Abnormal Detection Model of Energy Consumption Data in Beneficiation and Metallurgy Enterprises based on Transfer Learning

Zhuchao Yu¹, Gaixia Chu², Yuanzheng Zhang³, Siyuan Zhou⁴, Shaowei He⁵, Jing Li⁶

¹School of Business Administration, Northeastern University, Shenyang, China
²State Key Laboratory of Process Automation in Mining & Metallurgy, Beijing, China, Qinghai XKXX Information Technology Co., LTD, Xining, China
³College of Medicine and Biological Information Engineering, Northeastern University, Shenyang, China
⁴School of Business Administration, Northeastern University, Shenyang, China
⁵Qinghai XKXX Information Technology Co., LTD, Xining, China
⁶School of Business Administration, Northeastern University, Shenyang, China

Abstract—The energy consumption data of beneficiation and metallurgy enterprises has the characteristics of multi-dimensional and timing. An energy consumption data anomaly detection model based on transfer learning is proposed to help enterprises monitor the abnormal energy use. The combination of transfer learning and DTW algorithm is used to eliminate the adverse impact of timing on the model, and the ensemble learning is used to realize data detection. Finally, an efficient and accurate energy consumption data anomaly detection model is formed. The detection accuracy of the model is 91.6%, which can effectively meet the requirements of enterprise energy management.

1. INTRODUCTION

Since the reform and opening up, China's energy demand and energy consumption have been growing. The development and operation on a platform with higher energy efficiency and higher sustainability are of great significance to improve China's economic development. At present, the era of "big data" has come. If we can correctly use the enterprise's "big data" to predict the enterprise's energy consumption and carry out corresponding early warning, prediction, or optimal adjustment, we can further save energy and improve energy efficiency and operating benefits.

The beneficiation and metallurgy industry is an important part of China's national economic operating system. However, China's beneficiation and metallurgy industry has been excessively dependent on energy resource consumption, and the way and the amount of energy usage are lack of control [1]. This extensive way of energy use leads to the intensification of the contradiction among energy consumption, economic development, and environmental protection, and affects the sustainable development of the economy. Among them, the metallurgical industry, as a classic manufacturing enterprise, the consumption, and output of resources have a great impact on the energy use of the
manufacturing industry, which has been highly valued by the state and governments at all levels for a long time. Changing the extensive way of energy use and realizing the energy informatization of manufacturing industry is the inevitable requirement of China’s manufacturing industry transformation.

In energy informatization, energy anomaly detection is a key link to maintain the rational utilization of energy. The traditional extensive energy use mode does not care about the energy input at a certain time, nor can it form a data-based energy use. The energy detection system can monitor abnormal data in real-time based on the existing energy data, help enterprises adjust the energy input in time, solve energy problems, and finally realize high-precision digital energy management.

For manufacturing enterprises, energy consumption data shows the characteristics of diversity, timing, and small samples. Diversity means that the classification of energy consumption data involves many parameters. For a copper smelting enterprise, the energy consumption data includes general bituminous coal consumption, coke consumption, mining volume, comprehensive water consumption and other data, and the data types are diverse, so it is difficult to mine the data items that have a key impact on the change of energy consumption. Time-sequence refers to the data column recorded by the same indicator in chronological order. Generally speaking, the energy consumption data of the same enterprise should change periodically with the seasonal changes, and the data of the same period in different years are obviously comparable and in line with the characteristics of the time sequence. A small sample means that the sample data with the same characteristics is insufficient. For enterprises, the earlier an energy anomaly detection system is established, the more conducive it is to enterprise energy management.

Anomaly detection refers to the assumption that the intruder’s activities are different from normal activities. According to this concept, the database of the subject’s normal activities can be established and the current subject’s activities can be compared with the database [2]. Transfer learning refers to using old knowledge to obtain new knowledge. The main goal is to quickly transfer the learned knowledge to a new field. The deep learning method based on transfer learning has high requirements for the correlation between the data source domain and the target domain [3,4]. Energy consumption data does not have such a large data set for training as Imagenet data set in the field of computer vision and has strong timing, so it is not suitable for model fine-tuning and data enhancement methods. However, most of the data sets in the source and target fields provided by energy consumption data come from the same group, and the production mode is highly similar and remains unchanged for a long time. Therefore, they have high relevance, and transfer learning can be considered.

The research idea of this paper is as follows: firstly, the multivariate energy consumption data are processed by PCA to reduce the dimension and principal component analysis, to alleviate the problems of the unbalanced dimension of energy consumption data, and process the energy consumption data into characteristic data more suitable for in-depth learning method; DTW dynamic time warping algorithm is used to take into account the timing characteristics of energy consumption data, convolution neural network is used to train the sample data, and the idea of transfer learning is used to fine-tune the small sample characteristics of data; finally, an anomaly data classifier is constructed by ensemble learning, to find an anomaly detection model to solve small sample data.

2. Model
2.1. PCA dimensionality reduction algorithm
PCA (principal component analysis algorithm) is to reduce the data dimension by extracting the characteristic data of principal components[5]. It has good applicability in dealing with complex multi-dimensional data. Wang Yongjian and others have selected industrial time-series data of multi-region operation based on PCA algorithm to extract variables with high attention, which has laid a good foundation for further research [6]. There are two common implementations of PCA algorithm. One is based on an eigenvalue decomposition covariance matrix, but its data matrix is required to be a square matrix for calculation, which has great limitations; another implementation is SVD singular value decomposition covariance method. SVD is a special eigenvalue decomposition method for solving
nonsquare matrix data matrices. Compared with the eigenvalue decomposition method, SVD method has wider application and lower time complexity.

2.2. DTW dynamic time warping
Dynamic time warping (DTW) is a similarity or distance function based on time series data. It was first applied in the field of speech recognition. It is an algorithm to determine the similarity between two-time series. With the help of DTW, the similarity of two different time series can be extracted to remove the influence of time series fluctuation on the data. The basic principle of DTW is as follows:

The sample sequences \( X \) and \( Y \) are given respectively, and a point-to-point Euclidean distance function of the sequence is given as \( d(i, j) = f(x_i, y_j) \geq 0 \). The core of DTW is to solve the distorted curve, \( \phi(k) = (\phi_x(k), \phi_y(k)) \), in which the possible values of \( \phi_x(k) \) are \( 1, 2, ..., N \), and the possible values of \( \phi_y(k) \) are \( 1, 2, ..., M \), \( k = 1, ..., T \). Then, solve for the cumulative distance between two sequences:

\[
d_{w}(X, Y) = \sum_{i=1}^{T} d(\phi_x(k), \phi_y(k))
\]  

(1)

The goal of solving DTW is to find the most appropriate distortion curve \( \phi(k) \) to minimize the cumulative distance, which is the value of the last row and last column of the loss matrix, that is

\[
DTW(X, Y) = \min \phi d(\phi, X, Y)
\]  

(2)

2.3. Ensemble learning
The ensemble learning method completes learning tasks by constructing and combining multiple learners, sometimes referred to as multi classifier system, committee-based learning, and so on. The method of multi-model ensemble learning helps to describe the data set more accurately and improves the utilization efficiency of data for small sample data.

When the sample data set is small, there is not enough data to support the complete separation and training of the training set and the data set, so it is impossible to simply train the basic model on all the training data to produce an effective prediction. The training network of ensemble learning is simpler than the neural network used as basic training by transfer learning, and can effectively alleviate the phenomenon of easy overfitting in the training process of small sample data. Here, using the K-fold cross-training method, the training set, and the test set are divided into \( n \) mutually exclusive subsets, one of which is taken as the verification set, and the remaining \( n-1 \) is taken as the training set for \( n \) tests, and \( N \) results are obtained.

2.4. Anomaly detection model based on Transfer Learning
Transfer learning refers to training a basic network on the source data set, and then transferring the learned features (network weight) to the second network trained on the target data set, to achieve the purpose of inheriting the learning experience of the source data set.

To obtain the best source data set of the target data set, this paper proposes to use the dynamic time warping algorithm (DTW) to measure and quantify the similarity among the data sets, supplemented by the clustering algorithm to complete the screening of the source data set, to form an expanded new data set with a large number of samples relative to the source data.

In the process of selecting a basic training network, because the time series data set is very small, many classical neural networks are not applicable, and too complex networks are easy to produce overfitting phenomenon, resulting in huge accuracy differences on different data sets. Therefore, this paper selects the FCN network with the relatively simple structure, as shown in Figure 1.
In Figure 1, the first, second, and third layers are convolution layers with the rectified linear unit (ReLU) as the activation function. Each convolution layer takes a time series as input and converts it into a multivariate time series; the fourth layer consists of a global average pool. The layer structure accepts the input of the third convolution layer and averages each time series on the time axis. This averaging operation greatly reduces the number of parameters in the depth model and allows the use of class activation mapping to explain learning features. Finally, the output of the fourth layer is input to a Softmax classification layer with the number of neurons as the number of data set classes.

Because the training network convolution neural network used in transfer learning is still too complex, the direct use of convolution neural network training data will still produce overfitting. To ensure the prediction accuracy of the model and reduce the one-sidedness of the model, the ensemble learning is used to train the data and integrate the training data of multiple simple models. The model is shown in Figure 2. The model is a two-tier classification model architecture. In the first layer of the model, we use RandomForest, AdaBoost, ExtraTrees, GBDT, DecisionTree, KNN, SVM to train the data respectively.

3. DATA EXPERIMENT

3.1. Data Sources
This paper selects the energy consumption data of a beneficiation and metallurgy enterprise as the data source, which mainly includes two parts: the first is the 2018-2019 energy report data of a copper mine company; second, the local weather, temperature, wind, and other conditions of the mine. In addition, abnormal data in the data set is marked according to the energy management files provided by the company.

3.2. Data processing
Firstly, the production-related factors such as output and water consumption and external environmental factors such as weather and wind are treated by PCA. After dimensionality reduction and feature screening, it can be found that indicators such as product output, power consumption, comprehensive water consumption per unit product, comprehensive power consumption per unit product and standard
coal consumption are relatively important in the production process of the beneficiation and metallurgy enterprise.

The energy consumption data of beneficiation and metallurgy enterprises have obvious timing, which is a typical time series. Therefore, the problem of anomaly detection based on energy consumption data can be transformed into the construction of a time series classification (TSC) model. DTW algorithm is used to reduce the impact of data timing. Finally, according to the temporal correlation between the target data set and the source data set, the list of data sets is output from high to low.

3.3 Model Training

The target data set used in the model training is the data of a copper company from 2017 to 2018, and the source data set is the energy consumption-related data of other subsidiaries.

The model training steps are shown in Figure 3.

3.4 Result analysis

• Prediction accuracy of transfer learning model

The optimization effect of the transfer learning model is evaluated by the F1-score index. Precision is used to represent the accuracy of the classification model, recall is used to represent the recall rate of the classification model, and the calculation formula of the F1-score index is as follows

$$F_1 \text{- score} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

Finally, the accuracy of the extended data set based on the transfer learning output is shown in Table 1.

| Base learner | Rf  | Ada | Knn  | XGB | SNN |
|--------------|-----|-----|------|-----|-----|
| F1-score     | 0.846 | 0.875 | 0.761 | 0.833 | 0.916 |

• Comprehensive evaluation of results
Transfer learning can effectively alleviate the impact of insufficient sample data on model training. The best training effect can be obtained by using four basic learners and the Stacking framework. Based on this model, the energy consumption data from an enterprise are calculated, and the accuracy of anomaly detection results is 92%.

4. CONCLUSION
The abnormal detection model of energy consumption data in beneficiation and metallurgy enterprises based on PCA dimensionality reduction algorithm, transfer learning, DTW dynamic regularization algorithm, and ensemble learning proposed in this paper can effectively adapt to the diversity, and timing of energy consumption data and realize accurate abnormal detection of energy consumption.

However, the selection of source data sets will have a significant impact on the performance and generalization ability of the model. Therefore, we need to pay attention to the relationship between the target dataset and the source dataset. In addition, how to reasonably select baseline learners to improve the effect of ensemble learning is the next research orientation.

ACKNOWLEDGMENT
Our thanks to the 2019 National College Students Innovation and Entrepreneurship Training Program at Northeastern University, China. And, funded by the Open Foundation of State Key Laboratory of Process Automation in Mining & Metallurgy.

REFERENCES
[1] Blam I and Kovalev S 2013 Ecological management in ferrous and non-ferrous metals industrial enterprises Voprosy statistiki 6 22-30.
[2] G.E.P. Box and Davida Pierce 1968 Distribution of residual autocorrelations in autoregressive-integrated moving average time series models Journal of the American Statistical Association 65 (332) 1509-1526.
[3] Krizhevsky A, Sutskever I, Hinton G. ImageNet Classification with Deep Convolutional Neural Networks[C]// NIPS. Curran Associates Inc. 2012.
[4] Pan S J, Yang Q. A Survey on Transfer Learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(10):1345-1359.
[5] Wang Y, Yang K, Li H. Industrial Time-series Modeling via Adapted Receptive Field Temporal Convolution Networks Integrating Regularly Updated Multi-Region Operations based on PCA[J]. Chemical Engineering ence, 2020, 228:115956.
[6] Wang F, Wu S, Zhang W, et al. Emotion recognition with convolutional neural network and EEG-based EFDMs[J]. Neuropsychologia, 2020, 146:107506.