two months). It shows that the model can be used to successfully forecast fluctuation in industrial power consumption. In particular, we can filter the states “Prosperity and “Recession” through hidden Markov Model prediction model. Generally speaking, when the growth rate of power consumption after adjustments by various factors is bigger than 0, the economy is state “Prosperity” while if it is smaller than 0, it is state “Recession”. As the probability of state “Prosperity” in the model represents the extent of prosperity of the current industry, we can define the industrial property index with the probability of Prosperity. Figure 3 finally shows the probability distribution of Prosperity and Recession and their corresponding industrial power consumption fluctuation rate after adjustment. It can be discerned that the Prosperity and Recession probability for the industry changes over time and the adjusted growth rate of industrial power consumption corresponding to Prosperity is greater than that of Recession.

Figure 3. Prediction Results of Hidden Markov Model and Probability Distribution at Different States

Because the probability of Prosperity and Recession changes over time, Prosperity may dominate at one time step but transit to Recession at the other time step. The industrial prosperity index at given time is weighted by average to formulate monthly prosperity index and quarterly prosperity index. Figure 4 is the calculation results on the monthly and quarterly prosperity index for the industries of mining, metallurgy and construction special equipment manufacturing.

Figure 4. Monthly and quarterly prosperity index of mining, metallurgy and construction special equipment manufacturing.
We compare the industrial prosperity index for large-scale industries worked out by means of the equation (4) with the growth rate of output value for such enterprises released monthly by The Shanghai Statistics Bureau to verify the effectiveness of this method. The large scale industries cover not only the automotive industry, computer, communications and electronic equipment manufacturing industry with the largest industrial output value in Shanghai, but also Pharmaceutical manufacturing and special equipment manufacturing. See Figure 5 for the comparison results. Judged overall, the fluctuations may have slight difference at a specific time, but the overall tendency is consistent. However, if judged from perspective of statistics scope and internal economic theories, the prosperity index based on prediction of production by power consumption and growth rate of industrial output for industrial development based on values are not theoretically tantamount. The industrial output accounting does not only rely on accounting system but also is subject to several factors such as price adjustment, value transition, and extent of collaboration. In contrast, the prosperity index based on power consumption data is more timely and truthful.

![Figure 5](image_url)

Figure 5. Comparison between the prosperity index and growth rate of output value for large-scale industries.

5. Conclusion
This paper is mainly based on the big electric power data of industrial enterprises above designated size in Shanghai. It adopts a complex network to explore the correlations between the upstream and downstream industry chain. On the other hand, it also introduces the external factors to establish a hidden Markov model for prediction. By combining the two methods, the prosperity index for 164 medium-scale industries in Shanghai are constructed base on the probabilities of states under hidden Markov Model. Compared with growth rate of the output value released by the Shanghai Statistics Bureau, the
two statistics are highly consistent in trend, which therefore proves the effectiveness of industrial prosperity index for predicting economic operation. Compared with the growth rate of output value calculated by accounting system, the industrial prosperity index construction through this method is more timely and reliable without being limited by updating frequency. It therefore will play a significant role in helping the government improve refined management, strengthen economic monitoring and accurately grasp economic development trends.

Acknowledgement
This research is sponsored by the Science and Technology Project (B30900180068) of State Grid Shanghai Municipal Electric Power Company, 2018 Shanghai Artificial Intelligence innovation and development project (2018-RGZN-02056) of Shanghai Economic and Information Technology Committee and 2017 Shanghai Informational Development Project (201701048) of Shanghai Economic and Information Technology Committee.

References
[1] Chen, D. H., Li, H.Z., Yang, X.Y. (2003) Method Selection of Constructing Industrial Cycle Index and Its Demonstration Analysis. Systems Engineering, 23(04):72-76.
[2] Tian, C.B., Yin, Y., Jin, P., Wang, Y.L. (2015) Analysis and prediction of urban macroeconomic development based on power data, Electrical applications, 34:232-233.
[3] Wang, W.W. (2017) Correlation between data of enterprise electricity consumption information and macroeconomic forecast. In: Annual meeting of power industry informatization, pp. 276-279.
[4] Qiu, H.J., Li, B.S., Lin, H.D. (2018) Analysis of Industry Electricity Market Prosperity Index Using Electric Power Big Data Mining. Power information and communication technology, 16(11):24-28.
[5] Liu, Y.J., Song, K.H., Wang, X. (2020) Analysis of economic prosperity index based on big data of electricity. Telecom science, 36(06):166-171.
[6] Lin, B.Q. (2003) Structural Changes, Efficiency Improvement and Electricity Demand Forecasting. Economic Research Journal, 2003(05):57-65.
[7] Lin, W.B., Sun, J., Shi, Q.F. (2010) Economic growth, energy intensity and electricity Consumption-An analysis of the reasons for the deviation between electricity consumption and GDP growth rate. Economic Science, 2010(05):15-22.
[8] Yao, C. Z., Lin, J.N., Lin, Q.W. et.al. (2016) A study of causality structure and dynamics in industrial electricity consumption based on Granger network, Physica A. 462:297–320.
[9] Zhou, Y., Zhang, S.S., Wu, L.B., Tian, Y.J. (2019) Predicting sectoral electricity consumption based on complex network analysis. Applied Energy, 255:113790.
[10] Granger, C.W.J. (1969) Investigating Causal Relations by Econometric Models and Cross-spectral Methods. Econometrica, 37:424–438.
[11] Prim, R.C. (1957) Shortest connection networks and some generalizations. Bell Syst Tech J, 36:1389–1401.
[12] Aste, T., Di Matteo, T., Hyde, S.T. (2005) Complex networks on hyperbolic surfaces. Phys Stat Mech Its Appl, 346:20–26.
[13] Tumminello, M., Aste, T., Di Matteo, T., Mantegna, R.N. (2005) A tool for filtering information in complex systems. Proc Natl Acad Sci, 102:10421–10426.