

Enterprise Power Consumption Data and GDP Forecasting Based on Ensemble Algorithms

Yibin Xu^{1*}, Lu He², Ying Liang³, Jianhong Si⁴ and Yonglong Bao⁵

¹Marketing Service Center of State Grid Gansu Electric Power Company, Gansu, 730000, China

²Jiayuguan Power Supply Company of State Grid Gansu Province Electric Power Company, Gansu, 730000, China

³Tianshui Power Supply Company of State Grid Gansu Electric Power Company, Gansu, 730000, China

⁴Dingxi Power Supply Company of State Grid Gansu Electric Power Company, Gansu, 730000, China

⁵Jinchang power supply Company of State Grid Gansu Electric Power Company, Gansu, 730000, China

Abstract: This paper focuses on the development of regional GDP and proposes a method proposed for forecast of enterprise power consumption data and GDP based on ensemble algorithms. The enterprise power consumption data are used as independent variables and GDP data as dependent variables. A multiple linear regression model is selected as the primary learner for training and its outputs will be sorted into a new dataset of input features to train a secondary learner. The forecast of GDP is thus realized through ensemble learning.

1 Background

As one of the major indicators for economic development, gross domestic product (GDP) is referred to by the government as a significant basis to formulate development strategies and policies, and it is closely bound up with the development of regional enterprises, so the forecasting on regional GDP is of great value to the study on our national economy. Enterprises' power consumption data can be used to assess the changes in regional GDP and help figure out the economic growth trend in economic reforms and development transformation. GDP will be forecasted based on association analysis and forecasting on historical GDP data and various factors that influence GDP[1]. The methods employed to make such forecasting are generally ARIMA model and regression analysis. GDP forecasting based on regression analysis will analyze historical GDP data and the fitting functions of other impact factors, but it will only use overly complex data which are difficult to express and will process non-linear data. ARIMA model-based GDP prediction will presume GDP time series are stable, and then it will predict the time series data, but for unstable data, this model will be unable to capture changes[2,3,4].

The enterprise power consumption data and GDP forecasting based on ensemble algorithms, proposed in this paper, uses enterprise power consumption data as independent variables and GDP data as dependent variables. Multiple linear regression is selected as individual learner for training and the outputs of the linear regression will be applied as input features which will be collected as a new data set for the training of the meta-learner, thereby obtaining the forecasted GDP results by ensemble learning.

* Corresponding author: 17610027414@163.com

2 Algorithm Principle

2.1 Selection of Base Learner

Linear regression, as supervised learning, was utilized to predict the relationships between input variables and output variables. It was a regression analysis because the model was established based on the relationships between one or multiple independent variables and dependent variables by the use of least square function.

Among the given data set $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$, x indicates the enterprise' power consumption data, y the historical GDP data, with an attempt to learn a linear model to predict new GDP value, so the function is[5,6]:

$$f(x_i) = wx_i + b, s.t. f(x_i) \approx y_i \quad (1)$$

Loss function depicted the matching degree between models and training samples. Often, for regression issue, mean-square errors are used as loss function $E_{(w,b)}$. Finding solutions for w and b will minimize the loss function, the process of which is called the 'parameter estimation' of the least square. Take the derivative of w and b in the loss function, we get:

$$\frac{\partial E_{(w,b)}}{\partial w} = 2(w \sum_{i=1}^m x_i^2 - \sum_{i=1}^m (y_i - b)x_i) \quad (2)$$

$$\frac{\partial E_{(w,b)}}{\partial b} = 2(mb - \sum_{i=1}^m (y_i - wx_i)) \quad (3)$$

Setting the above formula equal to 0, optimal solutions for w and b would be:

$$w = \frac{\sum_{i=1}^m y_i(x_i - \bar{x})}{\sum_{i=1}^m x_i^2 - \frac{1}{m}(\sum_{i=1}^m x_i)^2} \quad (4)$$

$$b = \frac{1}{m} \sum_{i=1}^m (y_i - wx_i) \quad (5)$$

There into, $\bar{x} = \frac{1}{m} \sum_{i=1}^m x_i$ is the mean value of x .

There are deficiencies when putting simple linear regression into practical use, so adding penalty constraint to the least square estimation will make the estimation of some parameters minimal or equals to 0, thus, to a certain degree, avoiding the model to be ‘over-fitting’.

One problem with linear regression is that under fitting is possible because it is an unbiased estimate with a minimum mean square error. Obviously, if the model is under fitted, the best prediction effect will not be achieved.

If more features than the sample points, that is to say, the input data matrix x is not full rank matrix, a full rank matrix in inverse problems, so the introduction of ridge regression, in a nutshell, is a constant and units in the matrix and matrix, the constant, the greater the variance is smaller, the greater the deviation, by introducing the penalty term, can reduce the parameter is not important.

2.2 Stacking Ensemble

Ensemble learning methods will train and combine multiple learners to solve problems. Base learners are usually generated from training data by a base learning algorithm, and through ensemble learning can form generalization ability, more powerful than base learners. Common ensemble learning methods include bagging, boosting and stacking. The method used to combine individual learners together is called the combining strategy. For classification problems, we can use voting to select the class that outputs the most. For regression problems, we can calculate the average value of the output results of the classifier. Stacking was chosen to predict GDP after the practical application scenario was considered.

The basic idea of stacking was that first it used raw training data set to train individual learners, and then the output from individual learners were used as input features, plus the corresponding original markup as new markups, so a new data set were obtained to train the meta-learners. [7,8] To prevent the model from over-fitting, cross validation or leave-one-out cross validation was employed. The structure of Stacking is shown in the figure:

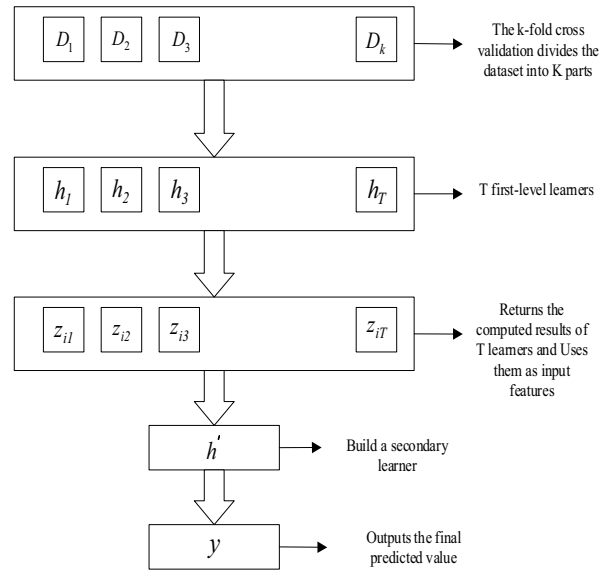


Figure1. Structure Diagram of Stacking

The steps are as follows:

Step 1: Split the training data set D into K folds. Randomly choose one as validation set and the rest testing sets.

Step 2: Choose individual learners. Linear regression is selected here, and fit the validation set and training sets to learners for training and getting outputs $z = \{(z_{i1}, z_{i2}, \dots, z_{iT})\}$.

Step 3: Treat the outputs from individual learners as new data set and input such data set to meta-learners for training.

Step 4: Outputs from meta-learners will be regarded as the final outputs.

3 Data Processing and Effect Analysis

Three-year GDP data and power consumption data from one region were selected. These numeric data were standardized to make their features fall within approximately same range of numeric values, thus realizing the uniform scaling of raw data. Acquired data are:

Table 1. Enterprise electricity data

COLUMNS	MEANS	DATA_TYP E	EXAMPLE
ID	Enterprise index	object	080899279 57
DATA_DAT E	Data date	date	2019-01-03
P	Load data	float	198.3
I	Current data	float	78.2
U	Voltage data	float	234.4
E	Power data	float	1897.2
...

After the preliminary data processing, corresponding base learners were selected, according to practical

scenario, for model learning and training. Firstly, A first-level learner is constructed by taking the electricity consumption data of the enterprise as the input feature and the GDP in previous years as the output feature. Secondly, the results of the first-level learner are taken as input features, and the corresponding original GDP marker is used as the new marker to form a new data set to train the second-level base learner. As you can see, the learner after the Stacking has a better performance and a higher fitting, as shown in the figure:

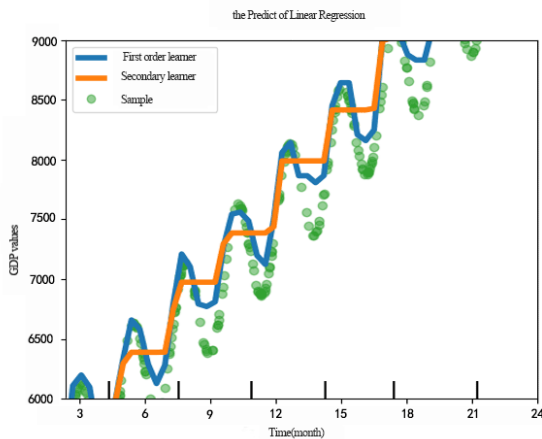


Figure 2. Linear Regression Prediction Chart

Linear regression will be applied to highly ill - conditioned matrices. For such matrices, a slight change in the target variable can cause huge variances in the calculated weights. In such cases, it is useful to set a certain regularization (alpha) to reduce this variation (noise). When alpha is very large, the regularization effect dominates the squared loss function and the coefficients tend to zero. At the end of the path, as alpha tends toward zero and the solution tends towards the ordinary least squares, coefficients exhibit big oscillations. In practise it is necessary to tune alpha in such a way that a balance is maintained between both. Each color represents a different feature of the coefficient vector, and this is displayed as a function of the regularization parameter. as shown in the figure:

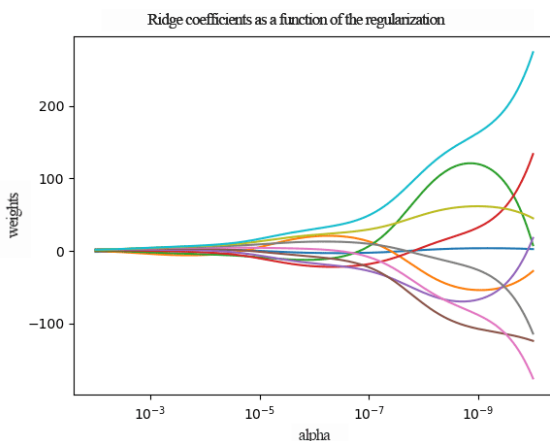


Figure 3. Ridge Coefficients as a function of the Regularization

Comparing to simple LR model, GDP forecasting model experienced noteworthy improvement of accuracy after Stacking ensemble, so this model learned the features from different power consumption data, and improved its accuracy by fusing these features to the model.

After linear regression analysis, we considered the Stacking learning machine model architecture built in combination with the strategy to perform a comprehensive analysis. The model architecture of the Stacking learning machine is:

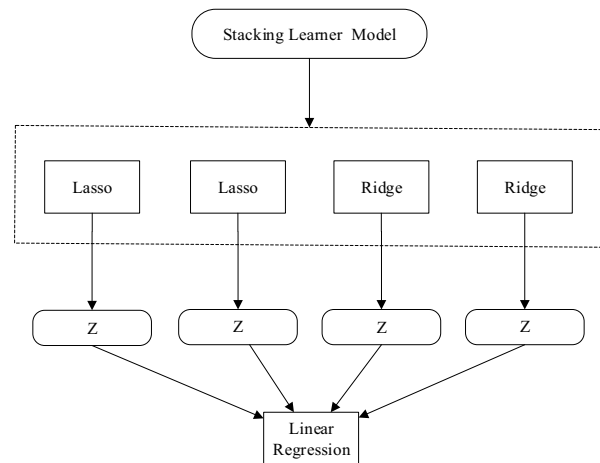


Figure 4. The structure diagram of Stacking Learner

The performance after linear regression and Stacking learner was aggregated, as shown in the figure below:

Table 2. Multiple results comparison table

Model	Precision	Recall	F1-sorce
LR	81.3%	84.2%	73.3%
Lasso	83.5%	94%	84.9%
Ridge	77.9%	83.7%	82.0%
Stacking Learner	92.0%	93.6%	94.2%

4 Conclusion

In this paper, the power of the enterprise data as independent variables, region's GDP data as the dependent variable, through linear regression, with L1 regularization Lasso, with L2 regularization of the Ridge and Stacking study comparing four kinds of methods, including Stacking of learning is to choose the best learning in linear regression, expressive best model study for Stacking.

Stacking indicates that learning is not from the effect of multi-layer stacks, but from the ability of different learners to learn various features. Forecasting the Enterprise' power consumption and GDP by ensemble learning receives better generalization ability from the model, compared to the deficiency in single forecasting model. The electricity consumption of enterprises reflects the business benefits of enterprises and the

economic development of urban areas, and provides decision support and auxiliary support for enterprises' development planning, formulation of urban macroeconomic policies and economic structural adjustment. GDP forecasting can provide a clearer idea on the GDP trend, so the economic development strategies can be timely adjusted in accordance with the results generated.

References

1. Wang Weiwei. Correlation between enterprise electricity Information data and Macroeconomic Forecast [A]. Power Informatization Professional Committee of Chinese Society of Electrical Engineering, 2017 Annual Conference proceedings of power Industry informatization [C]. Electric Power Informatization Committee of The Chinese Society of Electrical Engineering: Editorial Department of Telecom Science, Posts and Telecommunications Press, 2017:4.
2. Wang Fangfang, WANG Qian. Analysis and Prediction of GDP of Shaanxi Province based on ARIMA Model [J]. Industrial Innovation Research, 2020(19):22-24.
3. OFFICIAL NEWS; EBRD revised Kazakhstan's GDP forecast for 2020 further down [J]. Interfax: Kazakhstan Mining Weekly, 2020.
4. Zhang Yin, Zhan Qiwen, Hu Weian. GDP forecast based on NPP-VIIRS night Light Data -- A case study of Jiangxi Province [J]. Automation Applications, 2020(09):59-61.
5. Ethen Alpaydin. Introduction to Machine Learning [M]. Machinery Industry Press: Beijing, 2014.3:62.
6. Zhou Zhihua. Machine Learning [M]. Tsinghua University Press: Beijing, 2014.3:53.
7. Zhou Zhihua, LI Nan. Foundation and algorithm of Integrated Learning [M]. Electronic Industry Press: Beijing, 2020.8:76.
8. Zhuge Yue, calabash baby. Baima Machine learning algorithm engineer takes you to interview [M]. Posts and Telecommunications Press, Beijing, 2018.8:276.