Industry “Angiography”: an Industry Map System Based on Electricity Marketing Data

Qiuying Shen[1*], Wentao Zhang[2]
1405302529@qq.com[1],13951830329@139.com[2]

State Grid Suzhou Municipal Electric Power Company, Suzhou 210005, China

Abstract. Based on electricity marketing data, and integrating GIS, commerce data, this paper constructs an industry map to dynamically show and monitor the development status and trend of specific industries in specific regions under different granularity. In the dimension of electric power marketing data, through the statistical processing of basic electricity consumption information, electricity consumption level, electricity consumption fluctuation and electricity consumption trend of enterprises, the real operation status and trend of enterprises are presented; in the dimension of industry, the industrial label and industrial classification of enterprises are formed through the word division and clustering algorithm; in the dimension of geography, the geographical location of enterprises is marked with the help of GIS. Finally, with industry and GIS as the two-dimensional coordinates, we present the development status and trend of industries under different granularity through enterprises' electricity consumption data, and provide systematic support for the government's industrial development decisions and other scenarios.

Keywords: electricity marketing; big data; industry map

1 Introduction

With the integration of digital revolution and energy revolution, electric power data has become an indispensable production factor for the development of digital economy. 2020, the State Grid released the "Digital New Infrastructure" ten key construction tasks, aiming at tapping the value of electric power data through the construction of digital platform, energy data center and electric power data application. Through the digitalization platform of power grid, the construction of big data center and the application of big data of power, it aims to explore the value of big data of power, serve the government, enterprises and society, and help the economic and social development.

Electricity data has the characteristics of wide coverage, high real-time, high accuracy and high structured degree. The data on the marketing side of electricity, while possessing the above characteristics, concentrates on reflecting the level of electricity consumption of enterprises and its changing trends. The integration of this data with GIS, industrial and commercial data and other data sources will be able to show and monitor the development status of specific regional industries and their trends in different granularity in real time, providing scientific basis for industrial decision making of the government, relevant industrial organizations and commercial decisions of enterprises.

MSIEID 2022, December 09-11, Chongqing, People's Republic of China
Copyright © 2023 EAI
DOI 10.4108/eai.9-12-2022.2327573

2 Literature review

Electricity consumption, especially industrial electricity consumption, can partially reflect macroeconomic conditions, and the Keqiang Index is the most typical example. The Keqiang Index has also inspired many studies on the relationship between electricity consumption and macroeconomy, such as gross domestic product (GDP) forecasting based on electricity data, and electricity load forecasting considering economic factors [1-2]. The problem of predicting GDP from electricity data has been studied mainly with the help of classical statistical methods and artificial intelligence methods. The literature [3] uses gray correlation analysis to obtain electricity indicators with high correlation and then uses Bayesian network theory for GDP prediction. The literature [4] combined long short-term memory (LSTM) with differential autoregressive integrated moving average (ARIMA) model to achieve the present prediction of industrial value added by collecting real-time electricity data. In addition, some other studies focus on the relationship between industry economic indicators and industry electricity consumption at the meso level, for example, the literature [5] analyzed the development situation of ferrous metal industry and the trend of industry electricity demand in Anhui Province. Looking at the existing studies, the application of electricity data is still limited to the prediction and analysis of macro or meso economic indicators such as GDP and industries, and the level of analysis does not go deeper into the micro level such as the spatial characteristics of industries.

The spatial agglomeration of industries is the most prominent geographical feature of economic activities, and is a cross-cutting issue in the research fields of economic geography and industrial economics. At the same time, the spatial agglomeration degree of industry is also an important index to measure the competitiveness of regional industry, and has an important reference value to the government for industrial planning.

In CNKI, we found 60 journal papers and 4 dissertations with "industrial map" as the keyword. Only a few of the journal papers are serious academic papers, and the limited academic research on industrial maps is mainly focused on two directions: first, the design of industrial maps, mainly discussing the design principles, components and technical implementation paths and technical details of industrial maps [6-8]; second, using industrial maps as an analytical tool to analyze the development of industries [9-10].

From the existing literature, there are few studies related to industrial maps, which are still at a very preliminary stage, and the studies related to the construction of industrial maps from the perspective of power data have failed to be retrieved.

Based on electricity marketing data, this paper integrates GIS, industry and commerce data to build an industry map, which creates a angiography for industry and dynamically shows and monitors the development status and trends of specific industries in specific regions at different granularities.

3 Construction of industry labeling system in industry map

One of the most basic tasks in the industrial map is to label enterprises with industrial labels, which is based on a systematic industrial labeling system. At present, there are three major types of industry labeling systems: one is the National Economic Classification, which is widely used in the statistical system, including "sectors, major categories, medium categories and small categories", such as "B0711", in which "B" for mining, "07" for oil and gas extraction, "071" for oil extraction, "0711 " on behalf of onshore oil extraction. The labeling system is authoritative and has a wide range of applications, but the granularity is large; the second is the industry classification developed by various brokerage firms and index companies to better distinguish listed companies from the business level, such as the Shenyin Wanguo industry classification, which is based on the four classification principles of "earnings-driven, valuation clustering, physical form, and usage habits" and has constructed The latest version in 2021 includes 31 primary industries, 134 secondary industries and 346 tertiary industries. The above type of classification system is mainly applied to investment scenarios, which has the advantage of being tightly integrated with enterprise business and updated to accommodate new industries; thirdly, the industry classification created by local governments, mainly some emerging industries, which lacks a unified standard and has strong local characteristics.

The industry map system we designed is compatible with all the above three types of labeling systems. Firstly, there is a national economic industry classification code for each electricity-using enterprise in the electricity marketing data; secondly, we use the business scope in the industrial and commercial data as the text and Shenyin Wanguo's industry classification as the industry classification dictionary, and use the word separation technology to extract several industry labels for specific enterprises from the business scope of the industrial and commercial data, thus supporting the formation of associations between enterprises and Shenyin Wanguo's industry classification; finally, our system supports Manually create an industry labeling system and algorithm rules to tag electricity-using enterprises with specially designed industry labels.

We establish enterprise industry labels with the Shenyin Wanguo industry classification system as the reference system based on the following steps: Step 1: Based on the business scope text in the enterprise's industrial and commercial data, the business keywords of the enterprise are extracted by a neural network word separation algorithm using the Shenyin Wanguo three-level industry classification as the lexicon; Step 2: All business keywords of an enterprise constitute a vector space composed of dummy variables, and calculate the distance D between this vector The distance D between this vector space and the vector space composed of business keywords of each specific classification in the three-level industry classification of Shenyin Wanguo (calculated by the column table shown in Figure 1), D(X,Y)=(b+c)/(a+b+c). Where X represents a certain firm, Y represents an industry in Shenyin Wanguo, and a,b,c,d, represents the segmented business labels. the smaller the value of D, the closer the industry classification of the firm is to this specific Shenyin Wanguo industry.

		Y		
		1	0	sum
X	1	a	b	a+b
	0	c	d	c+d
	sum	a+c	b+d	p

Fig.1. Example of a column linkage for calculating business distance

4 Construction of enterprise power consumption labels in industry maps

The data on the power marketing side is mainly concentrated in the Power Marketing 2.0 system, which links enterprises and their metering points through the chain of "customer → contract account → contract → customer → metering point (meter)". Through the detailed research of Power Marketing 2.0 system, we sorted out and summarized the indicators of the system that are closely related to the enterprise's electricity consumption as shown in Table 1.

Table 1. Electricity consumption indicators for enterprises

Major categories of indicators	Subdivision indicators
Company Basic Information	Enterprise unified social credit code
	Household age
	Number of enterprise electricity accounts
	Industry Category
Electricity account information	User status (normal user, canceled account, etc.)
	Type of electricity consumption (high-voltage dedicated line, low-voltage)
	Payment method (fee-controlled, post-payment)
	Contract Capacity
	Number of aggregated measurement points
Electricity consumption and payment level of enterprises	Electricity consumption of enterprises (according to the summary of each metering point)
	Actual amount paid by enterprises for electricity consumption
	Electricity bill receivable from enterprises
Enterprise electricity consumption behavior	Electricity theft (monthly cycle statistics on the number of electricity theft and electricity theft charges)
	Violation of electricity consumption (monthly cycle statistics of the number of violations of electricity consumption and electricity charges for non-compliance)
Enterprise electricity consumption capacity	Yearly cycle of statistics for enterprises to apply for capacity changes

Based on the data and information shown in Table 1, combined with the industry label of each electricity-using enterprise and the geographic label provided by GIS, the electricity consumption of specific industries in a designated area can be counted in a two-dimensional coordinate (industry, geographic location) with different granularity, including electricity consumption trends and electricity consumption risks, so as to dynamically reflect, monitor and analyze the current situation of industrial development in the area. For example, the trend of industry can be characterized based on information such as capacity increase and electricity consumption growth, and the risk of industry development can be revealed by data such as electricity bill payment and electricity consumption behavior of enterprises.

In terms of methodology, an example of such a system dealing with industry-specific electricity consumption trends and characteristics in a given region is provided herein.

Example of the method: Based on the aggregated monthly actual electricity consumption time series of specific industries in the specified region, the U=U_Q×U_X×U_S model characterizing and predicting the monthly electricity consumption of specific industries in the specified region is constructed, which is expressed in the following form (Formula 1).

$$U = U\_Q \times U\_X \times U\_S \quad (1)$$

In the above formula:

U is the aggregated monthly electricity consumption series.

U_Q, U_X and U_S are the trend component, seasonal cycle component series and stochastic component series of U, respectively.

where the trend component is expressed in the form shown in Equation 2.

$$\omega_t = \begin{cases} (1-B)^d U\_Q_t \\ \emptyset_1 \omega_{t-1} + \emptyset_2 \omega_{t-2} + \cdots + \emptyset_p \omega_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + c \end{cases} \quad (2)$$

U_Q_t is the series of monthly electricity consumption trend components; d is the number of differences; B is the lag operator, i.e., $B^k \times U\_Q_t = U\_Q_{t-k}$; $\omega_t$is the smooth series formed by d differences of monthly electricity consumption trend components; c is a constant; $\emptyset_1, \emptyset_2, \cdots, \emptyset_p$are the autoregressive coefficients and p is the autoregressive order; $\theta_1, \theta_2, \cdots, \theta_q$ are the moving average coefficients and q is the moving average order; ε_t is a white noise series with mean 0 and variance σ^2.

The formal expression of the seasonal cycle component is shown in Equation 3.

$$U_{X_{i,j}} = \alpha U\_X_{i-1,j} + (\alpha - 1) U_{X_{i-2,j}} + \cdots + \alpha(\alpha - 1)^{n-1} U_{X_{i-n,j}} \quad (3)$$

$U_{X_{i,j}}$ is the seasonal cycle component of monthly electricity consumption in month j of year i.

The steps for parameter estimation and testing of the trend component model are as follows.

Step 1: Smoothness test. The ADF test is used to test the smoothness of the d-order difference series in U_Q in turn until the series is smooth, and the corresponding number of differences d is determined at this point.

Step 2: Model order determination (determine the orders P and q). Determine the orders P and q of the ARIMA model based on the truncated and trailing properties of P and q.

Step 3: Parameter estimation. Estimation of parameters other than p, d, q in the model by least squares.

Step 4: Model testing. The residual series ε of the constructed model is tested to see if it is a white noise series. This process is still based on the autocorrelation and partial autocorrelation coefficients. If the autocorrelation and partial autocorrelation coefficients tend to zero at the k-period lag, it indicates that the residual series is a white noise series and turn to step 5; otherwise, it is a non-white noise series and return to step 2 to reconstruct the model.

Step 5: Forecasting. After obtaining the appropriate model, apply the model to forecast the monthly electricity consumption trend components for the future period.

The steps for estimating and testing the parameters of the seasonal cycle component model are as follows: according to the principle of "far smaller and nearer", the historical seasonal cycle values are assigned certain weights and summed up, and the calculated values are taken as the seasonal cycle components of the current monthly electricity consumption. α is the weighting coefficient, which is taken between 0.1 and 0.5 when the seasonal cycle components do not fluctuate much, otherwise it is taken between 0.6 and 0.8. -The value of α is the weighting coefficient, and the value of this coefficient is between 0.1 and 0.5 when the seasonal cycle component does not fluctuate much, otherwise it is between 0.6 and 0.8.

The steps for parameter estimation and testing of the stochastic component model are as follows: the values of the historical stochastic components are averaged over the same period and their values are used as the stochastic components of the current monthly electricity consumption.

In the actual system, a model library for power marketing data processing is constructed to facilitate the adaptation of the required processing model to the application scenario.

5 Conclusion

This paper provides a design for building an industry map based on electricity marketing data. By establishing the industry labels of enterprises and combining GIS information, this design establishes a two-dimensional framework that can present the development status and trend of specific industries in a specified region from different granularity. Under this framework, we use the characteristics of broad coverage, real-time, accuracy and structured degree of power marketing data to summarize the status and trend of power consumption of specific industries in the designated area, so as to "angiography" the development status and trend of specific industries in the designated area, and provide scientific, systematic and real-time decision support for local government's industrial decision, industry organizations and enterprises' business decision.

The system can be further integrated with enterprise operation data, patent data, recruitment data, judicial data, etc. to enrich its data sources, reflect more comprehensively the development status and trends of specific industries in designated regions, and expand its application scenarios.

References

[1] Qin Meng, Tang Guangsheng, Zhang Yuchen, et al. Analysis of macroeconomic mixed-frequency forecasting in China based on "Keqiang Index"[J]. Statistics and Decision Making, 2021(13):4.

[2] Dong Yu, Ma Bing. Construction and empirical test of "Keqiang index" version 2.0 [J]. Economic and Management Research, 2015, 36(11):7.

[3] Tian Shiming, Gong Taorong, Huang Xiaoqing, et al. Forecasting regional E-GDP value using power big data[J]. Electric Power Automation Equipment, 2019, 39(11): 198–204.

[4] Peng Fang, Qi Yaru, Ren Junda, et al. Research on now casting industrial added value based on power big data--analysis based on LSTM[J]. Price: Theory & Practice, 2021(7): 110–114.

[5] Wang Bao, Yang Min, Li Zhou, et al. The relationship between economy and the electricity consumption and power forecasting in black metal smelting and rolling processing industry in Anhui

Province[J]. Electric Power, 2018, 51(5): 179–184.
[6] Ruan L., Long Y., Zhou T., et al. Design and implementation of an online atlas of urban industrial economy[J]. Mapping and geographic information, 2021, 46(1):4.
[7] Li G.D., Gu X.Y., Xu X.M., et al. Research on the visualization method of location information in industrial maps[J]. Journal of Beijing University of Information Science and Technology: Natural Science Edition, 2019, 34(5):6.
[8] Xing Xuejun. Elements of industrial map: points, chains, veins and prospects[J]. China Investment (English and Chinese), 2003(12):2.
[9] He Zhengchu, Wang Jiao, Wu Jingjing, et al. Geographical distribution of production capacity and output of China's automobile manufacturing industry[J]. Economic Geography, 2018(10):9.
[10] He Zhengchu, Wang Jiao, Cao Wenming. Industrial map of China's automotive manufacturing industry and factors affecting industrial layout[J]. Scientific Decision Making, 2018(5):29.